Whole-body magnetic resonance imaging is superior to skeletal scintigraphy for the detection of bone metastatic tumors: a meta-analysis

G. SUN¹, Y.-X. ZHANG², F. LIU³, N. TU⁴

¹Department of Radiology, Zaozhuang Hospital of Traditional Chinese Medicine, Zaozhuang, Shandong, China
²No. 2 Department of Orthopedics, Zaozhuang Hospital of Traditional Chinese Medicine, Zaozhuang, Shandong, China
³Zaozhuang Hospital of Traditional Chinese Medicine, Zaozhuang, Shandong, China
⁴Department of Magnetic Resonance, Zaozhuang Municipal Hospital, Zaozhuang, Shandong, China

Gang Sun and Yongxiang Zhang contribute equally to this work

Abstract. – OBJECTIVE: The meta-analysis aims to compare the diagnostic performance of whole-body magnetic resonance imaging (MRI) and skeletal scintigraphy (SS) for the detection of skeletal metastases.

MATERIALS AND METHODS: We searched Medline, Scopus, Embase and Cochrane library databases for identifying fifteen eligible studies with a total of 1939 participants, and the quality of these studies was assessed according to Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) guidelines. Sensitivities, specificities, diagnostic odds ratios (DOR), positive likelihood ratios (PLR), and negative likelihood ratios (NLR) were calculated. Summary receiver operating characteristic curves (sROC) were generated using bivariate models for whole-body MRI and skeletal scintigraphy.

RESULTS: Whole-body MRI had higher but comparable patient-based higher specificity compared to SS (99% vs. 95%). However, it had markedly higher sensitivity (94% vs. 80% respectively), DOR (966 vs. 82), and LPR (54.4 vs 17.1). LNR of whole-body MRI was <0.1 (0.06), while LNR of SS was >0.1 (0.22). The area under curves (AUC) for whole-body MRI and SS were 0.99 and 0.95 respectively.

CONCLUSIONS: We demonstrate that both whole-body MRI and SS have good diagnostic performance. However, MRI is superior for diagnostics of bone metastases, as it has higher sensitivity, higher diagnostic accuracy, and can be used for both confirmation and exclusion of metastatic bone disease.

Key Words: Magnetic resonance imaging, Skeletal scintigraphy, Bone metastatic tumors, Meta-analysis.

Introduction

Any type of metastatic cancer that spreads via the bloodstream can infiltrate the bone marrow and give rise to bone metastases. Certain types of cancers, such as prostate or breast cancer, are known for their ability to cause skeletal metastases, with a prevalence of up to 70%. There is high variability in the metabolic activity of different types of bone metastases. It is therefore important to choose the suitable diagnostic method among a variety of radiological and nuclear medical imaging techniques for the detection of bone metastases in patients with certain types of cancers.

Skeletal scintigraphy (SS) with labeled phosphonates (¹⁹⁸⁰TC-phosphonates) is routinely used for the detection of local bone metabolism in an early phase of some types of cancers. Therefore, SS is most effective in the visualization of metastases that are associated with reactive hypermetabolism of bone. This includes metastases of prostate and breast cancer, neuroendocrine tumors, and osteosarcomas. However, SS is relatively insensitive for tumors that are not hypermetabolic and can lead to false-positive results in cases of post-treatment bone matrix regeneration (flare phenomenon).

Magnetic resonance imaging (MRI) is quickly becoming a method of choice for detecting bone metastases due to its high soft-tissue contrast, high spatial resolution, and no requirements for intravenous contrast medium. In the 2011 me-
ta-analysis, Yang et al\textsuperscript{11} showed that MRI had 91% sensitivity and 95% specificity, being superior to planar skeletal scintigraphy. These findings were further confirmed in subsequent studies involving, among others, breast, prostate, lung and renal cancers\textsuperscript{3\textsuperscript{-}23}. However, these studies have certain limitations, such as a lack of pathological verification of skeletal metastases, selection bias, and not taking into consideration an improved SS with the use of more advanced single-photon-emission computerized tomography/combined with CT (SPECT/SPECT-CT) apparatus\textsuperscript{17,24,25}.

The purpose of this meta-analysis is to compare the efficiency of whole-body MRI and skeletal scintigraphy in the detection and characterization of skeletal metastases.

### Materials and Methods

**Eligibility Criteria**

The inclusion criteria were as follows: studies comparing the diagnostic performance of skeletal scintigraphy and whole-body MRI irrespective of study design employed; studies that report required statistics of the above-mentioned techniques or provide data to calculate these rates; full-text studies or published as conference abstracts were included while unpublished data, case reports, and studies with smaller sample size (fixed at 10 for the current review) were excluded.

Participants: patients with a primary malignant tumor in sites other than skeletal sites.

Index test: studies that used skeletal scintigraphy and whole-body MRI for the identification of bone metastasis.

Reference standards: studies where the diagnostic accuracy is compared with a definitive diagnosis of bone metastasis by histopathological or biopsy findings.

**Type of Outcome Measure**

Pooled sensitivity, specificity, positive likelihood ratio (PLR) and negative likelihood ratio (NLR) and diagnostic odds ratio (DOR).

**Search Strategy**

For this meta-analysis, we identified relevant studies by searching the following Medline, Scopus, Embase and Cochrane library databases. The following medical subject headings (MeSH terms) and free-text terms were used in PubMed in several combinations: “Validation Studies”, “Bone Metastasis”, “Skeletal Metastasis”, “Bone Scintigraphy”, “Skeletal Scintigraphy”, “Magnetic Resonance Imaging”, “Sensitivity”, “Specificity”, “Diagnosis”, and “Diagnostic Accuracy Studies”. Similar terms were also used in Cochrane library, Scopus, Embase for literature search of published studies. The time period was from inception to December 2019 without any language restrictions. Bibliographies of retrieved studies were searched, and eligible studies were included.

**Study Selection**

Two authors independently screened the title and abstract of the records identified during the literature search. The full-text article was retrieved for studies deemed relevant. The further full-text screening was again done by the two authors independently and those studies matching the inclusion criteria were finally included in our review. Disagreements between the two authors during this process were solved via consultation with a third investigator.

**Data Extraction Process**

Data extraction for the required characteristics from the included studies was done by the primary investigator. The data extracted were study design, setting, index test, reference standards (gold standard/comparator), comorbidities, the total number of participants, average age, inclusion, exclusion criteria, sensitivity, and specificity. The extracted data were transferred into STATA software.

**Risk of Bias Assessment In Included Studies**

The risk of bias was assessed by two authors independently using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Domains used for assessing the risk of bias were the selection of patients, characteristics of index test and reference standard, timing, and flow of assessments. The risk of bias was finally interpreted as low, high or unclear.

**Statistical Analysis**

Meta-analysis was done with the selected studies using STATA 14.2 software (StataCorp, College Station, TX, USA). Bivariate meta-analysis was done to obtain the pooled estimates of the diagnostic accuracy estimates like sensitivity and specificity, PLR and NLR, and DOR for MRI and Skeletal Scintigraphy. Summary Receiver Operator Characteristic curves (sROC) was made. The summary estimate obtained in the sROC was the area under the curve (AUC) with 95% confidence.
interval (CI). AUC value closer to 1 indicates the higher diagnostic performance of the imaging techniques.

Forest plot was used to graphically represent the study level and overall pooled diagnostic measures. Likelihood ratio (LR) scattergram was made to find the clinical significance of these imaging techniques. Fagan plot was constructed to demonstrate how much the result of MRI or Skeletal Scintigraphy changes the probability that a patient has a diagnosis of bone metastases.

Heterogeneity was evaluated via the following methods: graphical representation through bivariate box plot, Chi-square test for heterogeneity and F statistics to quantify the inconsistency. Potential sources of heterogeneity were explored by meta-regression using potential predictive covariates like quality-related factors (under QUADAS). A funnel plot was used to graphically represent the publication bias and tested by Deek’s test.

### Results

#### Search Results

We have conducted a systematic search to find studies reporting the diagnostic performance of MRI and SS for the diagnosis of bone metastases. Totally, 2834 records found, out of which 1189 studies from Medline, 835 from Scopus, 598 from Embase, and 212 from the Cochrane library (Figure 1). After the removal of duplicates, the remaining 2178 studies were subjected to the title, abstract and keywords screening. At that stage, 1981 studies were eliminated, due to different outcomes (n=1480) and irrelevant diagnoses (n=501). Full text of 197 relevant studies was reviewed for eligibility criteria. Of them, 111 studies were eliminated due to different outcomes, 21 studies described different diseases, and 50 studies were excluded after the bibliographic review. The remaining 15 potential studies

![Flowchart](image)
Whole-body MRI and detection of bone metastatic tumors: a meta-analysis

satisfied the eligibility criteria, with 1939 participants (Figure 1).

Characteristics of the Included Studies

Characteristics of the studies were described in Table I. Majority (12) of included studies are prospective studies\(^{14,16,17,21,26-33}\), 2 studies were retrospective\(^{34,35}\), and one study was a clinical trial\(^{36}\). The age of participants ranged from 2 to 82 years. In total, 1939 participants were found in the included studies with sample size varying from 22 to 1025, and the follow-up time ranging from 6 months to 2 years. All the included studies have documented histology or biopsy results as a reference standard. Majority of the studies (13) used 1.5T strength MRI\(^{14,16,17,21,26-29,31-34,36}\). The amount of \(^{99m}\)Tc ranged from 500 to 925 MBq (Table I).

The Methodological Quality of the Included Studies

Patient selection bias assessment demonstrated that almost 75% of the studies had a low risk of bias. Assessment of bias due to conduct and interpretation of the index test domain showed that 60% of the studies had a low risk of bias. 75% of the included studies had a low risk of bias for conduct and interpretation of reference standards, while only 5% of the studies demonstrated bias due to flow and timing of assessments (Figure 2).

Diagnostic Performance of MRI and SS

As summarized in Figure 3 and Table II, pooled sensitivity of MRI was 94% (95% CI: 87%-98%), pooled specificity was 98% (95% CI: 97%-99%) and DOR was 966 (95% CI: 264-3543). PLR was 54.4 (95% CI: 27.3-108.3) and NLR was...
Table I. Characteristics of the included studies (N=15).

<table>
<thead>
<tr>
<th>Study No</th>
<th>First author and year</th>
<th>Country</th>
<th>Study design</th>
<th>Sample size</th>
<th>Follow-up time &amp; Sequences</th>
<th>MRI Strength</th>
<th>Amount of 99MTC &amp; Delay time</th>
<th>Mean age (in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Daldrup-Link et al 2001</td>
<td>Germany</td>
<td>Prospective</td>
<td>39</td>
<td>&gt;10 months</td>
<td>1.5 T T1, T2, STIR</td>
<td>740 MBq 3-4h</td>
<td>2-19</td>
</tr>
<tr>
<td>2</td>
<td>Engelhard et al 2004</td>
<td>Germany</td>
<td>Prospective</td>
<td>22</td>
<td>&gt;12 months</td>
<td>1.5 T T1, T2, STIR</td>
<td>550 MBq 2-3h</td>
<td>53-87</td>
</tr>
<tr>
<td>3</td>
<td>Sohaib et al 2009</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>47</td>
<td>&gt;12 months</td>
<td>1.5 T T1, STIR</td>
<td>600 MBq 3h</td>
<td>29-79</td>
</tr>
<tr>
<td>4</td>
<td>Takenaka et al 2009</td>
<td>Japan</td>
<td>Prospective</td>
<td>115</td>
<td>&gt;12 months</td>
<td>1.5 T T1, STIR</td>
<td>555 MBq 3h</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>Venkitaraman et al 2009</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>39</td>
<td>&gt;667 days</td>
<td>1.5 T T1, STIR</td>
<td>640 MBq 3h</td>
<td>54-82</td>
</tr>
<tr>
<td>6</td>
<td>Balliu et al 2010</td>
<td>Spain</td>
<td>Prospective</td>
<td>40</td>
<td>&gt;12 months</td>
<td>1.5 T T1, STIR</td>
<td>925 MBq 2 hours</td>
<td>62.1</td>
</tr>
<tr>
<td>7</td>
<td>Kim et al 2009</td>
<td>Korea</td>
<td>Retrospective</td>
<td>134</td>
<td>2 years</td>
<td>3T T1, T2, STIR</td>
<td>750 MBq 3 hours</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>Lecouvet et al 2012</td>
<td>Belgium</td>
<td>Prospective study</td>
<td>100</td>
<td>6 months</td>
<td>1.5 T T1, STIR</td>
<td>Not specified</td>
<td>69</td>
</tr>
<tr>
<td>9</td>
<td>Altehoefer et al 2001</td>
<td>Germany</td>
<td>Retrospective study</td>
<td>81</td>
<td>11 months</td>
<td>3T T1, T2, STIR</td>
<td>500-650 MBq 2-4h</td>
<td>Not specified</td>
</tr>
<tr>
<td>10</td>
<td>Earnest et al 1999</td>
<td>United States of America</td>
<td>Prospective</td>
<td>29</td>
<td>12 months</td>
<td>1.5 T T1, T2, STIR</td>
<td>Not specified</td>
<td>67.2</td>
</tr>
<tr>
<td>11</td>
<td>Jambor et al 2015</td>
<td>Finland</td>
<td>Clinical trial</td>
<td>53</td>
<td>15 months</td>
<td>1.5 T T1, T2, STIR</td>
<td>670 MBq 3h</td>
<td>Not specified</td>
</tr>
<tr>
<td>12</td>
<td>Kumar et al 2008</td>
<td>India</td>
<td>Retrospective</td>
<td>208</td>
<td>16 months</td>
<td>1.5 T T1</td>
<td>2-4 hours</td>
<td>Not specified</td>
</tr>
<tr>
<td>13</td>
<td>Layer et al 1999</td>
<td>Germany</td>
<td>Prospective</td>
<td>33</td>
<td>Not specified</td>
<td>1.5 T T1</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>14</td>
<td>Ohlmann-Knafo et al 2009</td>
<td>Germany</td>
<td>Prospective randomized controlled trial</td>
<td>45</td>
<td>Not specified</td>
<td>1.5 T T1, T2, STIR</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>15</td>
<td>Xu et al 2008</td>
<td>China</td>
<td>Prospective</td>
<td>45</td>
<td>2 months</td>
<td>1.5 T T1, T2, STIR</td>
<td>2 hours</td>
<td>52.7</td>
</tr>
</tbody>
</table>

DOR: diagnostic odds ratio; LPR: positive likelihood ratios; LNR: negative likelihood ratio; AUC: area under sROC curve; 95% CI: 95% confidence interval.
**Figure 4.** Forest plot showing pooled sensitivity and specificity for SS.

**Figure 5.** Likelihood scattergram for MRI.
The pooled sensitivity of skeletal scintigraphy was 80% (95% CI: 68%-89%), pooled specificity was 95% (95% CI: 88%-98%) and DOR was 82 (95% CI: 27-248). PLR was 17.1 (95% CI: 6.6-43.9) and NLR was 0.21 (0.12-0.35) (Table II, Figure 4).

An LR scattergram was then generated to assess the clinical performance of whole-body MRI and SS diagnostic methods (Figures 5 and 6). For whole-body MRI, six of the included studies were plotted in the left upper quadrant of the scattergram\(^\text{14,16,21,26,27,29}\), while 6 additional studies were plotted to the upper right quadrant\(^\text{17,30,31,34-36}\), one study in the lower left quadrant\(^\text{28}\), and one study in the lower right quadrant\(^\text{32}\). A summarized PLR and NLR was localized to the left upper quadrant (95% CI) (Figure 5). For the SS, 6 of the included studies were plotted to the upper-right quadrant of the LR\(^\text{16,17,26,31,34}\), 6 to the lower-right quadrant\(^\text{14,21,28,30,32,35}\), 2 studies to the upper-left quadrant\(^\text{29,36}\), and one study to the lower left quadrant\(^\text{27}\).

We then constructed an sROC to assess the tradeoff between sensitivity and specificity. The AUC was 0.99 (95% CI, 1.00 to 0.00) for the MRI (Table II, Figure 7), and 0.95 (CI 95%, 1.00 to 0.00) for the SS (Table II, Figure 8).

Figures 9 and 10 show Fagan’s nomogram for LR, found post-test probabilities from various pre-test probabilities. Our results indicated that

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>DOR (95% CI)</th>
<th>LPR (95% CI)</th>
<th>LNR (95% CI)</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-body MRI</td>
<td>94% (87%-98%)</td>
<td>99% (97%-99%)</td>
<td>966 (264-3543)</td>
<td>54.4 (27.3-108.3)</td>
<td>0.06 (0.02-0.13)</td>
<td>0.99 (1.00-0.0)</td>
</tr>
<tr>
<td>Skeletal scintigraphy</td>
<td>80% (68%-89%)</td>
<td>95% (88%-98%)</td>
<td>82 (27-248)</td>
<td>17.1 (6.6-43.9)</td>
<td>0.21 (0.12-0.3)</td>
<td>0.95 (1.00-0.0)</td>
</tr>
</tbody>
</table>

DOR: diagnostic odds ratio; LPR: positive likelihood ratios; LNR: negative likelihood ratio; AUC: area under sROC curve; 95% CI: 95% confidence interval.
post-test probability for MRI (94% positive; 2% negative, Figure 9) and SS (84% positive, 6% negative, Figure 10) differed significantly from pre-test probability (23%).

**Assessment of Heterogeneity and Publication Bias**

There was substantial heterogeneity with the $I^2$ value of 99% and a chi-square test was significant ($p<0.001$). Bivariate box plot (Figure 11) shows that about 3 studies were out of the circle indicating the heterogeneity between the included studies. Meta-regression results indicated that patient selection and flow, and timing of the test standard was the potential sources of heterogeneity in the model ($p<0.05$) (Figure 12). Finally, publication bias was tested through the funnel plot (Figure 13) and asymmetry of the plot was assessed using Deek’s test. The symmetry of the funnel plot and the results of the Deek’s test suggest no significant publication bias ($p=0.09$).

**Discussion**

To our knowledge, our meta-analysis is the first study that clearly shows that whole-body MRI is of greater sensitivity and higher, but comparable specificity when compared to skeletal scintigraphy. MRI has higher diagnostic accuracy and can be used for both confirmation and exclusion, while SS can be used for confirmation only.

Recent advances in the treatment of many cancers mean that there is an increase in the life expectancy of many patients with metastatic disease and a greater chance of developing skeletal metastases. At present, no single imaging strategy is consistently superior for the early diagnosis and assessment of metastatic bone cancers.

While being widely used in diagnostics of bone metastases, skeletal scintigraphy’s main limitation is its sensitivity and specificity. The sensitivity of $^{99}$Tc scintigraphy has been reported to range from 62 to 89%, with a false-positive rate as high as 40%. SS relies on the osteoblastic response to skeletal destruction by tumor cells and the accompanying increase in regional blood flow.

Whole-body MRI is now feasible in scan times of less than 1 h, no need for contrast material, and can identify bone metastases early on, before the onset of host osteoblasts reaction. While several studies addressed the differences in the efficiency of SS and MRI in the diagnosis of skeletal metastases, the results were inconclusive. In the last meta-analysis by Wu et al, that included 7 studies, both methods had a comparable diagnostic performance for detecting bone metastatic tumors. However, the analysis was unable to determine what method is superior in detecting bone metastases. In our meta-analysis, we used summary estimates and SROC curves to assess the diagnostic accuracy of whole-body MRI and skeletal scintigraphy from 15 studies.
Whole-body MRI showed higher sensitivity than SS (an increase of 14%), and slightly higher, but comparable specificity (99% vs. 95%) in detecting bone metastases. The AUC is an index of the overall performance of a test with values ranging between 1 and 0, where 1 indicates a perfect test that correctly distinguishes between cases of disease and non-cases of disease, and value of 0 indicates a test that fails to diagnose. In our study, both SS and MRI had comparably high AUC values (0.99 and 0.95 respectively), indicating that these two methods are highly effective in diagnostics on skeletal metastases.

Fagan’s nomogram is a tool to determine diagnostic test characteristics (sensitivity, specificity, likelihood ratios) and/or determine the post-test probability of disease given the pre-test probability and test characteristics. In our study, for both methods pre-test probability (23%) differed significantly from the post-test one (94% positive; 2% negative for MRI, and 84% positive, 6% negative for SS).

Diagnostic odds ratio (DOR) was used to investigate multiple relationships between the chances of obtaining positive and negative results. DOR is a comprehensive evaluation index in diagnostic tests, as it shows the ratio of the odds of a positive test in a patient with disease relative to the odds of the positive test in a patient that does not have a disease. A good diagnostic test should have DOR>100, with LPR above 10 and LNR<0.1. We demonstrated that MRI performs better as a diagnostic imaging method, with DOR of 966, as compared to DOR of 82 for SS. We used LPR and LNR as measures of diagnostic accuracy, as higher LPR in combination with lower LNR. Our analysis shows that whole-body MRI has 3.18-fold higher LPR than SS (54.4 vs. 17.1 respectively) in combination with an LNR value of 0.06, which is below 0.1. This positions a summary LPR and LNR in the upper left quadrant of the likelihood scattergram, indicating that MRI can be used for...
Whole-body MRI and detection of bone metastatic tumors: a meta-analysis

**Figure 12.** Meta regression for sources of heterogeneity among the studies included.

**Figure 13.** Funnel plot for publication bias.
both confirmation and exclusion of bone metastases. At the same time, while the LPR value of SS was still above 10 (17.1), the LNR value of SS was 0.22, which is above 0.1. The likelihood scattergram clearly showed that SS can be used for confirmation only. Taken together, DOR, LNR and LPR values show that whole-body MRI has a better discriminating ability than SS. Moreover, as indicated by the LNR value of MRI, our analysis suggests that this method can be used alone to rule out bone metastases.

Our meta-analysis has some limitations. There was substantial heterogeneity between the included studies. It is well established that differences in study design and patient selection have a substantial impact on estimates of diagnostic accuracy. In studies, where patients were selected on the basis of whether they had been referred for the index test rather than on clinical symptoms, the diagnostic accuracy may be lower, while in retrospective studies, and in studies with nonconsecutive inclusion of patients, the accuracy may be higher\(^{44}\).

**Conclusions**

In summary, we showed for the first time a clear advantage of using whole-body MRI for the diagnosis of bone metastases. It has higher sensitivity than skeletal scintigraphy, as well as higher diagnostic accuracy, and can be used for both confirmation and exclusion of metastatic bone disease.

**Authors’ contributions**

GS and YZ designed the meta-analysis; FL and NT searched the literature; GS, YZ and FL analyzed the literature; GS and YZ wrote the manuscript; NT edited the manuscript.

**Conflict of Interest**

The Authors declare that they have no conflict of interests.

**References**


15) **Leclouet FE, El Moueden J, Collette L, Coche E, Danse E, Jamart F, Machels JP, Vande Bergh B, Omoumi P.**
Whole-body MRI and detection of bone metastatic tumors: a meta-analysis


37) Roberts CC, Daffner RH, Weissman BN, Bankcroft L, Bennett DL, Blesiea JS, Bruno MA, Fries I, Germano...


