Washington State Health Care Authority

Health Technology Assessment

Upright MRI

Effectiveness of upright MRI for evaluation of patients with suspected spinal or extra-spinal joint dysfunction

Date: May 11, 2007

Presented by:

SPECTRUM RESEARCH

Health Technology Assessment Program
P.O. Box 42712
Olympia, WA 98504-2712
http://www.hta.hca.wa.gov

Comprehensive
Evidence-Based Health Technology Assessment:

Effectiveness of upright MRI for evaluation of patients with suspected spinal or extra-spinal joint dysfunction

FINAL REPORT 5/11/2007

Provided by
Spectrum Research, Inc.

Andrea C. Skelly, MPH, PhD
Elya Moore, MS, PhD
Joseph R. Dettori, MPH, PhD

This technology assessment report is based on research conducted by Spectrum Research, Inc., as contracted by the Washington State Health Care Authority. This report is intended to be an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions contained herein are those of the investigators and authors who are responsible for the content. These findings and conclusions may not necessarily represent the views of the HCA/Agency and thus, no statement in this report shall be construed as an official position or policy of the HCA/Agency.

The information in this assessment is intended to assist health care decision makers, clinicians, patients and policy makers in making sound evidence-based decisions that may improve the quality and cost-effectiveness of health care services. **Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.**

Evidence-Based Technology Assessment: Effectiveness of upright MRI for evaluation of patients with suspected spinal or extra-spinal joint dysfunction

Table of contents

Evidence-Based Technology Assessment: Effectiveness of upright MRI for evaluation of patients with suspected spinal or extra-spinal joint dysfunction

I. Introduction

This technology assessment evaluates relevant published research describing the diagnostic accuracy, reliability and effectiveness of positional, standing and upright MRI and compared it with other currently available technologies. This assessment provides a basis for determining policy on the diagnostic value of this technology for those covered by state programs providing health care.

A. Overview of magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) is a relatively new imaging modality that has become widely used for evaluating musculoskeletal abnormalities as well and neurological disease. The basis of MRI imaging relies on the electro-magnetic properties of atom nuclei (specifically hydrogen in water molecules, i.e., protons) which makes them behave like tiny spinning magnets. When they are placed in a strong uniform magnetic field, they weakly align with the external magnetic field. When a short pulse of radio frequency (RF) energy (radio wave) is applied, the nuclei absorb a small amount of energy, change their alignment, and then gradually return to their previous positions, a process referred to as relaxation. This response to the radio wave generates a small electrical signal which can be detected, recorded electronically and used to create a computerized map of the radio signals generated by the human body. The differences between the signal generated by different organs' normal tissue and diseased tissue are assessed by computer to discriminate between normal and abnormal organs and tissues.

The information from MRI is recorded in three planes, the z-axis, y-axis and x axis. The longitudinal relaxation time, or T1, is the measurement of the changes in the z-axis during the relaxation pause. T2, or transverse relaxation time, relates to changes in the x-axis and the y-axis. Each tissue, normal or pathologic, has unique T1 and T2 values for a given MRI field strength. The inherent tissue differences between various T1 and T2 values give the visual

contrast seen between tissues on the MR image. MRI is particularly sensitive in assessing anatomical structures such as bones, organs and soft tissues for the detection and diagnosis of a broad range of pathological conditions such as disc herniation or spinal stenosis. As technology has progressed the resolution of MRI images has greatly improved, allowing for enhanced detection of specific tissues and disease entities with new imaging protocols that go beyond the standard T1 and T2 sequences.

The strength of a magnetic field is measured in a unit called a Tesla (T). The stronger the magnetic field, the greater the number of radio signals which can be elicited from the body's atoms and therefore the higher the quality of MRI images.[1] MRI systems used in human imaging are commercially available up to 3T. Low-field MRI is generally considered to be less than 0.5T with a medium strength MRI between 0.5-1.0T, and a high strength MRI > 1.0T. All of these are used in medical imaging. MRI does not use ionizing radiation.

B. Standard, recumbent MRI

Standard, recumbent MRI (rMRI) typically consists of a cylindrical superconducting coil surrounding the patient to generate a large, static magnetic field; auxiliary coils for generating the magnetic field gradients; radio transmitter/receiver coils in proximity to the patient; electronics for radio-frequency transmission and reception; and a computer to orchestrate the events and to reconstruct the spatial image of the anatomy.[2] The rMRI requires the patient be in a recumbent position, either prone or supine, lying flat and motionless during the imaging for periods ranging from a few seconds to 15 minutes at a time depending on the particular sequence, or type of image, required. Typically, the rMRI machine is a tube into which the patient is placed. There is limited space within the traditional machine. Patients may become claustrophobic or anxious within the machine; moreover, the machine may not be able to accommodate larger patients. Standard, recumbent MRI equipment generally employs a magnet of 1.0T or higher with 3.0T becoming more state-of-the-art.

C. Open MRI

Open and semi-open MRI systems have a variety of configurations wherein the patient is not completely surrounded by the magnet. Instead of a tunnel as with standard rMRI, common configurations are open along the sides and/or consist of a shorter tunnel such that only the portion of the body being imaged is surrounded by the magnet. Some designs have flared ends or two large discs separated by a pillar. Both are open on the sides, allowing for imaging in different patient positions and for axial loading. Some open and semi-open systems are high field systems (>1.0T) which allows for faster imaging an enhanced resolution. Such systems allow for guided interventions. Larger and claustrophobic patients may be more comfortable in open MRI systems. A number of open MRI scanners typically are "low field", < 0.5T, and therefore have lower resolution, smaller fields of view and longer scan times than the traditionally configured MRI scanners. Some low-field, open MRI scanners are dedicated to evaluation of extremities and joints, are almost totally open and allow for positional imaging. Systems with <0.5T magnets are not evaluated in this report.

D. Upright, standing or positional MRI

Upright, standing or positional MRI (uMRI) is a type of vertically open MRI that has been developed in recent years. Such systems are open at the front and top, with the magnetic poles placed on either side of the patient and allow for vertical (upright, weight bearing), horizontal (recumbent) positioning, and dynamic kinetic flexion and extension maneuvers. Current uMRI scanners generally use medium field magnets of 0.5T (e.g., GE Signa™ SP/i) or 0.6T (e.g., FONAR Upright™MRI). The GE Signa™ SP/I appears to marketed for interventional, intra-operative and research use.[3] There appear to be four clinical centers in the United States with the GE system, primarily in research settings. Across the United States, 112 FONAR units have been installed, almost exclusively in stand-alone outpatient clinics and a few in out-patient surgery centers.[4] Currently there is one center in Washington State with an additional center planned. Evaluation of the spine, particularly lumbar spine, appears to be the most frequent request for uMRI (>50% of patient referrals) as described by both FONAR and Upright MRI of Seattle. [4, 5] Evaluation of the extremities is currently not as common and as the spine. Standard CPT codes are apparently used. Total patient volumes

range from 1000 – 5000 per annum in the clinics with whom FONAR is familiar. It does not appear that this technology is currently considered the standard of care nor does it seem to be widely diffused based on this information.

For purposes of this technology assessment, uMRI will be defined any system of 0.5T or greater that allows for scanning in various positions, regardless of manufacturer. By comparison, the most advanced standard rMRI scanners have magnet strength of at least 1.0T and up to 3.0T allowing for the greatest resolution generally in a shorter amount of time. With 0.6T magnets, uMRI requires more time to obtain images with lower resolution. For example, Fonar Upright™ MRI system at 0.6T and gradient field strength of 12 mT/m, has an image time per slice of 0.1sec compared with 178 images per second for a Siemens Magnetom Trio™ system at 3.0T and field strength of 45mT/m with special resolution of 0.5mm for the uMRI compared with 10 micrometers for the 3.0T system.[6] In addition to magnet strength, coil selection and sequence parameters influence image quality. Slower imaging times with uMRI may create difficulty for patients who are unable to remain still while in a standing or sitting position; not comfortable secondary to pain; or are unstable in such positions. Longer exam times may also decrease the overall patient flow and volume of patients that can be accommodated.

The proposed advantages of uMRI are based on the ability to scan the spine (or joints) in different positions (including the position where clinical symptoms are more pronounced) and assess the effects of weight bearing, position and dynamic movement. A summary of some studies evaluating the effects of position and/or loading is presented in Appendix A. It is hypothesized that uMRI scanning in a variety of positions could help elucidate pathology that may be expressed more fully with positional changes or weight bearing. For example, in an evaluation of 50 patients with monosegmental clinical symptoms using a 0.5T open MR imaging system, Vitzthum and colleagues evaluated the mobility of lumbar vertebral bodies in the seated extension and flexion positions. The authors report that 32 images in the flexion and extension positions showed “important additional information” to the diagnosis made

from clinical exam.[7] Whether flexion or extension images would add information to images from rMRI was not tested in this study.

In addition to the lumbar spine, the position of the cervical spine may impact detection of clinical disease or structural abnormalities. Muhle et al., evaluated changes in the size of disc herniations, foraminal size, and cervical cord rotation or displacement in flexion, extension, and axial rotation positions compared with neutral position in 21 subjects. While they reported no change in the size of disc herniation comparing these positions to the neutral position, in five subjects they noted a change in cervical cord rotation or displacement when the subject was in the axial rotation position relative to the neutral position.[8] The same authors did a similar study on 46 patients and found an increase in spinal stenosis in 22 subjects in extension and 11 subjects in flexion versus neutral position. Both studies utilized a 1.5T whole body MR system attached to a patient table.[9] In separate study of 20 patients referred to an MRI center, assuming a flexion or extension position versus a neutral position resulted in a change in anterior compression, or anterior and posterior structures for 18 patients.[10]

Karadimas and colleagues studied 30 subjects with chronic degenerative low back pain who were wait-listed for surgery. They evaluated changes in mean end plate angles and disc height for all lumbar intervertebral levels in the supine position versus the seated neutral position.[11] They utilized a 0.2T open MRI for images in the supine position, and a 0.6T upright scanner for images in the seated position. They also assessed lumbar lordosis. The authors classified discs into four degrees of degeneration (healthy, mild, moderate, or severe). For degenerated discs and healthy discs below degenerated discs, there was a significant reduction in mean end plate angles ranging from -1.7° to -6.8° in the seated position relative to the supine position. The authors reported both increases and decreases in disc height for degenerated discs and healthy discs comparing the supine to the sitting position. There was no clear trend in these changes. Finally, no significant change in lumbar lordosis comparing the two positions was found.[11] The study by Karadimas et al., contributes to a greater

understanding of spinal kinematics; however, it does not address whether uMRI improves diagnosis of disc degeneration compared with rMRI.

Positional changes in the spinal anatomy have been evaluated in healthy volunteers. One study by Hirasawa et al., recruited 29 healthy male subjects to undergo MR imaging of the spine in the supine, standing, and seated (neutral, flexion, and extension) positions.[12] Changes in the mean cross-sectional areas and diameters comparing these positions were reported. The authors found significantly smaller mean dural sac cross-sectional areas at all spinal levels in the supine position versus the upright positions. This percent decrease was as large as 25.4% (supine versus seated extension at the L5/S1 level). Measurements of the mean dural sac diameter showed both increases and decreases comparing the different positions. This study utilized a 0.6T open MRI for all images.

Positional changes in mean cross-sectional area of the neural foramina were measured in smaller study by Schmid et. al. of 12 healthy volunteers.[13] Subjects were examined in the seated extension, flexion, and neutral positions, as well as the supine extension position. There was a statistically significant decrease of 22.4% in the overall (i.e., all spine levels) mean cross-sectional foraminal area in the seated neutral versus supine extension position. An overall significant decrease of 34.9% was seen comparing the seated flexion position to supine extension position. There was no overall significant change in mean cross-sectional area of the neural foramina comparing the upright extension position to the supine extension position. This study utilized a 0.5T open scanner for all images.

E. FDA Regulation

The FDA regulates uMRI systems as Class II Devices under the same classification as standard, recumbent MRI. Systems configured for uMRI have been approved via the 510(k) process and are considered "substantially equivalent" to standard rMRI for purposes of regulation. Two uMRI systems have received FDA clearance, the GE Signa™ SP/i system and the FONAR Upright™ MRI System.

F. Safety considerations

Whether standard, recumbent MRI, low-field open MRI or upright MRI is used, some standard contra-indications apply. The American College of Radiology (ACR) advises that the use of MRI in any form is precluded in patients with cardiac pacemakers, implanted defibrillators, cochlear implants, or ferromagnetic in situ metallic devices or clips. Care must be taken to ensure all ferromagnetic objects are removed from the patients such as watches, jewelry, piercing, etc.[14]

MRI contrast agents such as gadolinium may cause allergic reactions and patient response to contrast injection must be monitored. Gadolinium has also been found to cause acute renal failure in patients with preexistent renal disease, though it is much less toxic than iodinated contrast used in other imaging modalities such as computed tomography (CT). Prophylactic measures should be undertaken to prevent renal failure in susceptible patients.[15]

Failure to follow safety guidelines or use of outdated or inappropriate information on biomedical implants and devices accounts for most reported cases of MRI-related injury and death.[16] A review of biological effects, safety and patient care can provide the interested reader with additional information on these aspects of MRI.[16]

G. Overview of other relevant imaging technologies for the spine

Standard rMRI has commonly been used, when indicated, for the evaluation of low back pain (LBP) and various spinal disorders of the cervical, thoracic and lumbar spine. (See description of clinical guidelines below.) In theory, uMRI may elucidate a cause of back or neck pain as it allows for imaging in various positions and during weight bearing. However, there are a number of other imaging modalities that may or may not be considered competing technologies in the evaluation and treatment of back or neck pain including standard, rMRI with axial loading, CT-myelogram, flexion/extension radiographs, and discography. The role of uMRI and these other imaging modalities in the evaluation of LBP in particular is described below. The imaging modality of choice may be influenced by the patient's presenting symptoms, the time course of back pain and suspected abnormality.

1. Standard recumbent MRI with axial loading

Standard rMRI with axial loading is a method of recreating the effect of weight bearing while the patient is recumbent in the rMRI scanner. One method of axial loading is the use of a nonmagnetic vest placed on the chest/shoulders with straps that are then connected to footplates (DynaWell L-spine; DynaWell Int. AB, Billdal, Sweden). When the straps are tightened an axial load is created on the patient's spine, simulating standing.

Several clinical trials have attempted to assess axial loading in the recumbent position for a few spinal conditions. Willen and Danielson assessed 84 patients with sciatica or neurogenic claudication during CT myelography or rMRI in both the axially loaded (40% of the subject's body weight axially loaded in the supine, knee extended position with the assistance of a compression device) and the psoas relaxed supine positions. Dural sac cross-sectional area was measured. In 29 patients (35%), the image during axial loading revealed a relative or absolute spinal stenosis status compared with the psoas relaxed position , going from above 100mm^2 to below 100mm^2 (relative central stenosis) or from above 75 mm^2 to below 75 mm^2 (absolute central stenosis) at one or more sites (L2-S1). In 30 patients, there was a deformation of the dural sac during axial loading that was not there in during the MRI in the psoas relaxed position. In 11 patients (13%), there was a narrowing of the lateral recess. The authors concluded that a dynamic examination of the lumbar spine should be performed when the cross-sectional area of the dural sac is below 130 mm^2 in the psoas relaxed supine position or when there is suspected lateral recess stenosis.[17, 18]

Manenti et al., studied spinal stenosis, disc protrusion or herniation, and spondylolisthesis under axial loading conditions. After loading, spinal stenosis was seen in 18 patients (36%), discal protrusions in 10 patients (20%) and spondylolisthesis in six patients (12%).[19]

Hiwatashi and colleagues examined the effect of axial loading on treatment decisions in patients with spinal stenosis. From among a group of 200 patients with symptoms of spinal stenosis who underwent regular rMRI and axially loaded rMRI, 20 (10%) had

narrowing of the spinal canal identified on the axially loaded images. For these 20 patients, the clinical exam and the rMR images were then shown to three experienced neurosurgeons who were asked to recommend a course of treatment. Following the treatment decision, the neurosurgeons were given the axially loaded rMR image and asked to render a second treatment decision. After seeing the axially loaded rMR images, all three surgeons changed their treatment decision from conservative management to decompressive surgery for five patients, two surgeons changed their minds for two patients, and one surgeon's mind was changed for three patients.[20]

Clinical trials suggest that axially loading may effect the kinematics of the spine as viewed by imaging exams such as MRI and CT-myelogram. However, while the studies may suggest a possible role for imaging in the axially loaded spine in the recumbent or upright position, they do not address the impact of loading on diagnostic accuracy or therapeutic impact. The cost of axial loading standard rMRI is likely to be less than the cost related to acquiring a dedicated uMRI scanner.

2. Myelogram and CT-myelogram

A myelogram is an invasive procedure that requires a lumbar puncture for injection of dye into the spinal canal and around the nerve roots. It can assist with the detection of disc herniation but lacks diagnostic specificity when used alone and it is not able to detect far lateral disc herniations.[21] The addition of computed tomography (CT) enhances the diagnostic process. CT-myelogram is the term used when the myelogram is immediately followed by CT scanning. The contrast provides an outline of the soft tissue structures (spinal cord and nerve roots) that are otherwise not well seen on CT scan. Herniated discs and spinal stenosis may be well seen with this method. After injection of the dye, the patient is placed in a recumbent position for the CT scan. Compared with uMRI, CT-myelogram is unable to provide images taken while weight bearing or dynamic positions because the patient must be recumbent in the CT scanner. It is an invasive procedure requiring physician presence and a number of side effects are possible. CT-myelogram, however, requires only a standard CT scanner which may be more widely available than

uMRI or even rMRI. In addition, CT-myelogram is an alternative when MRI scanning is contraindicated, such as in patients with pacemakers or MRI incompatible metallic implants.

3. Plain and flexion/extension radiographs

Plain radiographs are not typically required during the first four weeks unless patient presentation includes trauma or suspicion of tumor or infection.[21, 22] Vertebral bodies, facet joints, disc space and the intervertebral foramen can been well seen in lateral views. Spondylolisthesis (slippage of a vertebra) of can be evaluated in lateral and oblique views.[21] Soft tissue visualization is not required for evaluation of spondylolisthesis. Flexion/extension radiographs are plain x-rays of the lumbar spine taken of patients while standing in alternately in flexion and in extension. This technique can detect lumbar bony instability by showing differential movement of lumbar segments. This technique is useful for assessing ligamentous and bony abnormalities in the axial plane including instability but is very poor in detecting soft tissue abnormalities.[21] Radiographs expose the patient to ionizing radiation. Costs for such radiographs are around $100 or less.

4. Discography

Discography is an invasive technique that involved injection of contrast material in to a dic to assess its internal structure and how pressure changes are tolerated. It is not considered a primary diagnostic tool.[21] The technique involves fluoroscopic placement of a 22- to 25-gauge needle into the intervertebral disc (IVD) and subsequent injection of 1 to 3 mL of contrast media and carries the risk of infection and neural injury. The intradiscal pressure is recorded, and an assessment of the patient's pain response to the injection is monitored. A morphologically normal disc will rarely produce pain after injection. The procedure may be used in patients who have failed conservative management and have had negative or equivocal standard noninvasive imaging such as plain radiographs, CT or MRI. A recent systematic review indicates that there was strong evidence for discography's utility as an imaging tool and for invoking pain. While the evidence supporting discography's role in identification of patients with chronic lumbar discogenic pain was considered strong, its role in identifying patients with chronic

cervical discogenic pain was only moderate and its role in identifying patients with chronic thoracic discogenic pain limited.[23]

H. Overview of extra-spinal joint MRI evaluation

Both uMRI and rMRI have been used for the evaluation of extra-spinal joint disorders such as patellofemoral pain and shoulder impingement syndrome. In theory, joint abnormalities may be more evident in the upright or weight-bearing position. Open uMRI systems, and weight-bearing techniques can be used for evaluating the knee, shoulder, wrist, and other extra-spinal joints.[24] The specifications of these techniques depend on the area imaged and goal of imaging. For example, imaging of the ankle requires very little imaging space, either open or closed, and high resolution views. A positioning device may be employed to guide the position of the ankle (Captain Plastic, Seattle, WA). Conversely imaging of the shoulder, or glenohumeral joint, may be evaluated in a large, vertically opened MRI system (0.5 or 0.6T) either in the upright or recumbent position. Different positions for imaging of the shoulder may include extension or flexion, abduction or adduction, and internal or external rotation. Recent advances in technology facilitate a physical examination of the shoulder guided by the MR system. Imaging of the knee, or patellofemoral joint, may be done in a conventional system in the axially-loaded supine position. Low-field open systems allow upright images with the knee in extension or flexion, or while the patient is squatting.[24]

I. Evaluation of new diagnostic tests --general concepts

Studies evaluating diagnostic tests are essential to guiding decisions pertaining to clinical care, however if done improperly they are subject to bias, and potentially misleading findings. In particular, if a diagnostic test is evaluated in a known diseased population, and then separately in a group of healthy normals, substantial overestimations of test accuracy may be found. The same may occur if consecutive enrollment of symptomatic patients is not employed, or if different reference tests were used within the same study.[25] This important issue is discussed in more detail in Appendices C and H.

In addition to attention to methodological rigor, evaluation of any new diagnostic modality requires first the establishment of the technical feasibility, including assessment of its reliability (reproducibility) and the precision with which it can correctly classify disease or characteristics of disease (validity). The validation of diagnostic accuracy requires that the test be compared with an appropriate "gold standard" in a broad population. It is necessary to evaluate the performance of the test across the range of clinical populations in which the test may be employed. Use of standardized common protocols across studies facilitates comparison between study populations. Evaluation of the diagnostic and therapeutic impacts (and thus the clinical utility) of a modality requires consideration of test performance in those who truly do and do not have the condition or disease (i.e., diagnostic accuracy) as well as the outcomes relevant to those who test positive and those who test negative. In the evaluation of a new technology, the primary focus should be on findings that are relevant to the further evaluation of the patient and pertinent treatment options. The relevance of "additional findings" from a new diagnostic test must be put in the context of the extent to which meaningful treatment options are available and whether the findings from the new test improve patient outcomes beyond the outcomes achieved using the old diagnostic test. This is difficult in situations where the treatment is suboptimal, such as in the case of spinal fusion for conditions addressed in this technology assessment. In order to fully understand the diagnostic and therapeutic impact in this case, a randomized clinical trial is needed where the "intervention" is the diagnostic test (old versus new), and the standard treatment is applied in both cases.[26] Furthermore, costs and safety associated with new technology need to be weighed against the benefits of the new technology.

J. Importance of this Health Technology Assessment

Conditions of the spine and extra-spinal joints are painful and debilitating, and unfortunately, extremely common. MRI has in many instances, become the diagnostic modality of choice for evaluating suspected causes of such pain. It is hypothesized that uMRI, by obtaining images in the axial-loaded condition, and/or while the patient is in a position which elicits symptoms, may facilitate the diagnosis of various abnormalities that cause the symptoms.

No formal technology assessments, systematic reviews or critiques of evidence quality related to the diagnostic accuracy or reliability of uMRI used in the evaluation of the spine and extra-spinal joints were found in the published, peer-reviewed literature. There are, however, several organizations that have formulated thoughtful assessments of the published scientific literature on uMRI.[3, 27-32] Within Washington State, the Department of Labor and Industries' technology assessment of uMRI included critique of clinical studies of vertically open uMRI on asymptomatic volunteers and patients with various abnormalities of the spine and foot. The primary conclusion of this report was that there were limited data on the accuracy and diagnostic utility of standing, upright, weight-bearing or positional MRI and that there was no evidence from well-designed clinical trials demonstrating the accuracy or effectiveness of this technology for specific conditions or patient populations.[31] In general, these conclusions are echoed in the other organizational reports.[27-30, 32] These reports have been used to inform coverage policies and are further described in the results section (Section V) of this technology assessment.

Spectrum's technology assessment updates the literature described in the above assessments and provides an independent, in-depth, formal evaluation of the strength of evidence for the accuracy and reliability of uMRI relative to other modalities currently used for diagnosing specific cervical, spinal, and extra-spinal conditions. It is based on systematic review of the published, peer reviewed scientific literature and methodological precepts described by the Agency for Healthcare Research and Quality (AHRQ). It provides a basis for determining policy on the diagnostic value of this technology for individuals covered by state programs providing or paying for health care.

II. Background

A. Lumbar spinal conditions

1. Overview

Low back pain refers to spinal and paraspinal pain in the lumbosacral region of the spine and is very common in the general population. According to the Quebec Task Force on Spinal Disorders, more than 80% of the population experiences some low back pain at some point in their life. The estimated prevalence of low back pain in the United States is approximately 18%, with an annual incidence of 15–20%.[33] Given this information, low back pain is one of the top five most common reasons for visiting a healthcare provider.[34, 35] Consequently, the economic cost of LBP is considerable, accounting for the largest economic burden with respect to annual costs totaling more than 50 billion dollars.[35]

LBP can originate in the bony structures of the spine (vertebrae), in the ligaments and tendons surrounding the spine, the muscles of the lower back, the intervertebral disc or from spinal cord/nerve root compression. The differential diagnosis of LBP covers a broad range of mechanical, nonmechanical and visceral conditions. Of the mechanical causes of LBP, the estimated prevalence of lumbar sprain or strain is around 70%, with degenerative processes of the disc and facet around 10% as the next most common and disc herniation (4%), spinal stenosis (3%) and spondylolisthesis (2%) at less than 10%.[36] The exact cause of symptoms is found in only 12–15% of patients.[33] It is important to note, also, that findings from MRI may have a high prevalence among individuals who do not have low back pain but may not be diagnostically or clinically relevant.[37]

LBP is considered acute if the pain last less than four to six weeks and can be secondary to trauma/fractures, intervertebral disc herniation, or lumbar muscle strain among the more common causes. The estimated prevalence for serious causes of acute low back pain is around 3% for spondylolisthesis, 4% for compression fracture and ranges from 1% to 3% for disc herniation.[38] Subacute LBP is up to 12 weeks and beyond 12 weeks is considered chronic.[38] Chronic LBP could be secondary to degenerative changes (spinal stenosis, spondylolisthesis), disc herniation, cancer, or infection among others.

2. Role of imaging in evaluation of LBP

Various diagnostic imaging tests are commonly used to evaluate a range of spinal conditions. Table 1 summarizes some of these commonly used imaging tests with respect to the spinal conditions of interest in this technology assessment. In the matrix below "y" indicates that the condition is commonly evaluated with that modality, "n" indicates that it is not likely to be evaluated with the modality, and +/- means that in certain cases, the diagnostic test is thought to be useful to evaluate the condition:

Table 1. Currently used imaging tests for certain spinal conditions

	Imaging test					
	uMRI	rMRI	CT myelogram	Regular radiographs	Flex/ext radiographs	Discography
Stenosis	y	y	y	n	n	n
Spondylolisthesis	y	y	y	y	y	n
Herniated disc	y	y	y	n	n	y
Instability	y	n	n	n	y	n
Sciatica	n	n	n	n	n	+/-
LBP +/- leg pain	n	n	n	n	n	+/-

LBP = low back pain, uMRI = upright MRI, rMRI = standard recumbent MRI, Disc = discography,

Clinical evaluation of patients presenting with low back pain includes a thorough history with special attention paid to onset of symptoms, trauma history, alleviating and exacerbating maneuvers, and history of constitutional symptoms such as fevers, weight loss or night sweats. The physical exam should focus on the neurological and musculoskeletal exam and symptoms that may indicate possible systemic disease involvement. According to the Institute for Clinical Systems Improvement (ICSI), features of the exam that would prompt urgent evaluation for cauda equina syndrome include sudden onset of loss of bowel/bladder control, bilateral leg weakness or saddle numbness. Exam findings that require evaluation within 24 hours includes fever greater than 100.4°, unrelenting night pain or pain at rest, new onset pain with distal (below the knee) numbness/weakness, leg weakness or progressive neurological weakness.[39]

Uncomplicated LBP, without other symptoms will often resolve with conservative management without the need for imaging or further treatment. The cause of back pain in most patients is considered benign and neurological impairment does not typically

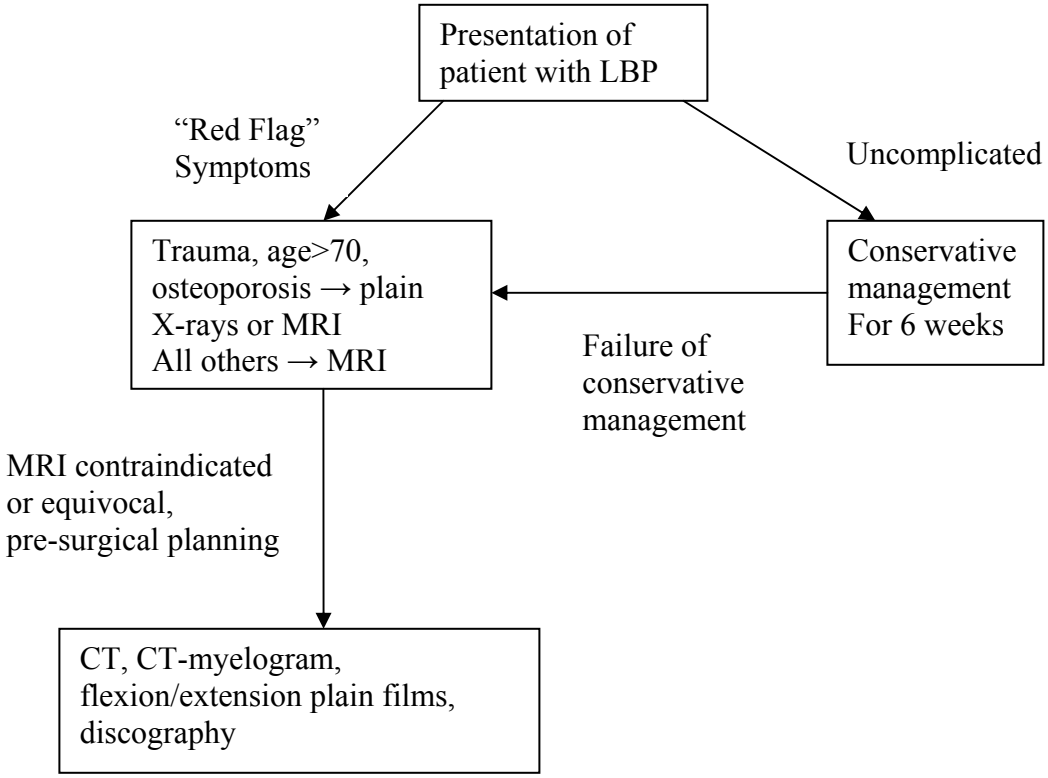
occur.[36] Conservative management includes rest and exercise. Other treatments include analgesics and anti-inflammatory medications.

Based on clinical guidelines (updated in 2005) from the American College of Radiology (ACR), imaging is indicated when pain lasts greater than six weeks with conservative management or when the patient presents with other symptoms that are red flags for more serious pathology.[22] These guidelines appear to be consistent with those from other organizations (e.g., ICSI) as found via the National Guideline Clearinghouse.

Red flags include: recent trauma, mild trauma in patients aged >50, unexplained weight loss, unexplained fever, immunosuppression, history of cancer, intravenous (IV) drug use, prolonged use of corticosteroids, osteoporosis, age >70, focal neurological deficit or progressive or disabling symptoms.[22] The ICSI[39] and an systematic review of diagnostic evaluation of low back pain by Jarvik et al.[36] recommend imaging for patient age>50, with complicated back pain, not just those with a trauma history. According to the ACR, plain films are a sufficient initial imaging modality for those patients with trauma at any age, patients age>70, and patients with osteoporosis, though MRI is considered the most appropriate exam. For patients with other red flag symptoms where there is a suspicion of cancer, infection, immunosuppression, spinal cord/nerve root compression or injury, or ongoing symptoms, MRI is considered the most appropriate initial exam. Plain CT or CT-myelogram is an appropriate exam if MRI is contraindicated. CT, CT myelogram, flexion/extension radiographs, and discography are often undertaken after MRI if MRI is equivocal or for pre-surgical planning. CT can be used as an alternative to MRI as an initial exam in some instances.[36, 39]

Figure 1 describes the general pathway for diagnostic imaging based on the current ACR guidelines.[22] Imaging is not advocated prior to six weeks in patients with uncomplicated back pain.[21, 22, 36] In patients with complicated back pain, the imaging modality of choice will depend on patient presentation and symptoms.

Figure 1. Diagnostic imaging pathway for evaluation of low back pain based on ACR clinical guidelines.



B. Extra-spinal conditions

Role of imaging in evaluation of extra-spinal joints

As is the case with the spine, the type of modality used for clinical diagnosis of extra-spinal joints is dependent on the suspected joint pathology. The American College of Radiology generally agrees that radiograph is the most appropriate for the diagnosis of non traumatic knee pain,[40] chronic wrist pain,[41] shoulder instability,[42] and chronic foot pain.[41] In the case of ankle instability,[43] MRI is suggested.

III. Report purpose, objectives and key question

The primary aim of this assessment is to systematically review and analyze research evidence comparing the use of uMRI with currently available diagnostic tests for the following

musculoskeletal conditions; degenerative spondylolisthesis, spinal or foramenal stenosis, radicular pain, non-specific back pain and extra-spinal joint pain/function loss.

Evaluation of uMRI included description and consideration of the prevalence, incidence and disease burden related to the various conditions as well as the potential role, advantages and disadvantages of uMRI for the conditions specified in the key questions in light of competing and complementary technologies. Relevant clinical guidelines for the use of uMRI were summarized as were relevant Centers for Medicare and Medicaid Services (CMS) policies and those from at least three representative bell weather coverage policies (e.g., Regence, Blue Cross/Blue Shield, Aetna).

A. Specific objectives:

1. Evaluate research describing uMRI test characteristics and ability of uMRI to detect clinically important findings compared with currently available diagnostic methods for specific spinal and musculoskeletal conditions. Included in this objective is consideration of uMRI performance in acute and sub-acute/delayed settings.
2. Evaluate studies describing the extent to which upright/standing MRI may impact clinical decision making related to the need and frequency for further diagnostic testing, as well as the impact it may have on treatment and pertinent outcomes of treatment in the above conditions. Included in this objective is consideration of testing in acute and sub-acute/delayed settings.
3. Evaluate and summarize any formal economic or cost-related studies involving uMRI for evaluation of conditions listed and provide information on anticipated costs of uMRI as available.

4. Identify and describe gaps in current research/evidence and recommend priorities and approaches for further research.

B. Specific key questions addressed.

Key questions were developed by the Washington State Health Technology Assessment Program. A conference call with Spectrum Research and representatives of the HTA program provided clarification of the questions, specification of conditions and appropriate gold standard for diagnostic accuracy studies of uMRI for spinal and joint evaluation.

Each of the key questions was addressed with respect to the following abnormalities/conditions:

1. Suspected degenerative spondylolisthesis (>25% slip)
2. Suspected spinal stenosis (moderate/severe central stenosis (>1/3 canal), lateral recess stenosis (displacing or compressing nerve root, disc extrusion)
3. Radicular pain (moderate /severe central stenosis, lateral recess stenosis, nerve root compression, disc extrusion)
4. Non-specific spine pain (moderate/severe central stenosis, lateral recess stenosis, nerve root compression, disc extrusion)
5. Extra-spinal joint pain/function loss (e.g narrowing, musculoskeletal only)

- Key Question 1:

  What is the evidence to describe the concordance (i.e., ability to detect clinically important findings associated with known conditions) of upright MRI compared with currently available diagnostic testing (e.g., standard MRI +/- loading, CT myelogram+/- upright, plain films [flexion and extension], discography, operative findings) in patients (including appropriate sub-populations) with conditions 1-5 above.

  If a reference standard is available for any of these conditions, what are the test characteristics, PPV (positive predictive value), NPV (negative predictive value),

sensitivity and specificity, of upright MRI compared with standard diagnostic testing?

- Key Question 2:

  What is the evidence to describe the reliability (i.e., test-retest, intra-reader, inter-reader performance) of upright MRI and how does this reliability compare with available diagnostic testing in patients with 1-5?

- Key Question 3:

  What is the evidence to describe the diagnostic impact (i.e., effect on additional diagnostic testing, effect on limiting the differential diagnosis) of upright MRI compared with available diagnosis testing in patients with conditions 1-5?

- Key Question 4:

  What is the evidence to describe the therapeutic and patient impact (i.e., effect on treatments received, efficiency of moving from diagnostic testing to treatment, outcomes [pain, function, adverse events] of test-directed treatment [operative and non-operative]) of upright MRI compared with available diagnostic testing in patients with conditions 1-5, (e.g., what is the likelihood that positive upright MRI findings accurately predicts favorable outcome following test-directed treatment?)

- Key Question 5:

  What is the evidence that upright MRI in the acute setting is more effective (diagnostic and therapeutic impact) than available diagnostic testing in the sub-acute/delayed setting in patients with conditions 1-5?

IV. Methods

Methods used in the development of this technology assessment followed those prescribed by the Cochrane Collaboration for formal systematic review of the literature and tenets described by AHRQ for technology assessments.

A. Data sources and search strategy

A systematic search was conducted to search bibliographic and other databases (e.g., MEDLINE/PubMed, EMBASE, CINHAL, PsycINFO) to identify studies comparing data on the accuracy, reliability and clinical utility of upright, standing MRI (uMRI) to currently available diagnostic tests. An attempt was made to identify all studies comparing uMRI to other diagnostic methods (e.g., systematic reviews, meta-analyses, randomized control trial, cohort, and case-control studies) for the conditions specified in the key questions. Searches were conducted using standard MeSH terms (controlled vocabulary) as well as specific free-text terms and combinations of terms related to the clinical conditions and diagnostic testing methods as well as economic evaluations. A complete description of databases searched, search strategies and listing of search terms used is found in Appendix B. Key articles that were identified from these strategies were explored further by using MEDLINE'S "Related Articles" feature. In addition, bibliographies of all retrieved articles were reviewed.

Databases related to health technology assessment (e.g., International Network of Agencies for Health Technology Assessment (INAHTA)) and evidence-based clinical guidelines were searched as were other potential sources of gray literature. Potentially relevant information was obtained from the website of one upright MRI manufacturer (FONAR) and its links, including presentation of scientific abstracts (e.g., RSNA meetings). Google and the websites of bell-weather payers were searched for coverage policies and supporting information. Appendix B contains more complete detail other databases searched.

Agency-related cost information was obtained from the published assessment of uMRI done by the Washington State Department of Labor and Industries[31] and background material received from the agency for this assessment. Publicly available cost (billing) data focuses on inpatient sources (e.g., CHARS, HCUPS). Washington State does not collect outpatient cost data. Since uMRI is generally done in an outpatient setting, actual cost data from billing information was not available from these public sources. The National Association of Health Data Organizations, a private, not-for-profit organization, was contacted to explore possible sources of outpatient data. The following companies were contacted in an attempt to obtain

general information about costs and billing: FONAR, Ambulatory Services Corporation and Upright MRI of Seattle.

Ideal studies for the evaluation of validity (diagnostic accuracy) and reliability (reproducibility) would contain comparisons of uMRI and currently available modalities to the same "gold standard" which represents the true presence or absence of disease (e.g. surgical findings). For purposes of this TA, upright myelogram in combination with CT-myelogram was considered the "gold" or appropriate reference standard for spine studies. Radiographs, rMRI, findings at surgery and exam under anesthesia were considered appropriate for extra-spinal conditions (e.g. knee, shoulder and ankle conditions).
In the absence of such studies, concordance between currently available technologies was assessed. Appendix B provides an overview of the study selection algorithm used.

B. Inclusion and exclusion criteria

Criteria for inclusion and exclusion were determined *a priori*, based on the goals of this technology assessment and key questions identified.

- Inclusions:

  All studies comparing uMRI with a currently available diagnostic modality in patients with suspected spine-related or joint related conditions were considered for inclusion. In addition all studies explicitly designed to evaluate reliability (e.g., test-retest, etc.) in clinical populations or formal economic analyses (e.g., cost-effectiveness studies) specific to uMRI were considered for inclusion. There was no restriction on publication date. Studies were considered for inclusion if:

  1. They were published in a peer-reviewed journal and written in English and
  2. uMRI was compared with one or more currently available diagnostic modality(ies) for one (or more) of the conditions listed in the key questions, namely:
     - suspected degenerative spondylolisthesis

- suspected spinal stenosis or clinically important lateral recess stenosis
- radicular pain
- non-specific spine pain
- extra-spinal joint pain/function loss (e.g narrowing, musculoskeletal only)

3. They were the basis for a relevant completed dissertation

- Exclusions:

The following types of studies or literature will be excluded:

1. Reviews, editorials, case reports, letters to the editor, commentaries
2. Studies written in languages other than English
3. Studies with fewer than 5 patients
4. Animal, *in vitro* or cadaver studies
5. Meeting abstracts that have not resulted in peer-reviewed publication
6. White papers
7. Unpublished studies
8. Clinical guidelines that do not contain an appropriate evidence-based evaluation
9. Studies of spine trauma or fractures (trauma)
10. Studies of trauma-related fractures involving joints
11. Studies on the use of uMRI for conditions other than those related to the conditions listed previously. Specifically excluded conditions:
    a. Cancer or tumor related evaluations
    b. Visceral or non-mechanical causes of back pain (pelvic organ problems, renal problems, aortic aneurysm, gastrointestinal problems, neoplasia, infection osteochondrosis, Paget's disease)
    c. Inflammatory or rheumatoid arthritis
12. Studies of functional MRI, dynamic or kinematic MRI (e.g ,brain evaluation, perfusion, supine MRI with patients in different positions), low-field MRI, rapid MRI or cine; contrast with MRI, MRI with axial

loading *unless* there is explicit comparison with upright/standing MRI for conditions specified above.

Studies were independently assessed for relevance by two reviewers using standardize procedures and criteria based on the key questions and associated inclusion/exclusion criteria at both the abstract level and the full text levels. Disagreements were negotiated and in cases where reviewer differences were not resolvable based on the discussion of the abstract, the full article was included for review. At the abstract level, studies that didn't compare uMRI with another diagnostic test and/or did not address the specific conditions listed for inclusion were excluded.

Additional exclusions were made after review of full articles if: (See Appendix D)

- uMRI was not compared with another currently available diagnostic method
- the study did not include one or more of the conditions listed in the key questions
- uMRI and its comparator were not used for diagnostic evaluation of one of the conditions

C. Assessment of evidence quality: Level of Evidence (LoE) Rating

Standardized abstraction forms and guidelines were used to determine the Level of Evidence (LoE) for each study included in this assessment. Separate methodologies are used for studies of diagnostic accuracy (i.e., validity) and reliability studies. Details of the LoE methodology are found in Appendix C. Two abstractors independently assessed the LoE at the level of full article review. When discrepancies in assessment occurred, they were resolved by negotiation. Each article chosen for inclusion as a diagnostic accuracy or concordance study was given a LoE rating based on the criteria listed in Table 2. Reliability studies were assessed based on the criteria listed in Table 3.

Table 2. Definitions of the different levels of evidence (LoE) for diagnostic test accuracy and validity studies.

Level	Study type	Criteria
I	Good quality prospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
II	Moderate quality prospective study	• Violation of any one of the criteria for a good quality prospective study (LoE I)
	Good quality retrospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
III	Poor quality prospective study	• Violation of any two or more of the criteria for a good quality prospective study (LoE I)
	Moderate quality retrospective study	• Violation of any one of the criteria for a good quality retrospective study (LoE II)
IV	Poor quality retrospective study	• Violation of any two or more of the criteria for a good quality retrospective study (LoE II)
	Case-Control Study	

Table 3. Definitions of the different levels of evidence (LoE) for reliability studies

Level	Study type	Criteria
I	Good quality study	• Broad spectrum of persons with the expected condition • Adequate description of methods for replication • Blinded performance of tests, measurements or interpretation • Second test/interpretation performed independently of the first
II	Moderate quality	• Violation of any one of the criteria for a good quality study
III	Poor quality study	• Violation of any two of the criteria
IV	Very poor quality study	• Violation of all three of the criteria

D. Assessment of Overall Strength of Evidence- Overall SoE criteria

After the LoE for each individual study was assessed and the body of evidence for a given topic was evaluated, the overall strength of evidence for that topic was determined based on the following definitions, criteria and the likely impact of additional research on the topic.

Table 4. Criteria for meeting each AHRQ Domain

Domain	Definition/Criterion
Quality	• At least 80% of the studies are LoE I or II
Quantity	• There are at least three studies which are adequately powered to answer the study question
Consistency	• Study results would lead to a similar conclusion (similar values, in the same direction) in at least 70% of the studies

Table 5. Criteria for determination of the overall "Strength of Evidence" (SoE)

SoE	Description	Further Research Impact	Domain Criterion Met		
			Quality	Quantity	Consistency
1	**High**	Very unlikely to change confidence in effect estimate	+	+	+
2	**Moderate**	Likely to have an important impact on confidence in estimate and *may* change the estimate	+	-	+
			+	+	-
3	**Low**	Very likely to have an important impact on confidence in estimate and *likely* to change the estimate	+	-	-
			-	+	+
4	**Very Low**	Any effect estimate is uncertain	-	+	-
			-	-	+
			-	-	-

E. Operational definitions

For purposes of this technology assessment, uMRI will be used to denote any vertically open MRI system of 0.5T or greater that allows for imaging while the patient is standing, sitting, weight-bearing or in various positions (e.g. flexion, extension, bending, rotation, etc.) in addition to scanning in the recumbent position regardless of manufacturer. Additional operational definitions for specific terms relevant to the key questions and an expanded glossary of terms are provided in Appendix G.

F. Data abstraction

A standardized approach was used to extract data from each of the included studies by one investigator and cross-checked for accuracy by at least one other investigator. The following information was abstracted from included studies:

Population characteristics (age, sex, ethnicity, etc.), diagnostic/clinical features and categories, eligibility/exclusion criteria, loss to follow-up information, diagnostic criteria and specifications, outcome assessment instrument(s), diagnostic test parameters (if reported), economic analysis components, and results for each primary outcome.

G. Data analysis, synthesis and review

Data from validation (diagnostic accuracy) studies are needed for calculation of test performance characteristics such as sensitivity, specificity and predictive value. The methods for calculating these are presented in Appendix H, which also includes a description of reliability study evaluation.

In the absence of validation studies, the concordance (percent agreement) was determined. Since this calculation does not take into account agreement that may be expected purely by chance, the kappa statistic (κ) was calculated, where there were adequate data, to correct for chance agreement according to the formula described in Appendix H. The limitations of this statistic are also described in Appendix H. Kappa describes the amount by which the observed agreement exceeds what would be expected by chance alone.

H. Peer review process

External peer reviews by experts with clinical expertise and/or methodological/research expertise were done following preliminary analysis and synthesis of the report draft submitted for public comment. Comments were collated and entered into a spreadsheet database. Comments were addressed and incorporated to create the final report as appropriate. A list of peer reviewers, their qualifications and affiliations is in Appendix E. Comments from the public, agency medical directors and other interested parties were also included as appropriate in the final report. A general listing of comments and their disposition is found in Appendix I.

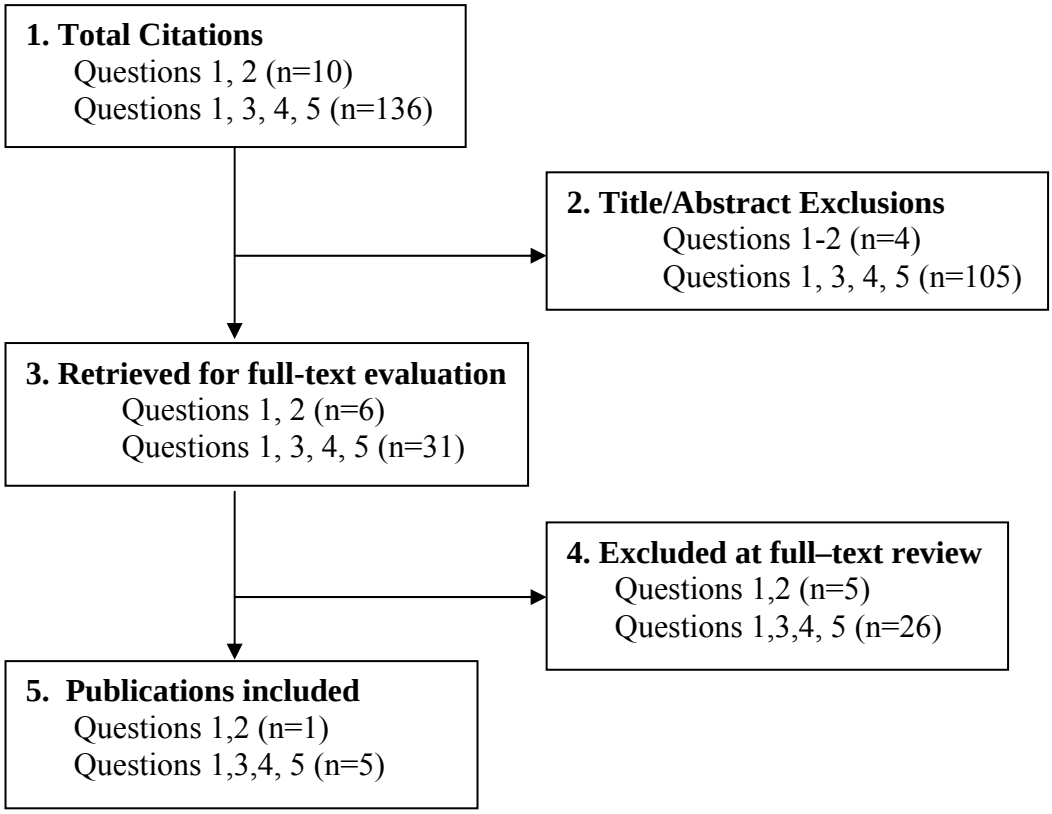
V. Results

A. Literature search and final study selection

Search results and overall quality of the literature comparing upright MRI with currently available diagnostic methods are described below. Detailed search strategy information and results are found in Appendix B. Details of included studies are found in Appendix F. Information on excluded studies is found in Appendix D. General results of the search strategies are outlined in Figure 5.

Our search strategy followed two primary approaches in order to identify studies addressing the key questions. First, we attempted to identify all validity and reliability studies evaluating uMRI, key questions 1 and 2. Of the four potential studies, one was applicable in that it tested inter-observer reliability for identifying foraminal stenosis as imaged by uMRI. The other three studies were excluded as they did not address any key question. Reasons for exclusion are detailed in Appendix D. Second, we broadened our search to identify any study using uMRI. From 26 possible studies, five made some comparison to either rMRI or myelogram. We found no studies that addressed diagnostic or therapeutic impact (key questions 3 and 4), and no studies addressing the impact in the acute or sub-acute/delayed setting (key question 5).

Figure 2. Results of systematic literature search



B. Overview of literature found

Based on systematic review of the literature, few studies met the inclusion criteria (based on the key questions) and quality of literature available to address the questions was poor.

For evaluation of the spine, a total of four studies were found relevant to key question 1 comparing uMRI with other currently available diagnostic methods allowing for evaluation of concordance. No studies comparing uMRI with the stated reference standard, upright myelogram combined with CT-myelogram, were found, thus calculation of test characteristics (e.g., sensitivity, specificity and predictive values) was not appropriate. Only percent agreement and where possible, kappa, were determined.

One of the concordance studies for the spine included a reliability component.

For evaluation of extra-spinal conditions, two concordance studies and no reliability studies were found. Percent agreement and kappa could not be calculated with the data as presented.

For key questions 3-5, no adequate data to determine diagnostic validity or accuracy for uMRI were found, and no specific studies addressing these key questions were identified.

C. Results by key question

1. **Key Question 1:** What is the evidence to describe the concordance of upright MRI compared with currently available diagnostic testing in patients with suspected degenerative spondylolisthesis (>25% slip), spinal stenosis (>1/3 canal), lateral recess stenosis (displacing or compressing nerve root), radicular pain, non-specific spine pain or extra-spinal joint pain/function loss?

SPINAL CONDITIONS

The methodological shortcomings of the four studies which compared uMRI with another currently available technology were considered substantial with each study considered to be the lowest quality of evidence (LoE IV) (Table 6).[44-47] All were retrospective with significant potential for selection bias since enrollment was not consecutive or those included in the study were not a random selection of eligible patients from a relevant clinical population. Only one study appeared to have enrolled a broad spectrum of patients and sought to decrease the possibility of verification bias by performing the comparison test independently from uMRI.[44] Significant potential for verification bias was present in the other three studies. Potential for interpretation bias was present in all studies as none reported blinded interpretation of results from both tests. These biases may have the potential to over-estimate the concordance of uMRI with rMRI (standard rMRI >1.0T or the uMRI in recumbent position, 0.5T) or with myelography in these studies. Sample sizes in three studies were small (≤30).[45-47] Although the fourth study contained 89 patients, the number of patients in some diagnostic groups (e.g., spondylolisthesis) were small.[44] As stated previously, none of

the studies compared uMRI with the appropriate reference standard selected *a priori.* Although percent agreement were calculated for these studies and appear to be high, they should be interpreted cautiously given the methodological shortcomings of these studies.

While each study was considered to have met the criteria for having adequate description of uMRI and the comparative test based on reporting of study population characteristics, equipment specifications and protocols, description of explicit criteria used to diagnose or categorize disease was inadequate particularly in studies by Zamani[47] and Ferreiro Perez,[44] possibly precluding replication of results.

Table 6. Assessment of LoE for individual studies of diagnostic test concordance for evaluation of specific spine conditions

Methodological Principle	Weishaupt (2000)	Zamani (1998)	Ferreiro (2007)	Wildermuth (1998)
Study Design				
Prospective cohort design				
Retrospective cohort design	■	■	■	■
Case-control design				
Broad spectrum of patients with expected condition			■	
Appropriate reference standard used				
Adequate description of test and reference for replication	■	■	■	■
Blinded comparison with appropriate reference				
Reference standard performed independently of test			■	
Evidence Level	**IV**	**IV**	**IV**	**IV**

* Blank box indicates criterion not met, could not be determined or information not reported by author. For more details refer to Appendix F, Table F2.

A summary of concordance findings for the various pathologies evaluated in studies included in this review is found in Table 7.

Disc Pathology

There was similar agreement comparing rMRI with uMRI in identifying disc pathology in both the cervical or lumbosacral spine. In the cervical spine, Ferreiro Perez noted 27 posterior disc herniations in 44 symptomatic patients (61%) on rMRI compared with 31 posterior disc herniations (70%) on uMRI. In the lumbosacral spine, there were 22 (31%) posterior disc herniations seen with the rMRI compared with 24 using uMRI

(45%).[44] Zamani et al., reported complete agreement when comparing rMRI with uMRI in the seated neutral position in the qualitative determination of posterior disc bulge (percent agreement =100%).[47] Likewise, the diagnosis of disc form (normal, bulging, protrusion, or sequestration) was similar when comparing supine neutral with seated flexion and extension positions (percent agreement, 95% and 91%, respectively).[45] It should be noted, however, in this study six of 36 patients (17%) could not finish the seated uMRI exam as a result of severe pain in that position.

Forminal Stenosis

Percent agreement comparing rMRI with uMRI for the detection of foraminal stenosis ranged from 84% to 100%. Zamani et al., reported complete agreement when comparing rMRI with uMRI in the seated neutral position in the qualitative determination of foraminal size (percent agreement, 100%).[47] Likewise, the evaluation of foraminal stenosis (graded as normal, slight foraminal stenosis, marked foraminal stenosis, and advanced stenosis) showed agreement when comparing the seated flexed and extended position with the supine neutral position (percent agreement, 84% and 86%, respectively).[45] Agreement was also seen in the score of foraminal stenosis in supine neutral, extension and flexion position comparing uMRI to myelography (percent agreement, 94% and 92%, respectively).[46]

Nerve Root Compromise

Nerve root compromise (graded as no compromise, contact without deviation, nerve root deviation, or nerve root compression) showed agreement when comparing seated flexion and extension with supine neutral positions (percent agreement, 74% and 77%, respectively).[45] In a study comparing uMRI with myelography, there was a substantial concordance (fourteen out of fifteen correlation coefficients ≥ 0.90) in mean sagittal diameter of the dural sac within each position (supine neutral, extension and flexion) at five separate intervertebral spaces.[46]

Spondylolisthesis

One paper evaluated lumbosacral spondylolisthesis in the recumbent and upright positions. The rMRI identified anterior spondylolisthesis seven times (15%) compared with uMRI which identified spondylolisthesis four additional times (n=11, 24%). Percent agreement comparing rMRI with uMRI was 91%.[44]

Table 7. Summary of studies measuring concordance between uMRI and currently available diagnostic testing in spinal conditions.

	Condition	Imaging Comparison	Percent agreement	LoE
Cervical Spine				
	Posterior focal disc herniation Ferreiro Perez et al. 2007[48]	rMRI vs. seated neutral MRI	DNA*	IV
Lumbosacral Spine				
Disc pathology	Posterior focal disc herniation Ferreiro Perez et al. 2007[48]	rMRI† vs. seated neutral MRI	DNA*	IV
	Disc Form Weishaupt et al. 2000[45]	rMRI† vs. seated flexion MRI	94.7	IV
		rMRI† vs. seated extension MRI	90.8	IV
	Posterior disc bulge Zamani et al. 1998[47]	rMRI‡ vs. seated neutral MRI	100	IV
Foraminal stenosis	Foraminal stenosis Weishaupt et al. 2000[45]	rMRI† vs. seated flexion MRI	84.2	IV
		rMRI† vs. seated extension MRI	85.5	IV
	Foraminal size Zamani et al. 1998[47]	rMRI‡ vs. seated flexion MRI	100	IV
	Foraminal stenosis score Wildermuth et al. 1998[46]	myelogram vs. seated flexion MRI	94	IV
		myelogram vs. seated extension MRI	92	IV
	Nerve root compromise Weishaupt et al. 2000[45]	rMRI† vs. seated flexion MRI	73.7	IV
		rMRI† vs. seated extension MRI	77.6	IV
	Spondylolisthesis Ferreiro Perez et al. 2007[48]	rMRI† vs. seated neutral MRI	DNA*	IV

*Data not available to calculate
†Recumbent supine neutral or extended ((knees slightly flexed or knees extended) not described
‡Recumbent supine neutral (knees slightly flexed)

Summary of results of studies of the spine

There is limited evidence (overall strength of evidence is low) to suggest that uMRI provides similar diagnostic information compared with rMRI with respect to disc pathology and foraminal stenosis of the lumbar spine. There is no evidence yet available to determine test characteristics of sensitivity, specificity and predictive values or likelihood ratios with the gold standard of upright myelogram combined with CT myelogram. The evidence for concordance between rMRI and uMRI is very low with respect to cervical disc herniation, lumbar nerve root compromise, and spondylolisthesis. There is some indication that the seated position in uMRI is not tolerated well in a number of patients with lumbar pathology, preventing the imaging exam from proceeding.

EXTRA-SPINAL CONDITIONS

The two included studies comparing uMRI with another method for evaluation of extraspinal conditions were somewhat stronger methodologically (LoE II and III, Table 8)[49, 50] than those for the spine in that they were prospective and one employed an appropriate reference standard. However both suffered from small sample sizes (< 20 patients). Each had other methodological shortcomings described below. A summary of concordance findings for these studies is given in Table 9.

Table 8. Assessment of LoE for individual studies of diagnostic test concordance for evaluation of specific extraspinal conditions

Methodological Principle	Weishaupt (2003)	Hodge (2001)
Study Design		
Prospective cohort design	■	■
Retrospective cohort design		
Case-control design		
Broad spectrum of patients with expected condition		■
Appropriate reference standard used	■	■
Adequate description of test and reference for replication	■	■
Blinded comparison with appropriate reference	■	
Reference standard performed independently of test		■
Level of Evidence	**III**	**II**

* Blank box indicates criterion not met or could not be determined or information not reported by author. For more details refer to Appendix F, Table F3.

Morton neuroma

One study assessed the effect of prone, supine, and upright weight-bearing positions on the visibility of Morton neuroma. Visibility was rated as good (good delineation of the Morton neuroma against the surrounding tissue, no motion artifacts), moderate (radiologist's judgment between good and poor visibility), poor (poor delineation of the Morton neuroma or severe motion artifacts) or none (Morton neuroma not visible). The prone, non-weight-bearing position was judged the best, showing good visibility of all 20 neuromas under study. Next was the supine, non-weight-bearing position, yielding the visibility rating of good for 60% of neuromas. Last was the weight-bearing position, in which only 50% of neuromas were given the good rating.[50]

The usefulness of this study is minimized by the fact that it included only those patients with a diagnosis of Morton neuroma as determined by the presence of a neuroma with a transverse diameter of 5 mm or larger by rMRI in the prone position. Visibility of the neuroma is presumed to be good in the prone position given that the inclusion criteria required such. Therefore, including only patients diagnosed by rMRI in the prone position prevents a valid comparison of visibility with other positions for this diagnosis. Furthermore, this study does not compare uMRI or rMRI with the referent standard (surgical confirmation) for the presence or absence of Morton neuroma.

Glenohumeral instability

Glenohumeral stability in one study was compared using uMRI with clinical exam under anesthesia (EUA), the referent standard to evaluate this condition. Patients were examined in the uMRI in a seated position. Grading of instability was determined by measuring the distance of shift of the humeral head from the glenoid center toward the rim. A shift less than 25% of the distance was assigned a Grade 1; a shift between 25% and 50%, a Grade 2; a shift greater than 50%, a Grade 3. Instability assessed by EUA was defined as Grade 0 (no translation), Grade 1 (mild translation of 0-1 cm), Grade 2 (moderate translation of 1-2 cm) and Grade 3 (severe translation, greater than 2 cm).

Grading of instability using uMRI underestimated instability compared with EUA in 7 of the 10 cases (70%) that underwent surgical repair.

A small sample size (N=10) prevents formal analysis of the diagnostic characteristics of uMRI during clinical exam compared with EUA.[49]

Table 9. Summary of studies measuring concordance between uMRI and currently available diagnostic testing in extra-spinal conditions.

Condition	Imaging Comparison	Percent agreement	LoE
Foot neuroma visibility score Weishaupt et al. 2003[50]	rMRI* vs. standing MRI	50	III
Shoulder instability Hodge et al. 2001[49]	Exam under anesthesia vs. seated MRI	30	II

*Prone position

Summary of results of studies of extra-spinal conditions

- Based on the extra-spinal joint conditions examined for this report, there is no evidence to suggest uMRI images contribute additional information to the identification of Morton neuroma or shoulder instability compared with existing diagnostic tests.
- No studies evaluating the diagnostic ability of uMRI in the evaluation of the hip, knee or ankle were found.

2. **Key Question 2:** What is the evidence to describe the reliability (i.e test-retest, intra-reader, inter-reader performance) of upright MRI and how does this reliability compare with available diagnostic testing in patients with suspected degenerative spondylolisthesis (>25% slip), spinal stenosis (>1/3 canal), lateral recess stenosis (displacing or compressing nerve root), radicular pain, non-specific spine pain or extra-spinal joint pain/function loss?

SPINAL CONDITIONS

One study (LoE II, Table 10) was identified that assessed the inter-observer reliability of uMRI in diagnosing lumbar foraminal stenosis.[46] In this study, foraminal stenosis (graded as normal, slight foraminal stenosis, marked foraminal stenosis, and advanced stenosis) was determined independently by two observers from uMRI images obtained with the patient in seated flexion and extension (Table 11). The kappa statistic calculated and reported by the authors was 0.62 (substantial agreement as defined by Landis and Koch[51]). However, it is not clear to which position (flexion or extension) this kappa applies. The value of these reliability results is limited by narrow spectrum of patients included in this study (those whose condition warranted a myelogram).

Table 10 Assessment of level of evidence (LoE) for reliability studies on uMRI for specific spine conditions*

Methodological Principle	Wildermuth (1998)
Broad spectrum of patients with expected condition	
Adequate description of methods for replication	■
Blinded comparison of tests/interpretations (interrater)	■
Evidence Level	**II**

* Blank box indicates criterion not met, could not be determined or information not reported by author

Table 11. Summary of studies measuring reliability for uMRI.

Condition	Observers	Kappa	LoE
Foraminal stenosis score Wildermuth et al. 1998[46]	Radiologists (n=2)	0.62	II

Summary of results of studies of the spine

There is a suggestion from one study that lumbar foraminal stenosis can be determined reliably between radiologists when seated uMRI is performed in patients whose symptoms are severe enough to warrant a myelogram. The extent to which these findings may extend to populations with different levels of stenosis severity is unknown. There is no evidence that uMRI is reliable in detecting degenerative spondylolisthesis, lateral recess stenosis, radicular pain or non-specific spine pain.

EXTRA-SPINAL CONDITIONS

There were no reliability studies evaluating extra-spinal joint conditions.

3. Key Questions 3, 4 and 5:

No published reports were found that address the diagnostic or therapeutic impact of uMRI on spinal or extra-spinal conditions overall or with respect to specific evaluation of acute or subacute/delayed conditions.

No studies of uMRI that speak to the issue of diagnostic or therapeutic impact were found.

D. Cost impact

Economic analyses and cost information sources

No peer-reviewed economic analyses of uMRI were found. No public sources of outpatient MRI costs or billing were found. General cost information was sought from FONAR, Ambulatory Services Corporation and Upright MRI of Seattle. Payer organization websites were surveyed for coverage and reimbursement information. Washington State agencies, Labor and Industries, submitted cost experience.

Unit and services charges

Anecdotal information from FONAR indicates that the uMRI system generally costs $1.55 million.[4] FONAR's two clinics and Washington's one clinic charge the global fee for basic uMRI of the spine in the neutral position, ranging from $1365 (New York) to $1600 (Florida)[4] to $1650 (Washington).[5] Additional extension or flexion views range from $350[4] to $1200[5] each. Charges submitted may also vary across clinics for both the basic exam and additional views.[4]

Washington agency reimbursement

The Washington State Health Care Authority, Uniform Medical Plan (UMP), currently does not cover upright or positional MRI because it is considered experimental and investigational after an internal evidence review. Total MRI spend for UMP for 2006 was $11 million, with

a billed amount of over $20 million. Prior to current coverage policy, in 2006 UMP did receive and pay for approximately 46 upright MRI claims paid at standard imaging rates, but did not pay claims for the additional views. The allowed amount for each service averaged $750; with a billed charge for one uMRI ranges from $1300-1650.

In 2006, Washington State Department of Labor and Industries (LNI) completed an evidence-based technology assessment of uMRI. Based on this review LNI does not cover standing, weight-bearing or positional MRI (effective July 1, 2006). Prior to this policy, LNI received and paid approximately 111 imaging claims for uMRI. The average number of positional, uMRI scans, per patient, completed and billed to LNI was 2.5.[31] In 2005, total costs for MR imaging of the spine on all LNI claims exceeded $10 million. Based on this claims experience, LNI estimated that uMRI could significantly increase total costs of MR imaging of the spine if widely adopted.

The Washington State Department of Social and Health Services (DSHS) currently has no coverage policy explicit to upright or positional MRI. DSHS does cover MR imaging and because upright MRI has no separate billing code, payment for imaging is paid under a "By Report indicator" (BR) policy in the Physician billing instructions. When services are verified as a BR, DSHS would pay at 45% of billed charges.

E. Payer coverage policies

CMS has no explicit coverage policy for upright or open MRI. Anecdotal information from manufacturers indicates the global fee in neutral position is within CMS's usual and customary charge and is reimbursable by CMS under standard MRI CPT billing codes.[4] Additional extension or flexion views billed under the miscellaneous CPT codes are not reimbursed by CMS.

As of April 20, 2007, the following third-party payers' policies do not include coverage of uMRI for evaluation of spine or extra-spinal conditions, based on lack of published data on diagnostic accuracy and efficacy: Regence, Premera/Blue Cross, and Cigna.[27-29] There may

be exceptions on a case-by case basis according to representatives from FONAR[4] and Upright MRI of Seattle[5]. Aetna considers any form of "open" MRI, including units which allow imaging in standing or sitting positions to be an acceptable alternative to standard rMRI. They do, however, consider scans in different positions (i.e., flexion, extension, rotation and lateral bending) to be investigational and do not cover these.[30] Anecdotally, "no fault" insurance providers (e.g., State Farm) generally do cover the additional views.[4] It appears that existing CPT codes for MRI are silent with regard to patient positioning. The few payers that cover this use miscellaneous or unlisted magnetic resonance procedure codes. This practice confounds the ability to gather data on cost and coverage. Below is a summary of payers and their stated policies and rationale:

Table 12. Summary of payers, stated policies and rationale

Payer	Policy	Stated Rationale
CMS[52]	• There is no National Coverage Determination about uMRI.	• No specific statement on uMRI. It may not be differentiated from standard rMRI for single images.
Premera-Blue Cross[28]	• MRI that is vertical, upright, positional, or dynamic is considered investigational	• There is a lack of evidence in published, peer-reviewed clinical studies
Regence[27]	• Positional or upright MRI for the diagnosis and management of any condition, including, but not limited to, cervical, thoracic or lumbosacral back pain is considered investigational	• Given the lack of data concerning analytical and clinical validity and clinical utility, it is not possible to reach conclusions regarding health outcomes and effects of positional MRI imaging in the diagnosis and management of patients with cervical thoracic or lumbosacral back pain.
Aetna[30]	• "Open" MRI units of any configuration, including units that allow imaging when standing (Stand-Up MRI) or when sitting, to be an acceptable alternative to standard "closed" MRI • Repeat MRI scans in different positions (such as flexion, extension, rotation and lateral bending) are considered to be experimental and investigational.	• The clinical value of standing MRI or position MRI imaging (e.g. flexion, extension, rotation and lateral bending) has not been systematically evaluated in clinical studies. • It has not been demonstrated in published prospective clinical studies that performing MRI in these various positions can consistently detect problems that cannot be detected with a standard MRI.
Cigna[29]	• Covers low-field MRI as medically necessary when used as guidance during interventional and intraoperative procedures. • Does not cover low-field strength MRI for ANY other indication because it is considered experimental, investigational or unproven. Cigna includes systems with magnet strengths of <1T as "low field" in the review for this policy	• Knee and shoulder: There is a lack of data clarifying what role low-field imaging should hold in the diagnostic algorithm of knee and shoulder injuries • Spine: The few small studies on positional MRI do not address the relevance, value or impact of positional MRI in the diagnosis, treatment or outcomes of patients with neck or back pain.
Washington State LNI[31]	• Due to lack of evidence addressing the diagnostic accuracy or diagnostic utility of standing upright, weight-bearing or positional magnetic resonance imaging is considered investigational and experimental	• There is limited scientific data available on the accuracy and diagnostic utility of standing, uparight weight-bearling or positional MRI.

F. Evidence-based clinical guidelines for uMRI

Search of the National Guideline Clearing House did not reveal specific clinical guidelines for the use of uMRI for the evaluation of back pain or extra-spinal conditions in general or for evaluation of the specific conditions delineated in the key questions. General guidelines for the use of imaging in evaluation of back pain and extremities do not specifically describe the use of upright MRI.

VI. Summary and conclusions: Evidence-based bottom line (Table 13)

There is a paucity of literature to validate its use as diagnostic tool or its reliability for diagnosing such conditions.

Since no studies validating the diagnostic accuracy of uMRI were found, estimates of concordance between uMRI and currently available diagnostic methods were reported. Concordance (or kappa if calculated) should not be interpreted as a measure of diagnostic accuracy, since the comparisons made were not against a "gold" or appropriate reference standard. While the results from individual studies described above seem to suggest that the agreement between uMRI and other available diagnostic methods may be high, the estimates of concordance from these studies may not be reflective of the true concordance and may not be stable estimates. The overall strength of evidence (SoE) reflects the level of confidence that the estimates from the studies on a given topic are likely to change with further research. The SoE is based on the following domains; the quality of the studies (individual study LoE), the number of studies that assess the same topic and the consistency of the estimates across studies of a specific topic. (Tables 4 and 5). Confidence in the stability of the estimate as further research is done is described by the following:

- **High**= Very unlikely to change confidence in effect estimate;
- **Moderate** = Likely to have an important impact on confidence in estimate and may change the estimate;
- **Low** = Very likely to have an important impact on confidence in estimate and likely to change the estimate;
- **Very Low** = Any effect estimate is uncertain

Based on these domains, confidence that the estimates will remain stable with future research is low to very low as described below for individual topic areas:

- There is limited evidence **(overall strength of evidence, SoE, is low**) to suggest that uMRI provides similar diagnostic information compared with rMRI with respect to demonstration of disc pathology and foraminal stenosis of the lumbar spine. There is no evidence yet available to determine test characteristics of sensitivity, specificity and predictive values or likelihood ratios based on comparison with the gold standard of upright myelogram in combination with CT-myelogram set *a priori.*

- The overall strength of evidence for estimates of concordance between rMRI and uMRI is **very low** with respect to cervical disc herniation, lumbar nerve root compromise, and spondylolisthesis. Again, no validation studies were found precluding determination of diagnostic test characteristics.

- Lumbar foraminal stenosis may be determined reliably between radiologists when seated uMRI is performed in patients whose symptoms are severe enough to warrant a myelogram **(overall strength of evidence is low**). The extent to which this reliability extends to other populations with a range of foraminal stenosis severity is unknown. There is no evidence that uMRI is reliable in detecting degenerative spondylolisthesis, lateral recess stenosis, radicular pain or non-specific spine pain.

- Based on the extra-spinal joint conditions examined in this report, there is no evidence to suggest uMRI images contribute towards the identification of Morton neuroma or shoulder instability compared with existing diagnostic tests.

There are no studies that assess the possible draw backs and benefits of uMRI or that assess a broader range of extra-spinal joint conditions. Based on this systematic review, additional conclusions are outlined below:

- There is some indication that the upright position (whether seated or standing) for uMRI exams is not tolerated well in a number of patients with lumbar pathology, preventing the continuation of the imaging exam.

- No studies evaluating the diagnostic ability of uMRI for the evaluation of the hip, knee or ankle were found or did not meet the inclusion criteria.

- There are no reports that address the diagnostic or therapeutic impact of uMRI on spinal or extra-spinal conditions overall or with respect to specific evaluation of acute or sub-acute/delayed conditions. Diagnostic accuracy and outcomes data are important and needed for such evaluations.

- No evidence-based clinical guidelines specific to the use of uMRI were found.

With regard to the potential economic impact of uMRI, there are no formal economic studies in the peer-reviewed literature and evaluation of costs may be challenging.

- It may be difficult to evaluate the cost impact of uMRI as standard CPT codes for basic MRI evaluation of the spine or extremities are used. Based on anecdotal information, the highest paying CPT codes appear to be most frequently used.

- The majority of exams are done in stand-alone out-patient clinics. Since publicly available data are comprised of hospital/inpatient data and some stand-alone surgery center data and only a few states report out-patient data, other sources of data would be needed to evaluate the cost impact of uMRI.

- The coverage of additional views (example extension or flexion views) appears to be an important policy issue, perhaps more so than coverage of the basic uMRI exam.

Table 13 on the following pages summarizes the evidence-based findings by key question for studies of uMRI included in this assessment.

Table 13. Summary of studies on the concordance, reliability and impact of uMRI for specific spinal and extra-spinal conditions

Key Question	Strength of Evidence*	Summary: Study Design and Imaging Comparisons	Conclusions/Comments
Key Question 1: Diagnostic Concordance in the Following Conditions			
Cervical disc herniation	Very Low	• 1 retrospective cohort study • uMRI vs. rMRI	• There is limited evidence (very low) to suggest that uMRI may provide similar diagnostic information as rMRI with respect to disc pathology of the cervical spine. • There is no evidence comparing uMRI to a diagnostic reference standard for cervical disc pathology.
Lumbar disc pathology	Low	• 3 retrospective cohort studies • uMRI vs. rMRI	• There is limited evidence (low) to suggest that uMRI may provide similar diagnostic information as rMRI with respect to disc pathology of the lumbar spine. • There is no evidence comparing uMRI to a diagnostic reference standard for lumbar disc pathology.
Lumbar foraminal stenosis	Low	• 3 retrospective cohort studies • uMRI vs. rMRI (2 studies) • uMRI vs. myelography (1 study)	• There is limited evidence (low) to suggest that uMRI may provide similar diagnostic information as rMRI or myelogram with respect to foraminal stenosis of the lumbar spine. • There is no evidence comparing uMRI to a diagnostic reference standard for foraminal stenosis.
Lumbar nerve root compromise	Very Low	• 1 retrospective cohort study • uMRI vs. rMRI	• There is limited evidence (very low) to suggest that uMRI may provide similar diagnostic information as rMRI with respect to nerve root compromise of the lumbar spine. • There is no evidence comparing uMRI to a diagnostic reference standard for nerve root compromise.
Lumbar spondylolisthesis	Very Low	• 1 retrospective cohort study • uMRI vs. rMRI	• There is limited evidence (very low) to suggest that uMRI may provide similar diagnostic information as rMRI with respect to spondylolisthesis of the lumbar spine. • There is no evidence comparing uMRI to a diagnostic reference standard for spondylolisthesis.

Table 13. Summary of studies on the concordance, reliability and impact of uMRI for specific spinal and extra-spinal conditions (CONTINUED)

Key Question	Strength of Evidence	Summary	Conclusions/Comments
Morton neuroma	Very Low	• 1 prospective cohort study • uMRI vs. rMRI	• There is no evidence to suggest uMRI images contribute towards the identification of Morton neuroma compared with existing diagnostic tests.
Shoulder instability	Very Low	• 1 prospective cohort study • uMRI vs. exam under anesthesia	• There is no evidence to suggest uMRI images add to the identification of shoulder instability compared with existing diagnostic tests.
Key Question 2: Reliability			
Spinal stenosis	Low	• 1 retrospective cohort study assessing inter-rater reliability	• There is limited evidence from one study that lumbar foraminal stenosis may be determined reliably between radiologists using seated uMRI in patients whose symptoms are severe enough to warrant a myelogram. • There is no evidence that uMRI is reliable in detecting degenerative spondylolisthesis, lateral recess stenosis, radicular pain, non-specific spine pain or extra-spinal conditions.

Key Question 3: Diagnostic Impact

- No studies of diagnostic impact were found.
- No determination can be made with respect to the effect uMRI has on the use of additional diagnostic testing or its effect on limiting the differential diagnosis using uMRI

Key Question 4: Therapeutic Impact

- No studies of therapeutic impact were found. Lack of data from included studies prevents one from drawing conclusions on the likelihood that positive upright MRI findings accurately predicts favorable outcome following test-directed treatment.

Key Question 5: Effectiveness in Acute vs. sub-acute/delayed setting

- No studies were found evaluating diagnostic or therapeutic impact in acute versus sub-acute or delayed setting.
- Lack of data addressing this issue precludes evaluation of the effectiveness of uMRI as a diagnostic imaging tool in these populations.

* Strength of Evidence expresses the confidence in the stability of the estimate as further research is done : High= Very unlikely to change confidence in effect estimate; Moderate = Likely to have an important impact on confidence in estimate and *may* change the estimate;Low = Very likely to have an important impact on confidence in estimate and *likely* to change the estimate; Very Low = Any effect estimate is uncertain

VII. Limitations of current literature

No studies which assessed the validity (diagnostic accuracy) of uMRI were found. This is the most significant limitation of the current literature since, without methodologically rigorous validation studies comparing uMRI to an appropriate "gold" or reference standard, characteristics which describe diagnostic accuracy, such as sensitivity, specificity and predictive values or likelihood ratios cannot be determined with any confidence. Furthermore, without a reasonably solid estimate of diagnostic accuracy, meaningful evaluation of the diagnostic and therapeutic impact of uMRI for various conditions or disease states may not be feasible.

While four studies compared uMRI with other available diagnostic methods for evaluation of spinal conditions and two comparative studies on evaluation of extra-spinal conditions met the inclusion criteria and were included in this report, the majority of published information on uMRI assessed anatomical changes or joint kinematics primarily among asymptomatic volunteers and/or consisted of small numbers of cases. These types of studies, while they may enhance understanding of how loading and position may facilitate the assessment of anatomy and pathology and provide some normative information, they do not provide information for determination of diagnostic accuracy and may not provide sufficient information to assess reliability in a broader range of patients with various conditions using consistent diagnostic criteria.

The overall quality of included concordance studies of uMRI is poor (class of evidence IV for spinal conditions LoE II/III for extra-spinal conditions). The six studies which allowed for assessment of concordance between uMRI and another diagnostic method had a number of methodological shortcomings that had potential to bias the results and thus, data on concordance should be interpreted cautiously. General methodological limitations of the studies summarized in this report include:

(1) Inadequate power (small sample sizes).
(2) Inadequate documentation of exam interpretation (blinding).
(3) Lack of independence related to performance of uMRI and method to which it was compared leading to verification bias and possible overestimation of concordance.

(4) Failure to include a broad spectrum of patients with the suspected condition(s) leading to selection bias.
(5) Use of poorly standardized or poorly validated measurement protocols.
(6) Limited ability to generalize results to populations with a broader spectrum of disease or condition than those represented in the individual studies.
(7) Poor reporting of and accounting for the number of patients and how they were enrolled and evaluated.

Available studies on the reliability of uMRI are limited in scope and number with only one included study measuring foraminal stenosis. The rest evaluated anatomy or morphology of the spine, the shoulder or the patellofemoral joint. Such studies may or may not be generalizable to populations with a range of pathologic conditions in a variety of clinical settings.

VIII. Recommendations for further research

To address some of the limitations noted above, methodologically rigorous studies which minimize common sources of bias for diagnostic test studies are needed. A description of such sources of bias is included in Appendix C. Specific recommendations include:

A. Validation studies of uMRI which compare it with an appropriate reference standard in a broad spectrum of patients and which follow methodologically sound protocols are needed in order to establish the diagnostic efficacy of uMRI. Such studies should allow for determination of diagnostic test characteristics (e.g., sensitivity, specificity and predictive values or likelihood ratios) that describe the diagnostic accuracy of uMRI.
B. Defining an appropriate reference standard for evaluation of uMRI for spinal conditions merits more in depth discussion.
C. Reliability studies which follow reproducible, sound protocols are needed. Both intra- and inter-rater reliability studies are important to evaluate the reproducibility-related accuracy. Such studies may also be beneficial in establishing “normal” and “abnormal” value ranges for various conditions and pathologies.
D. Studies designed to determine the extent to which uMRI findings correlate with patient symptoms and outcomes are necessary.

E. Diagnostic impact needs to be assessed in light of information on the diagnostic accuracy of uMRI, taking into account whether the condition evaluated has a known effective and safe treatment solution.

F. Studies which address the extent to which uMRI influences treatment decisions and the outcomes of those decisions are needed to address the issue of therapeutic impact.

G. If magnet strength is an important determinant of image resolution, it may be important to evaluate the extent to which axial loaded, standard rMRI (1.0-3.0T) may provided better resolution and accuracy compared with uMRI (0.5T) in the recumbent (+/- axial loading) position.

H. Since axial loading and body position could effect uMRI findings, it may be important to assess which factor contributes to any added information obtained from being upright. A study comparing supine axial loading with upright axial loading in different positions would help clarify the contribution of these factors.

IX. References

1. Vujovic LMD. A Short History of the Magnetic Resonance Imaging (MRI). Available at: http://www.teslasociety.com/mri.htm. Accessed April 20, 2007.

2. Chen MYM, Pope TL, Ott DJ. *Basic radiology*. New York: Lange Medical Books/McGraw-Hill; 2004.

3. GE Healthcare [website]. Magnetic Resonance Imagning Signa SP/i System. Brief description listing of clinical sites. Available at: http://www.gehealthcare.com/inen/rad/mri/products/spi/index.html.

4. FONAR Corporate Representative. Personal communication regarding system costs, general billing practices and reimbursement Melville, NY; 2007.

5. Upright MRI of Seattle. Personal communications with representative regarding charges and billing practices. Renton, WA; 2007:Phone conversations

6. MR-Tip. Magnetic Resonance Imaging MRI [Website]. April 9, 2007. Available at: http://www.mr-tip.com/.

7. Vitzthum HE, Konig A, Seifert V. Dynamic examination of the lumbar spine by using vertical, open magnetic resonance imaging. *J Neurosurg.* Jul 2000;93(1 Suppl):58-64.

8. Muhle C, Bischoff L, Weinert D, et al. Exacerbated pain in cervical radiculopathy at axial rotation, flexion, extension, and coupled motions of the cervical spine: evaluation by kinematic magnetic resonance imaging. *Invest Radiol.* 1998;33(5):279-288.

9. Muhle C, Weinert D, Falliner A, et al. Dynamic changes of the spinal canal in patients with cervical spondylosis at flexion and extension using magnetic resonance imaging. *Invest Radiol.* 1998;33(8):444-449.

10. Vitaz TW, Shields CB, Raque GH, et al. Dynamic weight-bearing cervical magnetic resonance imaging: technical review and preliminary results. *South Med J.* 2004;97(5):456-461.

11. Karadimas EJ, Siddiqui M, Smith FW, Wardlaw D. Positional MRI changes in supine versus sitting postures in patients with degenerative lumbar spine. *J Spinal Disord Tech.* 2006;19(7):495-500.

12. Hirasawa Y, Bashir WA, Smith FW, Magnusson ML, Pope MH, Takahashi K. Postural changes of the dural sac in the lumbar spines of asymptomatic individuals using positional stand-up magnetic resonance imaging. *Spine.* 2007;32(4):E136-140.

13. Schmid MR, Stucki G, Duewell S, Wildermuth S, Romanowski B, Hodler J. Changes in cross-sectional measurements of the spinal canal and intervertebral foramina as a function of body position: in vivo studies on an open-configuration MR system. *AJR Am J Roentgenol.* Apr 1999;172(4):1095-1102.

14. Kanal E. ACR Guidance Document for Safe MR Practices: 2007. *American Journal of Roentgenology.* 2007;188.

15. Sam AD, 2nd, Morasch MD, Collins J, Song G, Chen R, Pereles FS. Safety of gadolinium contrast angiography in patients with chronic renal insufficiency. *J Vasc Surg.* Aug 2003;38(2):313-318.

16. Shellock FG, Crues JV. MR procedures: biologic effects, safety, and patient care. *Radiology.* Sep 2004;232(3):635-652.

17. Willen J, Danielson B. The diagnostic effect from axial loading of the lumbar spine during computed tomography and magnetic resonance imaging in patients with degenerative disorders. *Spine.* Dec 1 2001;26(23):2607-2614.

18. Danielson BI, Willen J, Gaulitz A, Niklason T, Hansson TH. Axial loading of the spine during CT and MR in patients with suspected lumbar spinal stenosis. *Acta Radiol.* Nov 1998;39(6):604-611.

19. Manenti G, Liccardo G, Sergiacomi G, et al. Axial loading MRI of the lumbar spine. *In Vivo.* Sep-Oct 2003;17(5):413-420.

20. Hiwatashi A, Danielson B, Moritani T, et al. Axial loading during MR imaging can influence treatment decision for symptomatic spinal stenosis. *AJNR Am J Neuroradiol.* Feb 2004;25(2):170-174.

21. Humphreys SC, Eck JC, Hodges SD. Neuroimaging in low back pain. *Am Fam Physician.* Jun 1 2002;65(11):2299-2306.

22. Bradley WG Jr SD, Brunberg JA, Davis PC, DE La Paz RL, Dormont D, Hackney DB, Jordan JE, Karis JP, Mukherji SK, Turski PA, Wippold FJ, Zimmerman RD, McDermott MW, Sloan MA. Low Back Pain [online publication]: Reston (VA): American College of Radiology (ACR); 1996 (Revised 2005).

23. Buenaventura RM, Shah RV, Patel V, Benyamin R, Singh V. Systematic review of discography as a diagnostic test for spinal pain: an update. *Pain Physician.* Jan 2007;10(1):147-164.

24. Shellock F, Powers, CM. *Kinematic MRI of the Joints: Functional Anatomy, Kinesthiology, and Clinical Application*. Boca Raton: CRC Press; 2001.

25. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *Jama.* Sep 15 1999;282(11):1061-1066.

26. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med.* 2006;144(11):850-855.

27. Regence [website]. Positional Magnetic Resonance Imaging. Medical Policy No. 49. Feburary 6, 2007 Available at: http://www.regence.com/trgmedpol/radiology/rad49.html.

28. Premera Blue Cross [website]. Standing MRI. Corporate Medical Policy No. CP.MP.PR.6.01.513. February 14, 2006. Available at: http://www.premera.com/stellent/groups/public/documents/medicalpolicy/dynwat%3B5893_45048227_3076.pdf.

29. Cigna HelathCare [website]. Magnetic Resonancy Imaging MIR), Low Field. Coverage Position No. 0444. Last updated October 15, 2006. Available at: http://www.cigna.com/cgi-bin/search/search.cgi?search=Magnetic+Resonance+Imaging%2C+low+field&hitspercat=25&browseNav23.x=14&browseNav23.y=11.

30. Atena [Website]. Open Air, Low Field Strength, and Standing MRI Units. Policy No. 0093. Last reviewed March 30, 2007. Available at: www.aetna.com/cpb/medical/data/1_99/0093.html.

31. Washington State Department of Labor and Industries. Health technology assessment. Standing, weight-bearing, positional or upright magnetic resonance imaging. May 31. 2006. Available at: http://www.lni.wa.gov/ClaimsIns/Files/OMD/StandMriTAMay2006.pdf.

32. *Hayes, Inc. Hayes brief: Upright magnetic resonance imaging (MRI) for diagnosis of spinal disorders.* . Lansdale, PA April 1, 2007.
33. Skinner HB. *Current diagnosis & treatment in orthopedics*. 3. ed. New York: Lange medical Books/McGraw-Hill; 2003.
34. Jarvik JG, Haynor DR, Koepsell TD, Bronstein A, Ashley D, Deyo RA. Interreader reliability for a new classification of lumbar disk disease. *Acad Radiol.* Jul 1996;3(7):537-544.
35. Low back pain: Causes, symptoms, and diagnosis. *Harvard Men's Health Watch.* November 2006 2006;11(2):1-4.
36. Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med.* Oct 1 2002;137(7):586-597.
37. Jarvik JJ, Hollingworth W, Heagerty P, Haynor DR, Deyo RA. The Longitudinal Assessment of Imaging and Disability of the Back (LAIDBack) Study: baseline data. *Spine.* May 15 2001;26(10):1158-1166.
38. Atlas SJ, Deyo RA. Evaluating and managing acute low back pain in the primary care setting. *J Gen Intern Med.* Feb 2001;16(2):120-131.
39. ICSI. *Adult low back pain*: Institute for Clinical Systems Improvement (ICSI), Bloomington MN; 2006.
40. Pavlov H DM, Alazraki NP, Daffner RH, DeSmet AA, El-Khoury GY, KNeeland JB, Manaster BJ, Rubin DA, Steinbach LS, Weissman BN, Haralson RH III, Expert Panel on Musculoskeletal Imaging. Nontraumatic Knee Pain. *American College of Radiology.* 2005.
41. Dalinka MK DR, DeSmet AA, El-Khoury GY, Kneeland JB, Manaster BJ, Morrison WB, Pavlov H, Rubin DA, Schneider R, Steinbach LS, Weissman BN, Haralson RH III, Expert Panel on Musculoskeletal Imaging. Chronic Wrist Pain. *American College of Radiology.* 2005.
42. Steinbach LS DR, Dalinka MK, DeSmet AA, El-Khoury GY, Kneeland JB, Manaster BJ, Morrison WB, Pavlov H, Rubin DA, Weissman BN, Haralson RH III, Expert Panel on Musculoskeletal Imaging. Shoulder Trauma. *American College of Radiology.* 2005.
43. DeSmet AA DM, Daffner RH, El-Khoury GY, Kneeland JB, Manaster BJ, Morrison WB, Pavlov H, Rubin DA, Schneider R, Steinbach LS, Weissman BN, Haralson RH III, Expert Panel on Musculoskeletal Imaging. Chonic Ankle Pain. *American College of Radiology.* 2005.
44. Ferreiro Perez A, Garcia Isidro M, Ayerbe E, Castedo J, Jinkins JR. Evaluation of intervertebral disc herniation and hypermobile intersegmental instability in symptomatic adult patients undergoing recumbent and upright MRI of the cervical or lumbosacral spines. *Eur J Radiol.* Apr 3 2007.
45. Weishaupt D, Schmid MR, Zanetti M, et al. Positional MR imaging of the lumbar spine: does it demonstrate nerve root compromise not visible at conventional MR imaging? *Radiology.* Apr 2000;215(1):247-253.
46. Wildermuth S, Zanetti M, Duewell S, et al. Lumbar spine: quantitative and qualitative assessment of positional (upright flexion and extension) MR imaging and myelography. *Radiology.* May 1998;207(2):391-398.

47. Zamani AA, Moriarty T, Hsu L, et al. Functional MRI of the lumbar spine in erect position in a superconducting open-configuration MR system: preliminary results. *J Magn Reson Imaging.* Nov-Dec 1998;8(6):1329-1333.
48. Ferreiro Perez A, Garcia Isidro M, Ayerbe E, Castedo J, Jinkins JR. Evaluation of intervertebral disc herniation and hypermobile intersegmental instability in symptomatic adult patients undergoing recumbent and upright MRI of the cervical or lumbosacral spines. *Eur J Radiol.* 2007;12(47):27.
49. Hodge DK, Beaulieu CF, Thabit GH, 3rd, et al. Dynamic MR imaging and stress testing in glenohumeral instability: comparison with normal shoulders and clinical/surgical findings. *J Magn Reson Imaging.* 2001;13(5):748-756.
50. Weishaupt D, Treiber K, Kundert HP, et al. Morton neuroma: MR imaging in prone, supine, and upright weight-bearing body positions. *Radiology.* Mar 2003;226(3):849-856.
51. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* Mar 1977;33(1):159-174.
52. Centers for Medicare and Medicaid Services (CMS)[website]. NCD for Magnetic Resonance Imaging (MRI) (220.2). Updated September 10, 2004. Available at: http://www.cms.hhs.gov/mcd/viewncd.asp?ncd_id=220.2&ncd_version=1&basket=ncd%3A220%2E2%3A1%3AMagnetic+Resonance+Imaging+%28MRI%29.
53. Willen J, Danielson B, Gaulitz A, Niklason T, Schonstrom N, Hansson T. Dynamic effects on the lumbar spinal canal: axially loaded CT-myelography and MRI in patients with sciatica and/or neurogenic claudication. *Spine.* Dec 15 1997;22(24):2968-2976.
54. Edmondston SJ, Song S, Bricknell RV, et al. MRI evaluation of lumbar spine flexion and extension in asymptomatic individuals. *Man Ther.* Aug 2000;5(3):158-164.
55. Danielson B, Willen J. Axially loaded magnetic resonance image of the lumbar spine in asymptomatic individuals. *Spine.* Dec 1 2001;26(23):2601-2606.
56. Kimura S, Steinbach GC, Watenpaugh DE, Hargens AR. Lumbar spine disc height and curvature responses to an axial load generated by a compression device compatible with magnetic resonance imaging. *Spine.* Dec 1 2001;26(23):2596-2600.
57. Moffet H, Hebert L, Dufour M, Tardif J. Variation in the sub-acromial distance measured by magnetic resonance imaging during shoulder flexion and abduction movements. *Can J Rehabil.* Vol 11; 1998:265.
58. Beaulieu CF, Hodge DK, Bergman AG, et al. Glenohumeral relationships during physiologic shoulder motion and stress testing: initial experience with open MR imaging and active imaging-plane registration. *Radiology.* 1999;212(3):699-705.
59. Gold GE, Besier TF, Draper CE, Asakawa DS, Delp SL, Beaupre GS. Weight-bearing MRI of patellofemoral joint cartilage contact area. *J Magn Reson Imaging.* 2004;20(3):526-530.
60. Ward SR, Shellock FG, Terk MR, Salsich GB, Powers CM. Assessment of patellofemoral relationships using kinematic MRI: comparison between qualitative and quantitative methods. *J Magn Reson Imaging.* 2002;16(1):69-74.
61. Besier TF, Draper CE, Gold GE, Beaupre GS, Delp SL. Patellofemoral joint contact area increases with knee flexion and weight-bearing. *J Orthop Res.* Mar 2005;23(2):345-350.
62. Powers CM, Ward SR, Fredericson M, Guillet M, Shellock FG. Patellofemoral kinematics during weight-bearing and non-weight-bearing knee extension in persons with lateral

subluxation of the patella: a preliminary study. *J Orthop Sports Phys Ther.* Nov 2003;33(11):677-685.

63. Johal P, Williams A, Wragg P, Hunt D, Gedroyc W. Tibio-femoral movement in the living knee. A study of weight bearing and non-weight bearing knee kinematics using 'interventional' MRI. *J Biomech.* 2005;38(2):269-276.

64. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *Cmaj.* Feb 14 2006;174(4):469-476.

65. Phillips B, Ball C, Sackett D, et al. Levels of evidence and grades of recommendation. Available at: http://www.cebm.net/levels_of_evidence.asp. Accessed December 2, 2006, 2006.

66. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *Bmj.* Jun 19 2004;328(7454):1490.

67. West S, King V, Carey TS, et.al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center, Contract No. 290-97-0011): Agency for Healthcare Research and Quality, Rockville, MD; 2002.

68. Armstrong B, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*: Oxford University Press 1992.

69. Koepsell T, Weiss N. *Epidemiologic methods: studying the occurrence of illness*. New York: Oxford University Press, Inc; 2003.

70. Andreasen ML, Langhoff L, Jensen TS, Albert HB. Reproduction of the lumbar lordosis: a comparison of standing radiographs versus supine magnetic resonance imaging obtained with straightened lower extremities. *J Manipulative Physiol Ther.* 2007;30(1):26-30.

71. Jinkins JR. Upright, weight-bearing, dynamic-kinetic MRI of the spine: pMRI/kMRI. *Rivista di Neuroradiologia.* 2002;15:333-356.

72. Jinkins JR, Dworkin J. Upright, weight-bearing, dynamic-kinetic MRI of the spine: pMRI/kMRI. In: Kaech DL, Jinkins JR, eds. *Spinal Restabilization Procedures: Diagnostic and Therapeutic Aspects of Intervertebral Fusion Cages, Artificial Discs and Mobile Implants*. New York: Elsevier; 2002:73-82.

73. Jinkins JR. Upright, weight-bearing, dynamic-kinetic MRI of the spine - Review of the first clinical results. *J HK Coll Radiol.* Vol 6; 2003:55-74.

74. Jinkins JR, Dworkin J. Proceedings of the State-of-the-Art Symposium on Diagnostic and Interventional Radiology of the Spine, Antwerp, September 7, 2002 (Part two). Upright, weight-bearing, dynamic-kinetic MRI of the spine: pMRI/kMRI. *Jbr Btr.* 2003;86(5):286-293.

75. Jinkins JR, Dworkin JS, Damadian RV. Upright, weight-bearing, dynamic-kinetic MRI of the spine: initial results. *Eur Radiol.* 2005;15(9):1815-1825.

76. Koschorek F, Jensen HP, Terwey B. Dynamic studies of cervical spinal canal and spinal cord by magnetic resonance imaging. *Acta Radiol Suppl.* 1986;369:727-729.

77. Singh K, Samartzis D, Vaccaro AR, et al. Congenital lumbar spinal stenosis: a prospective, control-matched, cohort radiographic analysis. *Spine J.* Nov-Dec 2005;5(6):615-622.

78. Cicuttini F, Hankin J, Jones G, Wluka A. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. *Osteoarthritis Cartilage.* 2005;13(8):722-727.

79. Dufour M, Moffet H, Hebert LJ, Moisan C. *Kinematic MRI of the shoulder: Assessment in the seated position* 2001. 0-8493-0807-0.
80. Gedroyc WMW, Williams A. Kinematic MRI of the knee: Preliminary experience using the upright, weight-bearing technique. *Kinematic Mri of the Joints*; 2001:331-341.
81. Graichen H, Stammberger T, Bonel H, Karl-Hans E, Reiser M, Eckstein F. Glenohumeral translation during active and passive elevation of the shoulder - a 3D open-MRI study. *J Biomech.* 2000;33(5):609-613.
82. Merl T, Scholz M, Gerhardt P, et al. Results of a prospective multicenter study for evaluation of the diagnostic quality of an open whole-body low-field MRI unit. A comparison with high-field MRI measured by the applicable gold standard. *Eur J Radiol.* 1999;30(1):43-53.
83. Turner JA, Ersek M, Herron L, Deyo R. Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. *Spine.* 1992;17(1):1-8.
84. Weiss NS. *Clinical epidemiology : the study of the outcome of illness / Noel S. Weiss*. 3rd ed. ed. New York :: Oxford University Press; 2006.
85. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med.* May 2004;19(5 Pt 1):460-465.
86. Weiss NS, Koepsell, T.D. *Epidemiologic Methods: Studying the Occurrence of Illness*. 1st Edition ed. Oxford: Oxford University Press; 2003.

Appendix A: Overview of Studies Related to Position and Axial Loading

Table A1. Overview of excluded studies evaluating the effect of loading and/or body position on the spine

								Exam 1							Exam 2						
Author	Study type		Patient		Outcome				Axial load			Position				Axial load			Position		
year	reliab	concord	symp	asymp	diagnosis	anatomy	location	image	yes	no		stand	sit	supine	image	yes	no		stand	sit	supine
Muhle et al. 1998[9]		x	x		x		cervical	rMRI		x				n	rMRI	x					e
Muhle et al. 1998[8]	x	x	x		x		cervical	rMRI		x				n	rMRI		x				e,f
Vitaz et al. 2004[10]		x	x		x	x	cervical	uMRI	x				n		uMRI	x				f,e	
Willen & Danielson et al. 1997, 1998, 2001[17, 18, 53]		x	x		x		lumbar	CT-My/ rMRI		x				n	rMRI	x					e
Vitzthum et al. 2000[7]		x	x	x		x	lumbar	uMRI	x				f,e,r								
Edmondston et al. 2000[54]		x		x		x	lumbar	rMRI		x				n	rMRI		x				e
Hiwatashi et al. 2004[20]		x	x		x		lumbar	rMRI		x				n	rMRI	x					e
Manenti et al. 2003[19]		x	x		x		lumbar	rMRI		x				n	rMRI	x					e
Karadimas et al. 2006[11]	x	x	x			x	lumbar	rMRI		x				n	uMRI	x				n	
Danielson et al. 2001[55]	x	x		x		x	lumbar	rMRI		x				n	rMRI	x					e
Kimura et al. 2001[56]		x		x		x	lumbar	rMRI		x				n	rMRI	x					f
Hirasawa et al. 2007[12]		x		x		x	lumbar	rMRI		x				n	uMRI	x			x	n	
Schmid et al. 1999[13]		x		x		x	lumbar	rMRI		x				e	uMRI	x			x	n	

reliab = reliability, concord = concordance, symp = symptomatic, asymp = asymptomatic, e=extension, f = flexion, r = rotation, n = neutral, CT-My= CT myelogram

Table A2. Overview of excluded studies evaluating the effect of loading and/or body position on extra-spinal joints.

Author	Study type		Patient		Outcome		Joint or	Exam 1			Exam 2	
year	reliab	concord	symp	asymp	diagnosis	anatomy	body part	Image	Position		Image	Position
Moffet et al. 1998[57]	x			x		x	shoulder	uMRI	seated f, abd			
Beaulieu et al. 1999[58]		x		x		x	shoulder	uMRI	seated IR/ER, abd, add			
Gold et al. 2004[59]	x			x		x	patellofemoral	uMRI	squat 30º, 60º			
Ward et al. 2002[60]*	x		x	x		x	patellofemoral	uMRI	seated e 0-45º		uMRI	squat 0-60º
Besier et al. 2005[61]		x		x		x	patellofemoral	uMRI	squat 0º, 30º, 60º			
Powers et al. 2003[62]		x	x			x	patellofemoral	rMRI	supine e 0-45º		uMRI	squat 0-45º
Johal et al. 2005[63]	x			x		x	tibio-femoral	uMRI	squat 0-120º			

reliab = reliability, concord = concordance, symp = symptomatic, asymp = asymptomatic, e=extension, f = flexion, r = rotation, n = neutral, abd = abduction, add = adduction, IR = internal rotation, ER = external rotation

*Symptomatic and asymptomatic individuals analyzed together

Appendix B. Detailed listing of literature search strategies and terms

The clinical studies included in this report were identified using the algorithm below, Figure B1. We conducted the search in four stages. The first stage of the study selection process consisted of a comprehensive literature search using electronic means and hand searching. We then screened all possible relevant articles using titles and abstracts in stage two. This was done by two individuals independently. Those articles that met a set of a priori retrieval criteria were included. Any disagreement between screeners that were unresolved resulted in the article being included for the next stage. Stage three involved retrieval of the full text articles remaining. The final stage of the study selection algorithm consisted of the selection of those studies using a set of a priori inclusion criteria, again, by two independent investigators. Those articles selected form the evidence base for this report.

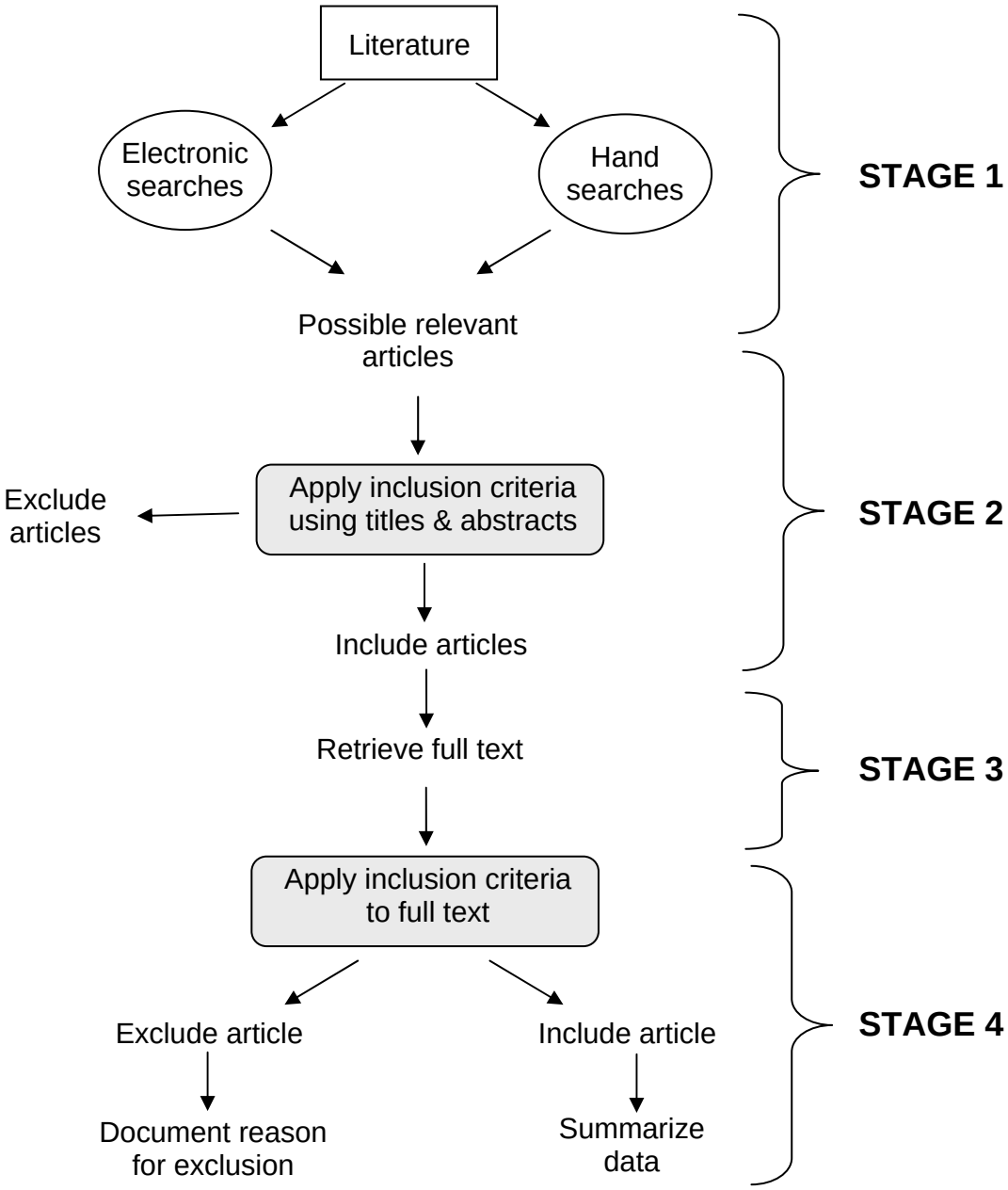
Electronic Database Searches

The following databases have been searched for relevant information:
Agency for Healthcare Research and Quality (AHRQ)
Cumulative Index to Nursing and Allied Health (CINAHL)
Cochrane Database of Systematic Reviews (through 2007, Issue 2)
Cochrane Registry of Clinical Trials (CENTRAL) (through 2007, Issue 2)
Cochrane Review Methodology Database (through 2007, Issue 2)
Computer Retrieval of Information on Scientific Projects (CRISP)
Database of Reviews of Effectiveness (Cochrane Library) (through 2007, Issue 2)
EMBASE (1985 through April 15, 2007)
PubMed (1975 through April 15, 2007)
Informational Network of Agencies for Health Technology Assessment (INAHTA)
NHS Economic Evaluation Database (Cochrane Library through 2007, Issue 2)
HSTAT (Health Services/Technology Assessment Text)
EconLIT

Additional Economics, Clinical Guideline and Gray Literature Databases

AHRQ- Healthcare Cost and Utilization Project
American College of Radiology (ACR)
Canadian Agency for Drugs and Technologies in Health
Centers for Medicare and Medicaid Services (CMS)
FONAR corporate website
Food and Drug Administration (FDA)
Google
IDEAS and RePEc-working papers and articles http://ideas.repec.org/
Institute for Clinical Systems Improvement (ICSI)
ISI Web of Science New York Academy of Medicine Library – Gray Literature Producing Organizations
http://www.nyam.org/library/pages/gray_literature_producing_organizations
National Association of Health Data Organizations http://www.nahdo.org/default.aspx
National Guideline Clearinghouse
RAND
Trip (Turning Research into Practice) Database: http://www.tripdatabase.com/index.html

Figure B1 Study Selection Algorithm


Literature
Electronic searches
Hand searches
Possible relevant articles
Exclude articles
Apply inclusion criteria using titles & abstracts
Include articles
Retrieve full text
Apply inclusion criteria to full text
Exclude article
Include article
Document reason for exclusion
Summarize data
STAGE 1
STAGE 2
STAGE 3
STAGE 4

The search strategies employed a number of free text keywords that included the following:

Magnetic Resonance Imaging
"Magnetic Resonance Imaging"[MeSH]
Magnetic Resonance Imaging
MRI
Dynamic
Vertical
Upright
Stand-up
Standing
Seated
Open
Position*
weight-bearing
functional imaging

Spine
Cervical vertebrae[MeSH]
Cervical myelopathy
cervical spine
Cervical spondylotic myelopathy
Dural sac
Facet
Herniation
Instability
Intervertebral disc
Intervertebral disk displacement[MeSH]
Intervertebral disk[MeSH]
Kyphosis
Lordosis
Low back
Low back pain[MeSH]
Lumbar
Lumbar stenosis
Lumbar vertebrae[MeSH]
Neck pain[MeSH]
Neck[MeSH]
Neck
Radicul*
Radiculopathy[MeSH]
Sciatica
Sciatica[MeSH]
Scoliosis
Spinal
Spinal curvatures[MeSH]
Spinal osteophytosis[MeSH]
Spinal stenosis
Spinal stenosis[MeSH]
Spine
Spondylolisthesis
Spondylolisthesis[MeSH]
Spondylosis
Thoracic vertebrae[MeSH]
Whiplash injuries[MeSH]

Extraspinal Joints
Joints[MeSH]
Foot
Feet
Knee*
Hip
Hips
TMJ
Temporomandibular
Shoulder*
Elbow
Wrist*
Hand
Hands

Study Types
Reproducibility of Results[MeSH]
Validation Studies[Publication Type]
Reliability
Valid*
Accuracy

Economic
Economics [MeSH]
economics [Subheading]
economic*
Cost*
Costs and Cost Analysis[MeSH]
Cost AND MRI

The detailed strategy below is presented in PubMed syntax. Parallel strategies were used to search the Cochrane Library and EMBASE. Keyword searches were conducted in the other listed resources.

PubMed Search Strategy
(1975 – April 15, 2007)
Limited to English language, human population

Search Strategy for Key Questions 1 and 2

#1	Search (dynamic [TI] OR vertical [TI] OR upright [TI] OR stand-up [TI] OR standing [TI] OR seated [TI] OR open [TI] OR position* [TI] OR weight bearing [TI])
#2	Search ("Magnetic Resonance Imaging"[TI] OR MRI [TI])
#3	Search #1 AND #2
#4	Search "dynamic MRI" [TI] OR "dynamic magnetic resonance imaging" [TI] OR "vertical MRI" [TI] OR "vertical magnetic resonance imaging" [TI] OR "upright MRI" [TI] OR "upright magnetic resonance imaging" [TI] OR "stand-up MRI" [TI] OR "stand-up magnetic resonance imaging" [TI] OR "standing MRI" [TI] OR "standing magnetic resonance imaging" [TI] OR "seated MRI" [TI] OR "seated magnetic resonance imaging" [TI] OR "open MRI" [TI] OR "open magnetic resonance imaging" [TI] OR "position* MRI" [TI] OR "position magnetic resonance imaging" [TI] OR "weight bearing MRI" [TI] OR "weight bearing magnetic resonance imaging" [TI]
#5	Search "Low Back Pain"[MeSH] OR "Intervertebral Disk Displacement"[MeSH] OR "Sciatica"[MeSH] OR "Radiculopathy"[MeSH] OR "Spondylolisthesis"[MeSH] OR "Spinal Stenosis"[MeSH] OR "Intervertebral Disk"[MeSH] OR "Lumbar Vertebrae"[MeSH] OR spine[TI] OR dural sac[TI] OR facet[TI] OR "low back"[TI] OR "intervertebral disc"[TI] OR sciatica[TI] OR radicul*[TI] OR spondylolisthesis[TI] OR "spinal stenosis"[TI] OR lumbar [TI] OR "cervical vertebrae"[MeSH] OR "neck"[MeSH] OR "neck pain"[MeSH] OR "cervical myelopathy" OR "cervical spondylotic myelopathy" OR "radiculopathy"[MeSH] OR "thoracic vertebrae"[MeSH] OR "spinal curvatures"[MeSH] OR neck[TI] OR "cervical spine" [TI] OR scoliosis[TI] OR kyphosis[TI] OR lordosis[TI] OR "spinal osteophytosis"[MeSH] OR spondylosis [TI] OR "Whiplash Injuries"[MeSH]
#6	Search #3 AND #5
#7	Search #4 AND #5
#8	Search #6 OR #7
#9	Search ("Reproducibility of Results"[MeSH] OR "Validation Studies"[Publication Type])
#10	**#8 AND #9**
#11	Search ("Joints"[MeSH] OR foot OR feet OR knee* OR hip OR hips OR tmj OR temporomandibular OR shoulder* OR elbow OR wrist* OR hand OR hands)
#12	Search #3 AND #11
#13	Search #4 AND #11
#14	Search #12 OR #13
#15	#9 AND #14
#16	**Limit: NOT (letter OR editorial)**

Search Strategy for Key Questions 1, 3, 4 and 5

#1	Search (dynamic [TI] OR vertical [TI] OR upright [TI] OR stand-up [TI] OR standing [TI] OR seated [TI] OR open [TI] OR position* [TI] OR weight bearing [TI])
#2	Search ("Magnetic Resonance Imaging"[TI] OR MRI [TI])
#3	Search #1 AND #2
#4	Search "dynamic MRI" [TI] OR "dynamic magnetic resonance imaging" [TI] OR "vertical MRI" [TI] OR "vertical magnetic resonance imaging" [TI] OR "upright MRI" [TI] OR "upright magnetic resonance imaging" [TI] OR "stand-up MRI" [TI] OR "stand-up magnetic resonance imaging" [TI] OR "standing MRI" [TI] OR "standing magnetic resonance imaging" [TI] OR "seated MRI" [TI] OR "seated magnetic resonance imaging" [TI] OR "open MRI" [TI] OR "open magnetic resonance imaging" [TI] OR "position* MRI" [TI] OR "position magnetic resonance imaging" [TI] OR "weight bearing MRI" [TI] OR "weight bearing magnetic resonance imaging" [TI]
#5	Search "Low Back Pain"[MeSH] OR "Intervertebral Disk Displacement"[MeSH] OR "Sciatica"[MeSH] OR "Radiculopathy"[MeSH] OR "Spondylolisthesis"[MeSH] OR "Spinal Stenosis"[MeSH] OR "Intervertebral Disk"[MeSH] OR "Lumbar Vertebrae"[MeSH] OR spine[TI] OR dural sac[TI] OR facet[TI] OR "low back"[TI] OR "intervertebral disc"[TI] OR sciatica[TI] OR radicul*[TI] OR spondylolisthesis[TI] OR "spinal stenosis"[TI] OR lumbar [TI] OR "cervical vertebrae"[MeSH] OR "neck"[MeSH] OR "neck pain"[MeSH] OR "cervical myelopathy" OR "cervical spondylotic myelopathy" OR "radiculopathy"[MeSH] OR "thoracic vertebrae"[MeSH] OR "spinal curvatures"[MeSH] OR neck[TI] OR "cervical spine" [TI] OR scoliosis[TI] OR kyphosis[TI] OR lordosis[TI] OR "spinal osteophytosis"[MeSH] OR spondylosis [TI] OR "Whiplash Injuries"[MeSH]
#6	Search #3 AND #5
#7	Search #4 AND #5
#8	Search #6 OR #7
#9	**Limit: NOT (letter OR editorial)**
#10	Search ("Joints"[MeSH] OR foot OR feet OR knee* OR hip OR hips OR tmj OR shoulder* OR elbow OR wrist* OR hand OR hands)
#11	Search #3 AND #10
#12	Search #4 AND #10
#13	Search #11 OR #12
#14	Search "Arthritis, Experimental"[MeSH] OR "Arthritis, Infectious"[MeSH] OR "Spondylarthritis"[MeSH] OR "Arthritis, Rheumatoid"[MeSH]
#15	**Search #13 NOT #14**

Search Strategy for Economic or Cost Evaluation

#1	Search ("Magnetic Resonance Imaging"[TI] OR MRI [TI])
#2	Search #1 AND "Economics" [MeSH]
#3	Search (DYNAMIC [TI] OR VERTICAL [TI] OR UPRIGHT [TI] OR STAND-UP [TI] OR STANDING [TI] OR SEATED [TI] OR OPEN [TI] OR POSITION [TI] OR WEIGHT BEARING [TI])
#4	Search #2 AND #3
#5	Search "dynamic MRI" [TI] OR "dynamic magnetic resonance imaging" [TI] OR "vertical MRI" [TI] OR "vertical magnetic resonance imaging" [TI] OR "upright MRI" [TI] OR "upright magnetic resonance imaging" [TI] OR "stand-up MRI" [TI] OR "stand-up magnetic resonance imaging" [TI] OR "standing MRI" [TI] OR "standing magnetic resonance imaging" [TI] OR "seated MRI" [TI] OR "seated magnetic resonance imaging" [TI] OR "open MRI" [TI] OR "open magnetic resonance imaging" [TI] OR "position* MRI" [TI] OR "position magnetic resonance imaging" [TI] OR "weight bearing MRI" [TI] OR "weight bearing magnetic resonance imaging" [TI]
#6	#5 AND "Economics" [MeSH]
#7	Search "Costs and Cost Analysis"[MeSH] AND "Low Back Pain"[MeSH]
#8	Search "Costs and Cost Analysis"[MeSH] AND "Magnetic Resonance Imaging"[MeSH] AND "Low Back Pain"[MeSH]

Appendix C. Level of Evidence Determination

Introduction:

Studies which evaluate the accuracy and reliability of diagnostic tests are subject to a number of biases which may provide inaccurate assessment of its characteristics and clinical utility.[25, 64]. Parameters related to diagnostic accuracy (validity) and reliability are described in Appendix H.

Methods for critical appraisal and level of evidence assessment

Spectrum Research's (SRI) methods for assessing the quality of evidence of individual studies as well as the overall quality of evidence incorporates aspects of rating scheme developed by the Oxford Centre for Evidence-based Medicine,[65] precepts outlined by the Grades of Recommendation Assessment, Development and Evaluation (GRADE) Working Group[66] and recommendations made by the Agency for Healthcare Research and Quality (AHRQ).[67] We believe that taking into account features of methodological quality and sources of bias that are important and our LoE method combines epidemiologic principles with characteristics of study design.

Our method incorporates the essential five domains and related elements delineated by AHRQ,[67]as described in the following table, in addition to considering whether the study was prospectively or retrospectively designed.

Table C1. Overview Spectrum Research's LoE Assessment Based on AHRQ Domains

AHRQ Domain	Spectrum Research LoE Assessment
1. Study Population	• Was a broad spectrum of persons with the expected condition was used?
2. Description of Test	• Are the technical features, measurements performed, planes of section, diagnostic criteria, etc. described for both the test and the reference standard with sufficient detail to permit replication?
3. Appropriate Reference Standard	• Based on the pathology/condition being evaluated, is the test compared with the current "best" standard that is likely to correctly classify patients according to disease status?
4. Blinded Comparison of Test and Reference Standard	• Interpretation of test reference standard must be done without knowledge of the results of other?
5. Avoidance of Verification Bias	• Reference standard must be performed independently of test?

Reproducibility studies are those that evaluate the extent to which measurements can be replicated on subject/patient. Grading the quality of evidence for reliability studies has not been well reported in the literature. SRI's method is based on epidemiologic methods for validation (degree to which measurements reflect the truth) and reliability (reproducibility) studies.[68] This

system takes into consideration pertinent study design features and methods that may induce bias.

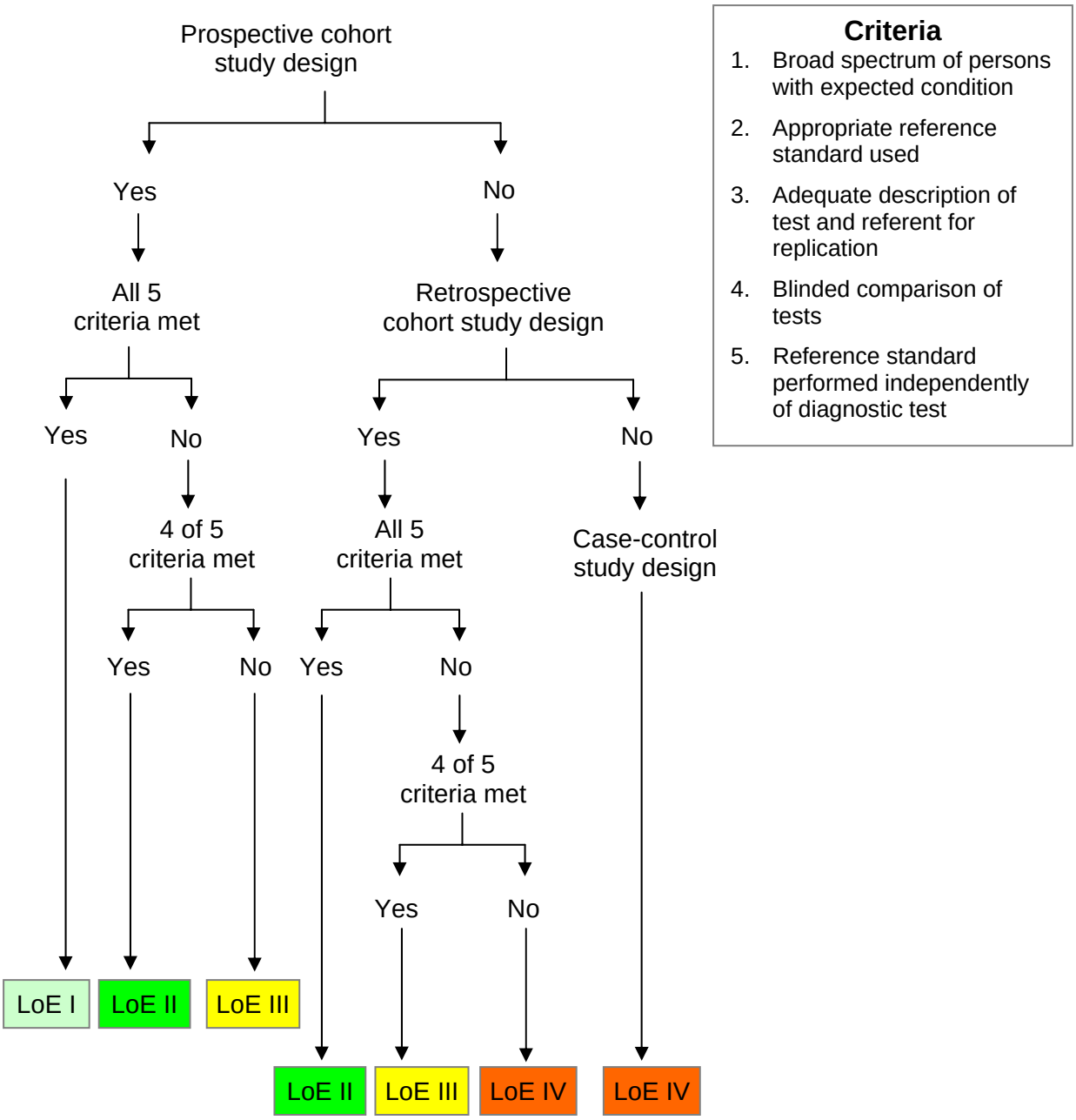
Levels of Evidence for Diagnostic Test Studies (Test Characteristics)
Table C2 and Figure C1 outline Spectrum Research's methodology for evaluating the quality of evidence for diagnostic studies and criteria used to determine the Level of Evidence (LoE). The procedure that follows describes specific considerations used to determine whether or not the various criteria were met. This method takes into account the primary sources of bias for such studies.

Each included study was evaluated independently by two investigators based on the criteria below and a LoE assigned to each article, initially at the abstract level and confirmed when the full articles were reviewed. Discrepancies in LoE determination were resolved by discussion until consensus was achieved.

Table C2. Definitions of the different levels of evidence for diagnostic test accuracy/validity studies.

Level	Study type	Criteria
I	Good quality prospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
II	Moderate quality prospective study	• Violation of any one of the criteria for a good quality prospective study (LoE I)
	Good quality retrospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
III	Poor quality prospective study	• Violation of any two or more of the criteria for a good quality prospective study (LoE I)
	Moderate quality retrospective study	• Violation of any one of the criteria for a good quality retrospective study (LoE II)
IV	Poor quality retrospective study	• Violation of any two or more of the criteria for a good quality retrospective study (LoE II)
	Case-Control Study	

Figure C1. Level of Evidence Algorithm –Accuracy/Validity Studies


Prospective cohort study design
Yes
No
All 5 criteria met
Retrospective cohort study design
Yes
No
Yes
No
4 of 5 criteria met
All 5 criteria met
Case-control study design
Yes
No
Yes
No
4 of 5 criteria met
Yes
No
LoE I
LoE II
LoE III
LoE II
LoE III
LoE IV
LoE IV
Criteria
1. Broad spectrum of persons with expected condition
2. Appropriate reference standard used
3. Adequate description of test and referent for replication
4. Blinded comparison of tests
5. Reference standard performed independently of diagnostic test

Procedures for determining adherence to LoE criteria
The following describes the method for determining whether or not a given study has met the specific individual criterion used to assign the LoE. Table C3 provides a template for indicating whether the individual criterion is met or not. A blank for the criterion indicates that the criterion was not met, could not be determined or was not reported by the author.

1. Determine if the study is **prospective or retrospective.**

   Accuracy of diagnostic tests is best assessed using a prospective study of consecutive series of patients from a relevant patient population (i.e. study designed for prospective collection of data using specific protocols). Ideally, a consecutive series of patients or random selection from the relavant patient population should be prospectively studied. Retrospective collection of data or evaluation of patients who have had the diagnostic test and reference test previously may be more subject to bias.

   If it is cannot be determined whether a prospective or retrospective approach was taken, no credit will be given for this criterion having been met.

2. Was a **broad spectrum of persons with the suspected condition** used to evaluate the diagnostic test and reference standard?

   The study population must be comprised of those with a broad spectrum of suspected disease who are likely to have the test now or in the future. A broad spectrum would include patients with mild as well as more severe cases, those presenting early as well as late and those whose differential diagnosis may be commonly confused with the condition of interest. Subjects from specialty referral sources may be more likely to have a specific abnormality/condition than those presenting to a general family practice clinic. Overestimation of diagnostic accuracy may occur if a population with known disease is compared with a group of normal individuals instead of those from the relevant patient population.

3. Was an **appropriate reference standard** used to compare the diagnostic test being evaluated?

   Ideal reference standards are termed “gold” standards and in theory, provide the “truth” about the presence or absence of a condition or disease. Such standards provide a basis for comparing the accuracy of other tests and allow for the calculation of characteristics such as sensitivity, specificity and predictive values.

   In most instances, the reference standard does not perfectly classify individuals with respect to the presence or absences of disease, but may reflect the current “best” reference and/or one that can be practically applied. It should be “likely” to classify patients according to disease status. A reference measure can be performed at the time of the testing. It may be an anatomical, physiological or pathological state or measure or a specific outcome at a later date.

The reference standard should be reproducible and the description of both the referent standard and the test should be explicit enough for replication, validation and generalization.

For purposes of this technology assessment relative to evaluation of the spine, a combination of upright myelography and supine CT-myelography would be considered an appropriate reference standard recognizing that it may be less sensitive than upright/standing MRI. Relative to the evaluation of joints, either supine MRI or radiographs will be considered an appropriate reference standard.

4. Are the **details of the test and the reference/gold standard sufficient to allow study replication**?

Are the technical features of the test and protocols used to collect information about test results, any measurements performed, planes of section evaluated, diagnostic criteria used, etc. sufficient that other investigators could duplicate the conditions and reproduce the findings in a similar population?

5. Was there **blinded comparison of the tests with the appropriate reference standard**?

Interpretation of the reference standard must be done without prior knowledge of the test results and the test must be interpreted without knowledge of the results of the reference test. This is necessary to avoid bias. It must be clear from the text that both tests were interpreted without knowledge of the results of the other.

6. Was the **reference standard performed independently of the diagnostic test**?

The reference standard must have been applied objectively or blindly to all patients without the results of test influencing use of the reference. If the "test" affects the reference (or referral to the reference test) or is part of the reference standard, this does not constitute independent performance of the test.

Table C3. Assessment of LoE for individual studies of diagnostic test evaluation

Methodological Principle	Author 1 (1999)	Author 2 (2002)	Author 3 (2004)	Author 4 (2005)
Study Design				
Prospective cohort design			■	
Retrospective cohort design	■	■		■
Case-control design				
Broad spectrum of patients with expected condition	■	■		■
Appropriate reference standard used	■	■		
Adequate description of test and reference for replication				
Blinded comparison with appropriate reference	■		■	
Reference standard performed independently of test	■	■	■	■
Evidence Level	**II**	**III**	**III**	**IV**

* Blank box indicates criterion not met, could not be determined or information not reported by author

Levels of Evidence for Diagnostic Test Studies –Reliability Studies

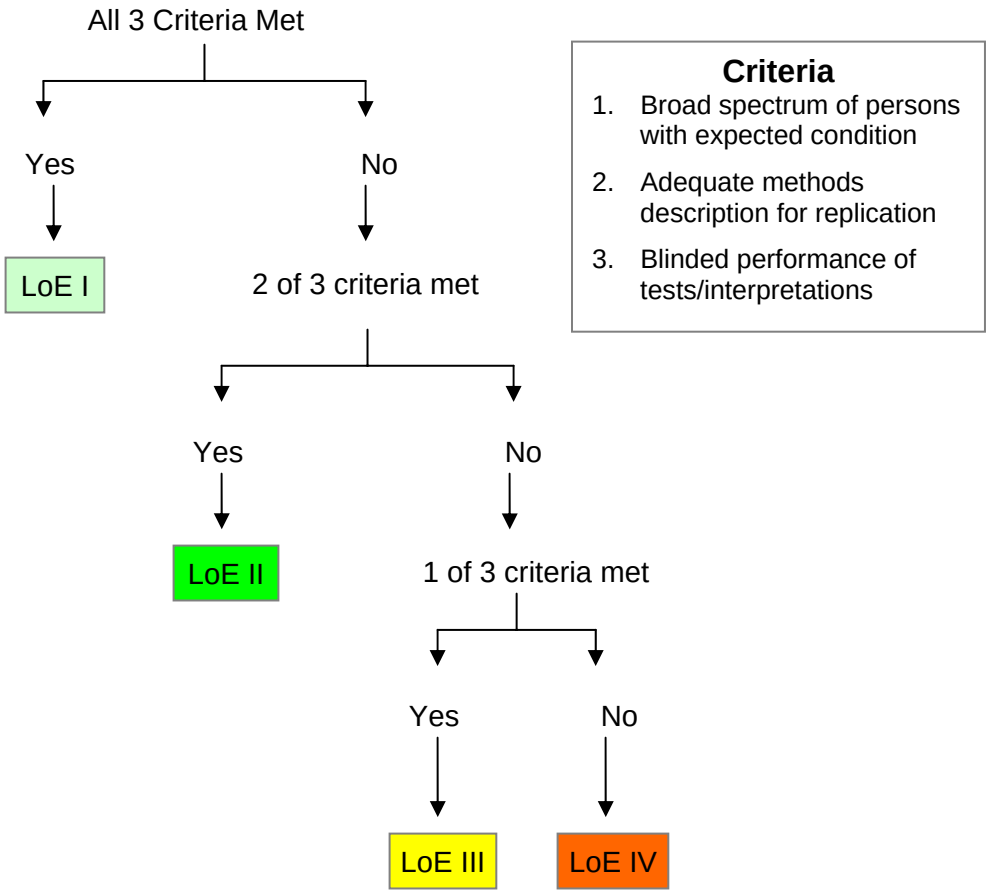
Methods for assessing the quality of evidence for reliability studies have not been well reported in the literature. Spectrum's determination of quality for such is based on epidemiologic methods for evaluating validity and reliability.[68]

The following describes the method for determining whether or not a given study has met the specific individual criterion used to assign the LoE. Table C4 provides a template for indicating whether the individual criterion is met or not. A blank for the criterion indicates that the criterion was not met, could not be determined or was not reported by the author.

Table C4. Definitions of the different levels of evidence for reliability studies

Level	Study type	Criteria
I	Good quality study	• Broad spectrum of persons with the expected condition • Adequate description of methods for replication • Blinded performance of tests, measurements or interpretation • Second test/interpretation performed independently of the first
II	Moderate quality	• Violation of any one of the criteria for a good quality study
III	Poor quality study	• Violation of any two of the criteria
IV	Very poor quality study	• Violation of all three of the criteria

Figure C2. Level of Evidence Algorithm – Reliability studies



Procedures for determining adherence to LoE criteria: Reliability studies
For these studies, the first performance or interpretation of the text is usually considered the "reference" and the second performance or interpretation the "test". Typical reliability studies are done using the same method (e.g., supine MRI) and include test-retest, inter- and intra-rater reliability. Statistical analysis is based on whether the same method or different methods are compared, the types of variables measured and the goal of the study.[68] In general, the degree (%) of concordance does not account for the role of chance agreement and is not a good index of reliability.[69] Different types of *kappa* (κ) or statistical correlation are frequently used to evaluate the role of chance.

Determination of the LoE involves evaluation of the following questions:

1. Was a **broad spectrum of persons with the suspected condition** used to determine reliability?

   The study population must be comprised of those with a broad spectrum of suspected disease who are likely to have the test now or in the future. Since differences in gender, age, body habitus and other characteristics may influence measurements and the ability to reproduce the results, the range of patients used for reliability studies is important.

Ideally a random sample of patients from the relevant clinical population would be used but may not be feasible, depending on the study. A broad spectrum would include patients with mild as well as more severe cases, those presenting early as well as late and those whose differential diagnosis may be commonly confused with the condition of interest. Reproducibility studies in a population with known disease may give different results compared with studies on a group of normal individuals and may not give an accurate picture of overall reproducibility. (If the goal of the study is to evaluate the potential for differential measurement error or bias, the separate analyses on "normal" and "diseased" populations should be done to evaluate the extent of such bias.[68] If it is a test-retest design, the test administrations should be on the same population. If it is an inter- or inter-rater reliability study the object (e.g., radiographs) should be the same for each reading/interpretation, (e.g., the same patients' radiographs are read twice).

2. Are the **details of the methods sufficient to allow study replication**?

   Is the description of the methods, i.e. the protocols used to collect information, measurements taken, planes of section, diagnostic criteria used, etc. sufficient that other investigators could duplicate the conditions and reproduce the findings in a similar population? Are the methods used for each part of the replication consistent?

3. Was there **blinded/independent performance of the repeat test administrations or interpretations**?

   The second administration of the test or second interpretation of results should be done without influence of the first test/interpretation. This is necessary to avoid bias. It must be clear from the text that both tests were interpreted without knowledge of the results of the other. Examples of when the administration would not be considered blinded or independent could include:
   - Interpretation of the second test is be done without prior knowledge of the test results or the first interpretation.
   - The timing of the second test administration or reading/interpretation of the results is not done such that sufficient time has elapsed between them to avoid influence of the first test/interpretation on the results of the second. In the case of re-administration of the test, the timing should not be so far apart that the stage/period of disease is different from the first administration.

Table C5. Assessment of level of evidence (LoE) for reliability studies

Methodological Principle	Author 1 (1999)	Author 2 (2002)	Author 3 (2004)	Author 4 (2005)
Broad spectrum of patients with expected condition	■	■		
Adequate description of methods for replication	■	■		
Blinded/independent comparison of tests/interpretations	■		■	
Evidence Level	**I**	**II**	**III**	**IV**

Determination of Overall Strength of Evidence

Following the assessment of the quality of each individual study included in the report, an overall "strength of evidence for the relevant question or topic is determined. Methods for determining the overall strength of evidence for diagnostic studies are variable across the literature and are most applicable to evaluation of therapeutic studies.

SRI's method incorporates the primary domains of quality (LoE), quantity of studies and consistency of results across studies as described by AHRQ.[67]

The following definitions are used by SRI to determine whether or not the body of evidence meets the criteria for each domain:

Table C6. Overall Strength of Evidence Domains

Domain	Definition/Criterion
Quality	• At least 80% of the studies are LoE I or II
Quantity	• There are at least three studies which are adequately powered to answer the study question
Consistency	• Study results would lead to a similar conclusion (similar values, in the same direction) in at least 70% of the studies

Based on the criteria described above, the possible scenarios that would be encountered are described below. Each scenario is ranked according to the impact that future research is likely to have on both the overall estimates of an effect and the confidence in the estimate. This ranking describes the overall "Strength of Evidence" (SoE) for the body of literature on a specific topic. The method and descriptions of overall strength are adapted for diagnostic studies from system described by the GRADE Working Group[66] for the development of clinical guidelines.

Table C7. Assessment of overall strength of evidence

SoE	Description	Further Research Impact	Domain Criterion Met		
			Quality	Quantity	Consistency
1	**High**	Very unlikely to change confidence in effect estimate	+	+	+
2	**Moderate**	Likely to have an important impact on confidence in estimate and *may* change the estimate	+	-	+
			+	+	-
3	**Low**	Very likely to have an important impact on confidence in estimate and *likely* to change the estimate	+	-	-
			-	+	+
4	**Very Low**	Any effect estimate is uncertain	-	+	-
			-	-	+
			-	-	-

The generalizability (or directness) of the study(ies) to various population is considered and addressed via narrative where applicable.

Appendix D. List of Excluded Studies and Rationale

The following tables describe studies that did not meet the inclusion criteria following review of the full-text.

Table D1. Studies that did not meet inclusion criteria

Author(citation)	Reason for Exclusion	Key*
Spine		
Andreasen et al 2007[70]	rMRI of the lumbar spine; no comparison with uMRI	ICNM (#2)
Danielson et al 1998[18]	Axial loading of the lumbar spine in the supine position; no comparison with uMRI.	EC (#12)
Hirasawa et al. 2007[12]	uMRI of the lumbar spine in individuals without spinal conditions of interest.	ICNM (#2)
Hiwatashi et al. 2004[20]	Axial loading of the lumbar spine in the supine position; no comparison with uMRI.	EC (#12)
Jinkins 2002[71]	Multiple case reports of select patients with fewer than 5 patients with any single diagnosis.	EC (#3)
Jinkins et al. 2002[72]	Same series of cases as Jinkins 2002; multiple case reports of selected patients with fewer than 5 patients with any single diagnosis.	EC (#3)
Jinkins 2003[73]	Same series of cases as Jinkins 2002; multiple case reports of selected patients with fewer than 5 patients with any single diagnosis.	EC (#3)
Jinkins et al 2003[74]	Same series of cases as Jinkins 2002; multiple case reports of selected patients with fewer than 5 patients with any single diagnosis.	EC (#3)
Jinkins et al. 2005[75]	Same series of cases as Jinkins 2002; multiple case reports of selected patients with fewer than 5 patients with any single diagnosis.	EC (#3)
Karadimas et al. 2006[11]	Comparison of anatomical changes and kinematics with uMRI vs. rMRI; no comparison of diagnosis.	ICNM (#2)
Kimura et al. 2001[56]	Axial loading of lumbar spine in supine position, no comparison with uMRI. Comparison of anatomical changes and kinematics with uMRI vs. rMRI, no comparison of diagnosis.	EC (#12)
Koschorek et al. 1986[76]	Supine dynamic MRI of the cervical spine, no comparison with uMRI.	EC (#12)
Manenti et al. 2003[19]	Axial loading of the lumbar spine in the supine position, no comparison with uMRI.	EC (#12)
Muhle et al. 1998[8]	MRI of the cervical spine in the supine position, no comparison with uMRI.	EC (#12)
Muhle et al. 1998[9]	MRI of the cervical spine in the supine position, no comparison with uMRI.	EC (#12)
Schmid et al. 1999[13]	uMRI of the lumbar spine in individuals without spinal conditions of interest.	ICNM (#2)
Singh et al. 2005[77]	rMRI of the lumbar spine, no comparison with uMRI.	ICNM (#2)
Vitaz et al. 2004[10]	uMRI of the cervical spine, no comparison with a current available diagnostic test.	ICNM (#2)
Vitzhum et al. 2000[7]	Comparison of anatomical changes and kinematics with uMRI vs. rMRI, no comparison of diagnosis.	ICNM (#2)
Willen et al. 1997[53]	Axial loading of the lumbar spine in the supine position; comparison with myelography but not with uMRI.	EC (#12)
Extremity		

Beaulieu et al. 1999[58]	uMRI of the shoulder in individuals without pathology.	ICNM (#2)
Cicuttini et al. 2005[78]	rMRI of the knee; no comparison with uMRI.	ICNM (#2)
Dufour et al. 2001[79]	uMRI of the shoulder assessing anatomy. No comparison to rMRI.	ICNM (#2)
Gedroyc et al. 2001[80]	uMRI of the knee assessing anatomy and kinematics. No comparison to rMRI.	ICNM (#2)
Gold et al. 2004[59]	Reliability of identifying anatomical features of the patella femoral joint in individuals without pathology.	ICNM (#2)
Graichen et al. 2000[81]	rMRI of the shoulder; no comparison with uMRI.	ICNM (#2)
Johal et al. 2005[63]	Reliability of identifying anatomical features of the tibio-femoral joint in individuals without pathology.	ICNM (#2)
Merl et al. 1999[82]	rMRI of the various anatomical joints; no comparison with uMRI.	ICNM (#2)
Moffet et al. 1998[57]	Reliability of identifying anatomical features of the shoulder joint in individuals without pathology.	ICNM (#2)
Powers et al. 2003[62]	uMRI assessing kinematics of the patella femoral joint under two conditions; no assessment of diagnosis.	ICNM (#2)
Ward et al. 2002[60]	Reliability of identifying anatomical features of the patella femoral joint in individuals without pathology.	ICNM (#2)

*ICNM = inclusion criteria not met (criteria #), EC = exclusion criteria met (criteria #). The criteria and associated numbers can be found in the text, pages 27 and 28.

Appendix E. Peer Reviewers

Peer Reviewer	Areas of expertise
Jens Chapman MD Professor, Dept of Orthopedic Surgery, University of Washington School of Medicine	• Surgical treatment of spinal disorders • Artificial disc replacement • Spinal Outcomes
Jennifer Mayfield, MD, MPH Primary Care and Preventative Medicine VA-Spokane	• Clinical diabetes care • Quality assessment and improvement • Chronic disease registries • Electronic medical records • Primary care • Health Services Research
Curtis P. Langlotz, MD, PhD Associate Professor, Dept of Radiology University of Pennsylvania School of Medicine Associate Professor of Epidemiology, Department of Biostatistics and Epidemiology, University of Pennsylvania	• Health Services Research • Imaging informatics

Appendix F. Evidence Tables for Included Studies-uMRI compared with other currently available diagnostic studies

Table F1. Overview of included studies of uMRI compared with other currently available imaging modalities.

							Exam 1								Exam 2						
Author	Study type		Patient		Outcome			Axial load			Position					Axial load			Position		
year	reliab	concord	symp	asymp	diagnosis	anatomy	image	yes	no		stand	sit	supine		image	yes	no		stand	sit	supine
Wildermuth et al 1998[46]	x	x	x		x		myel		x				n		uMRI	x				f,e	
Weishaupt et al. 2000[45]		x	x		x		rMRI		x				n		uMRI	x				f,e	
Ferreiro Perez et al. 2007[44]		x	x		x		rMRI		x				n		uMRI	x				n	
Zamani et al. 1998[47]		x	x	x	x	x	rMRI		x				n		uMRI	x				f,e	

reliab = reliability, concord = concordance, symp = symptomatic, asymp = asymptomatic, e=extension, f = flexion, n = neutral

Table F2. Description of concordance and reliability studies of the spine comparing upright MRI to other imaging modalities

Author (year)	LoE	Study Design	Demographics*	Patients	uMRI	rMRI	Protocol	Dx Criteria	Methods Concerns
Weishaupt (2000)	**IV**	Retrospective cohort Concordance	N=30 (n=76 disks) Mean age: 38 yrs (20-50) Male: 57% Race: NR	Low back and leg pain, not responsive to nonsurgical treatment, surgery not indicated or not urgent, disk protrusion ± extrusion without compression of neural structures required	n=30; **Specs:** 0.5T; T2 (4100/95, ETL=12) weighted; Matrix 256x192; ST= 4 mm	n=30; **Specs:** 1.0T; T1 (700/12, ETL=3) and T2 (5000/130, ETL=15) weighted; Matrix 512x210; ST = 4mm	**uMRI:** Seated, Flexion/Extension; Loaded **rMRI:** Supine; psoas relaxed **Assessed:** Quantitative:Dural sac diameter; Qualitative: Disk generation, Nerve root compromise, foraminal stenosis; pain (VAS) **Interp:** 2-3 radiologists **Blinding:** NR	**Disk abnormalities:** Grade 0: normal Grade 1: bulging Grade 2: protrusion Grade 3: extrusion Grade 4: sequestration **Nerve root compromise:** Grade 0: no contact Grade 1: contact, no deviation Grade 2: deviation Grade 3: compression **Foraminal size:** Grade 0: normal Grade 1: slight stenosis and epidural fat deformity, remaining fat surrounding root Grade 2: marked stenosis, epidural fat partially surrounding root Grade 3: advanced stenosis, obliteration epidural fat	• Narrow spectrum of patients, recruited after positive MRI • Six patients were excluded due to severe pain during uMRI

Table F2. Description of concordance and reliability studies of the spine comparing upright MRI to other imaging modalities (CONTINUED)

Author (year)	LoE	Study Design	Demographics*	Patients	uMRI	rMRI	Protocol	Dx Criteria	Methods Concerns
Zamani (1998)	**IV**	Retrospective cohort Concordance	N=15 pts (Including 4 healthy normals) Mean age: NR (22-79) Male: 64% Race: NR	Back pain, radicular pain, and claudication	n=15; **Specs:** 0.5T;T2 (3000/100, ETL=8) weighted; matrix= 256x192; ST= 4 mm	n=15; **Specs***:* same as uMRI, unclear if also same machine	**uMRI:** Seated, Flexion, and Neutral **rMRI**: Supine uMRI and sMRI performed ~2 days apart **Assessed:** Qualitative estimate of change in posterior disk bulge, foraminal and central canal size **Interp:** 2 neuroradiologists **Blinding:** NR	Qualitative evaluation of posterior disc bulge and neural foramina size	• Small sample size • Mixture of healthy asymptomatic and symptomatic individuals with no clear demarcation in reporting of results

Table F2. Description of concordance and reliability studies of the spine comparing upright MRI to other imaging modalities (CONTINUED)

Author (year)	LoE	Study Design	Demographics*	Patients	uMRI	rMRI	Protocol	Dx Criteria	Methods Concerns
Ferreiro Perez (2007)	**IV**	Retrospective cohort Concordance	N=89 Mean age: NR (20-60) Male: NR Race: NR	Low back or cervical spine pain, radiculopathy Exclusion: Subjects in whom accurate measurements could not be obtained	n=89; **Specs**: 0.6T; **1)** TI (350/20, ETL=NR) weighted, Matrix= 256x512, ST=4 mm; **2)** T2- (2000/120, ETL=NR) weighted; Matrix= 224x256; ST= 4 mm	n=89; **Specs**: same machine and specs as uMRI	**uMRI:** Neutral seated **rMRI**: recumbent **Assessed:** Posterior focal disc herniation, Anterior and posterior spondylolisthesis **Interp:** NR **Blinding:** NR	Quantitative measurement of area and linear dimensions of posterior disc herniations and residual patent central spinal canal. Grading and cut offs not reported.	• Patient motion artifact inhibited accurate measurements in 20% of images

Table F2. Description of concordance and reliability studies of the spine comparing upright MRI to other imaging modalities (CONTINUED)

Author (year)	LoE	Study Design	Demographics*	Patients	uMRI	rMRI	Protocol	Dx Criteria	Methods Concerns
Wildermuth (1998)	**IV**	Reliability	N=30 Mean age: 58 (27-84) Male: 43% Race: NR	Referred for lumbar myelography. spondylolisthesis, instability, segmental stenosis, persistent symptoms without diagnosis, and difficult postoperative situation; Exclusions: MR not available, patient underwent surgery immediately, patient unable to travel	n=30; **Specs:** 0.5T; T2 (3000/85, ETL=8) weighted; Matrix= 256x256; ST= 5 mm; Coil= body	**1) Lumbar myelography** n=30; **Specs**: Injection at L2-3 or L3-4. Fluoroscopic guidance to inject15 mL of iopamidol into spinal canal. Radiographs obtained with fluoroscopic guidance in lateral decubitus, prone, left and right PA oblique projections. **2) rMRI** n=30; **Specs**: Same as uMRI	**uMRI:** Seated, Flexion/Extension **Lumbar myelography:** Left lateral position, Flexion/Extension **sMRI:** Supine; rMRI images taken immediately prior to uMRI images. **Assessed:** Dural sac diameter, Foraminal size, Interobserver reliability **Interp:** 2 radiologists in six sessions two weeks apart **Blinding:** NR	**Foraminal size:** Grade 1: normal Grade 2: slight stenosis and epidural fat deformity Grade 3: marked stenosis, epidural fat partially surrounding root Grade 4: advanced stenosis, obliteration epidural fat	• Excluded patients unable to travel • Unable to quantitatively assess foraminal size due to scoliosis, severe pain, and changes due to sequence positioning rather than true changes • Patient recruited after results obtained from myelography (reference)

Table F3. Description of concordance studies of extraspinal joints comparing upright MRI to other imaging modalities

Author (year)	LoE	Study Design	Demographics*	Patients	uMRI	rMRI	Protocol	Dx Criteria	Methods Concerns
Weishaupt (2003)	**III**	Prospective cohort Concordance	N=18 Mean Age: 49.6 (25-72) Male: 6% Race: NR	All subjects had at least 1 Morton neuroma ≥ 5 mm in diameter	n=20 MNs, **Specs**: 0.5T T1 (500-600/19, ETL= NR) weighted, Matrix=256x 224; ST= 3mm	**rMRI (prone):** n= 20 MNs; 1. **Specs**: 1.0T; T1 (600/15, ETL=7) weighted; matrix=256x2 56; ST= 3mm; 2. **Specs:** 0.5T; T2 (4500/96, ETL=7) weighted; Matrix=256x2 56; ST= 3mm **rMRI (supine):** n=20 MNs; **Specs**: same as uMRI	**uMRI:** Standing **rMRI:** Prone (foot in planter flexion), supine (dorsiflexion of foot) **Assessed**: Visibility of neuroma **Interp:** 2 radiologists **Blinding:** Blinded to patient information	**Visibility score of MN**: 0=none 1=poor 2=moderate 3=good	• Small sample size • Limited to those subjects with a neuroma ≥ 5 mm in diameter

Table F3. Description of concordance studies of extraspinal joints comparing upright MRI to other imaging modalities (CONTINUED)

Author (year)	LoE	Study Design	Demographics	Patients	uMRI	rMRI	Protocol	Dx Criteria	Method Concerns
Hodge (2001)	**II**	Prospective cohort Concordance	N=11 Mean age= 25 (16-32) Male: 73% Race: NR	Unilaterally symptomatic or unstable shoulders (11 symptomatic shoulders and 8 asymptomatic shoulders)	n=11; **Specs**: 0.5T; weighted NR; 19.8msec/7.2msec; ETL= NR; Matrix= 256x128; ST= 7mm	**Clinical examination:** (n=11) **Examination under anesthesia:** (EAU) (n=10)	**uMRI:** Seated, each shoulder examined during abduction/adduction, internal/external rotation, and MR stress testing **EUA**: No information provided **Assessed:** Instability **Interp**: NR **Blinding:** NR	**MR instability:** Grade 1: humeral head shift within 25% of distance from glenoid center to rim; Grade 2: shift between 25-50%; Grade 3: shift greater than 50%. **Clinical grading:** Grade 0: No translation Grade 1: mild translation (0-1 cm) Grade 2: moderate translation (1-2 cm) Grade 3: severe translation (> 2 cm)	• Small sample size • Primary purpose to evaluate kinematics of asymptomatic and symptomatic individuals

Table F4. Recumbent and upright MRI visualization of pathology in 44 patients with cervical spinal symptomatology. Results: Ferreiro-Perez 2007.

	Patients with pathology n (%)
Posterior focal disc herniations identified by uMRI	31/44 (70.5)
Posterior focal disc herniations missed by rMRI	4/31 (12.9)
Posterior focal disc herniations underestimated by rMRI	21/31 (67.7)
Posterior focal disc herniations overestimated by rMRI	5/31 (16.1)

**rMRI= recumbent MR, uMRI=Upright MRI*

Table F5. Recumbent and upright MRI visualization of pathology in 45 patients with lumbar spinal symptomatology: Ferreiro-Perez 2007.

	Patients with pathology n (%)
Posterior focal disc herniations identified by uMRI	24/45 (53.3)
Posterior focal disc herniations missed by rMRI	2/24 (8.3)
Posterior focal disc herniations underestimated by rMRI	14/24 (58.3)
Posterior focal disc herniations overestimated by rMRI	4/24 (16.7)
Anterior spondylolisthesis missed by rMRI	4/13 (30.8)
Anterior spondylolisthesis underestimated by rMRI	7/11 (63.6)
Posterior spondylolisthesis overestimated by rMRI	2/2 (100)
Total pathology missed by recumbent-only MRI	6/37 (16.2)

**rMRI= recumbent MR; uMRI= Upright MRI*

Table F6. Visibility of Morton Neuroma on MR Images Obtained at Different Body Positions. Results: Weishaupt 2003.*

Visibility Score¥	Prone†	Supine‡	Weight-bearing
		n(%)	
0	0(0)	0(0)	0(0)
1	0(0)	0(0)	2(10)
2	0(0)	8(40)	8(40)
3	20(100)	12(60)	10(50)

** n(%) are Morton neuromas.*

† Plantar flexion of the foot

‡ Dorsiflexion of the foot

¥ Visibility Score: 0= None; 1=Poor; 2=Moderate; 3=Good

Prone vs. weight-bearing: p value =0.002; Supine vs. weight-bearing: p value =0.005; Prone vs. supine: p value=0.90 with Wilcoxon signed rank test

Table F7. Changes in disc form, nerve root compromise, and foraminal stenosis by position. Results: Weishaupt 2000.

Change in grading	Neutral vs. flexion	Neutral vs. extension
	n (%)	n (%)
Disk form		
Higher grade	0 (0.0)	6 (7.9)
Same grade	72 (94.7)	69 (90.8)
Lower grade	4 (5.3)	1 (1.3)
Overall change	5.3%	9.2%
Nerve root compromise		
Higher grade	32 (21.1)	26 (17.1)
Same grade	112 (73.7)	118 (77.6)
Lower grade	8 (5.3)	8 (5.3)
Overall change	26.4%	22.4%
Foraminal stenosis		
Higher grade	6 (3.9)	16 (10.5)
Same grade	128 (84.2)	130 (85.5)
Lower grade	18 (11.8)	6 (3.9)
Overall change	15.7%	14.5%

**Disk abnormality grade:s 0= normal, 1= bulging, 2= protrusion, 3= extrusion, 4= sequestration.*
**Nerve root compromise grade: 0= no contact, 1= contact, no deviation, 2= deviation, 3= compression*
**Foraminal stenosis grade: 0= normal, 1= slight stenosis and epidural fat deformity, remaining fat surrounding root, 2= marked stenosis, epidural fat partially surrounding root, 3= advanced stenosis, obliteration of epidural fat*

Table F8. Summary of effects of positional change on qualitative assessment of posterior disc bulge, foraminal size, and central canal size. Results: Zamani 1998.

	Supine Vs. Neutral
	N
Posterior Disc Bulge	
No change	15
Increased	0
Decreased	0
Unclear	0
Foraminal Size	
No change	15
Increased	0
Decreased	0
Unclear	0

** Data limited to those subjects who underwent both the uMRI and rMRI*
† NR= Not reported

Table F9. Foraminal scores for various positions. Results: Wildermuth 1998.

	Supine Neutral		Upright Flexion		Upright Extension	
Score	Observer 1	Observer 2	Observer 1	Observer 2	Observer 1	Observer 2
1	137	136	153	155	145	142
2	86	82	66	68	73	79
3	16	24	11	10	10	12
4	17	19	21	23	23	23
Total	256	261	251	256	251	256

Data are number of foramina considered to be adequately visible.
Foraminal score:
Grade 1- normal foramina
Grade 2- slight foraminal stenosis and deformity of the epidural fat
Grade 3- marked foraminal stenosis, with epidural fat only partially surrounding nerve root
Grade 4- advanced stenosis with obliteration of the epidural fat

Table F10. Correlation of mean sagittal diameters of dural sac: MR vs. Myelography. Results: Wildermuth 1998. *

Intervertebral Space	Supine Neutral	Flexion	Extension
L1-2	0.97	0.91	0.96
L2-3	0.97	0.86	0.90
L3-4	0.96	0.91	0.92
L4-5	0.93	0.94	0.96
L5-5	0.90	0.94	0.81

** Numbers represent correlation coefficients (reported by authors) comparing diameter as measured by MR with myelography*

Table F11. Changes in foraminal scores for comparing various positions. Results: Wildermuth 1998.

	Neutral vs. Flexion		Neutral vs. Extension	
Change in score	Observer 1	Observer 2	Observer 1	Observer 2
Higher	10	8	14	17
No change	242	249	235	236
Lower	4	4	2	3

Data are number of foramina considered to be adequately visible.
Foraminal score:
Grade 1- Normal foramina
Grade 2- slight foraminal stenosis and deformity of the epidural fat
Grade 3- marked foraminal stenosis, with epidural fat only partially surrounding nerve root
Grade 4- Advanced stenosis with obliteration of the epidural fat

Table. F12. Difference in mean sagittal diameters obtained using MR imaging by position, Results: Wildermuth 1998.*

Intervertebral Space	Supine vs. Flexion	Supine vs. Extension	Flexion vs. Extension
L1-2	NS	NS	NS
L2-3	NS	NS	NS
L3-4	S	NS	S
L4-5	S	NS	S
L5-5	S	NS	NS

** S=Paired student t test p value <0.05; NS= p value > 0.05*

Appendix G – Operational Definitions and Glossary of other Terms

Specific terms and operational definitions used in this technology assessment:

1. MRI Technology definitions
    - **MRI in the upright position** is referred to in this report as uMRI. The terms Upright™ MRI and Stand-up™ MRI are trademarks of FONAR Corporation. In the literature MRI performed in an upright position may also be referred to as positional MRI or standing MRI. These units have a magnet strength of 0.5T or 0.6T and allow for MRI in upright (sitting or standing) and weight-bearing states as well as in a recumbent position. Any open system with field strength of 0.5T or 0.6T, regardless of manufacturer is included under the heading of uMRI.
    - **Standard or conventional recumbent MRI** is referred to as rMRI in this report and have magnet strength of 1.0 to 3.0T. Exams are performed while the patient is lying down within a traditionally configured MRI unit where the bed does not allow for scanning in any position but supine or prone. To distinguish between exams done on a standard recumbent MRI and those where an Upright™ MRI or similar system is used to do an exam in the recumbent position, the magnet strength will be specified.

2. Condition definitions, clinical terms and diagnostic categories/classifications for comparison:
    - **Degenerative spondylolisthesis** refers to slippage of one vertebrae of the spine in relation to the adjacent vertebrae. A significant spondylolisthesis for the purpose of this systematic review is one where the slippage is >25%.
    - **Spinal stenosis** is the narrowing of the central spinal canal. For the purposes of this systematic review, moderate/severe central stenosis is defined as narrowing of the canal by $\geq$1/3.
    - **Lateral recess stenosis** is narrowing of the lateral recess or the intervertebral foramen. It is often accompanied by radicular symptoms in a specific dermatomal pattern; and pain at rest, at night, and with the Valsalva maneuver. It tends to be found more in patients who are younger (mean age 41 years) than patients with central canal stenosis (mean age 65 years).[83]

- **Radicular pain** is pain along the dermatome of a nerve due to pressure on the nerve root.
- **Non-specific spine pain** refers to pain not associated with neurological symptoms or signs.
- **Extra-spinal joint pain/function**. for the purposes of this systematic review, is defined as dysfunction of the appendicular skeletal system and its associated neuromuscular tissue.

3. Definitions pertaining to position:
   - **Supine neutral** refers to the supine position with a pillow under knees (psoas relaxed).
   - Supine extended position is defined as the supine position with the knees extended (psoas taut).
   - **Seated flexion** refers to the position where the patient is seated with a cushion placed on the legs, over which the body is bent.
   - **Seated extension** occurs when the patient is seated and leans back against an adjustable vertical back support that allows for differing degrees of extension.
   - **Seated neutral** occurs when the patient is seated upright without flexing or extending.
   - **Axial load** is a force exerted on the skull, spinal column, sternum or ribs
4. Diagnostic test characteristics and terms related to accuracy and reliability (Also see Appendix H):
   - **Sensitivity** is the proportion of patients testing/screening positive for the disease who actual have the disease.
   - **Specificity** is the proportion of patients testing/screening negative for the disease who are actually disease free.
   - **Positive predictive value** refers to the proportion of patients who have the disease of those who test/screen positive.
   - **Negative predictive value** is the proportion of patients who do not have the disease of those who test/screen negative.

- **Kappa** is a measurement of agreement beyond chance. Primarily used for intra-observer or inter-observer agreement, but can be used to as a measure of agreement between two tests when the reference test is not a gold standard.
- **Correlation coefficient** is a statistical measure of the interdependence of two or more continuous variables.
- **Concordance** refers to the degree of agreement.
- **Percent agreement** is the proportion of instances of agreement between two observers, not taking into account the role of chance.

5. Other definitions:

- **Diagnostic Impact**: The extent to which diagnostic test results influence the use of other diagnostic technologies. For example the test could replace another technology. Alternatively, the test could lead to the need for additional testing or obviate the need for additional testing. In order to determine the true diagnostic impact, information on the proportion of false positive and false negatives from validation (accuracy) studies is needed.
- **Therapeutic impact**: The extent to which diagnostic findings influence the selection and delivery of treatment. Information the diagnostic accuracy of the test for specific disease entity and the aspects common treatment options as well as patient outcomes are needed provide meaningful assessment of therapeutic impact.

GLOSSARY

Abduction - Movement away from the midline of the body.
Acute pain - Pain that generally lasts from a few days up to six weeks.

Adduction - Movement towards the midline of the body**.**
Axial load - Force exerted on the skull, spinal column, sternum or ribs

Contrast - The difference in signal intensity between two discrete areas of an image.
Contrast agent Any drug or material that is introduced to change the contrast between two tissues. MR contrast agents shorten the T1 and/or T2 relaxation times of tissue, thus improving the contrast-to-noise ratio of abnormal tissue.
Coronal plane The imaging plane that bisects the body into front and back parts.

Coupled movement The association of one motion (i.e., translation or rotation about an axis of rotation) with another motion about a second axis of rotation.

Cross-sectional area - The area of a material on a plane perpendicular to its longitudinal axis.

Dorsiflexion - Movement of the foot towards the anterior surface of the tibia while bending the ankle.

Dynamics - The study of forces acting on a body in motion.

Echo planar imaging (EPI) - A specialized MRI imaging technique or pulse sequence that is capable of producing images at rapid rates.

Echo time (TE) - The time between the center of the 90 degree pulse and the center of the spin-echo.

Extension - A straightening of a limb in which the bones making up the joint move to a more nearly parallel position.

Fast spin echo (FSE) A multiple echo spin-echo sequence that records different regions of k-space with different echos. Typically, a long repetition time (TR) multispin-echo pulse sequence where each echo is separately phase encoded.

Field of view (FOV) - The distance across an image, typically indicated in centimeters or millimeters; the size of the anatomical region that is imaged. The field of view in the frequency and phase encoding directions for an MR image may be different (the dimensions may be square or rectangular).

Flexion - The bending of a joint (i.e., the distal segment rotates toward the proximal segment).

Functional MRI - MRI technique used to evaluate or monitor physiological, anatomical, or metabolic processes.

Functional spinal units (FSU) - Two neighboring vertebrae and the interconnecting soft tissue, devoid of musculature.

Gadolinium - A lanthanide metal that has seven unpaired electrons. This paramagnetic metal is often used in MR contrast agents.

Gradient echo - A form of magnetic resonance signal produced by the refocusing of transverse magnetization caused by the application of a specific magnetic field gradient.

Gradient recalled echo sequence; gradient recalled echo in the steady state (GRASS) - An MRI pulse sequence that produces signals called gradient echoes as a result of the application of a refocusing echo. This type of pulse sequence is typically used to improve temporal resolution.

Inferior - The direction towards the feet in an anatomical coordinate system.

Inversion time (TI) The time between the inversion pulse and the sampling pulse(s) in an iversion recovery or STIR sequence.

Kinematic MRI - Any magnetic resonance imaging (MRI) technique used to assess joint function, including imaging the joint through a specific range of motion, during stress, or under loading condition.

Kinematics The branch of mechanics that deals with the motion of a body without reference to force or mass.

Magnetic resonance A phenomenon that results in the absorption or emission of electromagnetic energy by nuclei or electrons in the presence of a magnetic field after excitation by a resonance frequency pulse.

Magnetic resonance imaging (MRI) - The use of the magnetic resonance phenomenon to produce images of hydrogen or other protons.

Misregistration - The incorrect spatial mapping of an acquired MR signal. This act may be secondaty to motion, chemical shift, or wrap-around.

Number of excitations (NEX) - The number of signal averages used during the acquisition of an MR image.

Oblique plane A plane of imaging not perpendicular to the xyz coordinate system.

Open-chain - A descriptive term referring to when the foot is off the ground (i.e., the tibia is "free" during non-weight-bearing activity).

Plantarflexion - Movement of the foot away from the anterior surface of the tibia (i.e., straightening of the ankle joint).

Posterior The direction toward the back in an anatomical coordinate system.

Pulse sequence A series of radiofrequency (RF) pulses and magnetic field gradients, and time intervals between pulses applied to a spin system to produce a signal representative of some property of the spin system. For example, a Tl-weighted pulse sequence is indicated with the designation of repetition time (msec)/echo time (msec) as TR/TE, 300/20.

Radiofrequency (RF) - A frequency band in the electromagnetic spectrum with frequencies in the millions of cycles per second; frequencies of electromagnetic radiation often used in radio and television transmissions. For MRI, the RF used for imaging is dependent on the field strength of the MR system and typically ranges from 0.8 to 85 MHz.

Radiofrequency (RF) coil A device used for transmission or transmission and reception of magnetic resonance signals. RF coils are used to increase signal-to-noise and resolution.

Repetition time (TR) The time between the beginning of one pulse sequence and the beginning of the succeeding pulse sequence at a specified tissue location.

Sagittal plane A tomographic imaging plane bisecting the body into left and right parts.

Section thickness; slice thickness - The thickness of a slice of an MR image, usually indicated in millimeters.

Spin echo imaging In MR imaging, a spin echo is formed by the sequence of RF pulses and gradient reversals; an MRI sequence whose signal is an echo that results from the refocusing of magnetization after the application of 90 degree and 180 degree RF pulses.

Spoiled gradient echo; spoiled GRASS - Heavily T1-weighted gradient echo MR imaging technique, typically used for kinematic MRI examinations that require good temporal resolution.

Static magnetic field - The constant magnetic field of an MR system, usually indicated in Tesla.

Statics - The study of forces acting on a body in equilibrium.

STIR (short tau inversion recovery, short inversion-time recovery, short

T1 inversion recovery) - Inversion recovery MR imaging technique in which the TI- and T2-dependent contrasts are additive. This imaging technique is typically used to suppress signal from short T1 tissues (e.g., fat), thus reducing ghost artifacts and improving conspicuity of tissue that has increased fluid content.

Stress - Load per unit area that is produced on a plane surface within a structure in response to an externally applied load.

Subacute/delayed pain - Pain that persists for more than 3 months

Superior - The direction towards the head in an anatomical coordinate system.

Surface coil - A receive-only RF imaging coil that typically fits against the surface of the object being imaged. The use of receive-only RF coils facilitates imaging by improving signal-to-noise and resolution.

Synovial fluid - The fluid in a synovial joint.

TI-weighted image - An MR image obtained using a short repetition time and short echo time (short TR/TE) where the contrast is predominantly dependent on the T1 relaxation time of the tissue. Thus, this pulse sequence is commonly used to distinguish between tissues with differing T 1 relation times.

T2-weighted image - An MR image obtained using a long repetition time and long echo time (long TR/TE) where the contrast is predominantly dependent on the T2 relaxation time of the tissue.

T2* (T-two-star) - The spin-spin relaxation time composed of contributions from molecular interactions and inhomogeneities in the magnetic field. Contrast in gradient echo MR imaging depends on the T2* value.

TE - See Echo time (TE).

TR - See Repetition time (TR).

Tesla (T) - The SI unit of magnetic flux density. One Tesla equals 10,000 gauss (gauss is a cgs unit).

Torsion - A loading mode whereby the load is applied to a structure in a manner that causes it to twist about an axis, subjecting it to a combination of tensile, shear, and compressive forces.

Transverse plane - The imaging plane that bisects the body into top and bottom portions.

Appendix H. Overview of Diagnostic Test Validation and Reliability

Evaluation of validity and reliability studies

The accuracy of a diagnostic test consists of two general components: the accuracy of classifying patients with respect to their disease status (validity), and the degree to which repeated measures yield the same results (reliability). However, regardless of how accurate or predictive a test may be, health policy and public health perspectives assert that a diagnostic test should only be performed if it leads to the use of interventions that, on average, are likely to improve patient outcomes or it prevents the use of interventions that are not likely to improve outcomes.[84]

Validity and test accuracy

Validation of a measure refers to comparison of that measure against the true value. Validity is the degree to which a test ***accurately*** measures what it is intended to measure. Technically, an error free comparison method (i.e., true gold standard) is required in order to directly measure validity. For diagnostic tests, evaluation of the test against the "truth" allows the determination of how accurately the test classifies patients with and without disease. The accuracy of classification can be expressed by first accounting for the results as described in the following 2 x 2 table:

- True Positive (TP) results (cell a) = number of individuals with a disease who test positive
- False Positive (FP) results (cell b) = number of individuals without a disease who test positive
- False Negative(FN) results (cell c) = number of individuals with a disease who test negative
- True Negative (TN) results (cell d) = number of individuals without a disease who test negative

		True Classification	
		Disease present (+)	**Disease absent (-)**
Diagnostic Test	**Disease present(+)**	a = TP	b = FP
	Disease absent(-)	c = FN	d = TN
		a + c	b + d

The number of patients who truly have the disease is given by a + c and the number who truly do not have disease is given by b + d.

A true "gold standard" should be the definitive "truth" about the presence/absence of a condition or disease. Since an error-free method is not always available, a comparison of a diagnostic test to an appropriate reference standard, which may not be error-free, is commonly done. This referent method which is not always error free, may be better termed inter-method reliability.[68] An *appropriate* reference standard should be able to correctly classify patients with respect to the presence and absence of disease and be reproducible. However, variability in the test influences the ability to correctly classify patients according to disease status.

Sensitivity and specificity are the traditional measures of diagnostic tests used in validation to describe the accuracy of classification. They do not, however, describe the probability that a patient actually has the disease if the test is positive or does not have it if the test is negative.

Term	Definition		Calculation
Sensitivity =	% of patients with the disease who test positive	=	a/(a + c) x 100
Specificity =	% of patents who do NOT have disease who test negative	=	d/(b + d) x 100

The sensitivity and specificity are not fixed properties of a test. Instead, they reflect how the test performs among those with and without disease in a given population when administered in a specific manner. Sensitivity and specificity may appear to vary across populations, but do not directly depend on the prevalence of the condition.[69] Sensitivity and specificity form the basis of a receiver operating characteristic (ROC) curve which plots the relationship between the proportion of true positives (sensitivity) and the proportion of false positives (1-specificity) as a function of the diagnostic cut-off level for a disease.

When a true gold standard or appropriate reference standard is used, ***and*** the study population has a frequency of disease that approximates the frequency of disease in the population to which the results are to be applied (or the frequency of the disease in the population to which the test is to be applied is known), two additional measures of test accuracy can be used. These are the predictive value of a positive test (PPV) and the predictive value of a negative test (NPV) and are described as follows:

Term	Definition		Calculation
PPV =	% of patients with a positive test who have the disease	=	a/(a + b) x 100
NPV =	% of patents with a negative test who do NOT have the disease	=	d/(c + d) x 100

The PPV and NPV estimates are only accurate and meaningful if the actual proportion of true positives in the relevant population is represented by (a + c)/n. In other words, the actual prevalence of disease in the relevant population must be accurately estimated by the study population or it must be known for the population that is to be tested; otherwise, the predictive values are misleading.[69, 84] If the test is done in a population with a very low frequency of disease, for example, the PPV is quite low, even if the sensitivity and specificity are high.

Like PPV and NPV, most estimates of "overall accuracy" as an estimate of test validity vary with the prevalence of the disease or condition and can frequently lead to a distorted impression of a test's accuracy and validity.[84, 85] In addition, such measures do not fit into the decision making process as do PPV and PVN.[84] For these reasons, its (what does its refer to?) use is to be avoided.

Other measures of test performance include positive and negative likelihood ratios and the area under the receiver operator (ROC) curve. These measures are based on sensitivity and specificity and do not vary with disease prevalence even though they may vary across populations.[85]
Likelihood ratios (LR) are clinically useful and can be used to consider which test may be better for identifying the presence/absence of a disease. [ruling a disease in or out (in or out?)- not sure

what you were trying to say]. LR can also be valuable for comparing the accuracy of several tests to a gold standard. The LR is the ratio of the probability of a given test result in those with disease to the probability of that test result in people without the disease.

The likelihood ratio of a positive test (LRP) provides information about how well a positive test performs when a disease or condition is present compared with when disease is absent. The LRP describes how much the odds of disease *increase* when the test is positive. The likelihood ratio of a negative test provides information about how much the odds of disease *decrease* when the test is negative.

Term	Description		Calculation
LRP =	how much odds of disease *increase* with *positive* test	=	sensitivity/(1-specificity)
LRN =	how much odds of disease *decrease* with *negative* test	=	(1-sensitivity)/specificity

Likelihood ratios (LR) are combined with the pre-test odds of disease to determine the post-test odds of disease. The pre-test disease odds are based on disease prevalence, the patient population and individual patient characteristics. (The odds of disease can be determined from the probability of disease using Bayes' Theorum). The pretest odds are equal to the probability of having the disease divided by the probability of not having it.

LR provide insight into the extent to which doing the test is worthwhile in changing the odds of disease given the pre-test odds. The post-test odds, which represents the chance that the patient has the disease, thus incorporate disease prevalence, patient population information and patient-specific risk information via the pre-test odds as well as test performance information via the likelihood ratio as follows:

Post-test odds = Pre-test odds X likelihood ratio

If the test does not change the post-test odds of disease (e.g., a LR of 1), it is not likely to be helpful for ruling in (raising the post-test odds) or ruling out (lowering the post-test odds) disease. A test with a ***high LR*** is best to ***rule in*** a disease or condition while a test with a ***low LR*** is best to ***rule out*** a disease or condition.

While LR do not rely on disease prevalence, they are based on sensitivity and specificity of the test and therefore reflect how the test performs among those with and without disease in a given population when administered in a specific manner.

In the absence of validation studies, the concordance (percent agreement) was determined. Since this calculation does not take into account agreement that may be expected purely by chance, the kappa statistic (κ) was calculated where there were adequate data to correct for chance agreement according to the following formula:[69]

$$\kappa = (P_o - P_e) / (1 - P_e)$$

P_o is the observed concordance = (a+d)/N

P_e is the concordance expected by chance based on row and column totals:

$$P_e = \left[\frac{(a+b)(a+c)}{N} + \frac{(b+d)(c+d)}{N} \right] \Big/ N$$

Kappa describes the amount by which the observed agreement exceeds what would be expected by chance alone. While it can assist in putting results in perspective, there are several caveats that must be borne in mind. First, it is partly dependent on the true prevalence of the disease or characteristic in the population and declines as prevalence approaches 0 or 1.[86] Thus, it should not be viewed as a consistent property of the test comparison. In addition, although kappa is often used to adjust for the role of chance in studies that compare different methods (i.e. inter-method reliability studies), this is not the original intent for the application of the kappa statistic. It is most appropriately used in intra-method reliability studies described below. Guidelines for interpretation of kappa are provided by Landis and Koch. [51]

Reliability

The accuracy of a test also depends on its reliability. The purpose of reliability studies is to evaluate the reproducibility of a measure. That is, how well a measure can be replicated on, for example, a given patient, or imaging film, etc. under the same conditions. Even though a measure may be reproducible, it may still not be valid.

There are two general types of reliability studies:

Intra-method Reliability

- Test-retest reliability refers to the agreement when a test is done with the same instrument on the same subjects at two or more different times. Intra-rater reliability is test-retest reliability. This gives the upper limit of the extent to which the measure correlates with the truth or, $\rho_{XT,}$ where ρ is the correlation between the measure, X, and the truth, T.
- Inter-rater reliability refers to the agreement between two or more raters using the same instrument on the same subjects.

Inter-method reliability

- Refers to the agreement between two different instruments measuring the same underlying factor to yield similar results on the same subjects. Some refer to this as "validity" but technically, a "perfect" comparison is needed to determine validity. In certain circumstances, inter-method reliability studies can provide some information about the validity of a measure.[68]

Analysis of Reliability Studies

The following is an overview of common and appropriate statistical methods for reliability studies. There are two basic factors to consider when determining the appropriate analysis or statistical method: the type of study (i.e., intra-method or inter-method), and the type of variable (e.g., categorical). Additional information is found in Armstrong, White and Saracci.[68]

For dichotomous or categorical measures, in an intra-method study, percent agreement, and kappa are appropriate. For inter-method or validity studies where the categorical measure is nominal, a misclassification matrix is appropriate and for dichotomous variables (assuming reasonable ability to measure the true status), sensitivity and specificity can be determined. While kappa is sometimes used for these types of studies, it may not be an appropriate use of kappa. Continuous measures in an inter-method study are evaluated using Pearson's correlation coefficient.

Ordered categorical variables used in inter-method reliability or validity studies may be evaluated by the following methods:

- Misclassification matrix
- Spearman Rank correlation coefficient
- Pearson product moment correlation coefficient
- Either Spearman or Pearson on underlying variable from which variable was created

Continuous variables in intra-method reliability studies are generally evaluated using intra-class correlation coefficients. Cohen's kappa is used for evaluation of nominal or binomial variables while weighted kappa is most appropriate for ordered categorical variables.

Appendix I. Peer Reviewer Comment Summary and Disposition

Section	Comment	Disposition
Introduction	Comprehensive overview provided	No action required
	Topic is most relevant to allocate valuable healthcare resources to important areas and minimize unnecessary expenses.	No action required
Background	Objectives are clearly and fairly stated	No action required
Methods	Methodology appears well conceived and covered	No action required
Results	Results are comprehensively stated and readily identifiable	No action required
Conclusions	Conclusions are valid	No action required
Overall Presentation and Relevancy Comments	Review is well structured and presented in an organized fashion. The topic itself is important and timely in its discussion. Use of uMRI may have a marginal role in evaluation of some conditions. Arguably MRI as such is too readily available and in itself has never been subject to rigid research testing as to its utilization and sensitivity and specificity towards a number of common medical conditions. The greater public is well served with a balanced and critical review of this new technology. Prior to larger scale implementation specific indications should be clearly identified (i.e. Rheumatoid Arthritis patients with atlanto-axial instability or basivertebral invagination, or degenerative spondylolisthesis).	No action required
Quality of Report	Superior	No action required
	In addition to the exhaustive review provided I would suggest that in the future an evaluation also consider source of studies categorized by funding and potential for author bias. I.e. are the known or suspected ties to industry, and how the respective studies were funded.	No action required

Washington State Health Care Authority

Appendix J. Public Comment Summary and Disposition

Section	Comment	Disposition
Introduction		
Background		
Methods		
Results	I believe that the most remarkable finding of the report is the lack of studies on this technology. Based on the limited information that is available, it would appear that generally the studies are concordant (rMRI and uMRI). We do not know if studies with uMRI are likely to be superior or inferior when evaluating certain conditions involving the spine and related structures.	No action required
Conclusions		
Overall Presentation and Relevancy Comments	The report appears to be detailed, thorough and fair.	No action required
	There may be certain clinical situations where a clinician perceives a utility in ordering a uMRI. I would hope that in these circumstances a physician could obtain a thoughtful consideration of a request for an uMRI.	No action required
	Lastly, those of us who see patients on a day to day basis often are faced with situations in which there are no guidelines, or the guidelines have been followed and yet the clinical problem remains refractory. We need some latitude to act using our best clinical judgment and the best interest of the patient in distress.	No action required