Advanced Multi-Axis Spine Testing: Clinical Relevance and Research Recommendations

Timothy P. Holsgrove, Nikhil R. Nayak, William C. Welch and Beth A. Winkelstein

Int J Spine Surg 2015, 9 ()
doi: https://doi.org/10.14444/2034
http://ijssurgery.com/content/9/34

This information is current as of December 14, 2018.

Email Alerts  Receive free email-alerts when new articles cite this article. Sign up at:
http://ijssurgery.com/alerts
Advanced Multi-Axis Spine Testing: Clinical Relevance and Research Recommendations

Timothy P. Holsgrove, PhD,1 Nikhil R. Nayak, MD,2 William C. Welch, MD,2 Beth A. Winkelstein, PhD1
1Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 2Department of Neurosurgery, University of Pennsylvania, Philadelphia, PA

Abstract

Back pain and spinal degeneration affect a large proportion of the general population. The economic burden of spinal degeneration is significant, and the treatment of spinal degeneration represents a large proportion of healthcare costs. However, spinal surgery does not always provide improved clinical outcomes compared to non-surgical alternatives, and modern interventions, such as total disc replacement, may not offer clinically relevant improvements over more established procedures. Although psychological and socioeconomic factors play an important role in the development and response to back pain, the variation in clinical success is also related to the complexity of the spine, and the multi-faceted manner by which spinal degeneration often occurs.

The successful surgical treatment of degenerative spinal conditions requires collaboration between surgeons, engineers, and scientists in order to provide a multi-disciplinary approach to managing the complete condition. In this review, we provide relevant background from both the clinical and the basic research perspectives, which is synthesized into several examples and recommendations for consideration in increasing translational research between communities with the goal of providing improved knowledge and care.

Current clinical imaging, and multi-axis testing machines, offer great promise for future research by combining in-vivo kinematics and loading with in-vitro testing in six degrees of freedom to offer more accurate predictions of the performance of new spinal instrumentation. Upon synthesis of the literature, it is recommended that in-vitro tests strive to recreate as many aspects of the in-vivo environment as possible, and that a physiological preload is a critical factor in assessing spinal biomechanics in the laboratory. A greater link between surgical procedures, and the outcomes in all three anatomical planes should be considered in both the in-vivo and in-vitro settings, to provide data relevant to quality of motion, and stability.

Introduction

Spine-related symptoms and conditions, such as back pain, affect a large proportion of the general population, but the spine’s complex structure makes it difficult to determine the exact source and/or cause of pain. The number of patients seeking treatment for spine-related problems was estimated to be nearly 33 million in 2005,1 with a nearly 15-fold increase in the number of complex spinal fusion procedures performed between 2002 and 2007 in the Medicare population.2 In a brief published by the Agency for Healthcare Research and Quality in 2014, spinal fusion was the 6th most common surgical procedure, with 488,000 cases performed annually.3 However, in terms of aggregate hospital costs, it represents the single-most expensive operative procedure, accounting for $12.8 billion per year.4 This large aggregate expense, along with the trend of increased utilization,1 has made spine surgery a leading target for cost containment.5,6

A fundamental problem in spine management is that much of the pre-clinical research and in-vitro testing of surgical instrumentation and devices, which has led to approval of a staggering number of operative choices, has not necessarily produced improved patient outcomes.7,8 Patient outcomes may be improved upon by gaining a more detailed understanding of the performance of such surgical devices, both in biomechanical laboratory tests, and in the clinical setting, and assist with the ultimate goal of improving patient
care.

Each level of the spine (from C2 to S1) comprises a triple joint construct with six degrees of freedom (DOF) (Figure 1). The interaction of these structures during normal activities requires complex techniques to understand and fully define the biomechanics of the spine, the effect of injuries and degeneration, and to identify the most effective of various treatment options. Wear and fatigue testing standards are well-established for most forms of spinal instrumentation. However, these standards do not assess the likely biomechanical performance of devices in-vivo. Although previous calls for standardized in-vitro spine testing methods have outlined the importance of different aspects of spinal testing, how protocols can be developed along standardized procedures and what key areas of research should be focused on using such testing protocols, the link between in-vitro testing and the in-vivo environment, and between in-vitro test methods and clinical practice, can often be disconnected.

This review provides an overview of clinical practice relating to spinal degeneration, and outlines key developments in multi-axis biomechanical testing relating to procedures and instrumentation used clinically. The link between in-vitro and clinical correlates is then highlighted with presentation of specific case studies, which form the basis for recommendations for clinically relevant research.

Anatomy & Kinematics

The spinal column consists of 33 vertebrae, fibrocartilaginous and ligamentous structures, and numerous muscular attachments. Degeneration or compromise of any of these elements may lead to pain and/or disability. Aside from the unique uppermost cervical vertebrae (C1 and C2), most spinal levels have similar anatomy (Figure 1), properties and essential functions. The vertebral body (VB) at each level increases in size from the cranial to caudal end of the spinal column to accommodate the increased loads present.

The bilateral facet joints and the intervertebral disc (IVD) are responsible for the articulations at each spinal level. On the dorsal aspect of the bony spine, the inferior articulating process of the vertebra above and the superior articulating process of the vertebra below are encapsulated by a ligament to form the bilateral facet joints. The majority of axial loading in the spine is transferred through the vertebral bodies and IVD, with facet load bearing estimated to be 10-20% of the total axial load. The orientation of the facet joints dictate their function, with the coronally-oriented facet joints of the cervical and thoracic segments resisting translation but allowing flexion, extension, and rotation, while the sagittally-oriented lumbar facets resist rotation but allow flexion and extension. The healthy IVD consists of a hydrated inner nucleus pulposus surrounded by the fibrocartilaginous annulus fibrosus. The IVD distributes axial loads, allows motion between vertebrae (axial compression/distraction, flexion/extension, lateral bending, axial rotation), and limits rotation and shear. The mechanical characteristics of spinal degeneration of the IVD have been simulated in-vitro, and such methods can be used to further understand the degeneration of the spine. Degeneration or trauma to these anatomic components can compromise the spine’s ability to guide normal motions or to limit abnormal motions.

The ligaments of the spine provide passive stabilization and serve as tension bands to prevent excessive motion. The paraspinal musculature also has a significant role in stabilizing the spine, in addition to maintaining posture and providing motion. The abdominal trunk muscles and multifidus muscles are important for lumbar spinal stability. The erector spinae muscles of the lumbar region facilitate lower back extension, and the semispinalis cervicis muscle, which
attaches to the C2 spinous process, has a considerable role in neck extension and prevention of cervical kyphosis. Weakness and atrophy of the dorsal lumbar musculature is thought to play a significant role in the development of post-operative failed back syndrome.

The physiologic spinal curves in the sagittal plane have important biomechanical contributions to normal spinal function. The balance of cervical lordosis, thoracic kyphosis, and lumbar lordosis permits normal standing posture without excessive strain on the paraspinal musculature and spinal joints. In the coronal plane, the spinal column is expected to be relatively straight, and the severity of coronal plane curves are measured using the Cobb angle technique. In evaluating degenerative coronal plane curves in adults, Cobb angles less than 10° fall within normal limits, whereas angles greater than 10° are deemed to be scoliotic, which occur primarily in the lumbar spine. Imbalances, in either the coronal or sagittal planes, may cause back pain and varying degrees of neurologic dysfunction, although in patients with multi-planar deformity, the degree of sagittal imbalance has been found to be the most reliable predictor of clinical symptomatology. When symptomatic, patients may compensate for their spinal imbalance by altering their posture and pelvic tilt, which may lead to further pain, fatigue, and accelerated spinal degeneration.

For clinical purposes, spinal motion can be considered as occurring in the sagittal (flexion and extension), axial (rotation), and/or coronal (lateral bending) planes. In reality, motion is far more complex due to coupling of motions, such as that between axial rotation and lateral bending. Translation in each plane imposes shear loading, although normal spinal anatomy serves to limit such motion. Flexion/extension range of motion (ROM) is age-dependent and highest in the cervical spine, with normal values of 45-60° of flexion and 60-80° of extension. Total thoracic flexion/extension is approximately 30°, and lumbar flexion and extension are approximately 50 and 20°, respectively in asymptomatic individuals. Aside from C1-C2, lateral bending is relatively consistent throughout the spine. Axial rotation is greatest at C1-C2 (>30° in either direction), but relatively constant in most of the cervical and thoracic spines, and very limited in the lower thoracic and lumbar regions. Although ROM is often a focus of biomechanical tests, clinical examination of ROM is relatively rudimentary, and focuses primarily on flexion/extension without formal measurements.

Degeneration in the Clinical Scenario

Although many classifications of clinical instability have been proposed, the most commonly cited definition is that provided by White and Panjabi: "the loss of the ability of the spine under physiological loads to maintain relationships between vertebrae in such a way that there is neither damage nor subsequent irritation to the spinal cord or nerve roots, in addition, there is no development of incapacitating deformity or pain due to structural changes." Spinal pathology can be degenerative, traumatic, infectious, neoplastic, or iatrogenic, and can lead to clinical instability, although patients may have significant back and neck pain without overt instability secondary to abnormalities in specific pain generators (e.g. the IVD or facet joint).

Degenerative processes in the IVD begin with the loss of proteoglycans that results in lower water-binding capacity and shock absorption. The loss of elasticity of the annulus fibrosus makes it more susceptible to tears and herniations. As the disc desiccates, disruption of the normal distribution of axial loading creates a degenerative cascade (Figure 2). That change in axial loading can also produce increased stress on the facet joints, which can lead to arthritic changes, which in turn may lead to displacement or crowding of the ligamentum flavum, and compression of the neural structures. Additionally, age-related losses in bone mineral density (BMD) may result in osteopenia or osteoporosis, which predispose individuals to compression fractures, which may alter load transfer through the spinal column and lead to degeneration.

Imaging & Diagnosis

Clinical evaluation of the patient with spine pain includes a history and physical examination focusing...
on the neurological examination and review of relevant imaging. Once any “red flags” (e.g. fever, progressive neurologic deficit, history of unintentional weight loss, bowel/bladder dysfunction) have been ruled out, which would suggest infection, neoplastic process, or the need for urgent surgery, magnetic resonance imaging (MRI) is the first line imaging modality for degenerative disease. Unexpected findings on plain x-rays are exceedingly rare; thus, x-rays are not recommended for routine evaluation of degenerative back and neck pain unless there is a strong suspicion for malignancy, inflammatory conditions, acute fracture, or infection. However, dynamic flexion/extension radiographs, or dynamic MRI, may be obtained to evaluate for motion indicative of instability. MRI, in particular, has become the modality of choice for evaluation of spine patients, since it provides detailed views of the IVD, ligaments, and neural structures. Assessment of the bony anatomy is superior with computed tomography (CT) scanning, but most bony degenerative pathology can be adequately evaluated with MRI.

The mainstay of determining degenerative radiographic instability is standing flexion/extension x-rays demonstrating abnormal motion or static deformity (e.g. spondylolisthesis). Traction/compression x-rays have been deemed to be of limited use. The primary criteria in evaluating flexion/extension films are sagittal plane translation (the distance between straight lines drawn along two consecutive posterior VBs), and sagittal plane rotation (the change in the angle formed between lines drawn along the endplates flanking a disc space) (Figure 3). Sagittal translation is often expressed as a percentage of VB anterior-posterior length to minimize technical differences between films. Spondylolisthesis is graded from I-IV based on the Meyerding scale of translation: grade I is up to 25%; grade II is 25-50%; grade III is 50-75%; grade 4 is 75-100%; more than 100% translation is considered spondyloptosis.

White and Panjabi defined lumbar instability as translation >4.5mm (or 15% of anterior-posterior distance) and rotation of >15° at L1-2, L2-3, and L3-4, >20° at L4-5, and >25° at L5-S1. However, Iguchi and colleagues evaluated 1,090 outpatients for translation and angulation and found a cutoff of 3mm of translation to be associated with more severe clinical symptoms, and that angulation does not play a significant role. Likewise, many radiologists currently use a dynamic slip >3mm, static slip ≥4.5 mm, or angulation >10-15° as the rule-of-thumb for radiographic instability in the lumbar spine. The variation in reported findings highlight the importance of clinical correlation, since many asymptomatic patients may...
have spondylolisthesis or radiographic instability, with sagittal angulation as high as 25° being reported in healthy volunteers. In the cervical spine, White and Panjabi proposed cervical instability as >3.5 mm of vertebral translation or >11° of rotation on flexion/extension x-rays, based on work with a cadaveric model; however, more recent studies have suggested 2mm as a possible cutoff value for translation.

If a patient’s symptoms and imaging findings do not correlate, further diagnostic studies, such as nerve conduction studies and electromyography, may be obtained. Additionally, invasive testing, like discography, may identify the pain generator in cases of multiple degenerative discs, and a positive response to injections in the epidural space, facet joint, or transfornaminal space may further localize the pain generator. Normal appearing discs on MRI should not generally be tested with discography. However, it may be required to obtain “normal” results for the purposes of validation, in which cases discography on healthy discs may be necessary. Unless patients have progressive neurologic deterioration or intractable pain with correlating imaging findings, most surgeons advise the patient to undergo non-surgical management via activity modification, physical therapy, oral analgesics, and/or further steroid injections prior to offering surgical treatment.

Surgical Treatments
Historically, the goal of surgery for spinal pain with associated clinical and/or radiographic instability has been bony fusion based on the theory that instability is due to increased motion. There are many variables in determining the appropriate surgical procedure and approach including the symptoms, primary pathology, global spinal alignment, and surgeon experience. In fusion surgery, instrumentation is often used, and this serves as a temporary stabilizing scaffold until bony arthrodesis occurs. Although the goal of arthrodesis is also to correct and/or prevent deformity, it may lead to imbalances in the normal physiological curves and this can weaken the other stabilizing structures, which thereby increases stress on adjacent segments. In fact, symptomatic adjacent segment disease (ASD) has been observed in up to 25% of patients with cervical fusions and 36% of patients with lumbar fusions.

Instrumentation is mostly composed of posterior screw-and-rod systems. The strength of these cantilever constructs is largely from anchorage into the pedicle and is proportional to the rigidity of the connected system. Pullout strength of pedicle screws is related to variables such as the length, diameter, thread count, trajectory, use of transverse connectors, and bone mineral density. In the subaxial cervical spine, posterior screw options include lateral mass, translaminar, and, particularly at C7, pedicle screws. In the thoracic spine, smaller pedicles and articulation with the rib heads have led to the adoption of many different screw trajectories. The extrapedicular technique encompasses the transverse process, rib head, pedicle, and VB which, in theory, may result in greater pullout strength compared to screws placed purely transpedicularly; but, biomechanical data comparing the two techniques is equivocal. Additionally, two primary sagittal trajectories are employed: the “anatomic” trajectory which follows the angle of the pedicle, and the “straight-forward” trajectory, in which the screw tip is aimed towards the superior endplate and is thought to provide greater pullout strength (Figure 4). In the lumbar spine, the trajectory in the axial plane is most influnential.
ential, as triangulated trajectories converging toward midline, when transversely connected, have greater pullout strength than straightforward screws aimed toward the lateral portion of the VB. Additionally, rigidity of the posterior construct is related to rod diameter and stiffness, as larger diameter rods lead to a more rigid construct.

There has been increased use of interbody approaches utilizing autograft bone, allograft bone, synthetic implants, or a combination. The anterior interbody approach is frequently used for cervical spine pathology with good results. Cervical interbody implants are generally buttressed with an anterior locking plate, which is believed to reduce graft migration, increase fusion rates, and act as a tension band. In the lumbar spine, anterior-only instrumentation is not as common as posterior approaches, with or without interbody grafts. As with posterior instrumentation, anterior constructs provide more stability in flexion and lateral bending than in extension and axial rotation. Combined anterior/posterior approaches are sometimes utilized in cases of severe pathology, although these cases are more common in traumatic fracture-dislocation injuries, significant VB destruction from tumors, or planned iatrogenic destabilization for neural decompression.

Lumbar interbody cages may be used to restore anterior column height, provide indirect decompression of the neural foramen, and house cancellous bone graft to facilitate fusion. Cages placed posteriorly can be bilateral (posterior lumbar interbody fusion) or unilateral (transforaminal interbody fusion), although there is limited data on whether outcomes differ between these two techniques. Regardless, it is important that the cage should not excessively shield the bone graft from stress, which is required to promote appropriate bone remodeling. Another risk of interbody cage insertion, both in the cervical and lumbar regions, is cage subsidence into the cancellous bone of the adjacent body. This risk may be minimized with judicious removal of the bony endplate, particularly at the periphery. Migration of lumbar fusion cages also presents a clinical complication, and this has been shown to be affected by both the cage shape, and the positioning of the device.

Alternatives to arthrodesis include non-rigid posterior stabilization such as dynamic pedicle screw fixation, interspinous process distraction, and disc arthroplasty (i.e. total disc replacement (TDR)). The goals of these systems are to restore physiologic ROM (via the TDR) or limit motion without a fusion (via interspinous process distraction or dynamic pedicle screws) in order to alleviate symptoms and minimize the risk of ASD. These devices also rely on preservation of the stabilizing structures (i.e. ligaments, muscle), so meticulous surgical exposure is required for optimal outcomes. Pre-clinical and in-vitro studies have not correlated well to improved patient outcomes. In the lumbar spine, both fusion and TDR devices are approved by the FDA for the treatment of back pain from degenerative disc disease, since the disc is believed to be a common pain generator. However, in a 2013 Cochrane review on TDR for chronic discogenic low back pain, Jacobs et al. reported that compared to fusion, TDR did not result in improvements above a clinically meaningful threshold for pain relief, quality of life, or disease-related disability. Additionally, ASD was found to be inadequately studied, and much of the research on TDR has been via clinical trials, where stringent patient selection limits the generalizability of findings. The authors concluded that there is a great need for higher quality studies with less conflict of interest in this area.

In the cervical spine, TDR is not recommended as a...
treatment for axial neck pain from degenerative disc disease, but rather is reserved for patients with neurologic dysfunction (radiculopathy and/or myelopathy) from single level disc compression. Therefore, both anterior cervical discectomy with fusion (ACDF) and TDR can be used to achieve the same goal. Clinically, results are evaluated through patient-reported outcomes and objective physical exam findings. Biomechanically, success is largely measured by maintenance of the physiologic parameters presented above (segmental lordosis, ROM, disc space height), with the ultimate goal reduced frequency of ASD, and, therefore, lower rates of adjacent level surgery.

A recent meta-analysis of eight randomized controlled trials on cervical TDR and ACDF concluded that TDR was equivalent, or superior, to ACDF based on levels of reported pain, neurologic improvement, and rates of reoperation. However, it should be noted that while differences in pain levels based on the subjective outcome of visual analog scales were statistically significant, the difference was small and may not have been clinically meaningful. Additionally, there was no significant difference between groups in neck disability index scores, a validated, objective measure of disability from neck pathology. A second meta-analysis did not find any differences in patient-reported outcomes between the two groups, but did find ACDF to have higher rates of subsequent surgery for ASD. Conversely, a more recent meta-analysis found no difference in the rates of ASD requiring surgery between the two procedures at two to five years follow up. Another difficulty in evaluating TDR is that virtually all studies to date have had limited follow up. Most published trials have used 24-month outcomes, although a recently published 48-month study maintained similar results. Biomechanically, given that the TDR makes use of a mechanical device, which is relied upon for maintained structural support, the need for long-term outcomes and reoperation rates is paramount. Laboratory studies using biological specimens are limited in the number of cycles a device can be interrogated, while the number of cycles experienced in a patient’s lifetime is unknown and likely many times higher. It is, therefore, important to collect long-term clinical data, in order to fully understand the implications of TDR compared to alternative treatments.

Given increasing economic pressures and proposed limitations to healthcare spending, it is more important than ever to ensure that new devices do not just meet current standards, but surpass the outcomes of their predecessors. The equivocal results between TDR and fusion for both low back pain and cervical pathology highlight the need to review and critically understand how laboratory-based multi-axis spinal testing is used in order to predict clinical success, and how it may be modified to better simulate the in-vivo environment in future studies.

### Multi-Axis Biomechanical Testing

The aims of basic research using in-vitro spine testing are to understand more about the biomechanics of the healthy spine, the effects of injury and/or degeneration, and to assess the effectiveness of new spinal devices. The biomechanics of spinal specimens is significantly affected by many factors in the laboratory, including the application of a physiological preload, the testing velocity, the specimen moisture condition, and the specimen temperature. It has also been shown that the exposure period, and the number of test cycles a specimen is subjected to can significantly alter its biomechanical response. In spite of the increased understanding of how in-vitro test conditions affect spine biomechanics, it is challenging, and often impossible, to compare different in-vitro studies because of these considerations. Further, translating the findings of in-vitro studies to the clinical environment can be problematic if in-vivo conditions have not been replicated fully. Moreover, there is a similar problem if the biomechanical responses from such testing is not considered in the context of the experimental conditions.

Efficacy testing is a critical addition to the current barrage of pre-clinical testing standards, and for such testing to have the greatest clinical impact, it is crucial that testing methodologies replicate all aspects of the in-vivo environment. Dynamic testing standards for spinal devices generally require that spinal devices be tested dynamically, in a test fluid at 37°C, with an axial preload. However, biomechanical testing is often performed at sub-physiological
speeds, without a physiological preload, and in temperature and moisture conditions that are not physiological. Wear and fatigue tests are not used to replicate complex spinal loading but instead to provide approximate conditions that are highly repeatable. Biomechanical testing should take into account aspects of standardization from the test standards and apply more clinically relevant conditions in order to better inform the clinical arena as to the likely outcomes as a result of degeneration, injury, or treatment. Currently, this link is not as strong as it could or should be.

Simulating In-Vivo Conditions

It is well-understood that the spine is subjected to large compressive loads due to the weight of the head and torso, combined with the effect of muscle forces that provide stability. The stiffening effect of a physiological preload on spinal specimens in-vitro has also been well-documented in the literature in the lumbar, thoracic, and cervical regions of the spine, and through the application of a preload via an axial force, a follower-load, and simulated muscle forces.

The method used to apply a preload has been shown to affect specimen biomechanics with unconstrained preloads generating large moment and low shear force artifacts, and constrained preloads doing the opposite. Therefore, it is important to consider the most appropriate method, with minimal “side effects” when designing a testing protocol. It has been suggested that whilst a physiological preload should be applied when possible, a fair comparison can be obtained without it, provided specimens are tested in a similar manner in the intact state and with spinal instrumentation. Indeed, many studies have adopted such a technique. However, if a primary concern of the spinal surgeon is stability, which may be affected by the magnitude of axial loading, the potential instability of spinal devices may go unnoticed without an appropriate physiological preload in-vitro. The transfer of load between the IVD and the facets is also significantly altered by the application of an axial preload, and by mechanically stimulated degeneration of the IVD. Both of these aspects are a key part of understanding the mechanical behavior of the spine, spinal degeneration, and therefore, treatment.

Different postures change the axial load that is transmitted through the spine in-vivo. Intradiscal pressure has been shown to increase in-vivo as a result of altered postures with forward bending approximately doubling the load through the disc compared to relaxed standing. This occurs due to increases in the lever arms of the upper body in relation to the center of rotation (COR) of the different levels of the spine, and the resulting increase in muscle activity that is established in order to resolve these changes, as well as possible alterations in the load transfer through the disc and the facets. These complex interactions of load transfer between the spinal structures should be considered when applying a preload in-vitro, and relate to the compromise between moment and shear force artifacts. Such artifacts may relate to the effect of muscle forces in-vivo, which are difficult to adequately replicate in the laboratory setting. Increased artifact moments may produce inaccuracies in ROM and the resulting stiffness/flexibility data, whereas increased shear forces may alter the COR and engagement of the facets. Either or both of these may lead to inaccuracies in the load-sharing between the anterior and posterior elements of the spine compared to the in-vivo environment. However, these limitations may still be advantageous over not applying any preload at all. Similarly, the length of load application during testing also affects the biomechanical response of individual tissues and the spine as a whole, given the possibility for creep and the effects of the fluid behavior in the IVD and soft tissues.

The stiffness of spinal specimens is also significantly affected by the moisture conditions. Pflaster et al. reported human lumbar isolated disc specimens (ISDs), comprising a functional spinal unit (FSU) with the facets and posterior structures removed, could be maintained at approximately post-mortem mass through submersion in saline solution with the application of a 445N preload, or sprayed with saline and wrapped in plastic. However, in that study, neither the stiffness nor the flexibility of the specimens was compared between the different moisture conditions. Wilke et al. demonstrated that spraying ovine FSUs with saline solution and wrapping them in plastic led to little change in flexibility (<10%) in axial rotation.
compared to air-exposed (~30%) or saline-submerged (~30%) specimens over 500 test cycles. However, this lack of effect may not reflect the complete picture with regards to different moisture conditions; since those tests were performed without a physiological preload, the facets would have contributed to the majority of the stiffness in axial rotation. In addition, the IVD stiffness in axial rotation is predominantly related to the elastic response of the annulus fibrosus, rather than a fluid response in the nucleus pulposus. Therefore, prolonged testing along different axes, such as flexion/extension, may result in greater differences in flexibility due to moisture conditions. Holsgrove et al. performed stiffness matrix testing of porcine FSUs and ISDs without a preload and also with a 500N preload after 30 minutes of equilibration, and then repeated the testing after a total preload application time of 60 minutes. This study demonstrated that the initial application of the preload had a large effect in all six axes for both types of specimens, but the increased application time also significantly changed the stiffness in all primary axes, with the exception of anterior/posterior shear. The largest differences were increases of 40-60% in flexion/extension and lateral bending, which are responses affected most by the fluid phase compared to other axes. Shear and axial rotational stiffness terms were reduced by between 0-4%, which related to the creep of the elastic tissues of the annulus fibrosus. This emphasizes that not only is it important to ensure appropriate moisture conditions are maintained during in-vitro tests, but also that the large effects of the fluid response that is a factor in prolonged loading must be considered when designing and implementing testing protocols.

Similar interactions of the solid and fluid phases of the human IVD have been reported by Costi et al. as a result of changes in loading rate from 0.001Hz to 1Hz in all six axes with the application of an axial preload in a fluid bath at 37°C. The stiffness in anterior/posterior shear, lateral shear, and axial rotation increased by 26-35%, which are responses that are primarily governed by the solid phase of the IVD; increases of 29-83% were reported for axial compression/extension, lateral bending, and flexion/extension, which are primarily governed by the fluid response of the nucleus pulposus. Similar increases in stiffness in the neutral zone of human FSU specimens were reported by Gay et al. due to an increase in test frequency from 0.5-6.0°/s in pure moment testing in the sagittal plane.

The temperature of specimens also has effects on the measured stiffness. Bass et al. reported that the stiffness of the ALL was 38% greater at 21.1°C compared to 37.8°C. Although some in-vitro testing has been completed at body temperature, most studies use room temperature due to the relative ease by which it can be achieved and maintained. The effects of temperature may be more reasonably extrapolated to the in-vivo situation compared to other test factors. However, as with the moisture condition, the temperature of spinal devices, such as UHMWPE bearings, or elastomeric devices, may behave in an altogether different manner at room temperature than at body temperature. The testing frequency and preload magnitude has also been shown to affect the sagittal bending properties of the elastomeric Cadisc TDR, highlighting the notion that replicating the in-vivo environment is critical to understanding not only the properties of the natural spine, but also its properties with spinal instrumentation, and ultimately, the effects of the instrumentation.

In addition to the testing conditions, the type of specimen used is an important factor in biomechanical studies and their interpretation. Both single-level and multi-level specimens are commonly used in in-vitro spinal testing. Single-level testing provides a useful means to assess the spinal structures, and may enable highly accurate positional data to be acquired directly through the testing apparatus. Multi-level testing requires additional measurement techniques to acquire the motion of individual vertebrae, most commonly in the form of a multi-camera and marker system.

Dickey and Kerr reported that although the stiffness of single-level L3-L4 specimens was not significantly different from the stiffness of that same spinal level when it was considered as part of a multi-level specimen, the neutral zone and ROM were significantly greater in single-level specimens and multi-level specimens with resected supraspinous and interspinous ligaments. Single-level specimens have
been tested in-vitro as both FSU and ISD specimens; comparisons between these specimens has shown that the facets and posterior ligaments provide substantial components of stability to the spine in all six DOF.\textsuperscript{91,95} Comparisons of ISDs with a fusion device or TDR will allow direct comparison of the intact structures and the effect of replacing those structures. Testing the same devices in multi-level spinal specimens with the facets and posterior elements make direct comparisons of the intact disc and the device more difficult, but provide important information on how the device performs in the whole spine.

It is important to consider the relevance of spinal specimens used in in-vitro testing, and how findings translate to clinical practice. Many in-vitro studies use human cadaveric specimens of advanced years, which may themselves have some degree of degeneration because of the natural history. It has been shown in-vitro that higher vertebral bone density relates to better stabilization,\textsuperscript{96} and that the level of disc degeneration significantly alters the rotational stiffness of FSUs about all three axes.\textsuperscript{98,99} Porcine\textsuperscript{91,95,100} and ovine\textsuperscript{81,101} specimens have both been commonly used as alternatives to human specimens. These species can provide similar quality of motion in many aspects of spinal testing, and there is much greater repeatability between specimens compared to human cadavers. However, care must be taken to choose an appropriate species for the testing purpose, to provide the most relevant translational value to the clinical setting.

Although the studies reviewed above present the possible confounding effects of individual aspects and factors of the in-vitro testing environment, there are few published studies that have completed dynamic, multi-axis testing that also simulate the physiological preload, temperature and moisture conditions that are representative of the in-vivo environment. This may be due to certain impracticalities related to the measurement of variables in such conditions, or because of more indirect factors such as time and expense. Of all of the individual factors, axial preload has the greatest effect on the biomechanics of spine specimens, with increases in stiffness of over 100\% in flexion/extension, lateral bending, and axial compression/extension.\textsuperscript{91,99} However, the application of a physiological preload is also one of the more difficult aspects to achieve in in-vitro testing, with different methods of application leading to different artifacts. As such, it is critical to carefully consider and report those methods.

Systems & Experimental Approaches

There are four major types of multi-axis testing machines that have been used for spinal testing (Figure 5): (1) translational platform and gimbal assemblies, which may have passive shear axes,\textsuperscript{77,102-104} may be fully active in all six axes,\textsuperscript{105-107} or may use clutches to operate with axes in either passive or active modes;\textsuperscript{25} (2) hexapod testing machines,\textsuperscript{108,109} (3) robotic-arm systems;\textsuperscript{110,111} and (4) pulley arrangements.\textsuperscript{85,97,121-114}

While no single testing machine is more appropriate than another, an appreciation of the advantages and disadvantages of each for the testing of specific spinal devices, along with appropriate documentation, will assist in the comparison of studies using different machines. Both position- and load-control methods have been used to test spine specimens, and both have advantages and disadvantages.\textsuperscript{112} Position control has been adopted for both quasistatic and dynamic tests. Load control generally requires a greater level of sophistication than position control, due to the unknown stiffness/flexibility of the specimen prior to its testing; as such, control methods developed to test spinal specimens in full six-axis load control have been limited in terms of applying dynamic loading.\textsuperscript{106,109,111} Hybrid control methods have also been adopted, using position control for the primary axis, and operating the non-primary axes in load control,\textsuperscript{116,117} though this method still requires a high level of computation during each control iteration compared to six-axes position control, and testing using this method has been slower than the 0.5-5.0°/s rate recommended by Wilke et al.\textsuperscript{18} However, if higher test rates can be achieved, the advantage of combining control modes for different axes is more beneficial than simply completing tests at physiological speeds. Utilizing position control along the primary axis enables greater consistency across multiple tests, by ensuring the same cycle velocity, thus minimizing viscoelastic effects whilst also minimizing artifact forces and moments through the use of load control in non-primary axes.
There are two predominant methods used for spinal testing in six DOF: the stiffness matrix method; and the flexibility matrix method. In the stiffness matrix method, defined translations and rotations are applied to a specimen in one axis at a time, with all other axes constrained, and the resulting forces and moments are measured in six axes. The flexibility matrix method requires the inverse, with defined forces and moments applied in one axis at a time, and the resulting unconstrained translations and rotations measured in six axes. Both methods use data from testing in each of the six axes to calculate either the stiffness or the flexibility in a 6x6 matrix.

Whilst these methods result in 36 terms, half are assumed to be zero due to sagittal plane symmetry, for example, the lateral shear force during flexion/extension would be expected to be negligible. The remaining 18 terms comprise the six principal terms, and 12 non-principal terms: The principal terms are those directly related to the test axis, for example the term calculated from the anterior/posterior translation and the resulting anterior/posterior shear force, or that calculated from flexion/extension rotation and the resulting flexion/extension moment; the non-principal terms relate to the coupled behavior of the specimen, for example, the stiffness due to anterior/posterior translation and the resulting flexion/extension moment.

The first studies to adopt stiffness and flexibility matrix testing methods assumed matrix symmetry based on the conservation of energy, However, the conservation of energy assumption is based around infinitesimal rather than finite displacements, and is not applicable over normal physiological ROM, due to the complex interaction of different spinal structures. The facets play a substantial role in guiding motion in all three anatomical planes, resulting in an asymmetric matrix. However, even with the facets and posterior elements removed, porcine ISDs have asymmetric stiffness matrices over normal ranges of motion, due to the geometry and the combination of elastic and fluid phases of the disc that govern the mechanical behavior.

Although stiffness and flexibility matrices are inversely related, the constraint under which tests are completed differs, with stiffness matrices using a fixed COR, and flexibility tests using a non-defined and unconstrained COR. Unconstrained bending moments have been shown to increase stiffness compared to constrained moments in ovine FSUs, which may be due to the structure of the facets and the soft tissues being predisposed to resist motion to a greater extent about the natural COR compared to circumstances of constrained loading. These differences, combined with other constraints that may be applied in the laboratory setting, such as the method of preload application, mean that it is often difficult to compare studies using different testing methods.

The stiffness matrix method has been used to characterize the mechanical properties of single-level motion segments statically, quasi-statically, and dynamically. However, although the stiffness matrix protocol characterizes the mechanical properties of a spinal specimen in six DOF it does not necessarily apply physiological motions, and is inappropriate for multi-level specimens. The advantage of the flexibility method is that the COR does not need to be defined, nor is it fixed during testing. It is more common for flexibility testing to focus solely on the application of moments, referred to as “pure moment testing,” and this provides a way of effectively...
testing multi-level specimens, which is important in understanding the spine and its overall and localized responses to loading, naturally or via injury and/or treatments. 

It is increasingly understood that the quality of motion, in addition to the quantity of motion, should be considered both in in-vitro testing and in-vivo. Such considerations relate to the non-linear response of spinal tissues under load, and the variable COR about which motion occurs. Quality of motion assessments are now relatively common for in-vitro spinal testing, and more in-vivo quality of motion assessments would provide valuable data in terms of investigating how the pre-clinical efficacy testing of new spinal devices translates to the clinical setting.

A key area of research for multi-level specimen testing is to investigate the adjacent segment behavior following arthrodesis or total disc replacement. Panjabi et al. developed the hybrid method to investigate adjacent segment effects, which consisted of applying a pure moment to multi-level specimens and using the resulting global ROM as the input criteria to test the specimen following implantation of spinal instrumentation. This method has been used to assess TDRs and fusion procedures in-vitro, but has generally been performed without a follower-load, which may have limited adjacent segment effects postoperatively compared to the clinical setting. O’Leary et al. demonstrated that while the ROM due to pure moments in the sagittal plane increased significantly both without and with a 400N follower-load as a result of implanting a Charité TDR, application of the follower-load increased the lordosis angle significantly from 12.6° in the intact condition to 20.7° with the TDR at L5-S1. Similar increases in lordosis at the operative level have been reported clinically, and should, therefore, be regarded as an important aspect of in-vitro testing under a physiological preload.

Combined and asymmetric loading relate to an increased likelihood of injury. However, combined loading is not commonplace in in-vitro spinal testing, which may be due to limitations in testing equipment and the difficulty in comparing different loading protocols. Nevertheless, it has been shown that combined loading behavior cannot be easily predicted from known behavior in individual axes and future testing protocols should account for this in addition to testing axes individually.

Both stiffness and flexibility protocols have limitations in terms of fully replicating in-vivo conditions. Pure moment testing allows test cycles to occur about an unconstrained COR, and although Wilke et al. demonstrated that pure moment testing without a preload replicates qualitative aspects of in-vivo loading, muscle force simulation was recommended to reproduce in-vivo loading more accurately. In-vitro research has provided valuable data by using muscle force simulation to minimize artifact moments and forces, though optimally applying generalized muscle forces in the laboratory is a challenging and complex issue, and simplifications in the application of muscle groups may lead to inaccuracies in replicating the in-vivo environment. Applying complex displacements in six DOF offers an alternative methodology, and current advances in imaging techniques means that the in-vivo kinematics of vertebrae can be obtained dynamically, and in three dimensions. However, there remains a similar difficulty in generalizing such kinematics in the laboratory setting, when large inter-subject variations may occur as a result of degenerative pathology or spinal injury.

The complexity of the spine compared to other joints, such as the hip and knee, means that pathology due to mechanical and/or degenerative factors often occurs in a multi-faceted manner, with a direct mechanism being difficult to determine. This, in turn, increases the difficulty in designing clinically relevant in-vitro studies in a standardized manner. However, a greater understanding of the three dimensional kinematics and loading of the spine in-vivo, and advances in the technology relating to multi-axis testing systems provide great potential for simulating in-vivo biomechanics of the spine in the laboratory more accurately and better than ever before. Indeed, for many cases, including degeneration and disease, it is possible to utilize in-vivo imaging of patient populations to gather more information especially since clinical imaging techniques are similarly improving with better spatial and temporal resolution. Increased collaboration between clinicians, scientists, and engineer...
entists, and engineers provides the opportunity to further develop appropriate standardized testing protocols in relation to the surgical practice, with the aim of improving patient outcomes driving the direction of future research.

Case Studies
Having presented the clinical and biomechanical perspectives and considerations earlier in this review, it is useful to highlight several examples that point to the tight connection between them. Here we briefly present three examples in which the coordinated or iterative efforts between basic science and clinical research would, or will, provide benefit to the ultimate success for patient care. Certainly, these examples are only intended to highlight different connections and disconnects and to provide a thought-provoking perspective.

Anterior Cervical Plate Failure
Anterior cervical plates are used to reinforce anterior cervical constructs during cervical fusion in cases of fusion across an IVD, or in cases of VB replacement. Although surgeons understand the risk of construct failure at the bone-screw interface, and the effect of screw-orientation and plate design on the screw-bone interface has been investigated in-vitro, the potential failure of the plate-screw assembly is less well understood. The ASTM 1717 standard is the pre-clinical testing protocol required for spinal implant constructs, in which the device is secured to standard polyethylene blocks and subjected to three static loading tests and a single cyclic fatigue loading exposure. The scope of the testing standard specifically states that “the results obtained here cannot be used directly to predict in vivo performance. The results can be used to compare different component designs in terms of the relative mechanical parameters.”

There is anecdotal evidence of plate failure at the screw-plate interface, and reported cases of screw failure. Human cadaveric testing, under physiological loading conditions prior to clinical use, may have provided valuable data regarding such failures, and allowed a greater understanding of how they may be avoided.

Dynesys Spinal System
Biomechanical testing protocols are further complicated when non-traditional systems are assessed. Specifically, polycarbonate urethane spacers used in conjunction with titanium pedicle screws and a polyethylene terephthalate tensioning cord in the Dynesys Spinal System (Zimmer Spine Inc., Warsaw, IN) has required a re-analysis of pre-clinical testing systems. In general terms, rigid metal spinal fixation devices do not change their stiffness over time, but biomechanical testing of the Dynesys demonstrated creep in the spacer, and stress relaxation in the tensioning cord, and changes in stiffness with different diameter spacers.

Clinically, the posterior non-rigid fixation systems were designed to function as motion preservation devices, and while initial clinical outcomes were promising, evidence regarding the stiffness and ROM at the instrumented level has led to these devices being re-categorized as posterior dynamic stabilization devices. Further research is required regarding the adjacent level effects compared to traditional posterior stabilization procedures.

Total Disc Replacement
Despite the emergence of TDR procedures as a viable alternative to fusion procedures over 30 years ago, the conclusive evidence is still lacking to suggest relevant improvements in clinical outcomes. Various in-vitro multi-axis studies have compared spinal specimens in the intact condition and after a TDR, but these have generally been completed without fully simulating the physiological situation. Increased lordosis observed clinically after lumbar TDR has also been documented when a 400N follower-load is applied in in-vitro tests. More data describing the performance of TDR devices under such conditions may better inform the scientific community of the efficacy of new devices prior to their clinical use. Likewise the anatomical placement of TDRs has been shown to significantly affect clinical outcomes, and more in-vitro data concerning intra-operative variables would be valuable in determining the sensitivity of a device to adverse conditions.

Recommendations &
Conclusions

The in-vivo spinal environment is difficult to fully replicate in an in-vitro setting, even for short-term testing. This is complicated further by the "real world" variables, such as obesity, poor healing conditions, aberrant spinal loading conditions, and iatrogenic factors that the surgeon faces with each case. Minimum pre-clinical testing standards for new spinal instrumentation are well-established in terms of wear, fatigue, and yield testing standards, but it is not within the scope of these standards to assess the efficacy of a new device, nor to subject a device to the entirety of in-vivo loading conditions and possible clinical scenarios. This disconnect can lead to the failure of commercially available and FDA-approved devices, but can also lead to limited clinical improvement of new devices over existing products over the long-term.

It is important that both communities (basic science and clinical) continue to refine efficacy testing protocols through standardized procedures, so that instrumentation is tested under a variety of physiologically relevant conditions, including “worst-case” scenarios. These conditions should reflect the potential clinical scenarios and possibilities; developments in imaging techniques and post-operative follow-up can also assist in identifying key variables with influence, such as improper instrumentation placement, incomplete instrumentation construction, excessive instrumentation preload, and use in non-approved conditions, for example, with other instrumentation types and/or procedures. Such sensitivity analyses will aid in understanding how certain instrumentation may be better suited to improve clinical results.

It is also critical to continue to assess new devices in terms of their efficacy. Such testing, however, must replicate all aspects of the in-vivo environment, among them the physiological preload. It is paramount, as with all details of laboratory experiments, that the method and magnitude of preload application be clearly documented and justified in order to allow both reliable replication and comparison with other studies, but also to provide appropriate context for interpreting findings with respect to the clinical setting. Furthermore, in-vitro testing should not be limited to assessing ROM, stiffness, or flexibility, but must compare the subtleties of the non-linear behavior of spine in order to provide a more robust translation of spinal biomechanics from the lab to the clinic.

Regardless, as technology continues to advance in both the laboratory and clinical arenas, it is important to continue to obtain increased data in six DOF and under dynamic conditions, in both specimens and human patients. Such efforts will undoubtedly provide a greater understanding of the spine, but also enable coordination between communities to better describe spinal biomechanics and understand the effect of degenerative pathology and treatments. Moreover, such clinical data will provide valuable inputs for in-vitro studies, particularly in relation to quality of motion, and improved laboratory testing conditions will also inform how to better manage spine disorder. In summary, improving the link between multi-axis biomechanical testing in-vitro and imaging studies and treatment of spine conditions will provide greater partnerships, improved translation of in-vivo to in-vitro data, and assist in the iterative development of future spinal devices and improved spine care.

Acknowledgements

The authors gratefully acknowledge the Catherine Sharpe Foundation for supporting this work.

References


95. Holsgrove TP, Gill HS, Miles AW, Gheduzzi S.


116. Gilbertson LG, Doehring TC, Kang JD. New methods to study lumbar spine biomechanics: Delineation of in vitro load-displacement characteristics by using a robotic/UFS testing system with hybrid con-


**Disclosures**

All authors declare no relevant financial disclosures.

**Corresponding author**

Beth Winkelstein, Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 210 South 33rd Street, Room 240 Skirkanich Hall, Philadelphia, PA 19104. winkelst@seas.upenn.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.