Measurement Performance of a Computer Assisted Vertebral Motion Analysis System

Reginald J. Davis, David C. Lee, Chip Wade and Boyle Cheng

This information is current as of May 17, 2025.

Email Alerts	Receive free email-alerts when new articles cite this article. Sign up at: http://ijssurgery.com/alerts

Measurement Performance of a Computer Assisted Vertebral Motion Analysis System

Reginald J. Davis, MD,[1] David C.Lee, MD,[2] Chip Wade, PhD,[3] Boyle Cheng, PhD[4]

[1]Department of Neurosurgery, Greater Baltimore Medical Center, Baltimore, MD, [2]Southern Neurologic and Spinal Institute, Hattiesburg MS, [3]Department of Industrial and Systems Engineering, Auburn University, Auburn, AL, [4]Department of Neurosurgery, Drexel University College of Medicine, Pittsburgh, PA

Abstract

Background

Segmental instability of the lumbar spine is a significant cost within the US health care system; however current thresholds for indication of radiographic instability are not well defined.

Purpose

To determine the performance measurements of sagittal lumbar intervertebral measurements using computer-assisted measurements of the lumbar spine using motion sequences from a video-fluoroscopic technique.

Study design

Sensitivity, specificity, predictive values, prevalence, and test-retest reliability evaluation of digitized manual versus computer-assisted measurements of the lumbar spine.

Patient sample

A total of 2239 intervertebral levels from 509 symptomatic patients, and 287 intervertebral levels from 73 asymptomatic participants were retrospectively evaluated.

Outcome measures

Specificity, sensitivity, negative predictive value (NPV), diagnostic accuracy, and prevalence between the two measurement techniques; Measurements of Coefficient of repeatability (CR), limits of agreement (LOA), intraclass correlation coefficient (ICC; type 3,1), and standard error of measurement for both measurement techniques.

Methods

Asymptomatic individuals and symptomatic patients were all evaluated using both the Vertebral Motion Analysis (VMA) system and fluoroscopic flexion extension static radiographs (FE). The analysis was compared to known thresholds of 15% intervertebral translation (IVT, equivalent to 5.3mm assuming a 35mm vertebral body depth) and 25° intervertebral rotation (IVR).

Results

The VMA measurements demonstrated greater specificity, % change in sensitivity, NPV, prevalence, and reliability compared with FE for radiographic evidence of instability. Specificity was 99.4% and 99.1% in the VMA compared to 98.3% and 98.2% in the FE for IVR and IVT, respectively. Sensitivity in this study was 41.2% and 44.6% greater in the VMA compared to the FE for IVR and IVT, respectively. NPV was 91% and 88% in the VMA compared to 62% and 66% in the FE for IVR and IVT, respectively. Prevalence was 12.3% and 11.9% for the VMA compared to 6.1% and 5.4% for the FE in IVR and IVT, respectively. Intra-observer IVR and IVT had a CR of 2.49 and 2.62, respectively. Inter-observer IVR and IVT had a CR of 1.99 and 2.81, respectively. Intra-subject (test/retest) CR were 2.49 and 3.11 for IVR and IVT, respectively.

Conclusions

The VMA system showed greater measurement performance in the detection of radiographic instability compared with FE radiographs.

Introduction

Segmental instability of the lumbar spine is suggested to be a major cause of low back pain and a significant cost within the US health care system.[1-3] Segmental instability of the spine has been studied *in vivo* with radiographic methods since 1944 yet continues to be controversial in its determination.[4] Over the years several quantitative definitions of lumbar instability based on radiograph-derived measurements have been proposed,[5-8] however current radiographic criteria for instability, and consequential surgical and non-surgical patient selection criteria, are not well defined.

As recent May 2014 the North American Spine Society (NASS) taskforce for developing treatment guidelines for lumbar fusion posited that "there are no accepted radiologic criteria by which the change in alignment on flexion-extension views can be considered instability". Additionally the taskforce suggest that the commonly-used criteria for clinical instability were developed in order to assist physicians in evaluating traumatic instability using static radiographs and were not intended for the determination of clinical instability in the degenerative population.[9]

Common coverage guidelines for the coverage of fusion surgery for the treatment of lumbar intervertebral instability (InterQual, McKesson, San Francisco, CA, USA) advocate intervertebral rotation (IVR) larger than 22° and intervertebral translation (IVT) greater than 3 mm as criteria for fusion. The American Medical Association (AMA) proposes a range of 4.5–5 mm or 10–15% of the vertebra body width in IVT, or IVR greater than 15° at L1-2, L2-3 or L3-4, 20° at L4-5 or 25° at L5-S1. As it relates to postoperative pseudarthrosis, the US Food and Drug Administration (FDA) indicates IVR of less than 5° and IVT less than 3 mm as a successful fusion.[10] Given the wide range of proposed quantified definitions of radiographic instability, a better understanding of spinal motion characteristics through standardized, accurate, and reliable functional analysis is clearly necessary.

While there is not a consensus on the quantification of radiographic instability, the diagnostic imaging modalities of standard roentgenograms (X-rays), computed tomography (CT), and magnetic resonance imaging (MRI) have widely been accepted in the evaluation of spinal pathologies. In evaluating spinal instability, dynamic end-range flexion/extension radiographs are employed to measure intervertebral range of motion (ROM), comprised of intervertebral rotation and translation (IVR and IVT), through manual identification of vertebral segments across multiple radiographic images using a ruler and protractor or computer-assisted imaging software. In recent years, there have been marked improvements to imaging modalities such as MRI and CT in capturing diagnostically-useful static images of the spine, yet these lack the capability of providing detailed information about spinal motion.

Recently, there have been improvements to the measurement evaluation techniques used for interpreting the flexion/extension radiographs (FE). Traditional manual measurements are performed by manually drawing lines or overlaying protractors on film radiographs.[11,12] Computer- assisted manual measurements are performed manually by the user by drawing lines or identifying landmarks in digital radiographs using computer software which utilizes this user input to derive measurements of ROM.[13-15] In general, the precision and reliability of measurements are improved when moving towards computerized measurements, as they standardize and minimize user interaction and therefore result in less measurement variability. [16-19]

The variability associated with the process for deriving measurements of ROM from radiographs is therefore well understood, and at least partially-addressable through computerized measurement techniques. In contrast, the variability associated with patient specific bending in FE radiographs is less detailed in the literature. FE methods can be characterized as uncontrolled, non-standardized across patients, and involving bending at a patient-selected rate and range.[20] There is an inherent level of variability involved when patients are uncon-

strained in how they bend, both across patients and within a patient over time. In contrast, it has been proposed that FE methods that involve controlled patient bending via the use of a handling device that assists patients through a standard arc of lumbar bending have the potential to reduce variability in patient bending during FE.[21-24] Concerns about this high level of variability have driven the various approaches intended to control bending and/or ensure consistent patient effort. Esola et al.[22] showed that there are measureable differences in pelvic rotation during flexion in healthy subjects compared with patients with back pain. The findings of Esola et al. (1996) support the use of a handling device for pelvic fixation, such as that used by Takayanagi et al. (2001), to avoid unintended confounding of measurements from varying pelvic motion.

Since the 1940's, bending motions for radiographic imaging has been categorized based on the bending posture (i.e. standing, seated, etc.) as well as the bending method (i.e. maximum voluntary bending, assisted bending, etc.) (Table 1). Researchers and clinicians have utilized a number of postures and methods to elicit segmental motion during imaging with varying results. These postures include a standing uncontrolled posture, a posture with hips flexed or seated, standing while the pelvis is fixed, lying on one's side, and traction and compression; whereas the methods of the bending include maximum voluntary flexion/extension, the patient pulling oneself through a ROM, a clinician pulling the patient through a ROM or a given standard bending range. Wood et al. (1994) acknowledged that studies may exaggerate stress on the lumbar spine beyond that expected under physiologic conditions in an attempt to maximize the chances of detecting potentially abnormal translational movement.

There is a breadth of research that describes standing uncontrolled bending protocols that have the patient standing upright and unrestrained, and attempting to bend only at the lumbar spine while keeping the hips from flexing.[10,26-30] In standing upright, the knees are kept in a fully extended position. It is possible for hip flexion to occur, but through only patient-voluntary efforts this can be minimized. Standing uncontrolled bending may be further classified as involving either:

(I) maximum voluntary end range angles, or (IV) standardized end range angles.

Hips flexed bending or seated protocols accomplish very large ranges of bending through combined hip and lumbar flexion, which may additionally involve flexion of the knees. They can be classified as: (I) having no pulling forces, (II) having pulling forces generated by the patient, or (III) having pulling forces generated by an imaging technician. Research has described these postures as being performed while sitting off the end of a tables edge,[22,31] while sitting in a chair and pulling oneself to a maximum range,[32] and while standing upright and being pulled or pushed in extension or flexion, respectively by a radiology technician.[24]

Bolstering the pelvis while the patient stands upright keeps the hips in full extension and the pelvis upright. This type of bending is utilized to elicit a lumbar motion while reducing pelvic girdle and hip motion during bending.[22] For the hips to flex, the pelvis has to rotate forward during bending. Using a bolster minimizes pelvic rotation (and therefore hip flexion), which maximizes the amount of trunk bending that

Table 1. Bending posture and bending method classifications for classifying studies reviewed as part of the review of medical literature.

Bending Posture		Bending Method	
A. Standing uncontrolled	Patient stands uprights and bends on their own without grabbing anything or being bolstered to anything.	I. Max Voluntary	Patient bends as far as they can move themselves without using their hands to pull into flexion.
B. Hips flexed	In flexion, bending can occur in the hips and lumbar spine. Knees may or may not be flexed.	II. Patient pulling	Patient pulls on a fixed object to go further into flexion than the patient could otherwise go.
C. Standing pelvis bolster	Upright standing with a bolster to minimize pelvis motion. Knees kept in full extension.	III. Rad Tech Pulling	A technician pushes and/or pulls the patient into a maximal spine bend.
D. Side lying	Lying in lateral decubitus position and bending forward and backward.	IV. Standard End Range	IVT and IVR are measured at a standardized overall trunk bending angle, short of the maximal end range position.
E. Traction & compression	Using external forces to add load (compression) or to distract (traction) the spine while in an upright standing posture		

derives from the lumbar joints (as opposed to the hips). Bolsters may involve padded devices or harnesses. Bronfort (1994) and Pearcy (1985) had subjects stand in an apparatus which limited the movement of the pelvis and hip. Okawa (1988) supported the patient based on an apparatus at the pubis and sacrum while bending; whereas Teyhen (2006) minimized pelvic and hip motion by restricting the subject in a "mountain climbing" type harness to restrict motion.

Side-lying flexion / extension has been reported in several studies as well.[28,30,36] Each of these studies examined flexion and extension in the recumbent position at maximum flexion extension points. Lying lumbar imaging, in general, is commonly used in current US medical practice (not necessarily with bending views) to determine the extent of spondylolisthesis reduction occurring between standing and lying lateral views.

Traction / compression was studied and reported in the 1980s and 1990s.[37-39] In the tension method patients hung from their hands from a bar perpendicular to the film plane, while compression was elicited by an upright posture while wearing a backpack full of 20kg of sand. This method found in increase in the rate of radiographic instability of 4.8%, however it never became widely adopted and there were no references to any studies involving traction/compression after the Pitkanen (1997) study.

Given the various methodologies of bending and associated postures during lumbar imaging, the evidence suggests a challenge in determining applicable methodologies for accurate lumbar imaging. Anecdotal evidence suggest many of the aforementioned methodologies do not fall within the standard clinical workflow or accepted modalities which are widely used today. More importantly these modalities may present a challenge to the clinician and patient to perform based on either logistical or safety concerns from the authors experience. To overcome the high measurement variability experienced in the FE imaging modality, other imaging methodologies have been proposed.

Non-quantitative dynamic fluoroscopy methods have been proposed for the dynamic visualization of segmental motion; however to the author's knowledge no studies examining accuracy and repeatability of these methods have been published. With respect to quantified approaches for assessing ROM, research has examined accuracy and reliability of quantitative fluoroscopy (QF) and digital video-fluoroscopy (DVF) and has demonstrated repeatability comparable to that of standard computer-assisted X-ray techniques.[24,40,41] Several additional studies have evaluated intra- and inter-observer reliability and find that computer-assisted processing methods significantly improve intervertebral motion measurements.[10,42] However there is a dearth of literature examining key aspects of the diagnostic efficacy of the clinical application of quantitative fluoroscopy (QF), including its diagnostic sensitivity and specificity in assessing radiographic instability as well as the test/retest (intra-subject) repeatability of ROM measurements produced via QF.

When assessing clinical efficacy, diagnostic tests should be reliable, accurate, and beneficial.[43] By definition a test is considered reliable if the identical test conducted on the same patient continuously derives then same outcome.[44] A test is accurate if it results in the correct outcome in identifying pathology when pathology is present. Sensitivity and specificity are the fundamental measures of accuracy. At present, lumbar spinal instability is diagnosed on the basis of findings on physical exam plus FE radiographs, but there is no consensus regarding the radiographic criteria for instability, and the clinical definition of lumbar spinal instability is therefore ambiguous.[8,9] Although many studies have described clinical examination measures for the diagnosis of lumbar spinal instability, few of them have investigated the sensitivity and specificity of using FE to detect radiographic instability. Fritz evaluated patients with low-back pain referred for FE radiographs due to suspicion of lumbar instability and reported lumbar flexion ROM measures were a significant predictor of instability and that lumbar intervertebral motion testing was beneficial for predicting instability.[45]

Despite the common acceptance of FE radiographs as a key clinical indicator of radiographic instability, the role of FE radiographs remains debated due to

several important limitations including poor diagnostic accuracy and reliability.[46] In a recent study, Yeagar et al.[10] demonstrated that reliability of intervertebral measurements by the Vertebral Motion Analysis (VMA) (Ortho Kinematics Austin, TX, USA) was significantly greater when compared with a digitized manual technique. This study assesses the reliability of the computer-assisted vertebral motion analysis system in comparison to FE through intra- and inter-observer reliability, and intra-subject reliability. This study further assesses the diagnostic sensitivity, specificity, prevalence, and NPV of the computer-assisted vertebral motion analysis system in comparison to FE, in detecting radiographic instability. All assessments in this study were produced using a prospective analysis of collected image data.

Materials and Methods

Image acquisition

The VMA system assesses intervertebral motion through the use of: (1) a patient handling device that assists patients through a prescribe-able arc of lumbar bending in both an upright (Figure 1) and recumbent (Figure 2) posture (flexion/extension arcs were 70 and 60 degrees as measured by the rotation of the VMA for upright and recumbent, respectively); (2) video fluoroscopy imaging of the lumbar spine during bending (capturing images at 8 frames per second) utilizing a standard 12-inch surgical C-Arm (OEC 9800 Radiographs, General Electric, Fairfield, CT, USA; 2006 Phillips BV Pulsera, Andover, MA, USA); (3) console mounted computer equipped with data acquisition hardware (Accustream Express As205A, Foresight Imaging, Chelmsford, MA, USA) to digitize fluoroscopic video signals and (4) proprietary image processing software capable of assisting the user in the semi-automatic frame-to-frame registration and tracking of vertebral bodies across the sequence of video-fluoroscopic images to derive measurements of intervertebral rotation and translation (Ortho Kinematics, Austin, TX, USA). Previous research has reported the accuracy and repeatability of the VMA system.[10,24]

The standardized imaging protocol consist of a standard FE radiographs of uncontrolled bending under fluoroscopy (Figure 3) and VMA radiographs captured during controlled bending with the patient positioned in the patient handling device in both the upright and recumbent configuration (Figure 1, Figure 2). The upright motion platform "guided" active lumbar bending, under normal physiologic weight-bearing conditions, while constricting flexion extension to the sagittal plane and bolstering the pelvis to minimize pelvic motion and maximize lumbar bending. The individual follows the VMA rotation on their own using the VMA as a proprioceptive guide. The recumbent platform "controlled" passive lumbar bending, which minimized the gravitational and muscular forces that present during standing radiographs. The individual lies on the VMA table and the table, through a servo motor driven actuator, actively moves the individual through the flexion extension motion. Flexion/extension angles spanned a range of 70º arc for standing and recumbent motion. The flexion/extension ranging was preset on the VMA hardware to allow for 20° of extension and 50° flexion for upright and 35° flexion and 25° of extension in recumbent of VMA motion. This compensated for the reduced capability of lumbar extension resulting from extended hips in an erect posture.

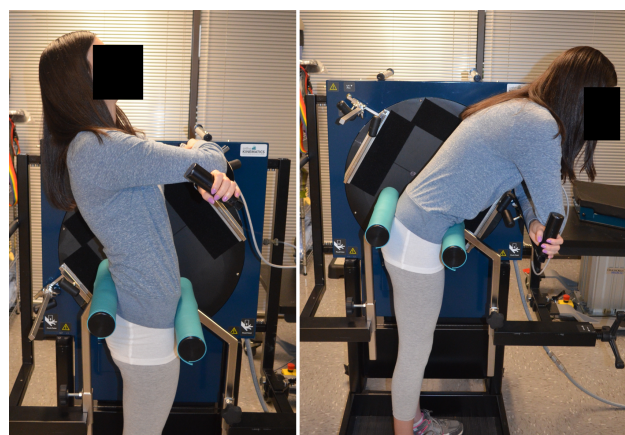Upon initiation of the test movement, a fluoroscopic

Fig. 1. VMA standing flexion/extension position.

Fig. 2. VMA recumbent flexion/extension position.

sequence of lumbar motion was captured at 8 pulses per second on high level flouro (i.e boost) and 4 frames of data per second were analyzed. Flexion and extension were captured as separate sequences, which began in a neutral position, progressed to the predefined maximum angle, and returned to neutral. In the upright position, a total of 28 seconds of data were collected (neutral to flexion-10 seconds, flexion to neutral-10 seconds, neutral to extension-4, extension to neutral-4) for a total of 112 frames of data. In the recumbent position, a total of 24 seconds of data were collected (neutral to flexion-7 seconds, flexion to neutral-7 seconds, neutral to extension-5, extension to neutral-5) for a total of 96 frames of data.

Experimental Design

This study combines image data from 4 separate neurosurgery facilities that utilized the VMA system. Five hundred and nine (509) symptomatic patients (44% female, 56% male; mean age, 53.2 ± 17.8 years) contributed to 2,239 intervertebral levels. One hundred and sixty six (166) asymptomatic individuals (43% female, 57% male; mean age, 48.1 ± 15.4 years) contributed to 705 intervertebral levels. All individuals completed all imaging sequences in a single visit. Image processing software generated IVR and IVT measurements. In the case of FE measurements, this was done via the manual land-marking of an image pair and the subsequent calculation of IVR and IVT by the software. In the case of VMA measurements, this was done via an automated tracking process synchronized to the fluoroscopic sequences.[10,24] The IVR were defined as the change in angle between the inferior endplate of the superior vertebra and the superior endplate of the inferior vertebra in the motion segment, while the IVT were expressed as a percentage of inferior vertebral body depth.[10]

The symptomatic patient set was collected via a consecutive case enrollment design, wherein all patients

Fig. 3. Flexion extension radiographs (FE) at a self-selected pace and position.

seen in the surgeon's clinic and prescribed VMA testing over a specific period of time were included. The exact timeframes varied by clinic, but overall the consecutive case series data were collected between January, 2011 and February, 2014. It is important to delineate the VMA as the "index test" and the FE radiographs as a commonly accepted "reference test". Problems with the reference standard based on a "reference test " are evident in diagnostic accuracy studies. The outcome of the reference standard may not be available in all patients, it may be unreliable, it may be inaccurate or there could be no acceptable reference standard at all.[47] Lord et al. (2006) reported that measurements of specificity of a new diagnostic test through a cross-sectional descriptive study design are satisfactory evidence to recommend substitution of the reference test only when the new test has a sensitivity similar to that of the reference test and when the new test is safer or is more specific than the reference test.[48]

Multiple data sets were constructed to assess reliability and accuracy of the VMA system. One distinct data set (Table 2) was constructed based on the complete sample of asymptomatic individuals and symptomatic patients to evaluate the system specificity, % change in sensitivity, negative predictive value (NPV), and the prevalence of radiographic instability while a second data set (Table 3) on only asymptomatic individuals was constructed to evaluate inter and intra observer agreement and intra subject (test/retest) agreement.

Statistical Methods

Specificity, Prevalence, and NPV

A standard methodology was employed to calculate specificity, NPV, and the prevalence of radiographic instability. In evaluating the diagnostic test performance every individual being tested either has or does not have radiographic instability. Overall the diagnostic test may or may not match the individual's actual status resulting in a true positive (TP), false positive (FP), true negative (TN), or false negative (FN); the current analysis focuses on TP, FP, and TN results. Individuals in the current analysis were either symptomatic or asymptomatic. Individuals in the current analysis were classified by both the VMA

and FE tests for radiographic instability based on measurements of IVR and IVT. For IVR, the thresholds used to define radiographic instability (excessive

Table 2. Average angles and translations measured for both VMA and FE imaging radiographs (N=2239 intervertebral levels).

Panel A. IVR measured in degrees (°)

Level	FE		VMA (Standing)		VMA (Lying)		Δ VMA (Standing)-FE		Δ VMA (Lying)-FE	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
L1-L2	5.9	2.3	7.6	4.3	8.6	3.6	1.7**	2.0	2.7**	1.3
L2-L3	6.8	2.4	11.2	4.8	10.1	3.7	4.4***	2.4	3.3***	1.3
L3-L4	7.7	2.6	10.1	5.1	9.2	4.1	2.4**	2.5	1.5*	1.5
L4-L5	9.1	3.3	11.9	7.1	8.3	4.4	2.8**	3.8	0.8	1.1
L5-S1	8.5	4.2	8.1	6.2	10.3	4.6	0.4	2.0	1.8*	0.4
Average	*7.6*	*3.0*	*9.8*	*5.5*	*9.3*	*4.1*	*2.3**	*2.5*	*2.0**	*1.1*

Panel B. IVT in percentage vertebral body depth (%)

Level	FE		VMA (Standing)		VMA (Lying)		Δ VMA (Standing)-FE		Δ VMA (Lying)-FE	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
L1-L2	4.4	1.4	5.7	3.8	5.4	2.3	1.3*	2.4	1.0	0.9
L2-L3	5.2	1.3	6.1	4.0	7.9	2.1	0.9	2.7	2.7**	0.8
L3-L4	5.9	1.5	6.6	4.5	9.0	2.7	0.7	3.0	3.1**	1.2
L4-L5	7.7	2.1	5.8	5.2	8.3	2.0	1.9*	3.1	0.6	0.1
L5-S1	6.9	3.3	6.4	3.7	2.8	1.1	0.5	0.5	4.1***	2.2
Average	*6.0*	*1.9*	*6.1*	*4.2*	*6.7*	*2.1*	*1.04**	*2.3*	*2.3**	*1.1*

Panel A reports IVR measured in degrees (o) and Panel B reports IVT in percentage vertebral body depth (%). Student t-test where ***p<0.01, **p<0.05, *p<0.1.

Table 3. Contributions of intervertebral levels comprising the dataset.

Level	n	Contribution
L1-L2	446	11
L2-L3	438	16
L3-L4	471	27
L4-L5	452	23
L5-S1	432	23
Total	2239	

rotation) were 22°. For IVT, the thresholds used to define radiographic instability (excessive translation) were 15% of the vertebra body width, which corresponds to a threshold of roughly 5.3 mm of IVT assuming a standard vertebral body depth of 35 mm. For all analyses presented herein, individuals within the symptomatic patient population were categorized as TP whenever there was evidence of radiographic instability according to the thresholds listed above. Similarly, subjects from within the asymptomatic patient population were categorized as FP when IVR or IVT was observed to be above these thresholds.

Inter-observer, Intra-observer, Intra-subject (test/retest) Reliability

Inter- and intra -observer agreement and intra-subject agreement was evaluated using the Bland-Altman method. The differences between each observer's first and second round measurements (A1–A2, B1–B2) for each intervertebral level were used to establish the intra-observer agreement. The differences between each pair of observer's first round measurements (A1–B1, A2–B2,) were used to establish the inter-observer agreement. Measurements were conducted with 1 week between measurements. Intra-subject (test/retest) agreement was measured by a single observer evaluating measurements (A_{t1}-A_{t2}) taken from 2 separate imaging sessions performed on the same subject 6 weeks apart. The mean (d) and standard deviation (SD) of each set of differences were calculated. The coefficient of repeatability (CR), upper and lower limits of agreement (LOA), and corresponding standard error of measurement (SEM) were defined as follows: Calculations were performed using SAS 9.3 (version 19; SAS Institutes, Cary, NC, USA).

Results

Reporting on the ROM values from the VMA compared to the FE radiographs, Table 2 and Table 3 present summary data and performance data respectively. Table 2 presents the average IVR and IVT by level for FE, standing VMA and lying VMA for 2,239 intervertebral levels separately between the VMA and the FE radiographs. A student t test univariate analysis was conducted to show statistical differences across modes and level. Table 3 provides a

level by level comparison which constituted the complete data set, additionally the contribution of each level to the overall motion of the lumbar body.

Table 4 details the accuracy measurements of specificity and NPV between the VMA and the FE radiographs for IVR and IVT measurements in detecting radiographic instability, and also provides the prevalence of radiographic instability within the symptomatic population as detected by both VMA and FE radiographs. Specificity was 0.3% (VMA 99.4%, FE 98.3%) and 0.9% (VMA 99.1%, FE 98.2%) greater in the VMA compared to the FE in intervertebral rotation and translation, respectively. Further, data indicates a 25% and 26% greater NPV in the VMA compared to the FE radiographs in IVR (VMA 91%, FE 66%) and IVT (VMA 88%, FE 62%) respectively.

Percent increases or decreases to the prevalence of radiographic instability as defined above of VMA relative to FE can be shown to be exactly equivalent to the magnitude of percent increases or decreases to the sensitivity of VMA relative to FE. This can be demonstrated algebraically. Sensitivity is defined as $TP / (TP + FN)$, and prevalence of radiographic instability is defined as $TP / (TP + FN + CN_{symptomatic})$, where $CN_{symptomatic}$ is defined as the number of "Condition Negative" patients within the symptomatic patient population (i.e. those who have pain but do not have radiographic instability). The sum $(TP + FN)$ is equivalent to the total number of "condition positive" patients within the symptomatic patient population, which along with $CN_{symptomatic}$, were constant in the patient population and therefore do not vary between VMA and FE, even though their sizes were not directly measurable in this study. Any percent increase or decrease in the prevalence of radiographic instability for VMA relative to FE is shown to be exactly equivalent to the percent increase or decrease in sensitivity for VMA relative to FE, as both are defined by the same variable quantity (TP) divided by a constant. As such, sensitivity in this study was 41.2% and 44.6% greater in the VMA compared to the FE in intervertebral rotation and translation, respectively. Although many studies have described clinical examination measures for the diagnosis of lumbar spinal instability, few have investigated the sensitivity and specificity of these examinations in detecting radiographic instability.

The reliability measurement comparison of the VMA to the FE radiographs is presented in Table 5 and Table 6. Table 5 presents the average IVR and IVT by level for FE, standing VMA and lying VMA for 705 (inter- and intra-observer agreement) intervertebral levels and the 287 (intra-subject agreement) intervertebral levels, respectively, separately between the VMA and the FE radiographs. A student t test univariate analysis was conducted to show statistical differences across modes and level. Table 6 provides a level by level comparison of the inter- and intra-observer dataset (n=705 levels) as well as the intra-subject dataset (n=287 levels).

Table 7 presents the summary statistics for inter- and intra-observer agreement and intra-subject agreement VMA measurement. Inter- and intra-observer variability of the VMA is comparable to that present by Yeagar et al.[10]. The inter observer CR of IVR and IVT is reports as 2.49 and 2.62 respectively, while the ICC was 0.974 and 0.926 respectively. Intra observer CR of IVR and IVT is reports as 1.99 and 2.81 respectively, while the ICC was 0.939 and 0.946 respectively. To our knowledge this study is the first to report intra-subject variability on a computer assisted vertebral motion analysis system. The intra-subject CR of IVR and IVT are reported as 2.49 and 2.57 respectively, while the ICC was 0.919 and 0.923 respectively.

Discussion

The current study demonstrates improved performance of the VMA radiographs compared to the standard FE radiographs in measurement performance. While there have been significant improve-

Table 4. Specificity and NPV of the VMA and FE lumbar imaging radiographs in detecting radiographic instability across levels, and the prevalence of VMA- and FE-detectable radiographic instability.

Method	IVR			IVT		
	Specificity	NPV	Prevalence	Specificity	NPV	Prevalence
VMA	99.4%	90.7%	12.3%	99.1%	87.4%	11.9%
FE	98.3%	72.4%	6.1%	98.2%	71.2%	5.4%
Difference	1.1%	18.3%	6.2%	0.9%	16.2%	6.5%

ments to imaging modalities and imaging analysis a lack of consensus around segmental and clinical in-

Panel A. IVR measured in degrees (°)

Level	FE Mean	FE Std Dev	VMA (Standing) Mean	VMA (Standing) Std Dev	VMA (Lying) Mean	VMA (Lying) Std Dev	Δ VMA (Standing)-FE Mean	Δ VMA (Standing)-FE Std Dev	Δ VMA (Lying)-FE Mean	Δ VMA (Lying)-FE Std Dev
L1-L2	6.3	1.8	9.1	4.2	10.3	3.3	2.8**	2.4	4.0***	1.5
L2-L3	7.1	2.2	12.2	5.1	9.8	3.5	5.1***	2.9	2.7**	1.3
L3-L4	7.9	2.1	11.6	3.8	10.2	3.6	3.7**	1.7	2.3**	1.5
L4-L5	10.2	2.7	12.3	4.6	9.7	3.8	2.1**	1.9	0.5	1.1
L5-S1	9.4	3.8	10.2	5.1	11.1	3.9	0.8	1.3	1.7*	0.1
Average	*8.2*	*2.5*	*11.1*	*4.6*	*10.2*	*3.6*	*2.9***	*2.0*	*2.2***	*1.1*

Panel B. IVT in percentage vertebral body depth (%)

Level	FE Mean	FE Std Dev	VMA (Standing) Mean	VMA (Standing) Std Dev	VMA (Lying) Mean	VMA (Lying) Std Dev	Δ VMA (Standing)-FE Mean	Δ VMA (Standing)-FE Std Dev	Δ VMA (Lying)-FE Mean	Δ VMA (Lying)-FE Std Dev
L1-L2	5.1	1.5	6.2	2.9	5.9	1.9	1.1*	1.4	0.8	0.4
L2-L3	5.8	1.1	6.4	3.7	7.4	2.2	0.6	2.6	1.6*	1.1
L3-L4	6.2	1.4	7.5	4.1	8.7	2.8	1.3*	2.7	2.5**	1.4
L4-L5	7.1	1.8	6.4	5.3	8.9	3.1	0.7	3.5	1.8*	1.3
L5-S1	7.8	2.8	6.9	3.2	3.2	2.2	0.9*	0.4	4.60***	0.6
Average	*6.4*	*1.7*	*6.7*	*3.8*	*6.8*	*2.4*	*0.9**	*2.1*	*2.3**	*1.0*

Panel A reports IVR measured in degrees (o) and Panel B reports IVT in percentage vertebral body depth (%).Student t-test where ***p<0.01, **p<0.05, *p<0.1.

Level	Intra/Inter Observer n	Intra/Inter Observer Contribution	Intra Subject n	Intra Subject Contribution
L1-2	446	11	-	-
L2-3	438	16	74	14
L3-4	471	27	71	23
L4-5	452	23	72	27
L5-S1	432	23	70	28
Total	705		287	

stability still exist. Panjabi (2003) suggested that the ambiguity in consistent definition of clinical instability partially a result of increased variability in the patient bending and the limited accuracy of *in vivo* methods for measuring motion. In this study, the VMA system addresses the first two primary concerns by providing a controlled patient bending ROM and accuracy inherent to the automatic frame-to-frame registration and tracking algorithm. The current study provides evidence that there is a marked improvement in measurement performance of the VMA compared to the FE radiographs while also exhibiting statistically greater reliability. This is especially notable because both modalities (VMA and FE) had relatively high specificity, and the specificity for VMA was if anything slightly higher than FE. The fact that there were increase in sensitivity with no corresponding decrease in specificity demonstrates that VMA testing detects radiographic instability to the same extent as that detected with FE.

Diagnostic misclassification errors have received in-

Measurement	Mean in Differences	CI	Upper LOA	CI	Lower LOA	CI	CR	ICC (3,1)	CI	SEM
Observer Variability Rotation	0.02	0.03 to 0.09	2.52	2.68 to 2.36	-2.49	-2.25 to -2.73	2.49	0.974	0.978 to 0.97	0.13
Intra-Observer Variability Translation (%)	0.06	0.05 to 0.13	2.69	2.91 to 2.47	-2.62	2.44 to -2.81	2.62	0.926	0.936 to 0.916	0.18
Inter-Observer Variability Rotation	0.05	0.06 to 0.19	1.97	2.16 to 1.78	-2.03	-1.81 to -2.25	1.99	0.939	0.946 to 0.932	0.23
Inter-Observer Variability Translation (%)	0.03	0.04 to 0.21	2.84	3.07 to 2.61	-2.91	-2.60 to -3.22	2.81	0.946	0.957 to 0.935	0.24
Intra-Subject Variability Rotation	-0.14	-0.15 to -0.26	2.46	2.71 to 2.22	-2.52	2.25 to -2.79	2.49	0.919	0.928 to 0.911	0.27
Intra-Subject Variability Translation (%)	-0.11	-0.09 to -0.21	2.67	2.87 to 2.34	-2.69	-2.34 to -2.96	2.57	0.923	09.38 to 0.913	0.25

This table reports the confidence intervals (CI), mean absolute error, upper and lower limits of agreement (LOA), interclass correlation (ICC), and standard error of the mean (SEM).

creasing attention as these errors constitute significant sources of liability claims as well as increased risk of morbidity and mortality.[29,48] Misclassification errors resulting from the methodology and reporting of radiological examinations are common place across modalities.[49] These errors can originate in multiple ways from testing to reporting. Misclassification errors can result from incomplete or inappropriate evaluation of medical history, physical examinations, or diagnostic testing; interpretation of diagnostic test results; or therapeutic or surgical decisions based on diagnostic test results.[49-50] The accuracy and test/retest reliability of diagnostic test can be a principal confounding factor is diagnostic error. The current study shows that compared to the FE radiographs the VMA is 31% to 35% more efficient in identifying radiologic instability, based on the current thresholds.

The reliability of the VMA system has been reported by Yeagar et al.[10] who evaluated inter- and intra-observer reliability between the VMA and digital manual measurements. Their study demonstrates that marked improvements in intra- and inter-observer precision are achieved by VMA analysis compared with a digitized manual technique. Further, Yeagar et al.[10] suggests that a physician making repeated measurements on the same image (intra observer) can be 95% confident they will fall within a range of ±1.53° and ±2.20% of the actual value in IVR and IVT respectively. Further they found that different physicians making a repeated measurement on the same image (inter-observer), the CR 2.15° and 3.90%, respectively. The current study extended the Yeagar et al.[10] analysis be volume of vertebral levels analyzed as well as by evaluating intra-subject reliability.

In comparing the current study's CR to that of Yeagar et al.[10] the results were quantitatively similar. The current findings indicate inter-observer reliability (95% confident) falls within a range of ±2.49° and ±2.62% of the actual value in IVR and IVT respectively. Further the current study reports a quantitatively similar inter-observer reliability CR of 1.99° and 2.81% in IVR and IVT, respectively. To our knowledge the current study is the first to evaluate intra-subject reliability in a computer assisted verte-

bral motional analysis system. The current findings extend those of Yeagar et al.[10] and provide evidence of the test/retest reliability of the VMA system.

Conclusion

Many clinical tests are used to confirm or refute the presence of a disease or further the diagnostic process. A primary clinical test for detecting spinal instability is the dynamic radiograph imaging modality and have been so for nearly a decade. However, there still seems to be a lack of consistency in the utilization of dynamic radiographs specifically in determining an applicable standard threshold for rotational and translational radiographic instability in the lumbar spine. The current technique (i.e. VMA) has the potential to provide greater measurement performance by utilizing the combination of controlled motion and a semi-automated measurement technique to enhance the clinician's ability to identify radiographic instability of the lumbar spine.

Ideally diagnostic tests correctly identify all patients with the disease, and similarly correctly identify all patients who are disease free. Ideally, a perfect test would never be positive in a patient who is disease free and would never be negative in a patient who is in fact diseased. Even though it remains controversial, FE radiographs are a principal diagnostic modality for identifying radiographic instability of the lumbar spine. More importantly the quantitative ROM thresholds for detecting radiographic instability have been debated for half a century with little consensus. Unfortunately there have been few advances in assessing for radiographic instability of the lumbar spine in recent years. Through computer assisted vertebral motion analysis as well as the use of patient bending control devices and fluoroscopy, the ability improve upon current methods for detecting radiographic instability is clinically relevant. The current study demonstrates the improved measurement performance of the VMA system as a diagnostic modality in assessing for and detecting radiographic instability of the spine both in rotation and translation.

Acknowledgements

tion of the following physicians as contributing investigators in collecting asymptomatic individual data: Morgan Lorio, Neurospine Solutions, Bristol, TN; Hyun Bae, The Spine Institute Comprehensive Spine Center, Santa Monica, CA; Antonio Castellvi, Florida Orthopedics Institute, Tampa, FL; and David Harris, Center for Healing and Regenerative Medicine, Austin, TX. Additionally, we acknowledge Will Malpica and Kristin Madl of Ortho Kinematics in assistance with data acquisition.

References

1.   Dagenais S, Galloway EK, Roffey DM. A systematic review of diagnostic imaging use for low back pain in the United States. The Spine Journal 2014; 14: 1036-1048.

2.   Haldeman S, Dagenais S. A supermarket approach to the evidence- informed management of chronic low back pain. Spine J 2008; 8: 1–7

3.   Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. Spine 2006;31: 2724–7.

4.   Panjabi MM. Clinical spinal instability and low back pain. Journal of Electromyography and Kinesiology 2003; 13: 371–379.

5.   Dupuis PR, Yong-Hing K, Cassidy JD, Kirkaldy-Willis WH. Radiologic diagnosis of degenerative lumbar spinal instability. Spine 1985; 10: 262–76.

6.   Farfan HF, Gracovetsky S. The nature of instability. Spine 1984; 9(7): 714–719.

7.   Pope MH, Frymoyer JW, Krag MH. Diagnosing instability. ClinOrthopRelat Res 1992; 279: 60–67.

8.   American Medical Association. Guides to the evaluation of permanent impairment. 5th ed. Chicago, IL: American Medical Association, 2000; 379.

9.   North American Spine Society (NASS). Lumbar fusion: defining appropriate coverage positions. 2014; 1-18.

10.   Yeager MS, Cook DJ, Boyle C. Cheng BC. Reliability of computer-assisted lumbar intervertebral measurements using a novel vertebral motion analysis system. The Spine Journal 2014; 14: 274–281.

11.   Diab K, Sevastik J, Hedlund R, et al. Accuracy and applicability of measurement of the scoliotic angle at the frontal plane by Cobb's method, by Ferguson's method and by a new method. Eur Spine J 1995; 4:291-295.

12.   De Carvalho A, Vialle R, Thomsen L, et al. Reliability analysis for manual measurement of coronal plane deformity in adolescent scoliosis. Are 30 x 90 cm plain films better than digitized small films? Eur Spine J 2007; 16: 1615-1620.

13.   Tanure M, Pinheiro A, Oliveira A. Reliability assessment of Cobb angle measurements using manual and digital methods. Spine J 2010; 10: 769-774.

14.   Cheung J, Wever D, Veldhuizen A, et al. The reliability of quantitative analysis on digital images of the scoliotic spine.Eur Spine J 2002; 11: 535-542.

15.   Stokes IAF, Aronsson DD. Computer-assisted algorithms improve reliability of King Classification and Cobb angle measurement of scoliosis. Spine 2006; 31:665-670.

16.   Gstoettner M, Sekyra K, Walochnik N, et al. Inter- and intraobserver reliability assessment of the Cobb angle: manual versus digital measurement tools. Eur Spine J 2007; 16:1587-1592.

17.   Shea KG, Stevens PM, Nelson M, et al. A comparison of manual versus computer-assisted radiographic measurement: intraobserver measurement variability for Cobb angles. Spine 1998; 23:551-555.

18.   Carman DL, Browne RH, Birch JG. Measurement of scoliosis and kyphosis radiographs: intraobserver and interobserver variation. J Bone Jt Surg Am 1990; 72:328-333.

19.   Mok JM, Berven SH, Diab M, et al. Comparison of observer variation in conventional and three digital radiographic methods used in the evaluation of patients with adolescent idiopathic scoliosis. Spine 2008; 33:681-686.

20.   Breen A, Muggleton J, Mellor F. An objective spinal motion imaging assessment (OSMIA): Reliability, accuracy and exposure data. BMC Musculoskeletal Disorders 2006; 7: 1-10.

21.   Knutsson F. The instability associated with disc degeneration in the lumbar spine. Acta Radiology 1944; 25: 593-608.

22.   Esola MA, McClure PW, Fitzgerald GK, Siegler S. Analysis of lumbar spine and hip motion during forward bending in subjects with and without a history of low back pain. *Spine (Phila Pa 1976)* 1996; 21:71–78

23.   Takayanagi K, Takahashi K, Yamagata M, Moriya H, Kitahara H, Tamaki T. Using cineradiog-

raphy for continuous dynamic-motion analysis of the lumbar spine. *Spine (Phila Pa 1976)* 2001; 26:1858–1865

24. Wood KB, Popp CA, Transfeldt EE, Geissele AE. Radiographic evaluation of instability in spondylolisthesis. *Spine (Phila Pa 1976)* 1994; 19:1697–1703

25. Muggleton J, Kondracki M, Allen R. Spinal fusion for lumbar instability: Does it have a scientific basis*?* J Spinal Dis 2000*;* 13: 200–204.

26. Boden SD, Wiesel SW. Lumbosacral Segmental Motion in Normal Individuals Have We Been Measuring Instability Properly? Spine1990; 15(6), 571-576.

27. Iguchi T, Kanemura A, Kasahara K, Sato K, Kurihara A, Yoshiya S, Doita M. Lumbar instability and clinical symptoms: which is the more critical factor for symptoms: sagittal translation or segment angulation? Journal of Spinal Disorders & Techniques 2004; *17*(4): 284-290.

28. Penning L. Functional pathology of lumbar spinal stenosis. Clinical Biomechanics 1992; 7: 3-17.

29. Stokes IA, Frymoyer J W. Segmental motion and instability. Spine 1987; 12(7): 688-691.

30. [30]Wood KB, Popp CA, Transfeldt EE, Geissele AE. Radiographic evaluation of instability in spondylolisthesis. Spine 1994; 19(15): 1697-1703.

31. Tallroth K, Ylikoski M, Landtman M, Santavirta S. Reliability of radiographical measurements of spondylolisthesis and extension-flexion radiographs of the lumbar spine. European Journal of Radiology 1994;18(3): 227-231.

32. Hipp JA., et al. Quantitative motion analysis of motion-preserving and fusion technologies of the spine. Motion Preservation Surgery of the Spine. Philadelphia: Saunders Elsevier (2008): 85-96.

33. Bronfort, G. Chiropractic spinal manipulative therapy and exercise versus pharmacological treatment and exercise for chronic low back pain. Proc Intern Conference on spinal manipulation, June. 1994, 35-37.

34. Pearcy M, Portek IA, Shepherd J. Three-dimensional x-ray analysis of normal movement in the lumbar spine. Spine 1984; 9(3): 294-297.

35. Teyhen DS, Flynn TW, Childs JD, Abraham LD. Arthrokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. Physical Therapy 2007; 87(3): 313-325.

36. Tanz SS. Motion of the lumbar spine; a roentgenologic study. The American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine 1953: 69(3): 399-412.

37. Friberg O. Lumbar instability: a dynamic approach by traction-compression radiography. Spine 1987; 12(2): 119-129.

38. Kalebo P, Kadziolka R, & Sward L. Compression-traction radiography of lumbar segmental instability. Spine 1990; 15(5): 351-355.

39. Pitkänen M, Manninen HI, Lindgren KA, Turunen M, Airaksinen O. Limited usefulness of traction-compression films in the radiographic diagnosis of lumbar spinal instability: comparison with flexion-extension films. Spine 1997; 22(2): 193-197.

40. Champain S, Benchikh K, Nogier A, et al. Validation of new clinical quantitative analysis software applicable in spine orthopaedic studies. Eur Spine J 2006; 15:982–91.

41. Schulze M, Trautwein F, Vordemvenne T, et al. A method to perform spinal motion analysis from functional X-ray images. J Biomech 2011; 44:1740–6.

42. Pearson AM, Spratt KF, Genuario J, et al. Precision of lumbar intervertebral measurements: does a computer-assisted technique improve reliability? Spine 2011; 36:572–80.

43. Newman T, Browner W, Cummings S. Designing studies of medical tests. In: Hulley S, ed. Designing clinical research. Philadelphia: Lippincott Williams & Wilkins, 2001:175–1931.

44. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Interpretation of diagnostic data, in Clinical Epidemiology. A basic science for clinical medicine. Boston/Toronto/London: Little Brown and Company, 1991. 69–152.

45. Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. European Spine Journal 2005; 14:743-750.

46. Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery—the casefor restraint. N Engl J Med 2004; 350: 722–6.

47. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006; 332: 1089–92.

48. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a di-

agnostic test, and when do we need randomized trials? Ann Intern Med 2006; 144:850–855

49.   Weinstein JN, Lurie JD, Olson PR, et al. United States' trends and regional variations in lumbar spine surgery: 1992-2003. Spine 2006; 31: 2707–14.

50.   Leone A, Guglielmi G, Cassar-Pullicino VN, et al. Lumbar intervertebral instability: a review. Radiology 2007; 245: 62–77.

51.   Newman-Toker D, Pronovost P. Diagnostic errors: The next frontier for patient safety. Journal of American Medical Association 2009; 301: 1060-1062.

52.   Tan E, Ting B, Jia K, Skolasky R, McFarland E. Examinations for orthopedic trauma patients. American Journal of Medical Quality 2012, 1– 9.

Disclosures

Chip Wade is a consultant for and owns stock in Or-
thokinematics; David Lee is a consultant for and sits on the Scientific Advisory Board for Orthokinematics. All other authors have no relevant financial disclosures.

Corresponding Author

Chip Wade, Department of Industrial and Systems Engineeeing, Auburn University, 3310 Shelby Center, Auburn, AL 36849. Lrw0002@auburn.edu.