
Positron Emission Mammography Image Interpretation for Reduced Image Count Levels

Lawrence R. MacDonald¹, Daniel S. Hippe¹, Leila C. Bender², Elizabeth W. Cotter², Pooja R. Voria², Paula S. Hallam², Carolyn L. Wang^{1,2}, David R. Haseley², Mary M. Kelly², Jay R. Parikh², J. David Beatty², and James V. Rogers²

¹Radiology Department, University of Washington, Seattle, Washington; and ²Swedish Cancer Institute, Swedish Medical Center, Seattle, Washington

We studied the effects of reduced ¹⁸F-FDG injection activity on interpretation of positron emission mammography (PEM) images and compared image interpretation between 2 postinjection imaging times. **Methods:** We performed a receiver-operating-characteristic (ROC) study using PEM images reconstructed with different count levels expected from injected activities between 23 and 185 MBq. Thirty patients received 2 PEM scans at postinjection times of 60 and 120 min. Half of the patients were scanned with a standard protocol; the others received one-half of the standard activity. Images were reconstructed using 100%, 50%, and 25% of the total counts acquired. Eight radiologists used a 5-point confidence scale to score 232 PEM images for the presence of up to 3 malignant lesions. Paired images were analyzed with conditional logistic regression and ROC analysis to investigate changes in interpretation. **Results:** There was a trend for increasing lesion detection sensitivity with increased image counts: odds ratios were 2.2 ($P = 0.01$) and 1.9 ($P = 0.04$) per doubling of image counts for 60- and 120-min uptake images, respectively, without significant difference between time points ($P = 0.7$). The area under the ROC curve (AUC) was highest for the 100%-count, 60-min images (0.83 vs. 0.75 for 50%-counts, $P = 0.02$). The 120-min images had a similar trend but did not reach statistical significance (AUC = 0.79 vs. 0.73, $P = 0.1$). Our data did not yield significant trends between specificity and image counts. Lesion-to-background ratios increased between 60- and 120-min scans ($P < 0.001$). **Conclusion:** Reducing the image counts relative to the standard protocol decreased diagnostic accuracy. The increase in lesion-to-background ratio between 60- and 120-min uptake times was not enough to improve detection sensitivity in this study, perhaps in part due to fewer counts in the later scan.

Key Words: positron emission mammography (PEM); injected activity; low-dose PET; free-response ROC

J Nucl Med 2016; 57:348–354

DOI: 10.2967/jnumed.115.165787

Motivated by remaining challenges in diagnosis, staging, and management of breast cancer, small, high-resolution PET scanners dedicated to breast imaging have been investigated since the 1990s (*1–13*). Dedicated breast PET systems (positron emission

mammography [PEM]) have spatial resolutions better than whole-body (WB) PET scanners by factors of 2–4 and are much smaller than WB PET scanners, allowing placement of the detectors close to the breast, thus increasing geometric detection efficiency for annihilation photons relative to WB PET. In theory, increased detector efficiency allows for lower injected activities or shorter scan times while maintaining a fixed image noise level.

Increased detector efficiency adds to the number of detected coincidence counts during a scan, which is the underlying metric determining inherent image noise. The activity injected into the patient (A_{inj}), uptake time (T_{up}), scan duration (T_s), and detector efficiency are the primary factors determining the number of PEM scan counts. The PEM Flex Solo II scanner (PEM Flex; CMR Naviscan) consists of 2 bar detectors (6×16 cm imaging area) that scan in unison along the 6-cm dimension to cover 16×24 cm. Photon detection efficiency is fundamentally limited by the small detector size, which counteracts the increased geometric efficiency obtained by proximity to the subject.

Recommendations for ¹⁸F-FDG injected activity have been established for WB PET in the United States at 370–740 MBq (*14*). Equivalent guidelines have yet to be established for PEM scanning. A protocol of 370-MBq injection, 45–60 min of uptake, and 10-min T_s was used in early studies that began with an earlier version of the scanner (PEM Flex Solo I) (*5*). The PEM Flex Solo II model uses thicker scintillation crystals, and, in consultation with the system's vendor, we at the Swedish Cancer Institute adopted a PEM protocol of $A_{inj} = 370$ MBq, 60-min uptake, and 7-min T_s . We were interested in whether this protocol provided advantages over the use of lower A_{inj} .

In a previous study, we found that noise in PEM Flex images decreased only slowly as the activity concentration was increased above 2 kBq/mL, which is roughly one-half of the concentration expected in a typical patient when using the standard injection and uptake protocol (*15*). That study also showed a slow decrease in detection sensitivity with lowered image count density, with sensitivity remaining above 90% for a lesion diameter of 7.8 mm or more for image noise corresponding to A_{inj} down to 100 MBq. Furthermore, specificity was 98% for all activities tested (46–370 MBq), suggesting that statistical noise texture alone does not lead to false-positive findings. One limitation of the phantom study was uniform activity in the background; normal breast ¹⁸F-FDG uptake varies with tissue type, creating a heterogeneous background (*16*).

In the present study, our goal was to evaluate the effect of lowering A_{inj} on the interpretation of PEM images. To study this, we investigated lesion detection sensitivity and specificity on PEM images with different image count levels corresponding to

For correspondence contact: Lawrence R. MacDonald, University of Washington Radiology Department, 1959 NE Pacific St., Rm. RR-215, P.O. Box 357987, Seattle, WA 98195.

E-mail: macdon@uw.edu

Published online Dec. 3, 2015.

COPYRIGHT © 2016 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

conventional and reduced A_{inj} . We repeated the analysis for images acquired at 2 postinjection time points.

MATERIALS AND METHODS

Patient Cohort

This study was approved by an institutional review board and was Health Insurance Portability and Accountability Act compliant. Consecutive patients who met the study criteria and provided informed written consent were imaged with the PEM Flex between August 2010 and October 2012. We enrolled 30 patients: 10 each with invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), and ductal carcinoma in situ (DCIS). For the first 5 patients of each pathology, we used an equivalent of our standard PEM protocol. The second half of the study used one-half of the injected activity. For this study, we used a lower dose equivalent alternative to our standard PEM protocol as described below.

Disease type was confirmed by core biopsy, and classification was based on routine pathologic histologic assessment of the biopsy samples using the standard care at our institution. PEM imaging was performed before surgical treatment.

Control images were taken from PEM scans of the disease-free contralateral breast of the study subjects (contralateral controls); however, not all study subjects had a contralateral control image. Additional control images were taken from a database of prior institutional review board–approved PEM studies (unmatched controls) (17). We included the additional controls to better balance the ratio of disease-to-normal scans. Normal breast parenchyma in controls was confirmed by negative physical examinations, negative mammograms, and negative contrast-enhanced MRI after 18–36 mo of follow-up. Eight contralateral controls were also negative on pathology after a prophylactic mastectomy.

PEM Scanner

The PEM Flex scanner is a limited-angle tomosynthesis system yielding 12 image slices, 16.3×24.0 cm with 1.2-mm pixel size (18,19). Slice thickness is one-twelfth of the distance between the 2 detectors, which is adjusted for each patient to immobilize the breast with mild compression. Spatial resolution is 2.4 mm in full width at half maximum on image slices reconstructed in the standard mode that we used in this work. The image reconstruction algorithm is maximum likelihood expectation maximization with no user-adjustable parameters. No corrections are made for scattered, attenuated, or accidental coincidence events in this system. Details of the PEM Flex hardware and performance characteristics are found in previous publications (15,18,19).

PEM Scanning

Study patients were imaged at 60 and 120 min after injection of ^{18}F -FDG. Imaging was performed in the mediolateral oblique (MLO) orientation only. Patient preparation requirements were fasting at least 12 h, blood glucose less than 150 mg/dL, no history of diabetes, and limited exercise for 24 h before scan.

The first half of patients for each pathology type received an A_{inj} of 185 MBq and a T_s of 14 min, which yielded image counts equivalent to our standard PEM protocol of 370 MBq for A_{inj} and 7 min for T_s by virtue of the same $A_{inj} \times T_s$ product and scanner counting linearity (15). We refer to this as the standard injected activity (A_{STD}). For the second half of the study, we used an A_{inj} of 92.5 MBq while keeping the same T_s of 14 min ($A_{STD}/2$).

Control images were acquired using different T_s and tracer T_{up} , resulting in different image noise characteristics relative to case images. To classify image noise levels, we used the effective injected activity (A_{eff}) for control and case images, by adjusting the actual A_{inj}

to match a protocol with 7-min acquisition duration and a 60-min T_{up} using the following equation:

$$A_{eff} = A_{inj} \left(\frac{T_s}{7 \text{ min}} \right) 2^{-(60 \text{ min} - T_{up})/T_{1/2}}, \quad \text{Eq. 1}$$

where $T_{1/2}$ is the ^{18}F half-life.

Contralateral controls were acquired after the 120-min ipsilateral scan and included both MLO and craniocaudal views. All images/views were separately randomized, and interpreters interpreted only a single view at a time. Interpreters were not made aware that disease cases were MLO only. Unmatched controls taken from the earlier study followed a protocol specified for that earlier study in which PEM scanning followed a clinically ordered WB PET/CT. That protocol specified $A_{inj} = 592$ MBq, PEM acquisition of $T_s = 7$ min, and variable T_{up} due to the PEM scan after a WB PET/CT examination (17).

Image Generation

From each PEM scan, we generated 3 PEM images with different count levels to represent different injected activities. This allowed us to study image interpretation differences in matched pairs differing only in the image count level. In a previously validated offline process (15), we subtracted events from the list-mode data file saved by the PEM Flex scanner to keep only 50% and 25% of the acquired data. Subtraction was done uniformly throughout the entire data file. The count-subtracted list-mode files were reconstructed in the same manner as the original file that contained all counts. The result was a triplet (100%, 50%, and 25% count levels) of the same patient image but with differing levels of statistical noise. Reduced-count images were assigned A_{eff} that were reduced relative to A_{inj} by the fraction of counts subtracted, resulting in 4 categories of A_{eff} (A_{STD} , $A_{STD}/2$, $A_{STD}/4$, and $A_{STD}/8$) among the 2 cohorts receiving $A_{inj} = A_{STD}$ and $A_{inj} = A_{STD}/2$.

The contralateral control images were processed using the same count-subtraction method to generate a triplet of contralateral controls. Raw list-mode data from the unmatched controls were not available for the count-subtraction procedure.

Images were randomly sorted into 3 groups with similar disease-to-control image ratios, with each group containing a similar distribution of image counts spanning all count levels. No 2 images from a triplet (100%, 50%, and 25%) were in a single group. Groups did contain both 60- and 120-min postinjection images of the same patient, each at a random count level.

Interpreter Study

Eight radiologists interpreted the PEM images. The 8 interpreters specialized in breast imaging—6 radiologists were Mammography Quality Standards Act (MQSA)–certified; 1 was an American Board of Radiology (ABR)–Nuclear Medicine and American Board of Nuclear Medicine physician with 1 y of PEM interpretation experience; and 1 was an ABR physician previously MQSA-certified, with PET/CT fellowship training and 3 y of PEM interpretation experience. The interpreters received training by one of the authors on PEM image interpretation and dedicated PEM viewing software (MIMVista). Interpreters practiced the scoring process on an independent training set of 16 PEM images, with interactive follow-up before initiating the study. The interpreters observed single PEM scans (consisting of 12 image slices) and were told to place a region of interest (ROI) on up to 3 ^{18}F -FDG foci that they considered suggestive for disease; with each ROI, interpreters provided a confidence score between 1 (almost definitely no lesion) to 5 (almost definitely a lesion present).

Each radiologist interpreted 3 groups of PEM images in at least 3 interpretation sessions on different days. No other patient information, imaging or otherwise, was available to the interpreters.

TABLE 1
Patient Demographics and Imaging Parameters

Variable	Cohort group		
	Cases, first half* (n = 16)	Cases, second half* (n = 14)	Unmatched controls† (n = 8)
Age (y)	58 (38–69)	57 (40–72)	56 (44–69)
Mass (kg)	72 (50–116)	70 (54–116)	74 (59–92)
Tumor type			N/A
DCIS	5 (31)	5 (33)	
IDC‡	6 (38)	5 (33)	
ILC‡	5 (31)	5 (33)	
Longest dimension (mm)			
DCIS	25 (18–59)	21 (4–61)	N/A
IDC‡	22 (3–24)	12 (12–26)	
ILC‡	30 (12–75)	18 (12–55)	
Acquisition time (min)			
Ipsilateral	13.9 (13.9–13.9)	13.9 (13.9–13.9)	N/A
Contralateral/unmatched control	7.0 (7.0–13.9)	9.9 (7.0–13.9)	7.0 (4.9–7.0)
A _{inj} (MBq)	200 (189–218)	96.2 (92.5–104)	599 (566–611)
A _{eff} (MBq) [§]			
Ipsilateral	399 (377–432)	193 (181–205)	N/A
Contralateral and unmatched control	197 (188–400)	133 (96–196)	571 (422–602)
Postinjection time (min)			
60-min image (ipsilateral)	59 (51–71)	60 (55–72)	N/A
120-min image (ipsilateral)	118 (90–131)	119 (116–123)	N/A
Contralateral controls	137 (131–180)	153 (135–174)	N/A
Unmatched controls	N/A	N/A	111 (79–153)
Counts (per 1,000)			
60-min image (ipsilateral)	1,673 (1,083–5,605)	955 (463–1,735)	N/A
120-min image (ipsilateral)	1,017 (710–3,980)	609 (243–987)	N/A
Contralateral controls	323 (247–680)	274 (156–494)	N/A
Unmatched controls	N/A	N/A	974 (674–2,390)

*12 of 30 (7 from 93-MBq group and 5 from 185-MBq group) had contralateral imaging.

†9 control breasts from 8 patients were imaged using older protocol with injected activity of 592 MBq; craniocaudal and mediolateral oblique views were acquired for each breast except for 1 patient with only MLO view.

‡One patient in 93-MBq group had both IDC and ILC in same breast.

§A_{eff}, Equation 1.

N/A = not applicable.

Values are median, with range in parentheses.

We created an answer key for the study by identifying the true lesion location on the PEM images by correlation with other imaging (mammography and MRI) and using the final, postsurgery pathology report to confirm locations and sizes.

Data Analyses

As an image noise metric we used the coefficient of variation, COV = (ROI SD)/(ROI mean), from multiple background ROIs placed in areas of relatively uniform ¹⁸F-FDG uptake away from lesions, using at least 1 ROI on each image slice. We calculated changes in image noise as the relative (%) change in COV for matched ROIs between 100% and 25% count-level images, and between 60- and 120-min uptake images, which were then averaged across image ROIs and patient images.

We calculated ¹⁸F-FDG uptake using the lesion-to-background ratio (LBR), defined as the maximum voxel value in the lesion divided by the mean of an adjacent background ROI. We calculated changes in LBR between 60- and 120-min images and used Student *t* tests to check for statistical differences between baseline LBR and changes in LBR across lesions of differing pathology.

To process interpreter interpretation data, the location of each interpreter ROI and its distance to a true lesion, if present, was considered. For truly matched controls, we used disease-free regions of the ipsilateral image. On each image with a lesion, the ipsilateral disease-free region was assigned the maximum score of ROIs not associated with any lesion, for example, false-positives. We also used each interpreter's maximum score per image with multiple ROIs on

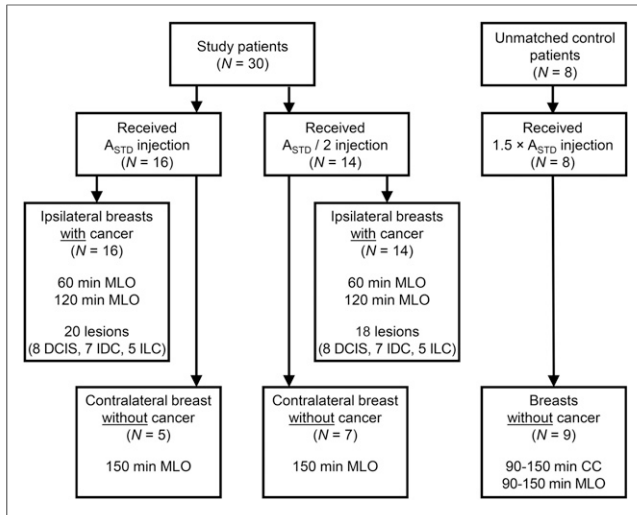


FIGURE 1. Study patient flowchart. CC = craniocaudal view; MLO = mediolateral oblique view.

control images. Light's κ was used to summarize interinterpreter agreement (20).

The overall sensitivity and specificity were estimated by classifying scores 3 or greater as a positive diagnosis and 2 or less as a negative diagnosis. Sensitivity and specificity were estimated using images grouped by A_{eff} . Trends in sensitivity by A_{eff} level were evaluated using odds ratios (ORs) per doubling of A_{eff} from conditional logistic regression to account for the matched design (each image was reproduced with 50% and 25% of the counts of the original image). The conditional logistic regression models were stratified by subject, lesion, and interpreter, so the ORs correspond to changes in sensitivity due to changes in A_{eff} for the same interpreter and the same lesion. The nonparametric bootstrap and the percentile method was used to calculate 95% confidence intervals (CIs) and P values by resampling patients to preserve the dependence between images of the same patient (21).

We used free-response receiver-operating-characteristic (FROC) curve methods, which summarize diagnostic performance while accounting for correct localization of a lesion (22). The matched diseased-free regions of the ipsilateral breast were used as the controls because all A_{eff} levels were available for these controls. The area under the FROC curve (AUC) acts as the figure of merit. The AUC from each curve was compared between A_{eff} levels using the nonparametric bootstrap.

All statistical calculations were conducted with the statistical computing language R (version 3.1.1; R Foundation for Statistical

Computing). Throughout, 2-sided tests were used, unless otherwise specified, with statistical significance defined as a P value of less than 0.05. P values were not adjusted for multiple comparisons.

RESULTS

The final patient cohort consisted of 30 breast cancer cases. Twenty-three patients had a single lesion, 6 had 2 lesions, and 1 had 3 lesions. There were 12 matched controls and 8 unmatched control patients, with 1 providing control images from both breasts. After adjusting for T_{up} and acquisition duration according to Equation 1, the 8 unmatched controls were categorized with $A_{\text{eff}} = A_{\text{STD}}$. Table 1 provides patient demographics and imaging parameters, and Figure 1 is a study flowchart.

A total of 232 distinct images were available for review. This included 180 breast cancer images, 35 contralateral control images, and 17 unmatched control images. Breast cancer images were generated from 30 patients, with 2 T_{up} and 3 count levels. Control images came from count-subtracted triplets of contralateral controls and dual-view unmatched control scans.

Eight interpreters produced 1,814 usable reviews of the 1,856 assigned. Forty-two (2.2%) could not be used because the interpreter did not leave a rating on the image, which occurred with approximately the same proportion for cases and controls. There was moderately good interinterpreter agreement on the presence/absence of lesions across all images ($\kappa = 0.56$; 95% confidence interval, 0.47–0.63).

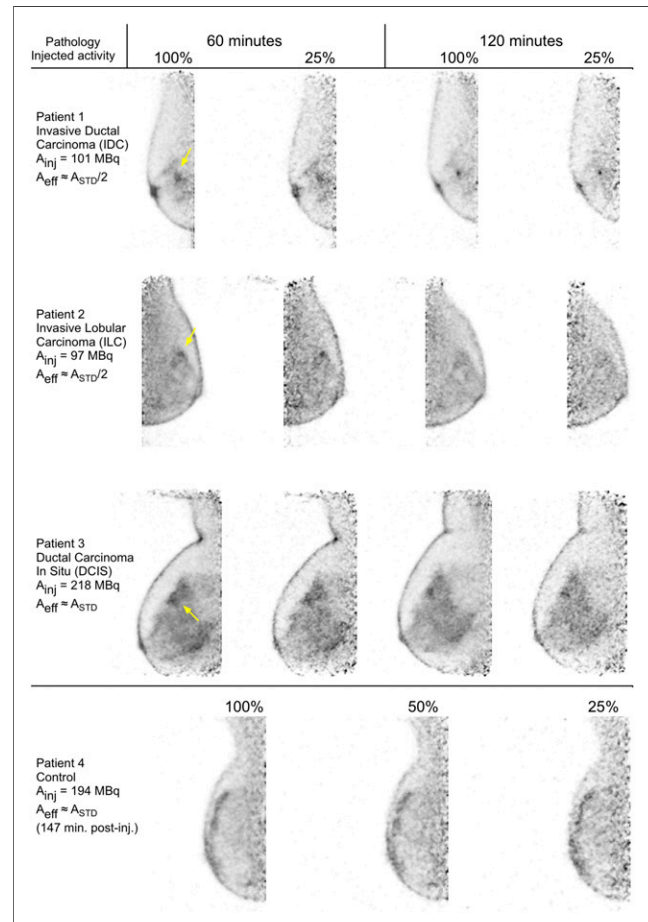


FIGURE 2. Example PEM images at different count levels, acquired 60 and 120 min after ^{18}F -FDG injection.

TABLE 2
LBR

Pathology	LBR: 60-min images		ΔLBR : $\text{LBR}_{120 \text{ min}} -$ $\text{LBR}_{60 \text{ min}}$		P^*
	Mean	SD	Mean	SD	
Cumulative	4.24	2.73	1.89	1.88	<0.001
DCIS	3.25	1.07	0.86	1.00	0.024
IDC	5.27	2.30	2.70	1.84	<0.001
ILC	4.10	3.92	2.04	2.24	0.018

*One-sample t test for $\Delta\text{LBR} = 0$.

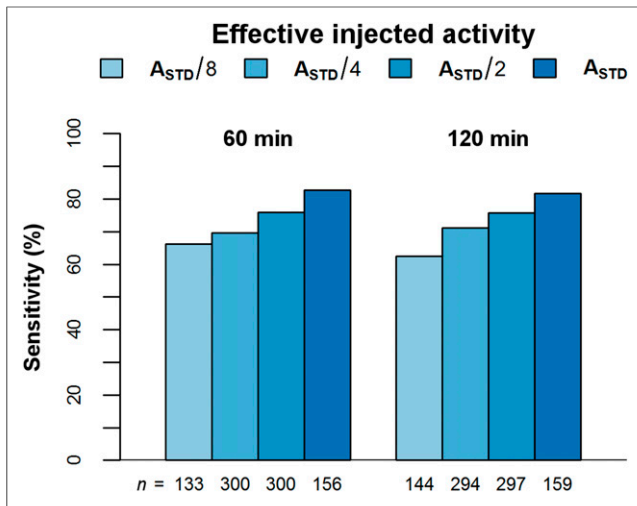


FIGURE 3. Overall sensitivity across interpreters by A_{eff} , based on 38 lesions from 30 patients. There were statistically significant trends between increasing dose and increasing sensitivity (Table 3). Below each bar is corresponding number of interpretations used in calculations (each image of each lesion was interpreted by up to 8 interpreters). Bootstrap 95% confidence intervals for each bar are approximately $\pm 11.2\%$, $\pm 12.0\%$, $\pm 11.9\%$, and $\pm 12.9\%$, respectively, for 60-min images and $\pm 18.8\%$, $\pm 12.9\%$, $\pm 10.3\%$, and $\pm 9.5\%$, respectively, for 120-min images.

LBR was seen to increase between 60- and 120-min scans for all lesion types (Table 2). The increase in LBR was significantly higher for IDC than for DCIS ($P = 0.01$), but no other pairs differed significantly. Two-sample t tests for differences in mean LBR for differing lesion pathologies showed statistical significance only between IDC and DCIS at both 60-min ($P = 0.02$) and 120-min ($P = 0.01$) time points.

Image noise as measured by background ROI COV followed the expected trends of increasing for images reconstructed using fewer counts and for later-time-point imaging. Image noise versus image counts in this study followed a nonlinear trend similar to the one seen with phantom image tests (15), in which COV changes slowly for higher A_{eff} ($A_{\text{eff}} > A_{\text{STD}}/2$), then begins to increase rapidly for lower image counts ($A_{\text{eff}} < A_{\text{STD}}/2$).

Figure 2 shows example images for each lesion pathology at different T_{up} and percentage count levels and a control image.

There was a statistically significant increasing trend in interpreter sensitivity for diagnosing the presence of a lesion with each

doubling of A_{eff} for both the 60-min T_{up} images (OR = 2.2 per doubling A_{eff} , $P = 0.01$) and the 120-min uptake images (OR = 1.9 per doubling A_{eff} , $P = 0.004$), as shown in Figure 3 and Table 3. These trends in sensitivity were not significantly different between the 60- and 120-min images ($P = 0.7$).

Diagnostic sensitivity by A_{eff} was also explored within subgroups defined by the different lesion types. Qualitatively, sensitivity was most strongly affected by A_{eff} for ILC with 120-min uptake (OR = 4.3 per doubling of A_{eff}) and least for IDC at 60 or 120 min (OR = 1.5–1.6) and DCIS at 120 min (OR = 1.6) (Table 3). Sensitivity tended to be lowest for DCIS at nearly all A_{eff} levels and T_{up} (Fig. 4). The trends with A_{eff} did not change when A_{eff} was further normalized by patient weight and compression thickness (attenuation correction).

Table 4 summarizes specificity by A_{eff} using different control images. The only type of controls that was available for each A_{eff} level was the ipsilateral disease-free region of the images with lesions present (ipsilateral controls). The images of the contralateral disease-free breasts were always acquired after the 120-min images, so A_{eff} was lower due to tracer decay. On the basis of the ipsilateral controls, the average specificity across A_{eff} levels was 73% (60 min) and 71% (120 min) without a clear trend across levels. Specificity was lower using the other types of controls, but the sample size of those controls was substantially lower and they did not cover all A_{eff} levels.

FROC curves are shown in Figure 5 for each A_{eff} level. The AUC was highest for A_{STD} at both 60 and 120 min of uptake. At 60 min, the A_{STD} AUC was significantly higher than the $A_{\text{STD}}/2$ AUC (0.83 vs. 0.75, $P = 0.02$). There was a similar trend at 120 min (0.79 vs. 0.73, $P = 0.1$). The AUCs at $A_{\text{STD}}/4$ and $A_{\text{STD}}/8$ were similar to that at $A_{\text{STD}}/2$ and ranged from 0.73 to 0.75.

DISCUSSION

This study showed trends of decreasing lesion detection sensitivity in PEM Flex images when image count levels were reduced from our standard PEM scan protocol. The FROC analysis also yielded the best results for the standard injection protocol via a significantly higher AUC for $A_{\text{eff}} = A_{\text{STD}}$, relative to lower A_{eff} . These results suggest that for the highest interpretation accuracy, A_{inj} should not be lowered below commonly used levels without a proportional increase in T_{s} using the PEM Flex Solo II. This result held for ^{18}F -FDG T_{up} of 60 and 120 min, without significant difference

TABLE 3
Conditional Logistic Regression Analysis of Sensitivity Trends for Doubling A_{eff} Within Subgroups

Group	60-min images			120-min images			60 vs. 120, P
	OR*	95% CI	P	OR*	95% CI	P	
All lesions	2.2	1.2–4.3	0.012	1.9	1.3–3.1	0.004	0.74
Lesion pathology							
IDC	1.5	0.5–4.3	0.30	1.6	1.2–9.5	0.005	0.86
ILC	2.2	1.0–9.5	0.042	4.3	1.7– ∞	0.001	0.43
DCIS	2.7	0.8–49.6	0.082	1.6	0.5–4.6	0.26	0.33

*OR corresponds to each doubling of A_{eff} .
CI = confidence interval.

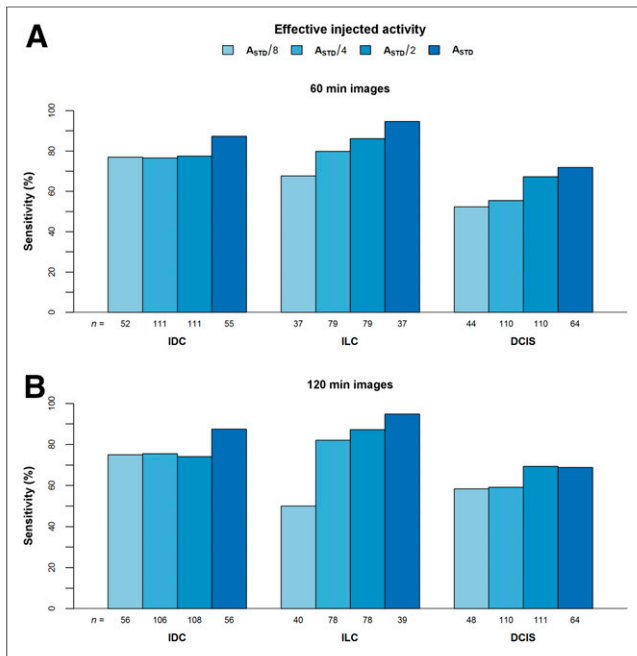


FIGURE 4. Detection sensitivity by pathology for 60-min images (A) and 120-min images (B). Table 3 shows tests of trends between sensitivity and A_{eff} . Calculations are based on 14 IDC, 10 ILC, and 14 DCIS lesions from 38 patients. Below each bar is corresponding number of interpretations used in calculations (each image of each lesion was interpreted by up to 8 interpreters). Because of small sample sizes in subgroups, width of bootstrap 95% confidence interval for each bar ranges from 22% to 48% (median, 37%) for IDC, 15% to 70% (median, 33%) for ILC, and 36% to 76% (median, 47%) for DCIS.

between the 2, despite the fact that 120-min images had higher LBR, which should increase lesion conspicuity.

The trends were similar when analyzed by lesion pathology type. The differences between lesion types observed here are consistent with prior studies of ^{18}F -FDG in breast lesions of differing pathology (17,23,24). In this study, the LBR was lower in ILC ($P = 0.4$) and in DCIS ($P < 0.02$), relative to IDC. Further,

we observed ILC detection sensitivity having the strongest dependence on A_{eff} and overall lower sensitivities for DCIS at all A_{eff} .

Matching the image count levels between contralateral control images and case studies is challenging when each breast must be imaged sequentially, as is the case with the PEM Flex scanner. The delay between scans changes the count level due to radioisotope decay. We corrected for this, and for differences in acquisition duration, by categorizing images according to A_{eff} (Eq. 1). In retrospect, a preferable study design would have acquired contralateral breast images between the 60- and 120-min scans of the ipsilateral breast. Instead, we acquired contralateral matched control images after the 120-min ipsilateral scan, which inhibited ROC analysis using contralateral controls due to insufficient matched case/controls in each A_{eff} category.

Variable patient weight and breast size influence how much activity is in the scanner field of view, thus leading to different PEM image count densities for a given A_{inj} . Although this variability makes it difficult to predict the number of counts that will be collected for a given A_{inj} , we did find a strong correlation (Pearson $\rho = 0.61$) between counts collected in the PEM images and A_{eff} . As noted, further adjustments to account for patient mass and photon attenuation did not change our results. Variable counts within the A_{eff} categories did not confound the comparisons in this study because the paired-image study design ensured that comparisons were made between images with known count differences.

No clear trends were seen in the relationship between specificity and A_{eff} , perhaps due to the limited sample size. An initial concern was that pronounced nonuniform noise textures in higher noise images could be mistaken for focal uptake of tracer, thus reducing specificity. Alternatively, if fewer lesions were identified on lower count control images, as was the case on images with disease, then specificity could improve.

Radiologists performing the study had different backgrounds and mixed experience interpreting PEM images. There was reasonable agreement between interpreters ($\kappa = 0.55$). A study by Narayanan et al. showed that experienced breast imagers interpreted PEM images with high performance after minimal training (25).

The geometry of the PEM Flex with 2 small bar detectors results in limited-angle data acquisition and consequently tomosynthesis image reconstructions with anisotropic spatial resolution. Breast PET

TABLE 4
Specificity of A_{eff} Using Each Type of Control Image

A_{eff}	60-min images, matched controls				120-min images, matched controls				Unmatched controls [†]	
	Ipsilateral		Contralateral*		Ipsilateral		Contralateral*			
	Specificity	95% CI	Specificity	95% CI	Specificity	95% CI	Specificity	95% CI	Specificity	95% CI
A_{STD}	78.2% (97/124)	69.1–87.1	—	—	75.6% (96/127)	66.9–84.0	—	—	54.5% (73/134)	31.4–71.1
$A_{\text{STD}}/2$	68.6% (162/236)	60.2–76.6	—	—	65.4% (153/234)	55.5–75.0	57.1% (32/56)	35.9–79.7	—	—
$A_{\text{STD}}/4$	76.4% (181/237)	69.3–82.9	61.1% (44/72)	45.0–78.1	72.1% (168/233)	63.8–80.0	78.3% (72/92)	68.4–87.9	—	—
$A_{\text{STD}}/8$	72.1% (75/104)	61.6–80.9	81.3% (74/91)	70.7–91.3	83.9% (94/122)	73.3–94.2	77.9% (67/86)	67.1–88.3	—	—

*Contralateral control images were decay-corrected to 60 or 120 min.

[†] A_{eff} of unmatched controls decay-corrected to 60 min.

CI = confidence interval.

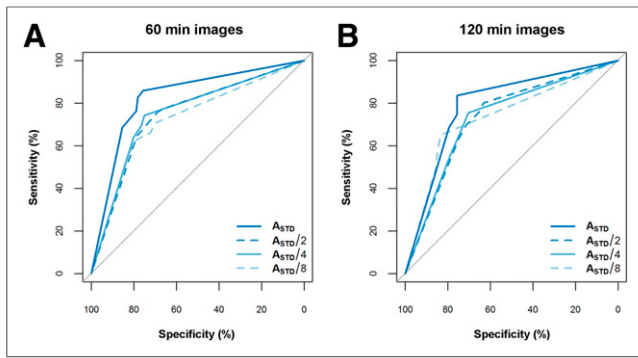


FIGURE 5. FROC curves for each A_{eff} (solid and dashed curves) using 60-min images (A) and 120-min images (B). AUC for A_{STD} was significantly higher than for $A_{\text{STD}}/2$ using 60-min images (AUC, 0.83 vs. 0.75, $P = 0.02$), with similar trend for 120-min images (AUC, 0.79 vs. 0.73, $P = 0.1$).

scanners whose detectors completely surround the breast (4,11,12) or that rotate to obtain complete 360° angular sampling (6,7) are capable of fully 3-dimensional, isotropic tomographic images. The impact of the anisotropic resolution on our results, and potential benefits of isotropic resolution for lesion detection, are not clear. Another consequence of tomosynthesis is that image quantification of tracer uptake is typically compromised (26), although there are approaches to overcome this limitation (27).

PET detectors that completely surround the breast provide much higher photon detection sensitivity than scanning-detector systems, and as a result, lower image noise would be expected from such systems for a given acquisition protocol. Consequently, such systems would be expected to reach their maximum diagnostic accuracy at a lower $A_{\text{inj}} \times T_s$ level than scanning systems. We continue to assert that there is a limiting $A_{\text{inj}} \times T_s$ level above which diagnostic accuracy will increase negligibly or not at all. The data in this study showed steady increases in detection sensitivity up to the highest $A_{\text{inj}} \times T_s$ that we tested (2.59 GBq-min), suggesting that we did not reach this limiting value or a level at which accuracy begins to plateau.

CONCLUSION

Our hypothesis that lowering PEM Flex image counts below what is obtained from our standard protocol would not change image interpretation was not substantiated by this study; diagnostic accuracy was lower on images with fewer counts than expected from an $A_{\text{inj}} \times T_s$ product of 2.59 GBq-min (at 60-min T_{up}). Hence, reducing injected activity without degrading image interpretation is limited by the capacity to increase T_s . The trend for lower lesion detection sensitivity with lower image counts may contribute to our observation that images acquired 120 min after injection did not have improved diagnostic accuracy despite the higher LBR in those images.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. This work was supported by the Swedish Foundation, Swedish Medical Center. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

We thank Angie King, Sarah Fanizzi, Julie Cleveland, Claire Buchanan, Patricia Dawson, James Hanson, and Christine Lee

for patient recruitment and Weidong Luo for assistance in data count subtraction.

REFERENCES

- Thompson CJ, Murthy KI, Weinberg IN, Mako F. Feasibility study for positron emission mammography. *Med Phys.* 1994;21:529–538.
- Doshi NK, Shao Y, Silverman RW, Cherry SR. Design and evaluation of an LSO PET detector for breast cancer imaging. *Med Phys.* 2000;27:1535–1543.
- Freifelder R, Cardi C, Grigoras I, Saffer JR, Karp JS. First results of a dedicated breast PET imager, BPET, using NaI(Tl) curve plate detectors. *IEEE Nucl Sci Symp Conf Rec.* 2001;3:1241–1245.
- Wang GC, Huber JS, Moss WW, et al. Characterization of the LBNL PEM camera. *IEEE Trans Nucl Sci.* 2006;53:1129–1135.
- Berg WA, Weinberg IN, Narayanan D, et al. High-resolution fluorodeoxyglucose positron emission tomography with compression ("positron emission mammography") is highly accurate in depicting primary breast cancer. *Breast J.* 2006;12:309–323.
- Raylman RR, Majewski S, Smith MF, et al. The positron emission mammography/tomography breast imaging and biopsy system (PEM/PET): design, construction and phantom-based measurements. *Phys Med Biol.* 2008;53:637–653.
- Bowen SL, Wu Y, Chaudhari AJ, et al. Initial characterization of a dedicated breast PET/CT scanner during human imaging. *J Nucl Med.* 2009;50:1401–1408.
- Berg WA, Madsen KS, Schilling K, et al. Breast cancer: comparative effectiveness of positron emission mammography and MR imaging in presurgical planning for the ipsilateral breast. *Radiology.* 2011;258:59–72.
- Schilling K, Narayanan D, Kalinyak JE, et al. Positron emission mammography in breast cancer presurgical planning: comparisons with magnetic resonance imaging. *Eur J Nucl Med Mol Imaging.* 2011;38:23–36.
- Berg WA, Madsen KS, Schilling K, et al. Comparative effectiveness of positron emission mammography and MRI in the contralateral breast of women with newly diagnosed breast cancer. *AJR.* 2012;198:219–232.
- Moliner L, Gonzalez AJ, Soriano A, et al. Design and evaluation of the MAMMI dedicated breast PET. *Med Phys.* 2012;39:5393–5404.
- Miyake KK, Matsumoto K, Inoue M, et al. Performance evaluation of a new dedicated breast PET scanner using NEMA NU4-2008 standards. *J Nucl Med.* 2014;55:1198–1203.
- Kalinyak JE, Berg WA, Schilling K, et al. Breast cancer detection using high-resolution breast PET compared to whole-body PET or PET/CT. *Eur J Nucl Med Mol Imaging.* 2014;41:260–275.
- Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with ^{18}F -FDG PET/CT 1.0. *J Nucl Med.* 2006;47:885–895.
- Macdonald LR, Wang CL, Eissa M, et al. Positron emission mammography (PEM): Effect of activity concentration, object size, and object contrast on phantom lesion detection. *Med Phys.* 2012;39:6499–6508.
- Narayanan D, Madsen KS, Kalinyak JE, Berg WA. Interpretation of positron emission mammography: feature analysis and rates of malignancy. *AJR.* 2011;196:956–970.
- Wang CL, MacDonald LR, Rogers JV, et al. Positron emission mammography: Correlation of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 status and ^{18}F -FDG. *AJR.* 2011;197:W247–W255.
- Macdonald LR, Edwards J, Lewellen TK, Haseley D, Rogers JV, Kinahan PE. Clinical imaging characteristics of the positron emission mammography camera: PEM Flex Solo II. *J Nucl Med.* 2009;50:1666–1675.
- Luo W, Anashkin E, Matthews CG. Performance evaluation of a PEM scanner using the NEMA NU 4 – 2008 small animal PET standards. *IEEE Trans Nucl Sci.* 2010;57:94–103.
- Conger AJ. Integration and generalisation of Kappas for multiple raters. *Psychol Bull.* 1980;88:322–328.
- Davison AC, Hinkley DV. *Bootstrap Methods and Their Application.* Cambridge, U.K.: Cambridge University Press; 1997.
- Chakraborty DP, Winter LHL. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology.* 1990;174:873–881.
- Buck A, Schirmeister H, Kühn T, et al. FDG uptake in breast cancer: correlation with biological and clinical prognostic parameters. *Eur J Nucl Med Mol Imaging.* 2002;29:1317–1323.
- Bos R, van der Hoeven JJM, van der Wall E, et al., Biologic correlates of ^{18}F fluorodeoxyglucose uptake in human breast cancer measured by positron emission tomography. *J Clin Oncol.* 2002;20:379–387.
- 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.* 2011;196:971–981.
- Springer A, Mawlawi OR. Evaluation of the quantitative accuracy of a commercially available positron emission mammography scanner. *Med Phys.* 2011;38:2132–2139.
- Surti S, Karp JS. Design considerations for a limited angle, dedicated breast, TOF PET scanner. *Phys Med Biol.* 2008;53:2911–2921.