Imaging bone metastases in breast cancer: evidence on comparative test accuracy

N. Houssami* & C. M. Costelloe

1Screening & Test Evaluation Program (STEP), School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia; 2Department of Diagnostic Radiology, University of Texas MD Anderson Cancer Center, Houston, USA

Received 8 May 2011; revised 12 July 2011; accepted 14 July 2011

Background: Numerous imaging modalities may be used to detect bone metastases (BM) in women with breast cancer.

Methods: Systematic evidence review, including quality appraisal, of studies reporting on comparative imaging accuracy for detection of BM from breast cancer.

Results: Eligible studies (N = 16) included breast cancer subjects who had imaging evaluation for suspected BM or for staging/restaging in suspected local or distant relapse. Median prevalence of BM was 34.0% (range 10.0%–66.7%). There was substantial heterogeneity in the quality of reference standards and in the prevalence of BM, which could account for some of the differences in reported comparative accuracy. Most frequently, bone scan (BS) was compared with newer imaging modalities in subjects selected to both tests; therefore, results could be affected by selection bias. There was some evidence that positron emission tomography (PET), and limited evidence that PET/computed tomography (CT), CT, and magnetic resonance imaging (MRI), may provide small increments in accuracy relative to BS as add-on tests; there was little evidence regarding single photon emission computed tomography or whole-body MRI.

Conclusions: There is some evidence of enhanced incremental accuracy for some of the above-mentioned tests where used as add-on in subjects selected to more than one imaging modality, with little evidence to support their application as a replacement to BS in first-line imaging of BM. PET/CT appears to have high accuracy and is recommended for further evaluation.

Key words: bone metastases, breast cancer, cancer staging, imaging, positron emission tomography, sensitivity and specificity

introduction

Bone is the most common site of distant metastases from breast cancer and is the first affected site in a substantial proportion of women with advanced breast cancer [1, 2]. Bone metastases (BM) may have implications for prognosis, quality of life, and local and systemic therapy. A review from Hamaoka et al. [1] has highlighted that imaging modalities visualise different aspects of osseous tissues (cortex or marrow) in terms of density, water content, vascularity, or metabolism [1]. Hence, the appearance of osteolytic, osteoblastic, or mixed BM may differ considerably depending on the imaging modality used, leading to variable detection capability for different imaging tests [1].

background to imaging BM

Recommendations for the imaging of breast cancer patients at risk of BM have stressed a systematic approach based on patient symptoms and the strengths and limitations of the various imaging modalities [3]. Bone scan (BS) provides a relatively sensitive and inexpensive evaluation of the entire skeleton in a single imaging examination [4] and is recommended for evaluation of patients with multiple sites of bone pain or for staging of patients at high risk of having metastases [1]. BS images osteoblastic lesions and sclerotic/mixed bone lesions and images the reparative bone formed by lytic lesions [5]; it may therefore be less sensitive in detecting lytic lesions that fail to induce reparative bone formation [6]. Traumatic or degenerative processes that commonly occur in bone can cause tracer uptake, as does metastatic tumour; BS lacks anatomic resolution which increases the difficulty in distinguishing tumour from non-tumour uptake and may lead to false-positive BS.

Equivocal foci of tracer uptake on BS can be further evaluated with radiography or with single photon emission computed tomography (SPECT) which may also be fused with
computed tomography (SPECT/CT) [3], depending on the capabilities of the scanner. Modern SPECT scanners are multifunctional devices that can carry out routine planar BS, SPECT (essentially, a cross-sectional BS), and can fuse the SPECT and CT datasets to produce hybrid images that can be displayed in the axial, sagittal, or coronal plane. SPECT has been reported to identify more metastases than planar BS by virtue of its cross-sectional nature [7], and its accuracy is enhanced by the fused CT [8, 9], but SPECT/CT may not be widely available. Radiographs are otherwise recommended to evaluate indeterminate or non-specific tracer uptake on BS: at least 30% of the bone must be destroyed for lytic metastases to be identified on radiography, and overlapping structures can further limit the sensitivity of radiography for BM [1]. When visible, abnormalities are typically diagnosed with high confidence and radiographs are relatively inexpensive [1]. In patients with one or few sites of skeletal pain, targeted radiographs may be used for initial imaging [4], and if radiography shows metastases, the remainder of the skeleton can then be examined with BS [3].

Magnetic resonance imaging (MRI) may be used for evaluation when initial bone imaging is non-diagnostic [3, 4]. The high soft tissue resolution of MRI can allow the medullary cavity of bone, where most BM arise, to be more clearly discerned than can X-ray-based technologies (radiography and CT), which have lower soft tissue resolution. MRI has been reported to detect more BM than radiography, CT, or BS [10, 11] but is relatively expensive [1]. Conventional MRI is carried out on a limited portion of the body due to time constraints and due to artefacts that increase with enlargement of the field of imaging; however, diagnostic-quality whole-body MRI is now clinically feasible [12].

Where the above tests yield equivocal results, metabolic evaluation with positron emission tomography (PET) or fused PET and CT (PET/CT) has been recommended [3]. [18F]2-fluoro-2-deoxy-D-glucose (FDG), the most commonly used PET tracer, is a glucose analogue whose phosphorylated molecular structure delays metabolism by the cell; FDG accumulation can therefore be detected in highly metabolic tissues, such as malignancies [13]. PET can be used to investigate bone lesions, however, PET alone may lack anatomic detail; this problem can be overcome by fusion of PET/CT images which potentially improves its diagnostic capability for metastatic disease [14]. As with BS, PET can give false-positive results due to various common benign bone processes, and PET is more sensitive for detecting lytic than blastic BM [1].

Given that various technologies exist for imaging BM, we report a systematic review that updates the evidence on comparative test accuracy for imaging of BM in women with breast cancer.

methods
A systematic search of the literature (January 2000 to February 2011, Appendix 1) was carried out to identify studies of imaging for detection of BM that met the following eligibility criteria: (i) subjects with breast cancer as the primary cancer (studies of various primary cancers were ineligible) and reporting data on detection of BM; (ii) reporting comparative sensitivity and specificity of imaging tests in the same group of subjects (or for the vast majority within the same group of subjects) or at minimum data for true-positive and false-negative detection; and (iii) described a reference standard that was not entirely composed of one of the tests under comparison. This review did not consider studies of imaging in monitoring response to therapy.

Data extraction included information on study characteristics and quality appraisal, as well as quantitative data on test accuracy. Eligible studies were reviewed by both authors and disagreement on extracted information was resolved by discussion and consensus. Items included in the quality appraisal and evidence tables were adapted from recommendations for design of studies of comparative test accuracy [15–18] including checklists used in appraisal of imaging studies, and international standards for reporting of studies of test accuracy [19].

Study-specific estimates of sensitivity and specificity were extracted or were calculated. Subject-based and/or lesion-based accuracy estimates were extracted as reported in each study. Because of the limited number of studies for various test comparisons, and due to heterogeneity in the source studies, pooled analysis was not considered appropriate (see also ‘Results’ section). Where a sufficient number of studies reported comparisons of the same imaging tests, descriptive analysis (median and range) was performed.

results
Sixteen studies [20–35] comparing bone-imaging tests met eligibility criteria for this systematic review: the characteristics and quality appraisal of these studies are summarised in Table 1; data on comparative sensitivity and specificity for imaging BM are summarised in Table 2.

study characteristics and quality appraisal
Table 1 shows characteristics of all studies, including the proportion with BM (underlying prevalence). Studies, in general, included subjects with breast cancer who had imaging for evaluation of suspected BM (due to symptoms, clinical findings, or abnormal imaging or tumour markers) or for staging or restaging in women with suspected or known local or distant recurrence (Table 1). Substantial heterogeneity was evident for two factors: the quality of the applied reference standard, and the prevalence of BM across studies (in part reflecting the level of selection of subjects in each series). These findings limited the potential for meaningful pooling of data, and, of note, these may account for some of the observed differences in test accuracy. Most studies provided information on the timeframe between the two imaging tests under comparison.

prevalence of BM
The median prevalence of BM (study-specific proportion is shown in Table 1) from 14 studies [20–24, 26–32, 34, 35] was 34.0% (range 10.0%–66.7%).

selection of subjects
The majority of studies were based on series of subjects selected on the basis of having had the two tests under comparison, rather than a clinically defined cohort; frequently, this was selection of subjects who had the conventional test (BS in many studies) and also progressed to the newer test; hence, studies may be affected by selection bias. The study from Fuster et al.
<table>
<thead>
<tr>
<th>Study [P or R]* (timeframe)</th>
<th>Subjects</th>
<th>Reason for imaging/ definition of eligible subjects</th>
<th>Quality appraisal</th>
<th>Timeframe between two tests</th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heusner [20] [P] (NR)</td>
<td>20</td>
<td>Initial staging in women with BC or restaging in clinically suspected metastatic BC</td>
<td>Consecutive subjects having whole-body staging (various imaging tests)</td>
<td>Yes (then re-read using consensus)</td>
<td>NR</td>
</tr>
<tr>
<td>Morris [21] [R] (2003–2008)</td>
<td>163</td>
<td>Suspected metastatic BC: majority with symptoms and/or abnormal clinical findings, or abnormal (blood or radiology) tests</td>
<td>Selected if had both BS and PET/CT within 30-day period</td>
<td>No (blinded re-reporting for equivocal results only)</td>
<td>≤30 days</td>
</tr>
<tr>
<td>Fuster [22] [P] (NR)</td>
<td>60</td>
<td>Staging of women with newly diagnosed large (&gt;3 cm) BC</td>
<td>Consecutive subjects with large cancers</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mahner [23] [R] (1996–2004)</td>
<td>119</td>
<td>Women with new diagnosis of LABC (69) or with clinical suspicion of metastases (50)</td>
<td>Consecutive subjects with breast cancer who were referred for FDG–PET</td>
<td>Yes</td>
<td>20 ± 30 days</td>
</tr>
<tr>
<td>Bristow [24] [P] (2004–2005)</td>
<td>77</td>
<td>Staging of new clinically suspected or confirmed metastatic BC</td>
<td>Consecutive women attending advanced cancer clinic</td>
<td>Yes</td>
<td>≤14 days</td>
</tr>
<tr>
<td>Schmidt [25] [NR] (NR)</td>
<td>33</td>
<td>Suspected BC local or distant recurrence, based on symptoms or abnormal imaging or elevated tumour markers</td>
<td>Based on subjects who had both PET/CT and MRI</td>
<td>Yes</td>
<td>≤16 days</td>
</tr>
<tr>
<td>Aslan [26] [R] (1997–2002)</td>
<td>98</td>
<td>Women with BC who had various BS results (suspicous or positive scan, or normal scan with back symptoms)</td>
<td>Selected on the basis of BS result and had X-ray and CT</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Study [P or R]* (timeframe)</td>
<td>Subjects</td>
<td>Reason for imaging/ definition of eligible subjects</td>
<td>Quality appraisal</td>
<td>Timeframe between two tests</td>
<td>Reference standard</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>-----------------------------------------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>Total no.</td>
<td>No. with BM; % Median age (range)</td>
<td>Included all consecutive clinical cases or selected subjects on basis of having had both tests?</td>
<td>Were tests interpreted independently of each other?</td>
<td></td>
</tr>
<tr>
<td>Abe [27] [R] (NR)</td>
<td>44</td>
<td>14; 31.8 56 (35–81)</td>
<td>Based on subjects who had both BS and PET</td>
<td>NR (blinded to clinical information)</td>
<td>Mean 11.5 (0–69) days</td>
</tr>
<tr>
<td>Nakai [28] [NR] (2003–2004)</td>
<td>89</td>
<td>55; 61.8 59 (29–83)</td>
<td>Selected if had both BS and FDG–PET with ≤ 30 days between tests</td>
<td>Yes</td>
<td>≤30 days (mean 17)</td>
</tr>
<tr>
<td>Uematsu [29] [P] (NR)</td>
<td>15</td>
<td>7; 47 NR</td>
<td>Selected if had both FDG–PET and SPECT within 49 days of each other</td>
<td>Yes</td>
<td>≤49 days (mean 20)</td>
</tr>
<tr>
<td>Engelhard [30] [NR] (NR)</td>
<td>22</td>
<td>12; 54.5 63 (53–87)</td>
<td>Selected only if referred for MRI after other findings and had prior BS</td>
<td>Yes (however inconclusive BS one of inclusion criteria)</td>
<td>≤21 days</td>
</tr>
<tr>
<td>Gallowitsch [31] [R] (1997–NR)</td>
<td>62: data for 38 cases</td>
<td>13; 34.2 58.5 (NR)</td>
<td>Surgically resected BC, suspected local or distant recurrence (suspicious/inconclusive imaging or elevated tumour markers)</td>
<td>NR (probably not)</td>
<td>NR</td>
</tr>
<tr>
<td>Dose [32] [R] (1996–1999)</td>
<td>50, 45 had BS 12; 24</td>
<td>55.1 (28–89)</td>
<td>Suspected metastases or further evaluation in women with known metastases, or staging</td>
<td>Consecutive women with breast cancer referred for PET</td>
<td>Yes</td>
</tr>
<tr>
<td>Yang [33] [R] (NR)</td>
<td>40</td>
<td>NC NR (38–67)</td>
<td>Selected if had both BS and FDG–PET with ≤ 7 days between tests</td>
<td>NR</td>
<td>≤7 days</td>
</tr>
<tr>
<td>Ohta [34] [R] (1994–1997)</td>
<td>51</td>
<td>9; 17.6 49 (29–79)</td>
<td>Based on subjects who had both BS and PET within 30-day period</td>
<td>Yes</td>
<td>≤30 days</td>
</tr>
</tbody>
</table>
was the only study of a consecutive clinically defined cohort (all subjects with cancers >3 cm), and three other studies included consecutive referrals to imaging or to an advanced cancer clinic [23, 24, 32].

**Comparative Test Accuracy**

PET or PET/CT versus BS. Seven studies compared FDG–positron emission tomography (FDG-PET) with BS [23, 27, 28, 31–34], and two studies compared integrated PET/CT with BS [21, 22]. The largest number of studies reported comparisons of PET with BS: sensitivity was generally similar for both tests in most studies [27, 28, 31–34]; Mahner et al. [23] reported higher sensitivity for PET, while Gallowitsch et al. [31] (lesion-based data) reported higher sensitivity for BS. The median sensitivity (based on seven studies) [27, 28, 31–34] for PET was 84% (range 77.7%–95.2%), and for BS, it was 80% (67.0%–93.3%).

Specificity of PET and BS was also generally similar in four of the studies [23, 27, 28, 32] but was higher for PET in three studies [31, 33, 34]. The median specificity (seven studies) for PET was 92% (88.2%–99.0%) and for BS 82.4% (9.1%–99.0%); however, this included a low (outlier) estimate from Yang et al. [33] for lesion-based data and should be cautiously interpreted. If all data are used from 10 studies reporting data on BS specificity compared with any other imaging test (excluding the study from Yang et al. [33]), then the median BS specificity is 85.5% (range 68.0%–100%).

Integrated PET/CT versus BS: Evidence was limited to two studies [21, 22]; Fuster et al. [22] reported 100% sensitivity and specificity for PET/CT, much higher than estimates for BS (Table 2). Absolute sensitivity and specificity could not be reliably calculated for Morris et al. [21] which focused on discordant imaging among 132 paired tests: while PET/CT and BS were concordant in the majority (81%) of subjects, in the subset of 12 cases with discordant imaging results and pathology-confirmed BM, PET/CT was positive in 10 and equivocal in 2, while BS was negative in 11 and equivocal in 1 case.

Other Imaging Comparisons. SPECT versus PET.

One very small study [29] reported much higher sensitivity for SPECT (85.0%) than for PET (17.0%), Table 2; however, data from this study should be interpreted with caution because lesion-based estimates were from 900 bone lesions in 15 subjects, with the majority of the 163 metastatic lesions occurring in 4 subjects (230 metastatic lesions each). This study reported that SPECT altered treatment in one patient [29].

CT versus BS: One study [24] showed similar sensitivity for both CT and BS, but CT had higher specificity (Table 2).

CT versus Radiography: One study showed similar (modest) sensitivity, and the same specificity, for CT and X-ray (Table 2).

MRI versus BS: Two studies compared conventional MRI [35] or whole-body MRI [30], respectively, with BS—both studies...
showed ~10% higher sensitivity for MRI relative to BS, but findings on specificity were inconsistent (Table 2). The study from Altehoefer et al. [35] indicated that each of BS or conventional MRI detected sites of BM (in the same or in alternate anatomic region) that were not detected by the other test, and in some cases, this modified local treatment. The small series comparing whole-body MRI with BS [30] reported that MRI detected additional non-osseous metastases (lung or liver) and in one subject, this directed change in chemotherapy; tumour destruction of the vertebral spine with cord compression in one subject was identified only with MRI (a false negative on BS).

**MRI versus PET/CT:** Heusner et al. [20] examined whole-body MRI with diffusion-weighted imaging compared with integrated PET/CT in 20 subjects and reported higher accuracy for PET/CT than for whole-body MRI, with 100% sensitivity/specificity for PET/CT (Table 2). Schmidt et al. [25] compared whole-body MRI with PET/CT [25]: MRI had slightly higher

### Table 2. Evidence on comparative imaging sensitivity and specificity for studies reporting on detection of bone metastases in breast cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>Accuracy based on( ^{b} )</th>
<th>Tests being compared</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuster [22]</td>
<td>60 subjects</td>
<td>PET/CT</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Mahner( ^{c} ) [23]</td>
<td>119 subjects (95 had BS)</td>
<td>PET</td>
<td>87</td>
<td>92</td>
</tr>
<tr>
<td>Abé( ^{d} ) [27] (region-based)</td>
<td>9 regions in 44 subjects( ^{d} )</td>
<td>PET</td>
<td>84</td>
<td>99</td>
</tr>
<tr>
<td>Nakai [28]</td>
<td>89 subjects</td>
<td>PET</td>
<td>80.0</td>
<td>88.2</td>
</tr>
<tr>
<td>Gallowitsch [31]</td>
<td>38 subjects</td>
<td>PET</td>
<td>92.3</td>
<td>92.0</td>
</tr>
<tr>
<td>Gallowitsch [31] (lesion-based)</td>
<td>135 lesions in 38 subjects</td>
<td>PET</td>
<td>56.5</td>
<td>88.9</td>
</tr>
<tr>
<td>Dose [32]</td>
<td>50 subjects (45 for BS)</td>
<td>PET</td>
<td>83.3</td>
<td>89.5</td>
</tr>
<tr>
<td>Yang [33]</td>
<td>127 lesions in 40 subjects</td>
<td>PET</td>
<td>95.2</td>
<td>91.0</td>
</tr>
<tr>
<td>Ohta [34]</td>
<td>51 subjects</td>
<td>PET</td>
<td>77.7</td>
<td>97.6</td>
</tr>
<tr>
<td>Uematsu [29]</td>
<td>900 lesions in 15 subjects</td>
<td>SPECT</td>
<td>85</td>
<td>99</td>
</tr>
<tr>
<td>Bristow [24]</td>
<td>77 subjects</td>
<td>CT (thorax, abdomen, pelvis)</td>
<td>97.7</td>
<td>100</td>
</tr>
<tr>
<td>Aslan [26]</td>
<td>98 subjects</td>
<td>CT</td>
<td>71.8</td>
<td>100( ^{f} )</td>
</tr>
<tr>
<td>Altehoefer [35]</td>
<td>54 subjects</td>
<td>MRI (axial)</td>
<td>98</td>
<td>100( ^{f} )</td>
</tr>
<tr>
<td>Engelhard [30]</td>
<td>22 subjects</td>
<td>MRI (whole-body)</td>
<td>87</td>
<td>100( ^{f} )</td>
</tr>
<tr>
<td>Heusner [20]</td>
<td>20 subjects</td>
<td>Whole-body MRI with diffusion-weighted images</td>
<td>PET/CT</td>
<td>100</td>
</tr>
</tbody>
</table>

\( ^{a} \) Detection data for two studies [21, 25] are also presented in ‘Results’ section; however, sensitivity and specificity could not be reliably estimated; hence, these studies are not included in table (see ‘Results’ section).

\( ^{b} \) This may be same number as subject numbers shown in Table 1 or may slightly differ or may be lesion based, depending on the data reported for accuracy measures in each study.

\( ^{c} \) Study from Mahner et al. [23] also reported on a subset of ≤34 subjects who had CT (sensitivity 67%, specificity 95%) or X-ray (sensitivity 57%, specificity 100%).

\( ^{d} \) The study from Abé et al. [27] classified whole-body bones into nine anatomic regions (such as skull, upper limbs, lumbar region): data are reported on a regional basis (45 metastases from 187 regions) in 14 subjects with metastases (from a total of 44 subjects).

\( ^{e} \) Studies from Aslan et al. [26] and from Altehoefer et al. [35] incorporated concordance of both tests under comparison in the reference standard and may have over-estimated specificity.

BS, bone scan/skeletal or bone scintigraphy (technetium-99m methylene diphosphonate bone scan or Tc-99m MDP bone scan); CT, computed tomography; MRI, magnetic resonance imaging; PET or FDG–PET, [18F]2-fluoro-2-deoxy-D-glucose–positron emission tomography; PET/CT, integrated PET/CT; SPECT, single photon emission computed tomography.
sensitivity and specificity (95% and 92%, respectively) than PET/CT (sensitivity 91% and specificity 86%). However, these lesion-based estimates were for 212 lesions that included distant metastases at various sites (majority in bone or liver) occurring in 20 of 33 subjects—bone-specific accuracy could not be calculated therefore estimates from this study are not shown in Table 2.

discussion

Breast cancer guidelines and consensus recommendations [3, 4,36–40] indicate that various imaging tests may be used for investigation of suspected BM or in staging and restaging of women with breast cancer. Although nuclear medicine BS (or bone scintigraphy) is the most consistently recommended imaging test in guidelines, other tests have also been recommended for imaging BM, including plain radiography (X-ray), MRI, CT, PET, SPECT, and related hybrid scans [3, 4,36–40]. The majority of comparisons do not specify which test should be preferentially used [36–40]. In this work, we reviewed the evidence on imaging of BM in breast cancer to systematically update and summarise the evidence on comparative imaging accuracy.

Despite some methodological limitations in the source studies identified in our quality appraisal (and further discussed below), the data show that in subjects who had both PET and BS [23, 27, 28, 31–34], the sensitivity of PET was similar to or slightly higher than BS (median sensitivity for PET 84.0% and for BS 80.0%). In these studies, the specificity of PET was mostly higher than that of BS (median specificity for PET 92.0% and for BS 82.4%), though study-specific estimates varied considerably. These findings should be interpreted with the knowledge that heterogeneity between studies on key quality issues may account for some of the observed differences in accuracy [16, 41] and that in some studies PET was likely to have been applied as an add-on test [15] after BS in selected subjects to further evaluate BS findings. Even studies that reported ‘consecutive’ subjects were essentially studies of subjects selected to have both tests (conventional and relatively newer tests)—so this may not give a complete reflection of the accuracy of the conventional test (in this scenario, BS). No data were reported, for example, on BS accuracy in subjects who did not progress to further testing—methods have been described to integrate such data in studies of comparative imaging accuracy [42]. Comparisons between integrated PET/CT and BS, similarly limited by the quality of the source studies and also limited to evidence from only two studies, suggest that PET/CT may offer improved diagnostic capability over BS in selected subjects, with one study reporting both 100% sensitivity and specificity for PET/CT [22]. Little to no data were available to judge the potential for PET or PET/CT as a replacement test [15] for BS. One study comparing PET/CT to whole-body MRI with diffusion-weighted images [20] also showed very high accuracy for PET/CT for detection of BM (but did not report data for BS).

For other imaging comparisons (Table 2), there was very limited evidence—one study of each paired comparison met our inclusion criteria; most studies consisted of small series and were affected by similar quality limitations as those outlined above for PET or PET/CT comparisons with BS. Therefore, it is not possible to make definitive recommendations on the superiority of one imaging modality over another. Furthermore, it appears likely that relatively ‘new’ imaging tests in these comparative studies were generally carried out in subjects selected for further investigation after BS or selected to have both tests, so current evidence provides little insight on the accuracy of imaging modalities such as MRI or CT as a replacement test [15] for BS in a consecutive clinical cohort. Rather, the estimates in Table 2 are mostly indicative of their contribution as add-on tests with BS or as complementary testing to further evaluate or characterise bone lesions (e.g. after an inconclusive BS) or bone symptoms (e.g. after a negative BS) and for further anatomic detailing. The studies involving either conventional MRI [35] or whole-body MRI [30] showed improved sensitivity relative to BS in subjects selected to both tests, which may have clinical implications in a modest proportion of subjects. Altheofer et al. [35] indicated that each of BS or conventional MRI detected sites of BM that were not detected by the other test; for example MRI directed the need for local therapy in regions that were negative on BS in several cases, and overall, combined conventional MRI and BS upstaged 9 of 54 subjects from stage II/III. The series comparing whole-body MRI with BS [30] showed that MRI detected additional non-osseous metastases (see ‘Results’ section) that directed change in chemotherapy in one subject. It also reported that tumour destruction of the vertebral spine with cord compression in one subject was identified only with MRI and directed initiation of radiation therapy [30]. A study that did not meet eligibility criteria for this review highlights the value of MRI in characterising spinal lesions identified on BS [43], and recommendations also specify a role for MRI in evaluation of single ‘hot spots’ on BS [4].

The two studies that did not include BS in comparative evaluation (SPECT versus PET [29], whole-body MRI versus integrated PET/CT [25]) should be interpreted with caution because each of these studies reported lesion-based accuracy based on a very large number of lesions from a small number of subjects. This approach ignores expected within-subject correlation of lesion-based detection for a given test and may bias estimates of accuracy, which could result in unusually low or high estimates. For example, the study of SPECT versus PET [29] (900 bone lesions in 15 subjects) reported low PET sensitivity (17.0%), an estimate which is inconsistent with the median PET sensitivity (84.0%) calculated from the much larger subset of studies (of PET versus BS) discussed above. The study comparing whole-body MRI with integrated PET/CT [25], which showed marginally higher sensitivity for MRI, was based on 212 lesions from various metastatic sites (majority were bone or liver) occurring in 20 subjects, and estimates for bone lesions cannot be calculated separately. Hence, there is insufficient evidence to make conclusions regarding the relative accuracy or roles of these imaging modalities in evaluation of BM in breast cancer. It is advisable that imaging studies (of any design) reporting lesion-based accuracy in the metastatic setting also declare subject-based data for estimates of accuracy.

We pointed out earlier that substantial heterogeneity was evident in quality appraisal with regard to two key factors, and while this heterogeneity was one of the reasons we did not pool data, the important consideration here is that sources of variability may (at least in part) account for some of the
The reported differences in accuracy across studies [16, 41]. The first factor is the quality of the reference standard, which consisted of variable and sometimes minimal ascertainment with biopsy and/or various follow-up duration and strategies, including the use of team consensus opinion as reference standard [21, 24] or incorporation of concordance between the tests being compared into the reference standard [26, 35]. While one should be cognisant of the challenges inherent in ascertainment of suspected distant metastases, and biopsy of each individual suspected bone metastasis is not feasible or ethical particularly in patients with numerous lesions, the heterogeneity (and in some studies inadequacy) of the applied reference standards across studies means that some subjects may have been misclassified, and this could have biased estimates of accuracy. The second factor is variability in the underlying prevalence of BM across studies, which was in the range of 10%–66.7% [21–24, 27–32, 34, 35], and also reflects the extent of selection to tests. Variability in underlying prevalence can further influence the magnitude to which accuracy estimates may be biased due to an imperfect or inadequate reference standard classification [16, 41]. Thus, the combination of variability in the prevalence of BM in these studies and possible misclassification by the applied reference standard raises the possibility that results of these comparative studies may not be widely generalisable.

Considering the high sensitivity and specificity for integrated PET/CT described earlier, this appears to be a promising imaging modality and warrants further evaluation, although the evidence comes from only three studies [20, 21, 22] (see Table 2 and see ‘Results’ section) and is limited by selection issues highlighted above. Of note, while our review focuses on detection of BM, the integration of high-quality CT with metabolic PET datasets allows an in-depth anatomic and metabolic evaluation of multiple organ systems in a single PET/CT study (Figure 1). PET/CT may therefore have a future role as first-line imaging in breast cancer patients considered at very high risk of distant metastases, by virtue of its comprehensive nature, and has other uses such as metabolic evaluation for therapeutic response [44]. As far as BM, our review suggests that PET/CT merits further research in studies that use better design and take into account the methodological issues highlighted in this review, including consideration of prospective paired evaluation of PET/CT and BS in a consecutive clinically defined cohort of breast cancer patients who require bone imaging or staging.

In summary, there is some evidence that PET, and in a limited number of studies, PET/CT, CT, and conventional MRI, may result in small increments in the accuracy of imaging bones relative to BS, where used for evaluation of suspected lesions and/or bone symptoms or in staging or restaging. These studies have generally compared imaging tests such as PET, PET/CT, CT, or MRI as add-on tests with BS; hence, the evidence is indicative of the role of these bone-imaging tests as complementary to BS in an imaging strategy where BS was likely to have been the initial or baseline test. There is little evidence on which to base recommendations regarding SPECT or whole-body MRI for BM. Currently, there is no definitive evidence supporting that any of the imaging tests discussed in this review can be used as a replacement to BS in first-line imaging for evaluation of bone lesions or symptoms, or in staging and restaging, in breast cancer.

**funding**

This work was partly funded by Australia’s National Health and Medical Research Council program grant 633003 to the Screening and Test Evaluation Program.

---

**Figure 1.** Staging of a 60-year-old female with newly diagnosed breast cancer. (A) Planar Tc-99m methylene diphosphonate bone scan (BS) is equivocal for metastases. Mild heterogeneity of tracer uptake is seen diffusely but no discrete, suspicious foci are identified. Uptake at the right costochondral junctions is likely traumatic in aetiology. (B) Coronal reformation of a single photon emission computed tomography/computed tomography (SPECT/CT) scan provides details that identify widespread bony metastases. The SPECT/CT scan was carried out immediately following the planar BS and with the same tracer dose. (C) Maximum intensity projection of the positron emission tomography (PET) dataset from an [18F]2-fluoro-2-deoxy-D-glucose–positron emission tomography (FDG–PET)/CT scan demonstrates widespread bony metastases and FDG uptake in the primary right breast tumour (arrow).
disclosure

The authors have declared no conflicts of interest.

references


**appendix 1. literature search strategy**

Potentially relevant publications were identified by searching Medline (January 2000 to February week 2, 2011), by combining the following: (i) explode subject heading ‘breast neoplasms’ (all subheadings); (ii) search for ‘Diagnostic Imaging/or Diffusion Magnetic Resonance Imaging/or Magnetic Resonance Imaging/or Molecular Imaging/or Perfusion Imaging/or imaging.mp./or Imaging, Three-Dimensional/or Radionuclide Imaging/or Echo-Planar Imaging/or Tomography, Emission-Computed/or Positron-Emission Tomography/or PET/CT/or PET imaging.mp. or Fluorodeoxyglucose/or radiography’; and (iii) search for ‘bone’ in the title. Abstracts identified based on the above strategy (n = 130) were checked against pre-defined criteria by one author (NH) and 15 papers were considered to be eligible; one additional paper was identified through other sources. A total of 16 papers were included in the evidence review.