

Lexicon for standardized interpretation of gamma camera molecular breast imaging: observer agreement and diagnostic accuracy

Amy Lynn Conners · Carrie B. Hruska ·
Cindy L. Tortorelli · Robert W. Maxwell ·
Deborah J. Rhodes · Judy C. Boughey · Wendie A. Berg

Received: 8 September 2011 / Accepted: 27 December 2011 / Published online: 31 January 2012
© Springer-Verlag 2012

Abstract

Purpose To determine interobserver agreement and diagnostic accuracy using a lexicon for standardized interpretation of molecular breast imaging (MBI) studies by breast radiologists.

Methods An MBI lexicon was developed, including descriptors of lesion type, background uptake, and associated findings by a consensus of experts. In an institutional review board-exempted protocol, six breast imaging radiologist observers without prior MBI experience attended a 2-h MBI interpretation training session, including definitions of lexicon terminology, case examples, and ten unknown cases with expert feedback. Following training, each radiologist observer interpreted an independent set of MBI images of 50 breasts, including 20 (40%) with malignancies with a median invasive tumor size of 1.7 cm (range 1.0 to 6.3 cm). The findings were described using the lexicon and each breast was given a final assessment of 1 to 5, paralleling

BI-RADS assessment categories. Sensitivity, specificity, positive and negative predictive values were determined with core or surgical pathology results or 1-year imaging follow-up as the reference standard. Interobserver agreement for lesion-type classification, lesion and background uptake intensity, and final assessments were determined using Cohen's kappa.

Results For the six observers, median sensitivity was 1.0 (range 0.90–1.0), specificity 0.88 (range 0.83–0.97), and AUC 0.94 (range 0.93–0.98). Fair interobserver agreement was shown for background uptake ($\kappa=0.31$). Agreement was substantial for lesion type ($\kappa=0.79$) and non-mass distribution ($\kappa=0.63$), and near-perfect for final assessment ($\kappa=0.84$).

Conclusion Dedicated breast imaging radiologists, newly trained to interpret MBI with the proposed lexicon, achieved high agreement and diagnostic accuracy.

Keywords Molecular breast imaging · Lexicon · Observer agreement · Training · Gamma camera · Scintimammography

A. L. Conners (✉) · C. B. Hruska · C. L. Tortorelli ·
R. W. Maxwell
Department of Radiology, Mayo Clinic,
200 1st St. SW,
Rochester, MN 55905, USA
e-mail: conners.amy@mayo.edu

D. J. Rhodes
Department of Internal Medicine, Mayo Clinic,
Rochester, MN 55905, USA

J. C. Boughey
Department of General Surgery, Mayo Clinic,
Rochester, MN 55905, USA

W. A. Berg
Department of Radiology, Magee-Womens Hospital,
University of Pittsburgh School of Medicine,
Pittsburgh, PA, USA

Introduction

Molecular breast imaging (MBI) is an FDA-approved breast imaging technique that relies on the detection of ^{99m}Tc sestamibi accumulation within the breast using a dual-head cadmium-zinc-telluride (CZT) gamma camera. ^{99m}Tc sestamibi has been shown to preferentially accumulate in breast malignancies as well as in some types of non-malignant tissue. MBI is a sensitive technique for detection of malignancy in the screening setting [1] and has demonstrated a similar high sensitivity in the diagnostic setting when a dual-detector system is used [2]. As more becomes known about

the utility of this method of imaging, it will become increasingly important to have a common language for the greater community when describing MBI findings. To our knowledge, there exists neither standardized terminology nor interpretive criteria for MBI or other gamma-camera breast imaging studies.

The Breast Imaging Reporting and Data System (BI-RADS), developed through the American College of Radiology, is highly effective in standardizing reporting and communication of breast imaging findings and recommendations [3]. Training in BI-RADS for mammography [4] has been shown to improve observer agreement in mammographic interpretation [5]. BI-RADS lexicons for mammography [4], ultrasonography [6] and MRI [7] also serve as important tools for researching outcomes of particular imaging findings to guide management. We have developed a lexicon for the description of MBI images based on familiar BI-RADS lexicon terminology in existence for mammography, ultrasonography and MRI, as well as on the proposed BI-RADS-type lexicon for positron emission mammography (PEM) [8]. The goal of the current study was to determine interobserver agreement and diagnostic accuracy using this lexicon for standardized interpretation of MBI studies by breast radiologists.

Materials and methods

Institutional review

This observer study was deemed by our Institutional Review Board (IRB) as exempt, meaning that the protocol was approved for use without ongoing IRB oversight.

MBI system and imaging

MBI images were acquired using one of two research MBI systems under evaluation, both of which utilize a small field-of-view (20×16 cm or 20×20 cm) dual-head CZT gamma camera mounted on a modified mammographic gantry. FDA-approved versions of these MBI systems are now commercially available. One system was equipped with two LumaGem detectors (Gamma Medica-Ideas, Northridge, CA; intrinsic spatial resolution 1.6 mm) and the second system was equipped with GE prototype CZT detectors (GE Healthcare, Haifa, Israel; intrinsic spatial resolution 2.5 mm). The detectors of both systems were fitted with system-specific registered collimators which were optimized for high count sensitivity in the setting of near-field imaging with a dual-head gamma camera [9]. These detectors have a very small dead-space at the detector edge, permitting the breast to be positioned in close proximity to

the detector, in positions analogous to those obtained with mammography.

MBI studies were selected from a database of over 3,000 examinations which were performed under a variety of IRB-approved research protocols between August 2005 and February 2011. Written informed consent had been obtained from each woman as a part of the applicable protocol. Intravenous injection of 296, 740 or 1,110 MBq (8, 20, or 30 mCi) ^{99m}Tc sestamibi was used as previously described [1, 2, 10, 11]. MBI imaging commenced 5 min after injection. A craniocaudal (CC) and mediolateral oblique (MLO) view of each breast was acquired over 10 min per view. For each image, the breast was positioned between the two detectors and gentle compression was applied to limit patient motion. Of note, the amount of axilla included in the MBI images varies from patient to patient and is dependent on patient anatomy and positioning.

Observers

Five radiologists who met all requirements of the Mammography Quality Standards Act (MQSA) for physicians interpreting mammography, worked a minimum of 0.40 full-time equivalents in a Division of Breast Imaging and Intervention, and who had a mean of 14 years experience in breast imaging (range 7–23 years) participated in a 2-h didactic training session on MBI. A sixth observer was an MQSA-qualified board-certified radiologist currently enrolled in a breast imaging fellowship at the same institution who had completed 6 months of fellowship at the time of the interpretation task. None of the radiologists had prior training or experience in interpretation of MBI, and all agreed to have their results analyzed as part of participation in the training.

MBI training session

All observers first attended a 2-h live didactic session conducted by three of the authors. Imaging technique, gamma camera physics, lexicon definitions, and examples of malignant and non-malignant findings and artifacts, were reviewed. Ten unknown cases were then presented, each followed by review of corresponding multimodality images and pathology results.

MBI lexicon

A lexicon, as given in Table 1, was developed for the interpretation of MBI by three fellowship-trained, dedicated breast radiologists (A.L.C., C.T. and R.M.) who had each interpreted at least 270 MBI studies (mean 401, range 273–600), together with input from a member of the BI-RADS committees (W.B.). The MBI lexicon was modeled after familiar terms used in BI-RADS for mammography, ultrasound, and MRI [4, 6, 7]. We referred to the proposed

Table 1 MBI lexicon

Indication: Describe clinical problems (if any), history of biopsies (date and results), risk factors, indicate if patient is pre- (last menstrual period [LMP] less than one month ago), peri- (LMP more than one month ago and less than 12 months ago), or postmenopausal (LMP at least 1 year ago), phase of menstrual cycle (if relevant), and any use of selective estrogen receptor modulators or medications with estrogenic or progestogenic activity			
Comparison: Prior breast imaging, including prior gamma camera breast imaging studies (if any) should be reviewed, with the dates and types of prior studies reported			
Technical Factors: Report dose (MBq) and type of tracer injected and duration of circulation phase (time from injection to imaging). If additional views beyond routine CC and MLO projections were obtained, these should be detailed			
Limitations: Describe any suboptimal positioning, motion, pixel dropout, “hot pixels”, electronic, or other artifacts which are felt to affect image interpretation			
Background	Describe degree of radiotracer uptake in background normal parenchyma, which may be uniform (homogeneous) or patchy (heterogeneous)		
	Photopenic	Less than subcutaneous fat	
	Minimal-Mild	Equal to or slightly greater than subcutaneous fat	
	Moderate	Visually greater than mild, but less than twice as intense as subcutaneous fat	
	Marked	Visually at least twice as intense as subcutaneous fat	
Findings: Categories and Terms			Description
Mass	Uptake which has convex outward borders, no interspersed normal uptake, and is seen on two projections (if location is amenable)		
Non-Mass Uptake	Uptake distinct from the surrounding tissue that does not fit criteria for a mass and which usually contains interspersed areas of normal glandular tissue		
	Distribution	Focal area	<25% of a quadrant or < 2 cm in diameter in a confined area
		Segmental	Uptake in linear or triangular region or cone with apex pointing toward nipple that suggests (but is not specific for) intraductal pathology
		Regional	Uptake in a large volume of tissue, ≥ 2 cm in diameter, not conforming to a ductal distribution; may be geographic
		Multiple regions	Uptake in at least two large volumes of tissue;

Table 1 (continued)

			more than one area of geographic uptake
		Diffuse	Uptake distributed throughout the breast
	Internal pattern of uptake	Homogeneous	Confluent, uniform uptake
		Heterogeneous/Patchy	Variable, nonuniform uptake
	Symmetry	Symmetric	Similar uptake pattern in both breasts
		Asymmetric	More uptake in one breast compared to the other
Associated Findings		Axillary uptake	Uptake in the axilla, usually thought to be a lymph node which may or may not be pathologic
		Nipple uptake	Radiotracer uptake within the nipple, a physiologic finding if not associated with other suspicious uptake
		Vessel uptake	Serpiginous linear uptake corresponding with a vessel
Location	Breast	Right, left or bilateral	
	In-breast location	Quadrant or clock-face location, or specifically in the subareolar or central breast or axillary tail	
	Depth/Distance from the nipple	Anterior, central or posterior third or measured distance from the nipple	Measurement is made from the center of the finding and recorded in centimeters
Qualitative intensity of uptake in lesion^a	Photopenic	Uptake in lesion is less than surrounding background parenchyma	
	Mild	Uptake which appears to be less than 50% of background	
	Moderate	Uptake which appears to be at least 50% of background but not twice as intense as background	
	Marked	Uptake which appears to be at least twice background uptake	
Lesion size	X	Longest measurement of the lesion, made on whichever image best depicts the lesion	

Table 1 (continued)

	Y	Measurement orthogonal to X, made using the same image used to define X
	Z	If the lesion is visible on both projections, Z should be an orthogonal measurement made on the projection (CC or MLO) not used to define X/Y
Assessment Categories		
Incomplete Assessment	0 - Incomplete	Additional imaging is needed before a final assessment can be rendered
Final Assessment	1 - Negative	No lesion found (routine follow-up)
	2 - Benign	No malignant features; e.g., photopenia (routine follow-up)
	3 - Probably benign	Very low probability of cancer (follow-up MBI examination is recommended in 6 months if targeted diagnostic mammogram and ultrasound are negative)
	4 - Suspicious	Intermediate probability of cancer (biopsy is recommended)
	4a-Low suspicion	Used for a finding which requires intervention but is of low suspicion for malignancy
	4b-Intermediate suspicion	Used for a finding which is judged to be of intermediate suspicion for malignancy
	4c-Moderate suspicion (but not classic)	Used for a finding which is judged to be of moderate suspicion for malignancy
	5 - Highly suggestive of malignancy	High probability of malignancy (biopsy is recommended)
	6 - Known biopsy-proven malignancy	Appropriate action should be taken

^a These are the definitions of lesion intensity provided to observers for use during the interpretation task. However, the authors recommend that lesion intensity be judged relative to subcutaneous fat (rather than relative to background uptake) for greater consistency.

lexicon for PEM [8] in order to unify (where possible) the lexicons for existing nuclear breast imaging techniques.

The lexicon includes classification of radiotracer uptake in both lesions and background parenchyma. Background uptake was defined as the degree of background radiotracer uptake in the parenchyma relative to subcutaneous fat and was categorized as photopenic, mild, moderate or marked. Lesions could be classified as either a mass or non-mass radiotracer uptake. Qualitative radiotracer uptake within a lesion was classified as either photopenic (lack of uptake),

mild, moderate or marked relative to surrounding parenchyma. For non-mass uptake, the distribution was categorized as a focal area, regional, multiple regions, segmental, or diffuse.

The location of MBI findings is specified in terms of the clock face or quadrant and distance from the nipple, as in mammography, though consistency of such descriptions was not evaluated in this task. The size of a finding on MBI is determined on the image(s) where the finding is most discrete with an additional measurement taken in the orthogonal plane, though lesions were not measured by

observers in this evaluation. Symmetry between breasts and associated findings of nipple uptake, axillary uptake or vessel uptake can be described, though these were not required for each case in this task.

Final assessment codes, modeled after those used in BI-RADS for other breast imaging modalities, but tailored to MBI, were recorded for each breast. The assessment codes and corresponding recommendations were as follows: 1 – negative, routine follow-up; 2 – benign, routine follow-up; 3 – very low likelihood of malignancy, follow-up in 6 months if targeted diagnostic mammogram and ultrasound are negative; 4 – suspicious, consider biopsy; 5 – highly suggestive of malignancy, take appropriate action. Categories 0 (incomplete, needs additional imaging), and 6 (known malignancy, take appropriate action) were not included in this evaluation.

MBI interpretation task

The MBI interpretation task included studies of 27 patients in which 50 breasts were imaged. Both views of both breasts were shown together for 23 patients and only one breast was shown for 4 patients. The studies were selected from our file of more than 3,000 MBI patients to represent the breadth of pathology as well as the breadth of imaging findings seen on MBI, and were not previously shown as part of the training. Biopsy-proven malignancy was present in 20 of the 50 breasts (40%, 20 patients), including invasive ductal ± intraductal carcinoma in 9 (45% of malignancies); ductal carcinoma in situ in 5 (25% of malignancies); invasive lobular carcinoma in 3 (15% of malignancies); and mixed ductal and lobular carcinoma in 3 (15% of malignancies). For the 15 invasive cancers with size available, the median tumor size was 1.7 cm (mean 2.4 cm, range 1.0 to 6.3 cm). Benign lesions were diagnosed in 11 breasts (22%) including three biopsy-proven lesions (two fibroadenomata and one papilloma) and eight lesions with benign status verified by follow-up (including two benign cysts, two benign intramammary lymph nodes, one breast with benign nipple uptake and three areas of diffuse benign background uptake). The remaining 19 breasts had no diagnosed lesions. As the reference standard, a breast was considered disease-positive based on histopathologic diagnosis of malignancy from core biopsy and/or surgical excision. A breast was considered disease-negative based on either benign histopathologic findings or benign imaging findings at more than 1 year after the MBI study. All MBI studies included upper and lower head detector images in both the CC and MLO projections for a total of four images per breast. Lesions were not marked, so that observers decided whether a lesion was present and how many were present in each breast.

Observers were asked to describe the degree of background radiotracer uptake as well as any MBI findings according to the lexicon, and were asked to give a final assessment ranging

from 1 to 5 for each breast based on the MBI findings. Observers were required to specify whether mass or non-mass-like uptake was present, the distribution of non-mass-like uptake (if relevant), and the intensity of uptake in each lesion described. Observers could comment on symmetry and additional findings (axillary uptake, nipple uptake, vessel uptake) but were not required to do so, and these data were not analyzed. Lesion location and size were not evaluated. After recording MBI findings and assessments, observers were then shown conventional imaging of low-resolution mammograms for comparison in all cases but one in which no mammogram was available. In four cases where relevant, an ultrasonogram or a prior comparison mammogram was also shown. Observers then gave a second final assessment based on the combined interpretation of MBI and conventional breast imaging. Observers were instructed not to attempt to give a primary interpretation of the mammogram.

Lexicon descriptor terms, as well as final assessments, were recorded in person by one of the study authors. The observers had access to the detailed BI-RADS style MBI lexicon to use as a reference during the interpretation task. A data recorder was allowed to remind the observers that the lexicon was available for their use, but otherwise did not guide the observers in interpretation.

Statistical analysis

Results were analyzed for MBI alone and again based on assessments of combined MBI and conventional imaging. A final assessment of 4 or 5 was considered test positive, while an assessment of 1, 2 or 3 was considered test negative. Sensitivity, specificity, positive predictive value and negative predictive value were calculated for each observer and for all observers as a group. A receiver-operator characteristic curve was generated for each observer, and the area under the curve for each individual observer was calculated. Agreement for each category was calculated at the per-breast level. In two cases in which two lesions in a single breast were present on the consensus read, analysis was performed using descriptors of the dominant lesion. Any additional lesions described by observers but not by the consensus readers were not included in the analysis of feature agreement.

Cohen's kappa was used to assess interobserver agreement among the six observers for background uptake, lesion type, distribution of non-mass uptake, intensity of lesion uptake and final assessment code [12, 13]. Grouped kappas were calculated for assessments 1 or 2, 3, and 4 or 5. Agreement between each observer and the expert consensus readers was also calculated. A kappa value of 1.0 indicates complete agreement, while kappa of 0 indicates no agreement beyond that expected by chance. According to Landis and Koch [14], a kappa value below 0.2 indicates slight agreement, between 0.21 and 0.4 fair agreement, between

0.41 and 0.6 moderate agreement, between 0.61 and 0.8 substantial agreement, and 0.81 or greater near-perfect agreement.

Results

MBI interpretive performance

Across the six observers, median sensitivity for MBI alone was 20/20 (1.0, range 0.9–1.0) and median specificity was 26/30 (0.88, range 0.83–0.97). At least one observer assessed 11 breasts as 3, including 3 breasts with malignancy. The median positive predictive value was 0.85 (range 0.80–0.95, 20/25 to 19/20) and the median negative predictive value was 1.00 (range 0.93–1.00, 27/29 to 27/27). The median area under the curve was 0.94 (range 0.93–0.98). Three lesions in three breasts were described by the observers in addition to those known by the consensus readers, and were not included in the analysis.

MBI lexicon agreement

Kappa values for interobserver agreement and agreement between the observers and the consensus readers are given in Table 2. There was only fair agreement on background parenchymal uptake with $\kappa=0.31$ (95% CI 0.23–0.39). Interobserver agreement for lesion type (mass or non-mass) was substantial with $\kappa=0.79$ (95% CI 0.73–0.85) and was substantial for lesion intensity with $\kappa=0.67$ (0.61–0.73). There was moderate to substantial agreement for each subcategory of intensity.

For distribution of non-mass-like uptake, interobserver agreement was substantial at 0.63 (0.51–0.75). Agreement was substantial for segmental and regional distributions, moderate for focal area, fair for diffuse uptake (of which there was only one breast) and slight for multiple regions (of which there were two breasts). Results for background and lesion features compared to the expert consensus were similar to those between observers (Table 2).

Interobserver agreement for final assessment of MBI alone (without the corresponding mammogram) was substantial, $\kappa=0.80$ (95% CI 0.73–0.87). Agreement between the observer final assessments and the expert consensus was also near-perfect, $\kappa=0.83$ (95% CI 0.69–0.89). There was near-perfect agreement for combined categories 1 and 2 as well as for combined categories 4 and 5 (Table 2). Agreement for assessment category 3 was slight, with $\kappa=0.15$.

MBI combined with conventional imaging

When interpreted in conjunction with mammography, agreement further improved to $\kappa=0.87$ (95% CI 0.79–0.95), indicating near-perfect agreement. Agreement for

assessment category 3 improved to $\kappa=0.3$ (fair agreement) when interpreted together with conventional imaging, with a wide confidence interval. When MBI was interpreted with correlating images, an assessment of 3 was given by at least one observer in six breasts, none of which showed a malignant lesion (Fig. 1). Sensitivity of MBI combined with mammography was 20/20 (1.0) for all observers.

Specific problem cases

There were a few specific cases for which agreement was only slight or fair. There were two breasts in which five observers agreed with the consensus and described a larger single area of heterogeneous non-mass-like uptake, while one observer picked out a brighter area within it and described the mass as a second/additional lesion (Fig. 2).

The lexicon includes axillary uptake as an additional finding, usually thought to represent a lymph node, which may or may not be pathologic. A clearly pathologic axillary finding in one breast was described by five observers as a distinct lesion (three of the observers also described the associated finding of axillary uptake). One observer described this as only the associated finding of axillary uptake.

In three breasts some observers described moderate or marked background uptake as diffuse non-mass-like uptake (Fig. 3). However, all were symmetric, and all observers agreed on a benign final assessment in these cases.

Discussion

Substantial to near-perfect agreement and high diagnostic performance were found when the proposed lexicon was used for MBI interpretation. These results were achieved by observers with no prior MBI interpretation experience who had received only 2 h of training. These findings support the use of the proposed lexicon for reproducible communication of MBI findings and suggest that MBI studies can be interpreted well by newly trained radiologists with experience in breast imaging.

Interobserver agreement for final assessment was greater in our study than that seen in validation studies of mammography [15], ultrasonography [16–18] or MRI [8, 19, 20], and was also slightly greater than seen with PEM [8]. Interobserver agreement for lesion type was similar to or slightly lower than that seen with mammography or ultrasonography [15, 17], but was similar to or higher than that seen with MRI [8, 19]. In general, interobserver agreement for feature analysis was similar to or higher than that seen in lexicon studies of mammography, ultrasonography and MRI [8, 15–21]. In these studies and ours, as expected, kappa values were generally lower when a larger number of options were available within a category, and

Table 2 Interobserver agreement for MBI descriptors among six dedicated breast radiologists newly trained in MBI interpretation, and agreement with expert consensus interpretations for 50 breasts. The data presented are κ values (95% CI)

	Number of breasts ^a	Interobserver agreement among six observers	Agreement between observers and expert consensus readers
Background parenchymal uptake		0.31 (0.23 to 0.39)	0.43 (0.46, -0.04 to 0.96)
Photopenic	10	0.56 (0.37 to 0.74)	
Mild	35	0.29 (0.21 to 0.36)	
Moderate	3	0.15 (-0.01 to 0.30)	
Marked	2	0.13 (-0.46 to 0.72)	
Lesion type		0.79 (0.73 to 0.85)	0.82 (0.84, 0.75 to 0.87)
No lesion	22	0.85 (0.77 to 0.92)	
Mass	9	0.76 (0.58 to 0.93)	
Non-mass	19	0.77 (0.66 to 0.87)	
Distribution of non-mass lesions		0.63 (0.51 to 0.75)	0.64 (0.63, 0.51 to 0.76)
Not applicable, no non-mass lesions	31	0.77 (0.66 to 0.87)	
Focal area	6	0.42 (0.06 to 0.78)	
Segmental	6	0.63 (0.34 to 0.91)	
Regional	4	0.64 (0.34 to 0.93)	
Multiple regions	2	0.15 (-0.57 to 0.87)	
Diffuse	1	0.39 (-0.33 to 1.10)	
Lesion intensity		0.67 (0.61 to 0.73)	0.77 (0.76, 0.66 to 0.97)
Not applicable, no lesion	22	0.84 (0.76 to 0.91)	
Photopenic	2	0.59 (-0.08 to 1.26)	
Mild	6	0.59 (0.31 to 0.87)	
Moderate	12	0.53 (0.37 to 0.68)	
Marked	8	0.65 (0.43 to 0.86)	
Final assessment			
MBI interpreted in isolation		0.80 (0.73 to 0.87)	0.83 (0.85, 0.69 to 0.89)
Negative or benign (assessment 1 or 2)	26	0.91 (0.84 to 0.98)	
Probably benign (assessment 3)	3	0.15 (-0.28 to 0.57)	
Suspicious (assessment 4 or 5)	21	0.84 (0.77 to 0.92)	
MBI interpreted with mammogram ^b		0.87 (0.79 to 0.95)	0.84 (0.84, 0.77 to 0.92)
Negative or benign (assessment 1 or 2)	26	0.90 (0.82 to 0.97)	
Probably benign (assessment 3)	2	0.30 (-0.30 to 0.89)	
Suspicious (assessment 4 or 5)	20	0.92 (0.85 to 0.99)	

^a Number of breasts with given feature by expert consensus.

^b MBI studies of 48 breasts (26 of 27 women) had accompanying mammograms.

agreement was lowest (with widest confidence intervals) for features which were infrequently seen (such as multiple regions of non-mass-like uptake). Interobserver agreement for final assessment was slightly higher in our study than in a prior MBI study in which two observers retrospectively interpreted screening MBI images in isolation [1]. The diagnostic accuracy in this series was also higher than in the prior study, possibly related to the artificially high number of positive examinations included in this study for the purpose of comparing lesion descriptions.

Our study was similar in design to lexicon validation studies performed for other modalities. It differs because in most of those studies, the lesions to be described were

specified on the image, whereas in this study they were not. Hence, it was possible for observers in our study to identify lesions in addition to those identified by the consensus readers. Of the 50 breasts included in the interpretation task, only three such additional lesions were identified.

This proposed MBI lexicon does necessarily differ from other breast imaging lexicons. Because of the relatively low resolution of MBI, finer morphologic details such as mass shape and mass margins cannot be assessed and are not included. Additional terms such as lesion uptake intensity and background uptake intensity need to be included given the functional nature of imaging with ^{99m}Tc sestamibi.

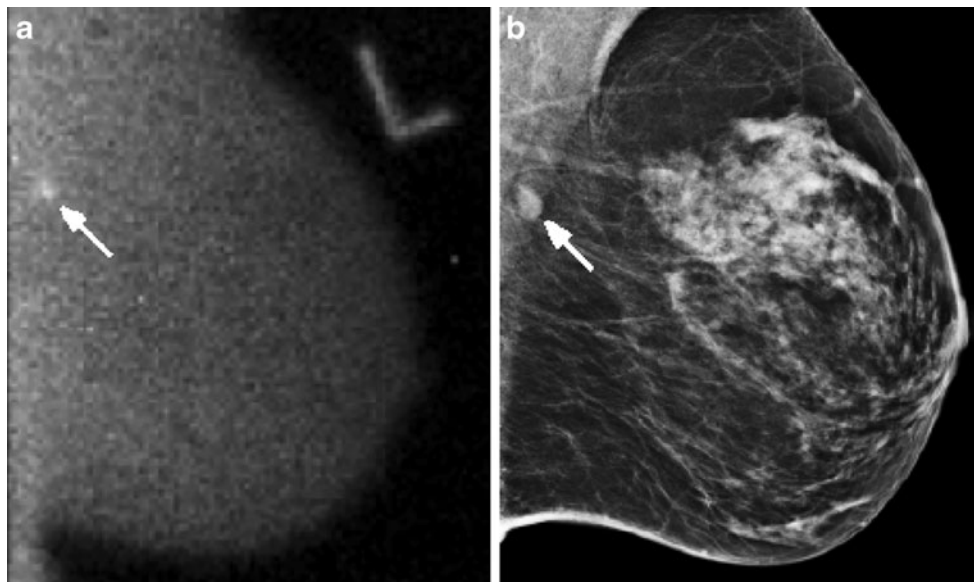


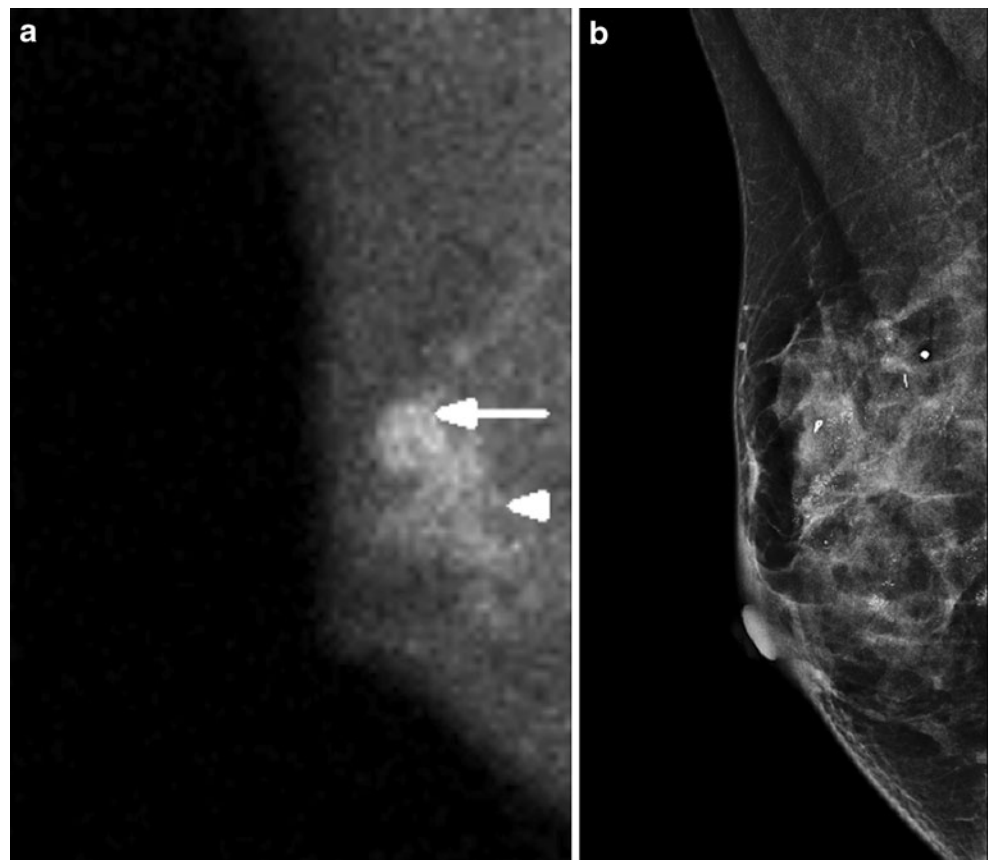
Fig. 1 MLO images of the left breast in a 64-year-old woman: MBI image following injection of 322 MBq of ^{99m}Tc sestamibi (**a**) and the mammogram (**b**) show a mild intensity focal area of non-mass-like uptake in the left axilla (*arrow*). This corresponds with a low axillary lymph node which had been mammographically stable for at least 4 years. Three observers (including consensus readers) who had given

a suspicious or probably benign assessment based on the MBI alone, changed to a benign assessment when the mammogram was provided. Two observers gave a suspicious assessment which did not change with mammographic comparison. One observer rated the finding as benign both before and after mammographic comparison

Because we did not prespecify which lesions should be described, a few potentially confusing diagnostic situations

were brought to light. In the case of consensus-determined marked background uptake, four observers described marked

Fig. 2 MLO images of the right breast in a 46-year-old woman: MBI image following injection of 1,099 MBq of ^{99m}Tc sestamibi (**a**) and mammogram (**b**). Three observers (as well as the consensus readers) classified this as heterogeneous non-mass uptake only (*arrowhead*), while three described it as non-mass uptake with a second finding of a mass (*arrow*). All observers as well as the consensus readers assigned assessment category 4 or 5. Final pathology showed multifocal invasive ductal carcinoma forming three adjacent masses



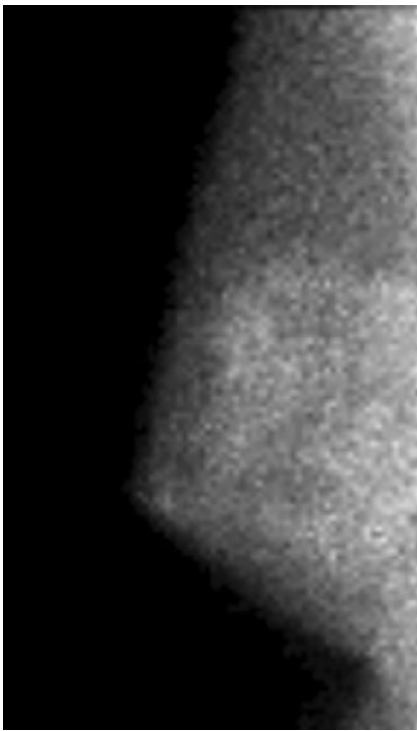


Fig. 3 MLO MBI image of the right breast in a 45-year-old woman following injection of 1,106 MBq of ^{99m}Tc sestamibi. Uptake in the right breast was classified by some observers (including the consensus readers) as moderate background uptake, and by other observers as diffuse non-mass-like uptake. All observers rendered a negative/benign final assessment. Pathology was benign at prophylactic mastectomy

background uptake and two described diffuse non-mass-like uptake. We propose that diffuse non-mass-like uptake only be reported when it is believed to constitute a lesion; diffuse uptake would be more likely to be a suspicious finding when asymmetric with the contralateral breast. In one study, some observers described an enlarged axillary lymph node as an additional finding of axillary uptake, while others described a mass located in the axilla. In this situation, comparison with mammography and/or ultrasonography would be especially helpful to determine whether the axillary uptake represents a pathologic finding (as it did in this case).

We noted relatively frequent use of the “probably benign” assessment category, with 11 of 50 breasts given such an assessment by at least one observer, including three malignancies. In the work-up algorithm presented during training, a negative targeted diagnostic mammogram and ultrasonogram showing the area of probably benign uptake would be necessary before short-term follow-up MBI could be recommended. Observers may have been more comfortable with category 3 because additional imaging is the initial recommendation. Of note, only four breasts were given an assessment of 3 after the mammogram was provided (no malignancies). There was one malignant MBI study where no mammogram was available at the time of this observer study, as the mammography had been performed at another

institution. Two malignant cases initially given an assessment of “probably benign” by at least one observer were appropriately upgraded to category 4 once the mammogram was shown. The use of category 3 has not been validated for MBI. Lesions designated as category 3 may have a higher rate of malignancy than the accepted rate of 2% for mammography [22–24], ultrasonography [25–27] and MRI [28, 29]; of note, a malignancy rate greater than 2% has been shown for category 3 lesions seen on PEM [8].

It is important that interpretive performance improved further when mammograms (and occasionally breast ultrasonograms) were shown to the observers. Similar results were seen for PEM interpretation [8], reinforcing the view that breast imagers should be involved in interpreting breast nuclear medicine examinations. Validation among nuclear medicine physicians has not been performed.

Additional studies at other sites will be necessary to show that the proposed lexicon is also useful at other institutions. We do not yet know how much risk is associated with each lexicon descriptor, such that the lexicon does not offer guidance to interpreters as to which assessment category to assign. Our hope is that the availability of this lexicon will lead to additional research which will further delineate the predictive value of each descriptor/combination of descriptors. Until that data are available, we determine our level of suspicion by factoring in lesion intensity, distribution, conspicuity, the pretest probability of malignancy, as well as findings on additional imaging, as recommended in the Society of Nuclear Medicine guidelines [30]. Higher final assessments are given when there is radiotracer uptake which represents a definite lesion distinct from background that is not accounted for by known benign findings.

This lexicon could likely be used to describe breast-specific gamma imaging (BSGI) and scintimammography findings as well, though validation for such uses is recommended.

Our study had a number of limitations. First, not all terms in the proposed lexicon were evaluated. Symmetry, internal enhancement pattern (homogeneous or heterogeneous) and additional findings (nipple, venous and axillary uptake) were not analyzed. Terms for the description of lesion size and location were also not studied, but are analogous to those in mammography. Quantitative measurement of uptake in breast lesions was not evaluated because of current software limitations, although this is technically possible [31, 32]. An additional limitation of our study was that all observers were from the same institution, which may in part account for the high feature analysis agreement. Cases were preselected in order to include discrete lesions which could be described for lexicon validation, which mimics a diagnostic population going for biopsy. This selection bias may have contributed to the high sensitivity observed. The artificial environment created in administering an interpretation

task may not represent how observers would perform in practice [33, 34]. Finally, there was no histological proof of benignity for 8 of the 11 benign lesions included. In some of these cases, follow-up was not longer than 1 year. It is possible that an underlying malignancy was present which did not present itself until after the follow-up period.

In summary, the lexicon used in this study was an easily adopted tool for describing findings on MBI studies. The proposed MBI lexicon is based on BI-RADS terms that are familiar to breast radiologists. Substantial to near-perfect agreement and high diagnostic performance were achieved with relatively little MBI-specific training.

Acknowledgments We gratefully acknowledge Drs. Darcy Adamczyk, Kathy Brandt, Tara Henrichsen, Gina Hesley, Katie Jones, and Marilyn Morton for participating as the radiologist observers in this study. Both Gamma Medica-Ideas and GE supplied the equipment used to acquire the images shown.

Conflicts of interest Dr. Hruska discloses a potential financial interest associated with technologies described in this article due to licensing arrangements between Mayo Clinic and Gamma Medica-Ideas. Dr. Berg is a consultant to Naviscan, Inc., has prepared educational materials for Gamma Medica-Ideas, is on the Medical Advisory Board of Philips, and has received travel support and been compensated for case review and manuscript preparation by SuperSonic Imagine.

The remaining authors declare that they have no conflicts of interest.

References

- Rhodes DJ, Hruska CB, Phillips SW, Whaley DH, O'Connor MK. Dedicated dual-head gamma imaging for breast cancer screening in women with mammographically dense breasts. *Radiology*. 2011;258:106–18.
- Hruska CB, Phillips SW, Whaley DH, Rhodes DJ, O'Connor MK. Molecular breast imaging: use of a dual-head dedicated gamma camera to detect small breast tumors. *AJR Am J Roentgenol*. 2008;191:1805–15.
- Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, et al. The ACR BI-RADS experience: learning from history. *J Am Coll Radiol*. 2009;6:851–60.
- D'Orsi CJ, Bassett LW, Berg WA, et al. *Breast Imaging Reporting and Data System, BI-RADS: mammography*. 4th ed. Reston: American College of Radiology; 2003.
- Berg WA, D'Orsi CJ, Jackson VP, Bassett LW, Beam CA, Lewis RS, et al. Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? *Radiology*. 2002;224:871–80.
- Mendelson EB, Baum JK, Berg WA, Merritt CRB, Rubin E. *Breast Imaging Reporting and Data System, BI-RADS: ultrasound*. 1st ed. Reston: American College of Radiology; 2003.
- Ikeda DM, Hylton NM, Kuhl CK, et al. *Breast Imaging Reporting and Data System, BI-RADS: magnetic resonance imaging*. Reston: American College of Radiology; 2003.
- Narayanan D, Madsen KS, Kalinyak JE, Berg WA. Interpretation of positron emission mammography and MRI by experienced breast imaging radiologists: performance and observer reproducibility. *AJR Am J Roentgenol*. 2011;196:971–81.
- Weinmann AL, Hruska CB, O'Connor MK. Design of optimal collimation for dedicated molecular breast imaging systems. *Med Phys*. 2009;36:845–56.
- Boughey JC, O'Connor MK, Hruska CB, Neal L, Jakub JW, Degnim AC, et al. Molecular breast imaging in the preoperative surgical workup of women with biopsy proven breast cancer (abstract). *Ann Surg Oncol*. 2010;17:S49.
- Rhodes DJ, Hruska CB, Tortorelli CL, Maxwell RW, Conners AL, O'Connor MK. Low-dose molecular breast imaging with Tc-99 m sestamibi for screening in women with dense breasts (abstract). *J Nucl Med*. 2011;52:671.
- Crewson PE. Reader agreement studies. *AJR Am J Roentgenol*. 2005;184:1391–7.
- Fleiss JL, Levin BA, Paik MC. *Statistical methods for rates and proportions*. 3rd ed. Hoboken: Wiley; 2003.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
- Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol*. 2000;174:1769–77.
- Abdullah N, Mesurolle B, El-Khoury M, Kao E. Breast Imaging Reporting and Data System lexicon for US: interobserver agreement for assessment of breast masses. *Radiology*. 2009;252:665–72.
- Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology*. 2006;239:385–91.
- Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol*. 2008;65:293–8.
- Ikeda DM, Hylton NM, Kinkel K, Hochman MG, Kuhl CK, Kaiser WA, et al. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *J Magn Reson Imaging*. 2001;13:889–95.
- Stoutjesdijk MJ, Futterer JJ, Boetes C, van Die LE, Jager G, Barentsz JO. Variability in the description of morphologic and contrast enhancement characteristics of breast lesions on magnetic resonance imaging. *Invest Radiol*. 2005;40:355–62.
- Baker JA, Kornguth PJ, Floyd Jr CE. Breast Imaging Reporting and Data System standardized mammography lexicon: observer variability in lesion description. *AJR Am J Roentgenol*. 1996;166:773–8.
- Sickles EA. Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. *Radiology*. 1991;179:463–8.
- Varas X, Leborgne JH, Leborgne F, Mezzera J, Jaumandreu S. Revisiting the mammographic follow-up of BI-RADS category 3 lesions. *AJR Am J Roentgenol*. 2002;179:691–5.
- Vizcaino I, Gadea L, Andreo L, Salas D, Ruiz-Perales F, Cuevas D, et al. Short-term follow-up results in 795 nonpalpable probably benign lesions detected at screening mammography. *Radiology*. 2001;219:475–83.
- Graf O, Helbich TH, Hopf G, Graf C, Sickles EA. Probably benign breast masses at US: is follow-up an acceptable alternative to biopsy? *Radiology*. 2007;244:87–93.
- Mainiero MB, Goldkamp A, Lazarus E, Livingston L, Koelliker SL, Schepps B, et al. Characterization of breast masses with sonography: can biopsy of some solid masses be deferred? *J Ultrasound Med*. 2005;24:161–7.
- Raza S, Chikarmane SA, Neilsen SS, Zorn LM, Birdwell RL. BI-RADS 3, 4, and 5 lesions: value of US in management – follow-up and outcome. *Radiology*. 2008;248:773–81.
- Eby PR, DeMartini WB, Gutierrez RL, Saini MH, Peacock S, Lehman CD. Characteristics of probably benign breast MRI lesions. *AJR Am J Roentgenol*. 2009;193:861–7.

29. Liberman L, Morris EA, Benton CL, Abramson AF, Dershaw DD. Probably benign lesions at breast magnetic resonance imaging: preliminary experience in high-risk women. *Cancer*. 2003;98:377–88.
30. Goldsmith SJ, Parsons W, Guiberteau MJ, Stern LH, Lanzkowsky L, Weigert J, et al. SNM practice guideline for breast scintigraphy with breast-specific gamma-cameras 1.0. *J Nucl Med Technol*. 2010;38:219–24.
31. Hruska CB, O'Connor MK. Quantitation of malignant and benign breast lesions with dual-head molecular breast imaging (abstract). Radiological Society of North America (RSNA) Annual Meeting. Chicago, IL; December 3, 2010.
32. Hruska CB, O'Connor MK. Quantification of lesion size, depth, and uptake using a dual-head molecular breast imaging system. *Med Phys*. 2008;35:1365–76.
33. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology*. 2008;249:47–53.
34. Rutter CM, Taplin S. Assessing mammographers' accuracy. A comparison of clinical and test performance. *J Clin Epidemiol*. 2000;53:443–50.