



Inter-laboratory variability in in vitro spinal segment flexibility testing

Daniel J. Wheeler^a, Andrew L. Freeman^{b,*}, Arin M. Ellingson^a, David J. Nuckley^a,
Jenni M. Buckley^c, Justin K. Scheer^c, Neil R. Crawford^d, Joan E. Bechtold^{a,b}

^a University of Minnesota, Minneapolis, MN, USA

^b Excelen Center for Bone and Joint Research, Minneapolis, MN, USA

^c Department of Orthopaedic Surgery, University of California, San Francisco, CA, USA

^d Spinal Biomechanics Laboratory, Barrow Neurological Institute, Phoenix, AZ, USA

ARTICLE INFO

Article history:

Accepted 29 June 2011

Keywords:

Laboratory comparison
Spine biomechanics
Pure moment
Flexibility testing
Lumbar spine range of motion

ABSTRACT

In vitro spine flexibility testing has been performed using a variety of laboratory-specific loading apparatuses and conditions, making test results across laboratories difficult to compare. The application of pure moments has been well established for spine flexibility testing, but to our knowledge there have been no attempts to quantify differences in range of motion (ROM) resulting from laboratory-specific loading apparatuses. Seven fresh-frozen lumbar cadaveric motion segments were tested intact at four independent laboratories. Unconstrained pure moments of 7.5 Nm were applied in each anatomic plane without an axial preload. At laboratories A and B, pure moments were applied using hydraulically actuated spinal loading fixtures with either a passive (A) or controlled (B) XY table. At laboratories C and D, pure moments were applied using a sliding (C) or fixed ring (D) cable-pulley system with a servohydraulic test frame. Three sinusoidal load-unload cycles were applied at laboratories A and B while a single quasistatic cycle was applied in 1.5 Nm increments at laboratories C and D. Non-contact motion measurement systems were used to quantify ROM. In all test directions, the ROM variability among donors was greater than single-donor ROM variability among laboratories. The maximum difference in average ROM between any two laboratories was 1.5° in flexion-extension, 1.3° in lateral bending and 1.1° in axial torsion. This was the first study to quantify ROM in a single group of spinal motion segments at four independent laboratories with varying pure moment systems. These data support our hypothesis that given a well-described test method, independent laboratories can produce similar biomechanical outcomes.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Spinal implant devices must demonstrate safety and efficacy before being introduced into clinical use. Spinal construct efficacy is currently assessed with a battery of tests that includes material biocompatibility, basic bench top mechanical testing, wear testing, animal studies, full-construct kinematic evaluations and failure testing. While each test method is designed to evaluate a specific characteristic of the device, the overall goal of biomechanical testing is to prove device efficacy in a model that is most representative of the final clinical construct. A significant portion of testing is therefore focused on the use of cadaveric motion segments to evaluate spinal kinematics under controlled loading conditions. Variation in loading conditions, test apparatuses, motion measurement techniques and data reduction algorithms used between laboratories has made the findings of different

research groups difficult to compare. This variability and the lack experimental standards make the demonstration of device efficacy difficult.

Over the past three decades, efforts have been made to standardize protocols for in vitro biomechanical testing of spinal implants, particularly in the quantification of specimen range of motion (i.e. flexibility testing). Non-constraining, pure moment loading in the three anatomic planes has been recommended, using either no preload or a compressive follower load system (Wilke et al., 1998b; Panjabi, 1988; Goel et al., 2006). It has been suggested that protocol standardization will enable new and existing devices to be compared in a laboratory-independent manner (Goel et al., 2006; Panjabi, 2007). For device comparison to be truly laboratory-independent, individual laboratories must adopt a common loading protocol and accurately apply and measure the agreed upon parameters. Given the high degree of specificity and variability in loading and motion measurement systems, consistency may not always be achieved. Conventional wisdom suggests that well-described in vitro test protocols (Wilke et al., 1998b; Goel et al., 2006; Panjabi, 2007) will mitigate any technical difference between

* Corresponding author. Tel.: +1 612 336 6612; fax: +1 612 336 6619.
E-mail address: afreeman@excelen.org (A.L. Freeman).

laboratories, but to our knowledge, there have been no attempts to confirm this assumption. In order to evaluate spinal implants on a laboratory-independent basis, it is first necessary to evaluate laboratories on a spine-independent basis. Testing the same specimens in separate laboratories enables the isolation of particular inter-laboratory differences and their effects on the biomechanical outcomes of in vitro testing of spinal segments.

In the present study, four independent laboratories capable of in vitro pure moment loading were provided with a single set of intact lumbar motion segments and instructed to test using a flexibility-testing protocol (Wilke et al., 1998b; Panjabi, 1988; Goel et al., 2006). Two of the laboratories applied pure moments using hydraulic actuators, whereas the other two utilized cable-pulley systems (Crawford et al., 1995; Eguizabal et al., 2010). These techniques have been individually recognized in the literature, but never have they been subjected to a side-by-side comparison. It was hypothesized that a well-described in vitro pure moment loading protocol without axial preload would result in reproducible ROM across the four laboratories.

2. Methods

2.1. Specimen preparation

Seven fresh-frozen human lumbar motion segments ($n=6$ L1–L2, $n=1$ L3–L4) were harvested (4 male, 3 female, age range: 45–75 years, mean age: 59.3 years). The specimens were screened radiographically for anatomic abnormalities and dissected of soft tissue, leaving all ligamentous and bony structures intact. Wood screws were placed into the exposed vertebral body endplates and specimens were embedded in a urethane resin (Smooth-On 300, Smooth-On Inc., Easton, PA) with the mid-disk plane aligned horizontally.

2.2. Pure moment testing protocol

The intact specimens were tested independently at four biomechanics laboratories using pure moment loading conditions based on recommendations in the literature (Wilke et al., 1998b; Panjabi, 1988; Goel et al., 2006; Niosi et al., 2006; Oxland et al., 1992). Each laboratory applied unconstrained pure moments of ± 7.5 Nm in flexion-extension, lateral bending and axial rotation. Specimens were tested without preload in order to maintain uniformity among laboratories and avoid potential confounding effects.

At laboratory A (Excels, Minneapolis, MN), pure moments were applied to the superior vertebral body by a hydraulically actuated gimbal mounted to a servohydraulic test frame (Fig. 1A; MTS 858 Mini Bionix, MTS, Minnetonka, MN). A 6 degree-of-freedom (DOF) load cell, located immediately above the specimen, was used to measure the applied loads and moments. The inferior vertebral body was allowed unconstrained movement on a passive X–Y table (resistance < 0.1 N),

while the actuator maintained a 0 N compressive load. Torsional moments were applied with the test frame actuator.

At laboratory B (University of Minnesota, Minneapolis, MN), pure moments were applied using three superior angular actuators on the crosshead of a servohydraulic load frame (Fig. 1B; 8821 Biopuls, Instron, Norwood, MA). Moment application and an axial preload of 0 N were controlled by a 6 DOF load cell (AMTI M4380, Watertown, MA) superior to the specimen. Transverse plane shear was minimized using a second inferior 6 DOF load cell and a controlled X–Y slide table.

At laboratory C (Biomechanical Testing Facility, San Francisco, CA), pure moments were applied using a uniaxial servohydraulic test frame (Fig. 1C; MTS 858 Mini Bionix, MTS, Minnetonka, MN) and a “sliding ring” cable-driven pulley for flexion-extension and lateral bending but not during axial rotation tests (Eguizabal et al., 2010). The sliding ring system allows freedom of movement for the pulley in the axial and anterior-posterior directions (Eguizabal et al., 2010). Inferiorly, the specimen was allowed unconstrained movement on a passive X–Y table.

At laboratory D (Barrow Neurological Institute, Phoenix, AZ), pure moments were applied using a uniaxial servohydraulic test frame (Fig. 1D; MTS 858 Mini Bionix, MTS, Minnetonka, MN) and fixed ring system of cables and pulleys (Crawford et al., 1995). A uniaxial load cell mounted to the actuator head was used to control cable tension (applied moment); a multiaxial load cell positioned beneath the inferior specimen cup was used to measure resultant moments at the base of the specimen.

In the laboratories using hydraulic gimbals (labs A and B), specimen loading consisted of three sinusoidal wave load-unload cycles to ± 7.5 Nm at 0.45 Nm/s. The first two cycles were used to precondition the specimen; ROM data were collected during the third cycle. In laboratories C and D loads were applied in 1.5 Nm increments and held for 45 s to a maximum load of 7.5 Nm. Laboratory C collected ROM data during the first cycle. In laboratory D, three preconditioning cycles to 7.5 Nm were applied for 60 s each. Data were collected after the third preconditioning cycle after a 60 s rest period.

Laboratories A and B utilized similar non-contact motion measurement systems to quantify ROM (Vicon, Oxford Metrics, UK). Three reflective markers were attached to each vertebral body and the three-dimensional position of the markers was measured using five infrared digital cameras. Experimental calibration found the system to accurately track displacements to 0.1 mm and rotations to 0.1°.

Laboratories C and D both used an Optotrak system (Optotrak 3020, NDI, Waterloo, Canada), with a reported resolution of 0.1°, to quantify ROM (Bozkus et al., 2004). A set of three noncollinear infrared diodes were attached to each vertebral body using screws (Lab C) or surgical guide wires (Lab D). The local coordinate system of the motion segment was established using a digitizing probe and customized software (Crawford and Dickman, 1997).

2.3. Quad-laboratory experimental design

After thawing, each specimen was tested on the same day at laboratories A and B. Specimens were wrapped in saline-soaked gauze, double-bagged, and placed in a protective, insulated plastic container prior to inter-laboratory transportation. Transit between labs A and B never exceeded 15 min. Each specimen was prepared and tested in less than 20 h. After testing, specimens were vacuum sealed to preserve moisture and re-frozen at -20 °C. Specimens were then packaged with dry ice and shipped overnight to laboratory C, where they were thawed, tested, re-frozen, and shipped to laboratory D. The testing process was repeated once again at laboratory D. Ultimately, the specimens were re-frozen and returned to laboratory A, where they were subjected to repeat testing. Laboratory A thus served as the control test site, in order to determine if

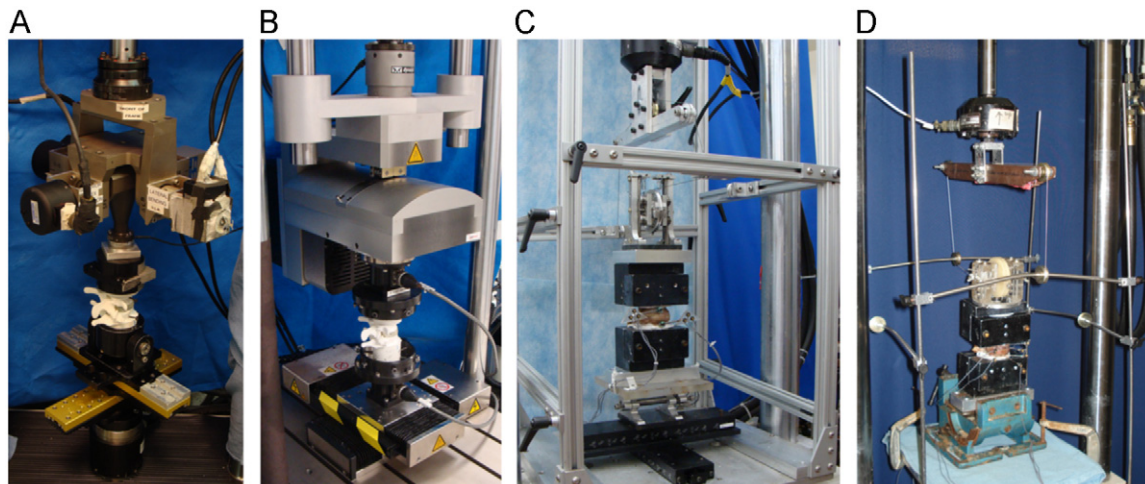


Fig. 1. Spinal testing apparatus for laboratories A, B, C and D. Laboratories A and B used hydraulically actuated gimbals while cable and pulley setups were used at laboratories C and D.

specimen physiology or motion characteristics had changed during previous testing. All testing occurred at room temperature ($22 \pm 5^\circ\text{C}$).

2.4. Outcome measures and statistics

For all test sites, coordinate systems were oriented with the positive x direction pointing left lateral, positive y superior and positive z anterior. At laboratories A and B, Euler angle calculations were performed in the R_y , R_z and R_x order to quantify specimen ROM. Laboratories C and D utilized the tilt/twist method to calculate specimen ROM (Crawford et al., 1999).

ROM was calculated as the difference between the peak positive and negative rotations. Using the maximum and minimum ROM values, the maximum ROM variability among donor specimens was calculated for each laboratory (“single-lab donor ROM variability”). The maximum variability in ROM for each individual donor specimen across all four laboratories was also calculated (“single-donor ROM variability”). In laboratory A, hysteresis load-deformation curves were created for each specimen during both initial and repeat testing. Hysteresis curves were not constructed at laboratories C and D due to the application of loads using a unidirectional, quasistatic loading technique.

The influence of angle calculation method (Euler or tilt/twist) in determining ROM was evaluated by converting 3D angular results from laboratory A from Euler to tilt/twist angles and assessing the discrepancy in ROM. Angles were converted by applying planar rotations to the identity matrix in the appropriate sequence to obtain a direction cosine matrix, and then calculating tilt/twist angles from components of the matrix (Crawford et al., 1999).

ROM data were compared statistically among laboratories using a one-way repeated measures ANOVA. Specific laboratory to laboratory contrasts were performed using paired Student’s t -tests with Shaffer’s S1 Bonferroni correction for multiple comparisons. Laboratory A and the repeat test were each considered different cases throughout this analysis. The base alpha acceptance was set at $p=0.05$.

3. Results

In all three loading directions, there was no significant difference in ROM among the four laboratories ($p=0.437$ for flexion-extension; $p=0.151$ for lateral bending; $p=0.069$ for axial rotation; Fig. 2).

Additionally, in all three loading directions the “single-lab donor ROM variability” was found to be greater than the “single-donor ROM variability” across the four laboratories (Fig. 3).

For flexion-extension and lateral bending, the greatest ROM variability among specimens was observed in laboratory B (Table 1). For axial rotation, the greatest ROM variability was seen in laboratory D. The maximum difference in average ROM between any two laboratories was 1.5° in flexion-extension, 1.3° in lateral bending and 1.1° in axial torsion.

A comparison of ROM data from the initial and repeat tests at laboratory A showed no significant difference in any of the three loading directions ($p=0.17$ in flexion-extension, $p=0.22$ in lateral bending, $p=0.14$ in axial rotation, Table 2). Hysteresis curves

Table 1
The average, minimum and maximum ROM’s at each laboratory.

Test direction	Laboratory ROM (deg.)			
	A	B	C	D
Flexion-extension				
Average	6.9	7.3	6.4	6.8
Minimum	4.4	4.4	3.9	3.0
Maximum	9.1	12.1	9.2	10.5
Lateral bending				
Average	7.5	8.7	7.2	8.3
Minimum	4.5	5.4	4.5	4.7
Maximum	11.2	14.9	11.0	14.1
Axial rotation				
Average	2.6	2.4	3.0	3.1
Minimum	1.1	1.3	1.2	1.2
Maximum	4.0	4.0	4.2	4.9

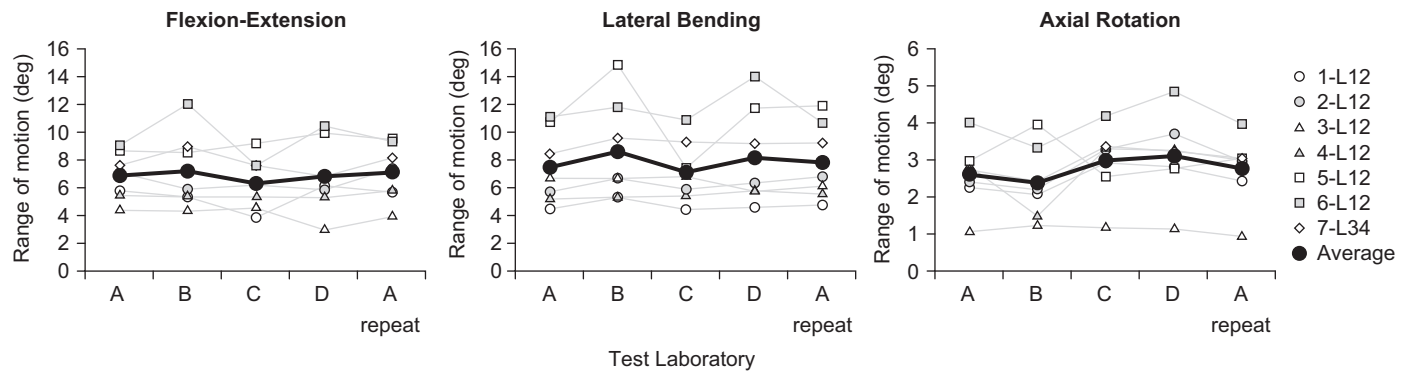


Fig. 2. Individual specimen and average ROM among laboratories A, B, C and D in each test direction. Laboratories A and B utilized hydraulic actuators, whereas C and D utilized cable/pulley systems.

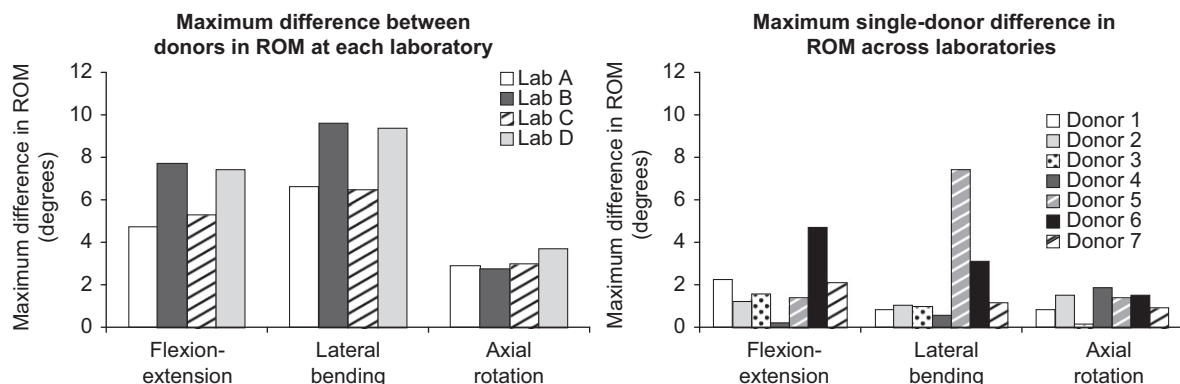


Fig. 3. (Left) Maximum difference in ROM between donors at each laboratory. (Right) Maximum difference in ROM for a single donor across all four laboratories.

developed from initial and repeat testing showed very good agreement in overall shape characteristics. Fig. 4 displays the most and least precise matches of initial versus repeat hysteresis plots during flexion-extension testing. Lateral bending and axial rotation overlays showed similar patterns.

The maximum discrepancy between the Euler angle-based ROM and tilt/twist-based ROM for any single specimen in laboratory A was 0.09° . The overall mean discrepancy was $0.02 \pm 0.03^\circ$ in flexion-extension, $0.01 \pm 0.02^\circ$ in lateral bending and $0.00 \pm 0.01^\circ$ in axial torsion.

4. Discussion

This study subjected a single group of lumbar motion segments to pure moment loading without axial preload at four independent biomechanics laboratories and found no significant differences in ROM between the laboratories. In all three anatomic planes, the ROM variability among donor specimens tested in a single lab was greater than the maximum single-donor ROM variability among the four laboratories. Thus, variability in individual donor specimens was a greater determinate of difference in ROM than the variability of testing apparatus or location. The results support our hypothesis that pure moment loading at independent laboratories can produce similar biomechanical outcomes.

The four participating groups in the current study are representative of the variety of test apparatuses that exist within the spine biomechanics community. The primary difference between the laboratories was the use of servohydraulic spinal loading fixtures to apply continuous pure moments versus cable-driven pure moment systems with stepwise unidirectional loading (Crawford et al., 1995; Eguizabal et al., 2010). It has been reported that continuous loading produces significantly less motion than stepwise loading in the cervical spine but no significant differences were observed in this

study (Goertzen et al., 2004). Other systematic differences between the loading apparatuses included unconstrained linear slides versus load-feedback controlled slides, fixed ring versus sliding ring cable-pulley setups, and passive versus powered markers for three dimensional motion measurement.

The method of ROM calculation also differed, with laboratories A and B utilizing the Euler method and laboratories C and D using the tilt/twist technique. The maximum discrepancy between the two techniques was only 0.09° in laboratory A. This indicates that angle calculation method is not a critical factor when comparing planar ROMs among laboratories. The angle calculation method should be more carefully considered when there is substantial coupling (Crawford et al., 1999).

Laboratory A was predetermined to act as a control test site to evaluate the integrity of specimen physiology over time. Although slight ROM increases of $0.2\text{--}0.4^\circ$ were observed after repeat testing, there were no significant differences between the initial and repeat tests in any of the three loading directions. The small ROM increases may be attributed to the multiple test and freeze-thaw cycles, the duration between tests and the repeatability of the load frame and camera system. Despite the minimal ROM increases, the repeat test results demonstrate that specimen condition had not significantly changed during the course of testing/transportation.

Flexibility testing was performed without compressive preload to maintain uniformity among laboratories. Since the introduction of the follower load by Patwardhan et al., numerous studies have investigated its value in the testing of in vitro spinal segments (Goel et al., 2006; Patwardhan et al., 1999; Rohlmann et al., 2001, 2009; Crompton et al., 2000; Wilke et al., 2003; Renner et al., 2007; Dreischarf et al., 2010). Proponents of the follower load argue that the system's application of pure compressive loads to the entire spinal segment more accurately simulates the physiologic effect of in vivo muscle forces (Goel et al., 2006; Patwardhan et al., 1999; Rohlmann et al., 2001). However, recent biomechanical and finite element modeling studies have shown that the specific method of follower load application has a significant impact on biomechanical outcomes (Rohlmann et al., 2009; Crompton et al., 2000; Dreischarf et al., 2010). According to Dreischarf et al. (2010), non-optimized follower load paths and ill-defined starting conditions can make comparisons between laboratories more difficult and lead to questionable conclusions.

This study has several limitations. First, our biomechanical outcome measures were limited to specimen ROM. Additional measures such as neutral zone, axis of rotation, stiffness, and intradiscal pressure were not incorporated as the goal of this study was to provide an initial investigation of inter-laboratory

Table 2

Average (\pm SD) ROM of specimens tested at laboratory A. Laboratory A served as the control testing site and performed both the initial and repeat ROM tests after the specimens had completed testing at the other three laboratories.

Direction	Laboratory A ROM (deg.)	
	Initial test	Repeat test
Flexion-extension	6.9 ± 1.7	7.2 ± 2.1
Lateral bending	7.5 ± 2.7	7.9 ± 2.7
Axial torsion	2.6 ± 0.9	2.8 ± 0.9

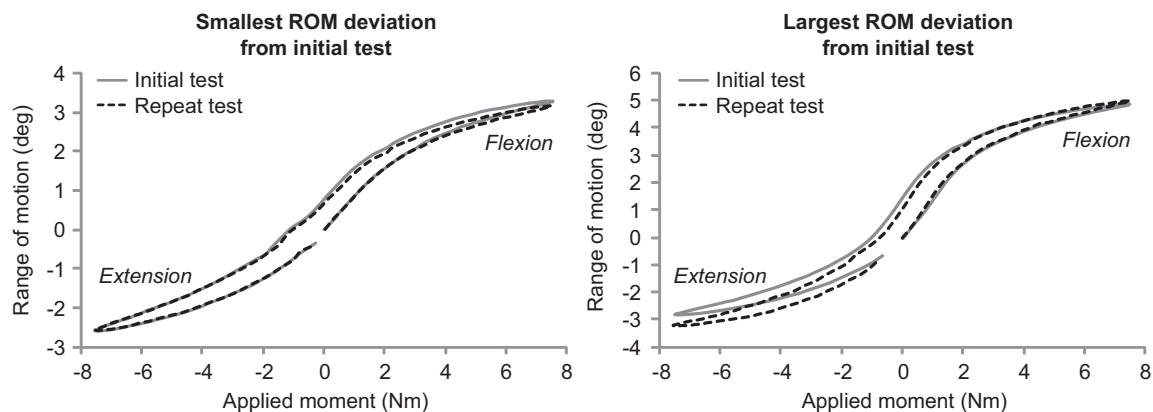


Fig. 4. Flexion-extension hysteresis curves for laboratory A showing the specimens that had the (left) smallest and (right) the largest changes in ROM after repeat testing. The hysteresis curves show very good agreement in overall shape characteristics.

variability based on specimen ROM. Future studies are warranted to examine the reproducibility of additional outcome measures. A second limitation of this study is that it tested single lumbar motion segments, rather than multi-segment specimens. It has recently been suggested that multiple segment units with one free segment on each side of the spinal construct may be preferable (Wilke et al., 1998b; Goel et al., 2006). The repeated measures ANOVA revealed that axial rotation might be a cause for concern in the comparisons between multiple laboratories. The smallest effect size was 0.1172 (Cohen's *d*), which would indicate a need of 27 specimens to identify a statistical difference in axial rotation using a power of 0.80 and multiple comparison correction. A multiple comparison correction was performed (Shaffer's *S1* Bonferroni) and the ROM in each direction did not approach statistical significance between any of the laboratories. This includes the lab A initial and repeat testing included as 'independent' measures to evaluate the repeatability of the results.

5. Conclusions

The results of this study show that spine ROM data generated using a pure moment testing protocol are comparable between biomechanics laboratories using a variety of different loading apparatuses and motion measurement systems.

Conflict of interest statement

The authors have no conflict of interest to report.

Acknowledgments

The authors thank their respective institutions for donating time to complete this study and the Minnesota Medical Foundation for research support.

References

- Bozkus, H., Chamberlain, R.H., Perez Garza, L.E., Crawford, N.R., Dickman, C.A., 2004. Biomechanical comparison of anterolateral plate, lateral plate, and pedicle screws-rods for enhancing anterolateral lumbar interbody cage stabilization. *Spine* 29, 635–641.
- Crawford, N.R., Brantley, A.G., Dickman, C.A., Koeneman, E.J., 1995. An apparatus for applying pure nonconstraining moments to spine segments in vitro. *Spine* 20, 2097–2100.
- Crawford, N.R., Dickman, C.A., 1997. Construction of local vertebral coordinate systems using a digitizing probe. Technical note. *Spine* 22, 559–563.
- Crawford, N.R., Yamaguchi, G.T., Dickman, C.A., 1999. A new technique for determining 3-D joint angles: the tilt/twist method. *Clinical Biomechanics* 14, 153–165.
- Cripton, P.A., Bruehlmann, S.B., Orr, T.E., Oxland, T.R., Nolte, L.P., 2000. In vitro axial preload application during spine flexibility testing: towards reduced apparatus-related artefacts. *Journal of Biomechanics* 33, 1559–1568.
- Dreischarf, M., Zander, T., Bergmann, G., Rohlmann, A., 2010. A non-optimized follower load path may cause considerable intervertebral rotations. *Journal of Biomechanics* 43, 2625–2628.
- Eguizabal, J., Tufaga, M., Scheer, J.K., Ames, C., Lotz, J.C., Buckley, J.M., 2010. Pure moment testing for spinal biomechanics applications: Fixed versus sliding ring cable-driven test designs. *Journal of Biomechanics* 43, 1422–1425.
- Goel, V.K., Panjabi, M.M., Patwardhan, A.G., Dooris, A.P., Serhan, H., 2006. Test protocols for evaluation of spinal implants. *The Journal of Bone and Joint Surgery. American* 88 (2), 103–109.
- Goertzen, D.J., Lane, C., Oxland, T.R., 2004. Neutral zone and range of motion in the spine are greater with stepwise loading than with a continuous loading protocol. An in vitro porcine investigation. *Journal of Biomechanics* 37, 257–261.
- Niosi, C.A., Zhu, Q.A., Wilson, D.C., Keynan, O., Wilson, D.R., Oxland, T.R., 2006. Biomechanical characterization of the three-dimensional kinematic behaviour of the Dynesys dynamic stabilization system: an in vitro study. *European Spine Journal* 15, 913–922.
- Oxland, T.R., Lin, R.M., Panjabi, M.M., 1992. Three-dimensional mechanical properties of the thoracolumbar junction. *Journal of Orthopaedic Research* 10, 573–580.
- Panjabi, M.M., 1988. Biomechanical evaluation of spinal fixation devices: I. A conceptual framework. *Spine* 13, 1129–1134.
- Panjabi, M.M., 2007. Hybrid multidirectional test method to evaluate spinal adjacent-level effects. *Clinical Biomechanics* 22, 257–265.
- Patwardhan, A.G., Havey, R.M., Meade, K.P., Lee, B., Dunlap, B., 1999. A follower load increases the load-carrying capacity of the lumbar spine in compression. *Spine* 24, 1003–1009.
- Renner, S.M., Natarajan, R.N., Patwardhan, A.G., Havey, R.M., Voronov, L.I., Guo, B.Y., Andersson, G.B.J., An, H.S., 2007. Novel model to analyze the effect of a large compressive follower pre-load on range of motions in a lumbar spine. *Journal of Biomechanics* 40, 1326–1332.
- Rohlmann, A., Neller, S., Claes, L., Bergmann, G., Wilke, H.J., 2001. Influence of a follower load on intradiscal pressure and intersegmental rotation of the lumbar spine. *Spine* 26, E557–561.
- Rohlmann, A., Zander, T., Rao, M., Bergmann, G., 2009. Realistic loading conditions for upper body bending. *Journal of Biomechanics* 42, 884–890.
- Wilke, H.J., Wenger, K., Claes, L., 1998b. Testing criteria for spinal implants: recommendations for the standardization of in vitro stability testing of spinal implants. *European Spine Journal* 7, 148–154.
- Wilke, H.-J., Rohlmann, A., Neller, S., Graichen, F., Claes, L., Bergmann, G., 2003. ISSLS prize winner: a novel approach to determine trunk muscle forces during flexion and extension: a comparison of data from an in vitro experiment and in vivo measurements. *Spine* 28, 2585–2593.