Bone Scans Are Reliable for the Identification of Lumbar Disk and Facet Pathology

Gregory M. Malham1 Rhiannon M. Parker2 Zita E. Ballok3 Ben Goss4 Ashish D. Diwan5 Juan S. Uribe6

1 Neuroscience Institute, Epworth Hospital, Melbourne, Australia
2 Greg Malham Neurosurgeon, Melbourne, Australia
3 Nuclear Medicine Department, Primary Healthcare Imaging, Epworth Hospital, Melbourne, Australia
4 NuVasive Australia and NZ Pty Ltd, Melbourne, Australia
5 Department of Orthopaedic Surgery, St. George Private Hospital, University of New South Wales, Sydney, Australia
6 Department of Neurosurgery and Brain Repair, University of South Florida, Tampa, Florida, United States

Introduction

Osteoarthritis of the facet joints is a feature of degenerative spinal pathology. The facet joint has been shown to be a possible contributor to spinal pain, with up to 96% of patients reporting pain relief from intra-articular or periarticular injection.1 Hence, facet joint pathology has historically been identified through diagnostic injection of anesthetic and steroid agents into the joint space, with a positive test being the immediate relief of pain.2 This dichotomous test

Abstract

Study Design Surgeon survey.

Objective To evaluate the reliability of bone single-photon emission computed tomography (SPECT) versus bone SPECT images co-registered with computed tomography (bone SPECT-CT) by analyzing interobserver agreement for identification of the anatomical location of technetium99m-labeled oxidronate uptake in the lumbar disk and/or facet joint.

Methods Seven spine surgeons interpreted 20 bone scans: 10 conventional black-and-white tomograms (bone SPECT) and 10 color-graded bone SPECT-CT scans. Each surgeon was asked to identify the location of any diagnostically relevant uptake in the disk and/or facet joint between L1 and S1. Reliability was evaluated using the free-marginal kappa statistic, and the level of agreement was assessed using the Landis and Koch interpretation.

Results Conventional bone SPECT scans and bone SPECT-CT scans were reliable for the identification of diagnostically relevant uptake, with bone SPECT-CT having higher reliability (kappa = 0.72) than bone SPECT alone (0.59). Bone SPECT and bone SPECT-CT were also reliable in identifying disk pathology, with kappa values of 0.72 and 0.81, respectively. However, bone SPECT-CT was more reliable (0.81) than bone SPECT (0.60) when identifying facet disease.

Conclusions For the identification of disk pathology, it is reasonable to use either conventional bone SPECT or bone SPECT-CT; however, bone SPECT-CT is more reliable for facet joint pathology.
remains the gold standard for the identification of diagnostically relevant facet arthropathy. The challenge remains to reliably identify a priori which facet(s) is the most likely cause of the symptoms.

A systematic review of grading scales used to assess both disk and facet degeneration by Kettler and Wilke found the interobserver reliability of grading lumbar facet joint pathology by plain film, magnetic resonance imaging (MRI), or computed tomography (CT) to be moderate at best. Plain film, MRI, and CT have all been used to classify and describe features of facet pathology, with each having advantages and disadvantages. However, none of these techniques are able to reliably categorize pathology or predict outcomes. Pathria et al classified facet pathology into four grades: grade 0 = normal; grade 1 = joint space narrowing; grade 2 = narrowing plus sclerosis or hypertrophy; grade 3 = severe osteoarthritis with narrowing, sclerosis, and osteophytes. They found kappa values of 0.26 and 0.46 for the interobserver agreement using oblique radiographs and CT, respectively. Coste et al used a simplified CT-based 3-point grading scale that distinguished between mild and severe osteoarthritis by the presence of osteophytes. However, this scale only showed poor to slight agreement with kappa values between –0.01 and 0.12, depending on the level. The grading scale of Pathria et al4 was adapted by Weishaupt et al5 for assessing the agreement between MRI and CT; this included more precise definitions about the amount of articulating cartilage remaining, bone erosion, and subchondral cysts. They found moderate agreement between the two techniques (kappa = 0.60).

There have been numerous efforts to correlate morphologic measurements on imaging modalities with symptom relief from intra-articular injection. It has been suggested that relatively minor changes in small joints are difficult to observe, even with current technology. The high specificity of facet joint injection is related to its biological activity, hence the hypothesis of this study was that an alternative biologically active test would be more reliable than tests that rely on morphologic interpretation alone.

Positive uptake in the facet joint by bone scan and single-photon emission computed tomography (SPECT) correlates with low back pain. Dolan et al correlated uptake on SPECT with the relief obtained from intra-articular facet injection of local anesthetic. They also proposed that SPECT was an appropriate diagnostic tool for facet arthropathy. Few studies have been performed to assess the reliability of SPECT in the detection of facet arthropathy.

Conventional bone scans using planar imaging and SPECT have high sensitivity but low specificity for spinal anatomy. Image fusion software enables bone scan images to be co-registered with high-definition (HD) lumbar CT, allowing anatomical localization of regions of increased radiotracer uptake. This has demonstrated utility in refining diagnosis and treatment in degenerative spine pathology.

The aim of this study was to evaluate the reliability of bone SPECT versus bone SPECT images co-registered with CT (SPECT-CT) by analyzing interobserver agreement for identification of the anatomical location of technetium99m-labeled oxidronate (Tc99m–HDP) uptake in the lumbar disk, facet joint, or both.

**Methods**

**Study Protocol**

Seven spine surgeons each interpreted 20 bone scans: 10 conventional black-and-white tomograms (bone SPECT) and 10 color-graded bone SPECT-CT. This represents a comparison between the standard modality of bone SPECT scan and image fusion technology available from select institutions. The surgeons did not have access to the report from the nuclear medicine specialist (Z.B.). The surgeons were from different institutions in Australia (n = 5), New Zealand (n = 1), and United States (n = 1) and had an average of 12 years of consultant experience (range 5 to 30). The base specialty of the spine surgeons was orthopedic surgery (n = 4) or neurosurgery (n = 3). Each surgeon was asked to identify the location of any diagnostically relevant uptake in the disk, facet joint, or both at every level between L1 and S1 on each image. Disk pathology was correlated with the radioisotope uptake that occurs during the remodeling of the vertebral end plates (Modic type 1 and 2). Facet pathology was similarly correlated with subchondral bone remodeling that occurs in early stage facet arthropathy.

**Patient Population**

Twenty consecutive patients with anterior column lumbar degenerative spine disease underwent either bone SPECT or bone SPECT–CT. Radioisotope was used as part of the standard clinical assessment, and the reliability study was a retrospective examination of interobserver agreement. The use of bone SPECT or bone SPECT–CT coincided with availability of imaging technology at the treating institution. No patient was studied primarily to evaluate facet joint degeneration.

**Imaging Protocol**

Triple-phase bone scans were performed using a dual-head gamma camera (e.cam, Siemens, Erlangen, Germany). Following a bolus intravenous injection of 800 megabecquerels (MBq) of Tc99m-HDP (Mallinckrodt, St. Louis, Missouri, United States), 120 perfusion phase images each of 1-second duration were obtained, immediately followed by blood pool images obtained for 2 minutes using a 256 × 256 matrix.

Delayed static images were obtained 3 hours postinjection using a 256 × 256 matrix for 4 minutes, followed by SPECT images of the same spinal region. SPECT images were acquired in a 128 × 128 matrix with 60 projections over 360 degrees for 20 seconds per projection. Axial, coronal, and sagittal black-and-white tomograms with a slice thickness of 4 mm were reconstructed using iterative reconstruction (four iterations and eight subsets). Raw data of HD CT (Somatom Definition Flash, Siemens) images of the same spinal region were exported to the workstation (e.soft, Siemens) for co-registration. Image fusion software (Syngo, Siemens) was used for co-registration of bone SPECT scans with CT to provide color-graded images using the Warm Metal color.
scale (Siemens) with white-yellow indicating high Tc<sup>99m</sup>-HDP uptake, pink-purple indicating medium uptake, and green-blue indicating low uptake. As part of standard clinical procedure, MRI and CT were also performed for all patients.

### Statistical Analysis

Interobserver reliability was calculated as a free-marginal kappa statistic based on the formulae outlined by Randolph. A kappa value $\geq 0.81$ represents almost perfect agreement, 0.61 to 0.80 represents substantial agreement, 0.41 to 0.60 represents moderate agreement, 0.21 to 0.40 represents fair agreement, 0.01 to 0.20 represents slight agreement, and $\leq 0$ represents poor agreement.

### Results

- **Table 1** details the demographics of the patients analyzed in this study, divided into bone SPECT and bone SPECT-CT patients. There were no significant differences between the two groups. The mean age of the patient cohort was 63.9 years (range 51 to 74), and 11 (55%) were women. The primary diagnoses were scoliosis in 6 patients (30%), spondylolisthesis in 6 (30%), degenerative disk disease in 5 (25%), and herniated nucleus pulposus in 3 (15%). Six patients (30%) had prior lumbar surgery.

All surgeons were able to interpret all the images. **Table 2** outlines the reliability for bone SPECT and bone SPECT-CT in detecting disk and facet pathology. For identification of any diagnostically relevant uptake, both conventional bone SPECT scans and bone SPECT-CT scans were reliable, with bone SPECT-CT showing substantial agreement (kappa = 0.72) and bone SPECT showing moderate agreement (kappa = 0.59). Bone SPECT scans and bone SPECT-CT scans were both reliable in identifying disk pathology, with kappa values of 0.72 and 0.81, respectively. However, SPECT-CT was more reliable (kappa = 0.81) than conventional bone SPECT scans (kappa = 0.60) in identifying facet disease bone. **Figs. 1 to 4** show illustrative examples of high- and low-agreement scans for bone SPECT and bone SPECT-CT.

## Table 1 Demographics of study population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Bone SPECT (n = 10)</th>
<th>Bone SPECT-CT (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, y (stdev) (range)</td>
<td>64.5 (6.9) (51–74)</td>
<td>63.3 (6.4) (52–69)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>5 (50%)</td>
<td>6 (60%)</td>
</tr>
<tr>
<td>Mean BMI (stdev) (range)</td>
<td>25.7 (3.1) (21.7–29.4)</td>
<td>26.9 (6.3) (19.6–39.0)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobacco use (%)</td>
<td>1 (10)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Prior lumbar spine surgery (%)</td>
<td>2 (20)</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Primary diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scoliosis (%)</td>
<td>4 (40)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Spondylolisthesis (%)</td>
<td>3 (30)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Degenerative disk disease (%)</td>
<td>3 (30)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Herniated nucleus pulposus (%)</td>
<td>0</td>
<td>3 (30)</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; SPECT, single-photon emission computed tomography; SPECT-CT, single-photon emission computed tomography images co-registered with computed tomography; stdev, standard deviation.

## Table 2 Reliability analysis of bone SPECT scans versus SPECT-CT in detecting disk and facet pathology

<table>
<thead>
<tr>
<th>Overall agreement (%)</th>
<th>Free marginal kappa</th>
<th>Agreement (Landis and Koch criteria&lt;sup&gt;15&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any diagnostically relevant uptake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All levels, all scans</td>
<td>83</td>
<td>0.65</td>
</tr>
<tr>
<td>All levels, bone SPECT</td>
<td>79</td>
<td>0.59</td>
</tr>
<tr>
<td>All levels, SPECT-CT</td>
<td>86</td>
<td>0.72</td>
</tr>
<tr>
<td>Uptake in disk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All levels, all scans</td>
<td>88</td>
<td>0.76</td>
</tr>
<tr>
<td>All levels, bone SPECT</td>
<td>86</td>
<td>0.72</td>
</tr>
<tr>
<td>All levels, SPECT-CT</td>
<td>90</td>
<td>0.81</td>
</tr>
<tr>
<td>Uptake in facet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All levels, all scans</td>
<td>85</td>
<td>0.71</td>
</tr>
<tr>
<td>All levels, bone SPECT</td>
<td>80</td>
<td>0.60</td>
</tr>
<tr>
<td>All levels, SPECT-CT</td>
<td>81</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Abbreviations: SPECT, single-photon emission computed tomography; SPECT-CT, single-photon emission computed tomography images co-registered with computed tomography.
Fig. 1 Bone single-photon emission computed tomography scan with a high level of agreement between surgeons.

Fig. 2 Bone single-photon emission computed tomography scan with a low level of agreement between surgeons.

Fig. 3 Bone single-photon emission computed tomography images co-registered with computed tomography scan with a high level of agreement between surgeons.

Fig. 4 Bone single-photon emission computed tomography images co-registered with computed tomography scan with a low level of agreement between surgeons.
The levels were not distributed widely enough in the patient group to enable the determination of differential reliability at specific levels.

**Discussion**

There have been many attempts to utilize noninvasive methods to identify and classify degenerative facets using oblique plain films, CT, and both CT and MRI. However, all of these methods show moderate reliability at best.

Initially, facet condition was graded using plain film, but the complex geometry of the spine makes detailed imaging of this anatomical feature difficult to reproduce. Table 3 shows the reported reliability of each technique and classification scale. An early grading scale assessed by Pathria et al showed only fair agreement using oblique films (kappa = 0.26) and moderate agreement for CT (kappa = 0.46). Coste et al developed a different CT grading scale that was proven to be unreliable with interobserver agreement ranging from kappa values of −0.01 to 0.12, depending on the level.

Weishaupt et al showed moderate agreement (kappa = 0.41 and 0.60) between observers for MRI and CT, respectively. In a later multirater reliability study, Stieber et al showed that MRI had only fair to slight reliability (kappa = 0.21 and 0.07) and CT had fair reliability (kappa = 0.33 and 0.27), depending on experience.

A recently developed CT-based classification for facet joint degeneration in the cervical spine showed almost perfect intraobserver reliability. However, in the lumbar spine this grading system was previously shown to have only fair reliability.

The major limitations of conventional imaging techniques for the assessment of facet condition involve subjective interpretation of morphologic changes in the joint; this manifests in poor interobserver reliability of grading scales reported by many authors. Bone scans have several theoretical advantages over plain film, CT, and MRI. The radioisotope is involved in the physiologic processes of joint degeneration and therefore may have an increased sensitivity. Bone scans (or bone scintigraphy) rely on the uptake of radiolabeled markers (Tc⁹⁹m-HDP) to identify areas of increased bone deposition. Contemporary image fusion software permits the anatomical location of radioisotope uptake to be registered on an HD CT scan, which makes more accurate location of the pathologic process possible. Tc⁹⁹m has a half-life of 6.02 hours and decays to the nonradioactive isotope Tc⁰⁰ which is eliminated through the urine and stool. Hypersensitivity reactions to Tc⁹⁹m-HDP are extremely rare. The estimated effective radiation dose to a patient undergoing a Tc⁹⁹m-HDP bone scan is 2.5 milli-Sieverts (mSv). This is comparable to the annual background radiation dose of 1 to 2 mSv. Bone scans are considered relatively safe procedures but have certain limitations. The lack of a grading scale may be a disadvantage compared with MRI or CT, although this is not strictly required as the scans only show diagnostically relevant uptake. In addition, uptake of Tc⁹⁹m-HDP can be variable depending on metabolic differences between patients. Bone scans without SPECT or image fusion have historically failed to identify several abnormalities that can be sources of postsurgery spinal pain, especially pseudarthrosis. However, the addition of SPECT improves utility of bone scans in detecting pseudarthrosis as well as spinal osteomyelitis. McDonald et al reported that SPECT co-registered with CT allowed definitive localization of increased radioisotope uptake compared with high-resolution SPECT alone.

Prior spinal surgery may cause variation in radioisotope uptake; however, because our study aimed to evaluate the interobserver reliability of the anatomical location of positive radioisotope uptake, the findings were independent of prior surgery. The characteristics of signal intensity and their relationship to primary diagnosis were not examined as this relationship is not yet well understood. It is unknown whether differences in primary diagnosis lead to varying signal intensity or if there is a difference in biological activity between pathologies.

Limitations of our study include the small sample size, which may pose a low chance of bias as the outcome is focused toward interobserver reliability. However, the involvement of surgeons from different specialties, from different countries, and with varying experience may strengthen interpretation of the interobserver reliability.

**Table 3** Reliability of oblique plain films, CT, and MRI reported in the literature

<table>
<thead>
<tr>
<th>Study</th>
<th>Diagnostic test</th>
<th>Interobserver reliability (kappa)</th>
<th>Agreement (Landis and Koch criteria)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathria et al⁴</td>
<td>Oblique plain films</td>
<td>0.26</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>0.46</td>
<td>Moderate</td>
</tr>
<tr>
<td>Coste et al⁵</td>
<td>CT</td>
<td>−0.01–0.12a</td>
<td>Poor–slight</td>
</tr>
<tr>
<td>Weishaupt et al⁶</td>
<td>CT</td>
<td>0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>Stieber et al¹⁶</td>
<td>CT</td>
<td>0.33b–0.27c</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td>MRI</td>
<td>0.21b–0.07c</td>
<td>Fair–slight</td>
</tr>
</tbody>
</table>

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging.

aOverall range based on left/right facet joints and L4–5 and L5-S1 levels.
bAttending.
cFellows.
A further limitation is that we describe a dichotomous scale where surgeons are asked whether radioisotope uptake is diagnostically relevant. This scale does not attempt to assess severity of facet arthropathy. Because this technique relies on radioisotope uptake, quantification of the severity of symptoms based on signal intensity may be impossible. Despite efforts to administer a consistent dose, the magnitude of radioisotope uptake in the anatomical target may be limited by uptake from other pathologies outside the imaged region. Additionally, the rate of radioisotope excretion may be affected by other unknown causes, such as kidney disease, which can lead to greater or reduced circulating Tc99m-HDP loads and therefore resultant signal. The nature of kappa statistics is such that the reliability of a dichotomous scale is generally less than a multipoint scale, as chance agreement is more likely with two variables compared with more than two.20

The strengths of this study are that seven surgeons (both neurosurgeons and orthopedic surgeons) of varying degrees of experience were surveyed, none of whom routinely use bone scans for the primary identification of facet arthropathy. Our study demonstrates bone scans can be reliably interpreted to determine the anatomical location of increased Tc99m-HDP uptake in disk and facet joint pathology.

Conclusions

Both conventional bone SPECT and bone SPECT scans co-registered with CT are reliable for the identification of disk pathology; however, bone SPECT-CT scans showed greater agreement for localized facet joint pathology. This study suggests that bone SPECT-CT is a reliable tool to assist in surgical planning and identification of levels that may benefit from intervention. Further investigations are required to provide definitive treatment recommendations from these measurements.

Acknowledgments

The authors wish to thank Drs. Neil Cleaver, John Ferguson, Terrence Hillier, Kevin Seex, and Leong Tan for completing the bone scan survey.

Disclosures

Gregory M. Malham, none
Rhiannon M. Parker, none
Zita E. Ballok, none
Ben Goss, Employment: NuVasive; Stock/stock options: NuVasive
Ashish D. Diwan, none
Juan S. Uribe, Consultant: NuVasive; Patents: NuVasive; Royalties: NuVasive

References

Bone Scans for Lumbar Disk and Facet Pathology

Malham et al.


30 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70(4):213–220