Measurement Performance of a Computer Assisted Vertebral Motion Analysis System

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

*Int J Spine Surg* 2015, 9 ()
doi: [https://doi.org/10.14444/2036](https://doi.org/10.14444/2036)
[http://ijssurgery.com/content/9/36](http://ijssurgery.com/content/9/36)

This information is current as of November 26, 2018.

---

**Email Alerts**  Receive free email-alerts when new articles cite this article. Sign up at: [http://ijssurgery.com/alerts](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, PhD⁴
¹Department of Neurosurgery, Greater Baltimore Medical Center, Baltimore, MD, ²Southern Neurologic and Spinal Institute, Hattiesburg MS, ³Department of Industrial and Systems Engineering, Auburn University, Auburn, AL, ⁴Department 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. Inter-observer IVR and IVT had a CR of 2.49 and 2.62, respectively. Intra-observer IVR and IVT had a CR of 1.99 and 2.62, 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.\textsuperscript{1-3} Segmental instability of the spine has been studied \textit{in vivo} with radiographic methods since 1944 yet continues to be controversial in its determination.\textsuperscript{4} Over the years several quantitative definitions of lumbar instability based on radiograph-derived measurements have been proposed,\textsuperscript{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.\textsuperscript{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 post-operative pseudarthrosis, the US Food and Drug Administration (FDA) indicates IVR of less than 5° and IVT less than 3 mm as a successful fusion.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\(^{22-24}\) 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,\(^ {22}\) 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>
<thead>
<tr>
<th>Table 1. Bending posture and bending method classifications for classifying studies reviewed as part of the review of medical literature.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bending Posture</strong></td>
</tr>
<tr>
<td>A. Standing uncontrolled</td>
</tr>
<tr>
<td>B. Hips flexed</td>
</tr>
<tr>
<td>C. Standing pelvis bolster</td>
</tr>
<tr>
<td>D. Side lying</td>
</tr>
<tr>
<td>E. Traction &amp; compression</td>
</tr>
</tbody>
</table>

---

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 the 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. 59 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. In a recent study, Yeager et al. 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.

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 constraining 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 70° 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
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. 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.

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

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 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\textsubscript{t1}–A\textsubscript{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 + CNsymptomatic), where CNsymptomatic 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 CNsymptomatic, 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. 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-
ments to imaging modalities and imaging analysis a lack of consensus around segmental and clinical in-

Table 5. Average angles and translations measured for both VMA and FE imaging radiographs in the agreement dataset.

**Panel A. IVR measured in degrees (°)**

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

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

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

Table 6. Contributions of intervertebral levels comprising the agreement dataset.

<table>
<thead>
<tr>
<th>Intra/Inter Observer</th>
<th>Intra Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>n Contribution</td>
</tr>
<tr>
<td>L1-L2</td>
<td>446</td>
</tr>
<tr>
<td>L2-L3</td>
<td>438</td>
</tr>
<tr>
<td>L3-L4</td>
<td>471</td>
</tr>
<tr>
<td>L4-L5</td>
<td>452</td>
</tr>
<tr>
<td>L5-S1</td>
<td>432</td>
</tr>
<tr>
<td>Total</td>
<td>705</td>
</tr>
</tbody>
</table>

Diagnostic misclassification errors have received in-

Table 7. Intra and inter observer variability and intra subject variability in VMA lumbar imaging radiographs.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Mean in Differences</th>
<th>CI</th>
<th>Upper LOA</th>
<th>CI</th>
<th>Lower LOA</th>
<th>CI</th>
<th>ICC (3,1)</th>
<th>CI</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer Variability Rotation</td>
<td>0.02</td>
<td>0.03 to 0.09</td>
<td>2.52</td>
<td>2.56 to 3.58</td>
<td>-2.49</td>
<td>-2.52 to -2.47</td>
<td>2.49</td>
<td>0.974 to 0.978</td>
<td>0.13</td>
</tr>
<tr>
<td>Intra-Observer Variability Translation (%)</td>
<td>0.06</td>
<td>0.05 to 0.13</td>
<td>2.69</td>
<td>2.67 to 2.71</td>
<td>-2.62</td>
<td>-2.64 to -2.60</td>
<td>2.62</td>
<td>0.926 to 0.936</td>
<td>0.18</td>
</tr>
<tr>
<td>Inter-Observer Variability Rotation</td>
<td>0.05</td>
<td>0.04 to 0.06</td>
<td>1.97</td>
<td>1.96 to 2.00</td>
<td>-2.03</td>
<td>-2.04 to -2.02</td>
<td>1.99</td>
<td>0.939 to 0.946</td>
<td>0.23</td>
</tr>
<tr>
<td>Intra-Observer Variability Translation (%)</td>
<td>0.03</td>
<td>0.02 to 0.04</td>
<td>2.84</td>
<td>2.83 to 2.85</td>
<td>-2.91</td>
<td>-2.92 to -2.90</td>
<td>2.81</td>
<td>0.946 to 0.957</td>
<td>0.24</td>
</tr>
</tbody>
</table>

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.\textsuperscript{29,48} Misclassification errors resulting from the methodology and reporting of radiological examinations are common place across modalities.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 $\pm 1.53^{\circ}$ and $\pm 2.20\%$ 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 $2.15^{\circ}$ and $3.90\%$, respectively. The current study extended the Yeagar et al.\textsuperscript{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.\textsuperscript{10} the results were quantitatively similar. The current findings indicate inter-observer reliability (95\% confident) falls within a range of $\pm 2.49^{\circ}$ and $\pm 2.62\%$ of the actual value in IVR and IVT respectively. Further the current study reports a quantitatively similar inter-observer reliability CR of $1.99^{\circ}$ 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 vertebral motional analysis system. The current findings extend those of Yeagar et al.\textsuperscript{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

The authors would like to acknowledge the participa-
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

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


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

Disclosures
Chip Wade is a consultant for and owns stock in Orthokinematics; 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 Engineering, Auburn University, 3310 Shelby Center, Auburn, AL 36849. Lrw0002@auburn.edu.

Published 17 July 2015.
This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2015 ISASS. To see more or order reprints or permissions, see http://ijssurgery.com.