2008

PET/CT detectability and classification of simulated pulmonary nodules using an SUV correction scheme

Andrew Nicholas Morrow
Louisiana State University and Agricultural and Mechanical College, andrew.n.morrow@gmail.com

PET/CT DETECTABILITY AND CLASSIFICATION OF SIMULATED PULMONARY NODULES USING AN SUV CORRECTION SCHEME

A Thesis

Submitted to the Graduate faculty of the
Louisiana State University and
Agricultural and Mechanical College
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in

The Department of Physics and Astronomy

by
Andrew Morrow
B.S., Louisiana State University, 2005
August, 2008

ACKNOWLEDGEMENTS

I owe some yard work to each of the observers (Dr. Bujenovic, Dr. Dahlin, Dr. Gibbons, Dr. Johnson, Dr. King, and Dr. Matthews), although not for the numerical one. She doesn't have a yard. My committee (Dr. Brown, Dr. Bujenovic, Dr. Gibbons, and Dr. Matthews) deserves thanks for the help they provided for my graduate work. The Board of Regents has my appreciation for funding me throughout my graduate career. I also thank Dr. Hogstrom and LSU for this wonderful opportunity to obtain a graduate degree in a great field.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Positron emission tomography (PET) and computed tomography (CT) together are a powerful diagnostic tool, but imperfect image quality contributes to false positive and false negative diagnoses by an observer despite experience and training. This work investigated a PET standard uptake value (SUV) correction scheme, based on partial volume effect (PVE), on the classification of lesions as benign or malignant in PET/CT images.

The correction scheme comprised several steps. The observer drew a region of interest (ROI) around the lesion using the CT dataset. The ROI was blurred with the assumed point spread function (PSF) of the PET scanner then re-sampled to the PET voxel size. The magnitude of the ROI-based PVE was used as a scaling factor to correct the lesion's tumor-to-background ratio (TBR), which was used as a surrogate for SUV in the PET images of the phantom. Computer simulations showed that the accuracy of the correction depends strongly on the accuracy of the ROI drawn on the CT images, especially for small lesions. Correction accuracy was affected slightly by mismatch of the simulation PSF to the actual scanner PSF. A receiver operating characteristic (ROC) study, using phantom data, was performed to evaluate the effect of the correction scheme on diagnostic performance. The correction scheme significantly increased sensitivity and slightly increased accuracy for all acquisition and reconstruction modes at the cost of a small decrease in specificity. Corrected TBRs more accurately represented actual TBRs than uncorrected TBRs. The observer study also found that, when using PET data alone, 3D ordered subset expectation maximization (OSEM) outperformed 3D filtered back-projection (FBP), 2D OSEM, and 2D FBP in terms of sensitivity, specificity, and area-under-the-ROC-curve values. However, when PET data was displayed with correlated CT data, with and without PVE correction, no combination of reconstruction algorithm and acquisition mode outperformed any other.

CHAPTER 1:
INTRODUCTION

The purpose of this study was to evaluate the effect of a standard uptake value (SUV) correction scheme, based on partial volume effect (PVE), on the accuracy of classification in terms of malignancy of single pulmonary nodules in positron emission tomography (PET) / computed tomography (CT) images. PET and CT together are a powerful diagnostic tool (1) (2) (3), but imperfect image quality resulting from noise, spatial resolution, PVE and other factors allows false positive and false negative diagnoses to be made by observers despite experience and training. PVE causes PET images to under report SUV in any sized region of interest (4). Corrections for PVE should allow for improved diagnostic performance (4).

The correction scheme comprised several steps. The observer drew a region of interest (ROI) around the lesion using the CT dataset. The ROI was blurred with the point spread function (PSF) of the PET scanner then re-sampled to the PET voxel size. The magnitude of the ROI-based PVE was used as a scaling factor to correct the lesion's tumor-to-background ratio (TBR), which was used as a surrogate for SUV in the PET images of the phantom. The correction scheme was characterized in terms of the assumptions that it made, such as lesion size, local PSF, and lesion positioning relative to the PET voxel grid. Simulated PET data were used to evaluate the correction scheme.

An observer study was conducted to obtain data to determine the effect of the correction scheme on lesion classification. Observers were asked to rate PET/CT phantom images according to the likelihood of malignancy of a lesion visible in the image, both with and without PVE correction. Ratings data were evaluated in terms of sensitivity, specificity and accuracy, as well with receiver operating characteristic (ROC) curve analysis. Lastly, diagnostic performance of observers was evaluated as a function of acquisition mode and reconstruction algorithm, using

PET data alone, PET/CT data, and PET/CT data with correction of PVE. Ratings used for this evaluation were acquired during the observer study.

The evaluation of the correction scheme focused on its use for classification of single pulmonary nodules. Simulations showed that the PVE correction accuracy depended strongly on the choice of lesion ROI, especially for small lesions. However, for the range of ROI boundary errors examined, the corrected value was always more accurate than the uncorrected value. Less error was observed for ROI boundaries that were too large than for boundaries that were too small. When the assumed ROI boundary closely matched the true lesion size, the correction scheme produced an accurate corrected SUV value. The correction scheme did not strongly depend on the accuracy of the assumed PSF.

The observer study concluded malignant lesion classification performance improved with correction for PVE. Using the correction scheme improved the observers' ability to classify malignant lesions correctly. This came at the cost of slightly reducing the observers' ability to classify benign lesions correctly. Corrected SUV values were more accurate compared to the true value than uncorrected SUV values 97% of the time.

Also, this study found that PVE correction had more impact on classification performance than the choice of acquisition mode and reconstruction algorithm. The observer study showed that for classification based on PET data alone, 3D OSEM images provided the best classification performance in terms of sensitivity, specificity, and area-under-the-ROC-curve values. However, when PET data was evaluated with its correlated CT, the selection of acquisition mode and reconstruction algorithm did not significantly alter classification performance, either with or without PVE correction.

1.1 OVERVIEW OF THESIS

This project, as summarized above, is described in the following chapters of the thesis. Chapter 2 discusses the goal, aims, hypotheses, and tasks of this work. Chapter 3 reviews PET/CT imaging and ROC analysis, and discusses why an observer study was used in this thesis. Chapter 4 describes the research methods, including image acquisition and processing, how the correction scheme works, and how the implementation of the observer study. Chapter 5 presents the results from the correction scheme characterization, the observer study, and the assessment of the correction scheme's impact on lesion classification. Chapter 6 summarizes the results and describes the conclusions drawn from the research.

CHAPTER 2:
HYPOTHESES AND MOTIVATIONS

2.1 PURPOSE

The purpose of this study was to show that an SUV correction scheme, based on a model of PVE, can improve the accuracy of classification of single pulmonary nodules in PET/CT images.

2.2 RESEARCH GOALS

This work had three primary goals.

1. To show that a model of voxelization, lesion size, and lesion location relative to the voxel grid can provide case-by-case correction of PVE in PET images.
2. To evaluate the effect of the correction scheme on the accuracy of classification of single pulmonary nodules.
3. To assess the influence of acquisition mode and reconstruction algorithm on the performance of the correction scheme.

2.3 HYPOTHESES

Two hypotheses guided this work. This work sought to show that an SUV correction scheme will improve classification of single pulmonary nodules, by showing that PET images corrected for the effects of PVE yield more accurate classification than uncorrected PET data. Acquisition mode and reconstruction algorithm were expected to affect the classification accuracy.

2.3.1 Hypothesis 1

The SUV correction scheme, for any choice of acquisition mode or reconstruction algorithm, will improve the quality of observers' classifications, as measured by an increased ROC area-under-curve (A_z), sensitivity, specificity, and accuracy.

An experienced observer should know that PET images misreport SUVs, especially for smaller lesions. However, experience is likely insufficient to allow the observers to correctly

estimate the true SUVs. The SUV correction scheme, by providing a quantitative estimate of the magnitude of error in the reported SUV, will help the observer to decide if an equivocal lesion should instead be classified as a concern.

2.3.2 Hypothesis 2

The observer study will show that for PET data, PET/CT data, and PET/CT data corrected for PVE, that ordered subset expectation-maximization (OSEM) reconstruction in 3D mode outperforms OSEM in 2D mode or filtered back projection (FBP) in either 2D or 3D mode, in terms of detection and classification of single pulmonary nodules.

OSEM reconstruction creates smoother, sharper images than FBP (5); studies have shown that OSEM outperforms FBP in terms of lesion detection (5) (6). The quantitative accuracy of OSEM is similar to that of FBP when the lesion is in a region of low background (7). However, because the OSEM reconstruction algorithm creates qualitatively better images than FBP, the OSEM reconstruction algorithm should outperform FBP in terms of malignant lesion detection and classification of single pulmonary nodules.

A 2D acquisition mode is less sensitive than a 3D mode, although it also has fewer scattered events and random coincidences. Because the 3D mode is more sensitive, it should be less noisy and able to detect smaller lesions; the 3D mode is quantitatively more accurate (8). Therefore, 3D mode should outperform 2D mode in terms of malignant lesion detection and classification of single pulmonary nodules.

2.4 LIST OF AIMS

This project was executed in three main aims. Chapter 4 describes the materials and methods used to complete these tasks. Chapter 5 presents the results.

1. Develop the SUV correction scheme (for Hypothesis 1)

    1a. Implement the correction scheme algorithm in software

1b. Characterize the behavior of the correction scheme using simulation data

2. Develop the materials and methods for the observer study (for both Hypothesis 1 and Hypothesis 2)

   2a. Acquire and process the PET/CT data of simulated solitary pulmonary nodules in an anthropomorphic torso phantom

   2b. Develop the presentation and data collection software for the observer study

3. Conduct the observer study (for both Hypothesis 1 and Hypothesis 2)

   3a. Recruit experienced observers

   3b. Obtain the observer performance data for classification of single pulmonary nodules with and without the correction scheme, for the combinations of acquisition mode and reconstruction algorithm studied in this project

   3c. Analyze the observer data with ROC analysis and in terms of sensitivity, specificity, and accuracy

   3d. Assess quantitative accuracy of the correction scheme, including the impact of acquisition mode and reconstruction algorithm

CHAPTER 3:
REVIEW OF PET/CT IMAGING AND ROC ANALYSIS

3.1 PET/CT IMAGING

PET/CT is a popular diagnostic imaging method. PET maps metabolic activity; CT shows anatomy and provides attenuation correction factors needed for PET imaging. Using these two modalities in combination reportedly improves diagnostic accuracy (9).

3.1.1 PET Imaging

PET is functional imaging. When the radiotracer fluorine-18 flourodeoxyglucose (F-18 FDG), a glucose analogue, is injected into a patient, the PET image maps glucose uptake. Glucose uptake is proportional to metabolic activity. Increased metabolic activity can imply a number of things, one of them being malignant growth. PET imaging has consequently become a popular tool for oncology imaging. The following paragraphs provide a synopsis of PET imaging, followed by a discussion of PVE in PET imaging. Reference (10) and other texts provide more extensive discussion about PET imaging hardware and methods.

PET images are emission images – that is, a PET image maps the origins of photons created within and emitted from the body. The photons are created when the radiotracer emits a positron which annihilates with a nearby electron, creating two 511 keV annihilation photons that travel in opposite directions from one another. If the PET scanner detects these two photons within a particular interval of time, called the coincidence window, it will record a line of response (LOR) that connects the points where the two photons were detected. The collection of LOR data is referred to as the projection data.

When acquiring PET data, two types of spurious events can occur. One type is when one or both of the annihilation photons can undergo Compton scattering before reaching a detector. A scattered photon no longer travels in the opposite direction from the other annihilation photon,

and thus the recorded LOR will not intercept the position where the annihilation event actually occurred. Another spurious event that can occur is called a random event. This occurs when two photons that did not originate from the same annihilation event are detected within the PET scanner's coincidence window. The machine will erroneously record a LOR for these events.

PET spatial image resolution is poor; a PET scanner's point spread function (PSF) has a full width at half maximum (FWHM) usually around 6.5 mm or greater. Because of poor spatial resolution, PET images typically have large voxels (relative to CT voxels), which cuts down on the number of voxels in an image and increases the speed of image reconstruction.

Photons are attenuated within the patient. The attenuation along each possible path the photons can take can be determined; for instance, the attenuation map of an object can be obtained from a CT scan. Attenuation correction compensates for loss of signal in the PET image. Attenuation has the largest effect on the central regions of the patient. Thus, attenuation correction makes lesions that are more central in the patient easier to find. The combination of the annihilation photon data and the CT scan (for attenuation correction) is used to create the final PET image.

Pet scanners acquire data in either 2D mode or 3D mode; some scanners are configured to utilize one mode exclusively, while other scanners are capable of using either acquisition mode. When acquiring data in 2D mode, the PET scanner places septa between the detector rings. These septa block photons that are travelling longitudinally (head-to-toe direction). While this decreases the sensitivity of the machine, fewer random and scatter events are recorded with the septa in place. In 3D mode, the PET scanner retracts these septa. With the septa retracted, the PET scanner can detect more photons travelling longitudinally than when the septa are not retracted. This increases the sensitivity of the machine at the expense of more recorded random and scatter events.

Once acquired, PET data can be reconstructed in a number of ways; two common methods are FBP and OSEM. For FBP, the projection data are filtered with a ramp filter to correct for the blurring caused by backprojection. The projection data are then added back, or back-projected, into a tomographic image (10). OSEM is an iterative approach to image reconstruction (10). For OSEM, an initial activity distribution within the patient is assumed. Projection data are calculated using the assumed activity distribution and compared to actual projection data. The assumed activity distribution is adjusted based on this comparison. This process is repeated until actual projection data and assumed projection data match, within a certain tolerance.

PET data are often displayed in units of SUV. SUV is ratio of the radiotracer density in a given region to the total body mean uptake in the patient, so

$$\text{SUV} = \frac{\text{Radiotracer uptake } [\frac{\text{MBq}}{\text{mL}}]}{\text{Administered activity [MBq]}/(1000 * \text{Patient weight [kg]})} \quad (1)$$

If the density of the patient is assumed to be the same as water (ρ = 1 g/mL), total body SUV is equal to one. SUV is higher for metabolically active tissues, such as the heart, brain, or tumors.

The combination of inherent system resolution and image voxelation, otherwise known as PVE, causes a misreporting of SUV in any region of interest, which may reduce a PET image's sensitivity for finding malignancies.

3.1.2 Partial Volume Effect

PVE occurs from voxelation of PET images and from finite spatial resolution inherent in a PET scanner. Unblurred images of two lesions are illustrated in the left column of Figure 1. These images have voxels that are very small relative to the size of the lesions they depict. In the top row, the lesion is centered on a voxel corner while the lesion in the bottom row is centered on the center of a voxel. The finite spatial resolution causes an apparent spillover of activity from one region into another (4), as seen in the middle column of Figure 1, where the lesions appear

blurred. Two sources of error arise from voxelation. One is that voxelation averages data over a volume, causing a loss of spatial information, as seen in the right column of Figure 1, where the blurred images of the lesions appear blocky. The other source of error is that an object will look different depending on its position relative to the center of a voxel, as seen between the top and bottom images in the right column in Figure 1. In the top right image in Figure 1, the lesion is positioned over a voxel corner, resulting in the highest value of the image being shared between the four voxels that share the center of the lesion. In the bottom right image in Figure 1, the highest value of the image is seen in only one voxel, because the lesion is centered inside a voxel.

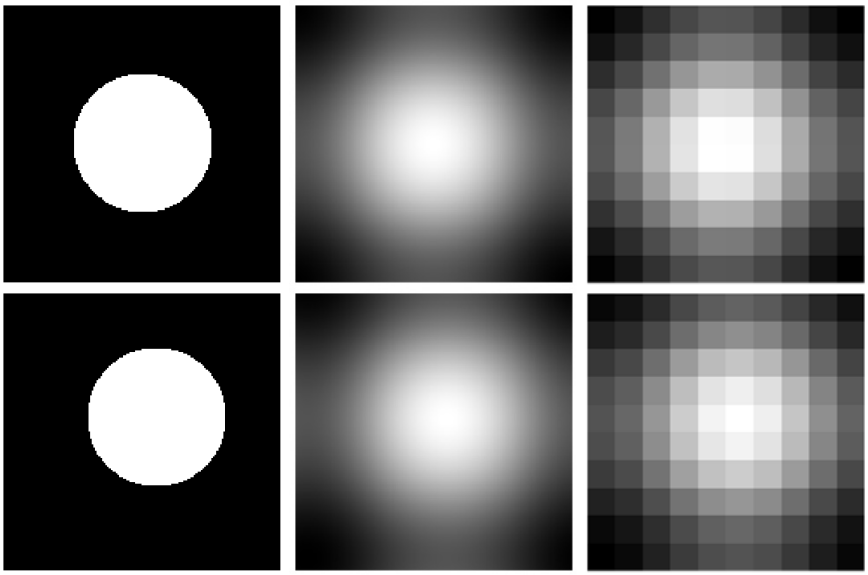

Figure 1: The effects of resolution and voxelation. Top row shows an object centered on the corner of a PET voxel and the bottom row shows an object centered on the center of a PET voxel. The left column shows the object unblurred and unvoxelated. The middle column shows the object blurred. The right column shows the effects of blurring and voxelation. The blurring function was a Gaussian with a FWHM equal to the diameter of the object.

There are a number of different methods to correct for PVE. Images can be corrected on a regional level (e.g. the average of an aggregate of pixels is corrected) or on a pixel-by-pixel basis. The values of individual pixels in PET images are highly variable and noise levels vary

between acquisition mode, reconstruction algorithm, patient size, and scan time length (4) (11) so pixel-by-pixel corrections are most useful for images with low noise (4). Three methods that correct for PVE on a regional level are the deconvolution method (12), the recovery coefficient method (13), and the geometric transfer matrix method (14). The deconvolution method iteratively deconvolves the region of interest in the PET image with the local PSF; it does not require anatomical data for delineation of the tumor boundary. The deconvolution method requires knowing the local spatial resolution to within 1 mm FWHM across the field of view (FOV) (8). Deconvolution is only appropriate for use with tumors with diameters greater than 1.5 times the local PSF FWHM (12). This method greatly amplifies noise. However, PET images corrected for PVE by deconvolution were more accurate than non-PVE corrected PET images even for sub-centimeter tumors (12).

The recovery coefficient method requires correlated anatomical data to find the actual tumor boundary. This method assumes homogenous uptake of FDG within the tumor. A recovery coefficient is found using the tumor region of interest (ROI). The recovery coefficient is multiplied by either the mean or maximum SUV within the ROI to give a PVE corrected value. There are several methods for creating a recovery coefficient. Assuming a spherical tumor, simple simulations of PVE can be done beforehand of spherical lesions, varying lesion diameter and scanner resolution to create a look up table. Once a tumor's volume has been determined, the proper recovery coefficient is obtained from the look-up table using a spherical volume equal to that of the tumor volume. This method has been commonly used in PET tumor imaging (4) and can increase PET's ability to find malignancies (13). Another way of finding the recovery coefficient is to do a simulation on-the-fly using a lesion ROI. This is more accurate for lesions that are not spherically shaped or for tumors with necrosis; it is identical to the look-up table method if the tumor and its correlated ROI are both spherical, and if the tumor is positioned in

relation to the voxel grid in the same way as the simulated lesion that was used to make the correction factor. For lesions only somewhat larger than the voxel size of the PET image, positioning with respect to the voxel grid may have a large impact on the accuracy of the recovery coefficient. Simulation based on the lesion ROI is the method used in this study.

The geometric transfer matrix method is an extension of the recovery coefficient method. Instead of assuming a single surrounding structure, as the recovery coefficient method does, the geometric transfer matrix method assumes that there are multiple different structures surrounding the ROI (4). Correlated anatomical data is required to delineate each surrounding structure. This method was not applicable for this study because lung is the only structure surrounding the simulated lesions in this study.

3.1.3 CT Imaging

CT is useful for anatomical imaging and can be used both for diagnosis and for attenuation correction of PET data. CT images are computed from transmission images – x-rays enter a patient from one side and those that emerge from the other side are detected. Data in a CT image thus represents x-ray attenuation. A single transmission measurement through an object made by a single detector is called a ray. Each set of rays that passes through an object at a given orientation of the x-ray source and detector is a projection. To create a CT image of a single slice of an object, many rays must be measured at many projection angles. Once these measurements are made, projection data are filtered to correct the blurring caused by the image reconstruction process. The filtered projection data are then added back, or back-projected, into a tomographic image. CT image spatial resolution is better relative to PET resolution. CT voxels are usually around 1 mm on a side, which are small compared to PET image voxels. The fundamentals of CT imaging are discussed in texts such as reference (10).

3.2 ASSESSING DIAGNOSTIC PERFORMANCE WITH OBSERVER STUDIES

While many studies have characterized PET and PET/CT scanners in terms of resolution, noise, and other characteristics (15) (16), the ultimate arbiter is diagnostic performance (17). Diagnostic performance is important because it assesses quality of decisions based on a system's images. Ultimately, an observer must make diagnoses based on a system's images. The quality of an observer's diagnoses is evaluated to give the diagnostic performance of a system.

3.2.1 Populations and Diagnostic Tests

A trait that may indicate disease can be shared between both healthy and diseased populations. A diagnostic test may be designed to measure the extent of this trait. In Figure 2, the distribution of healthy (left Gaussian curve) and diseased (right Gaussian curve) individuals with respect to the extent of the trait measured by a generic diagnostic test is shown. Notice that the test does not perfectly discriminate between healthy and diseased individuals – the two Gaussian curves overlap. Some healthy individuals show a higher amount of measured trait than some diseased individuals, for this particular example.

3.2.2 Metrics for Diagnostic Performance

Diagnostic performance can be assessed by measuring sensitivity, specificity, accuracy, or a number of other metrics. Sensitivity is the fraction of positive cases correctly identified as positive, specificity is the fraction of negative cases correctly identified as negative, and accuracy is the fraction of cases identified correctly, whether positive or negative (18).

3.2.3 ROC Analysis

Another method for assessing diagnostic performance of a test, person, or imaging system is ROC analysis. The ROC curve shows the tradeoff between true positive fraction (TPF or sensitivity) and false positive fraction (FPF) (18). The relationship of the populations to TPF and

FPF, as well as true negative fraction (TNF or sensitivity) and false negative fraction (FNF) are illustrated in Figure 2. FPF is the fraction of negative cases incorrectly identified as positive;

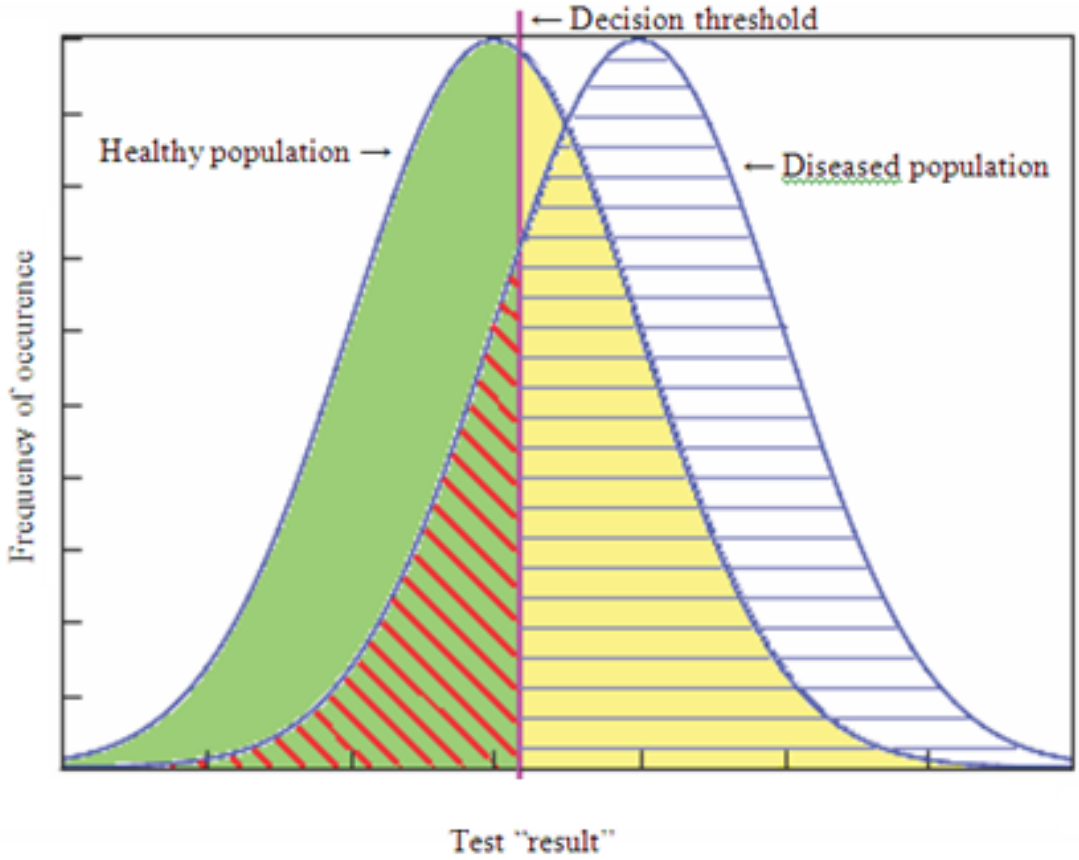

Figure 2: Distribution of healthy and diseased populations with respect to a test metric. The vertical purple line is the decision boundary above which cases are labeled positive. Displayed are true positives (blue hatch), true negatives (solid green), false positives (solid yellow), and false negatives (red hatch).

FPF is one minus the specificity. The ROC curve illustrates the impact of decision threshold placement. Each decision threshold corresponds to the value chosen to (arbitrarily) demarcate the states of health and disease using the test's measurements (Figure 2). An example of a decision threshold would be a particular SUV above which a test would label a lesion malignant. For the comparison of diagnostic tests, persons, or imaging systems, the ROC curve illustrates relative performance at specific decision thresholds.

Sensitivity, specificity, and accuracy only give diagnostic performance at a single decision threshold. However, an ROC curve maps TPF and FPF as a function of decision threshold,

showing the relative impact of decision threshold placement. Combining the knowledge of the tradeoffs between TPF and FPF at a given decision threshold and the knowledge of the relative costs (time, money, scheduling, etc.) of true and false positives can help with the selection of an optimal decision threshold for a given test (17) (19). Optimal placement of the decision threshold will maximize the number of true positives at an acceptable number of false positive mistakes.

The typical ROC curve plots TPF versus FPF as a function of decision threshold. To populate the points on this curve, diagnostic decisions are rated on a confidence scale. Ratings are used instead of a binary yes/no system to assess the performance of the system at different decision thresholds. Each rating level corresponds to a different decision threshold. At each threshold, positive images with measurements above the decision threshold are true positive results while negative images with measurements above the threshold are false positive results, as seen in Figure 2. The relationship between decision threshold (vertical purple line), the underlying populations of health (Gaussian distribution on the left) and disease (Gaussian distribution on the right), and the distribution of true positives (blue hatch), false positives (solid yellow), true negatives (solid green) and false negatives (red hatch) for a generic test is shown in Figure 2. The x-axis represents the extent of the measured trait and the y-axis represents the frequency of occurrence.

When attempting to fit a curve to the ROC rating data, a binormal population distribution may be assumed, with each normal distribution describing either the healthy or diseased populations (19). The independent variable is a diagnostic test result that correlates to the disease in question. The populations relate to the ROC curve through the parameters

(2) $\boldsymbol{a}$ = (distance peak-to-peak)/$\sigma_{\text{healthy population histogram}}$

and

(3) $\boldsymbol{b} = \sigma_{\text{diseased population histogram}} / \sigma_{\text{healthy population histogram}}$.

The ROC curve, as a function of ***a***, ***b***, and FPF, is

(4) $$\text{TPF} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{1}{\sqrt{2}}\left(\boldsymbol{a} + \sqrt{2}\,\boldsymbol{b} * \text{erf}^{-1}(2 * \text{FPF} - 1)\right)\right)\right)$$

An ideal test would maximize ***a*** and minimize ***b***, which would maximize TPF for all points on the ROC curve, as can be derived from equation (4). This would correspond to an increased A_z, the integral of the ROC curve, which is also called the area-under-curve. Maximizing ***a*** and minimizing ***b*** increases the separation of the distributions of the diseased and healthy populations with respect to the range of possible test results, and thus makes determining a diseased case from a healthy case easier. In the top left plot in Figure 3, the healthy and diseased populations are well separated and thus the fitted ROC curve has an ***a*** value that is large relative to its ***b*** value. Its plotted ROC curve, shown in the bottom right plot (dot-dashed line) of Figure 3, has a very high TPF for all decision thresholds and thus has a high A_z. A poor test cannot differentiate the populations from each other (top right plot of Figure 3), so it will have a fitted ROC curve with a small ***a,*** ***a*** large ***b,*** and a correspondingly low A_z (bottom right plot of Figure 3, dashed line). The bottom left plot of Figure 3 shows a typical situation where the distributions overlap. The fitted ROC curve will have moderate values for ***a*** and ***b***, resulting in the typical ROC curve shown in the bottom right plot of Figure 3 (solid line). The distributions in Figure 3 are normalized. However, ROC analysis does not require normalized distributions.

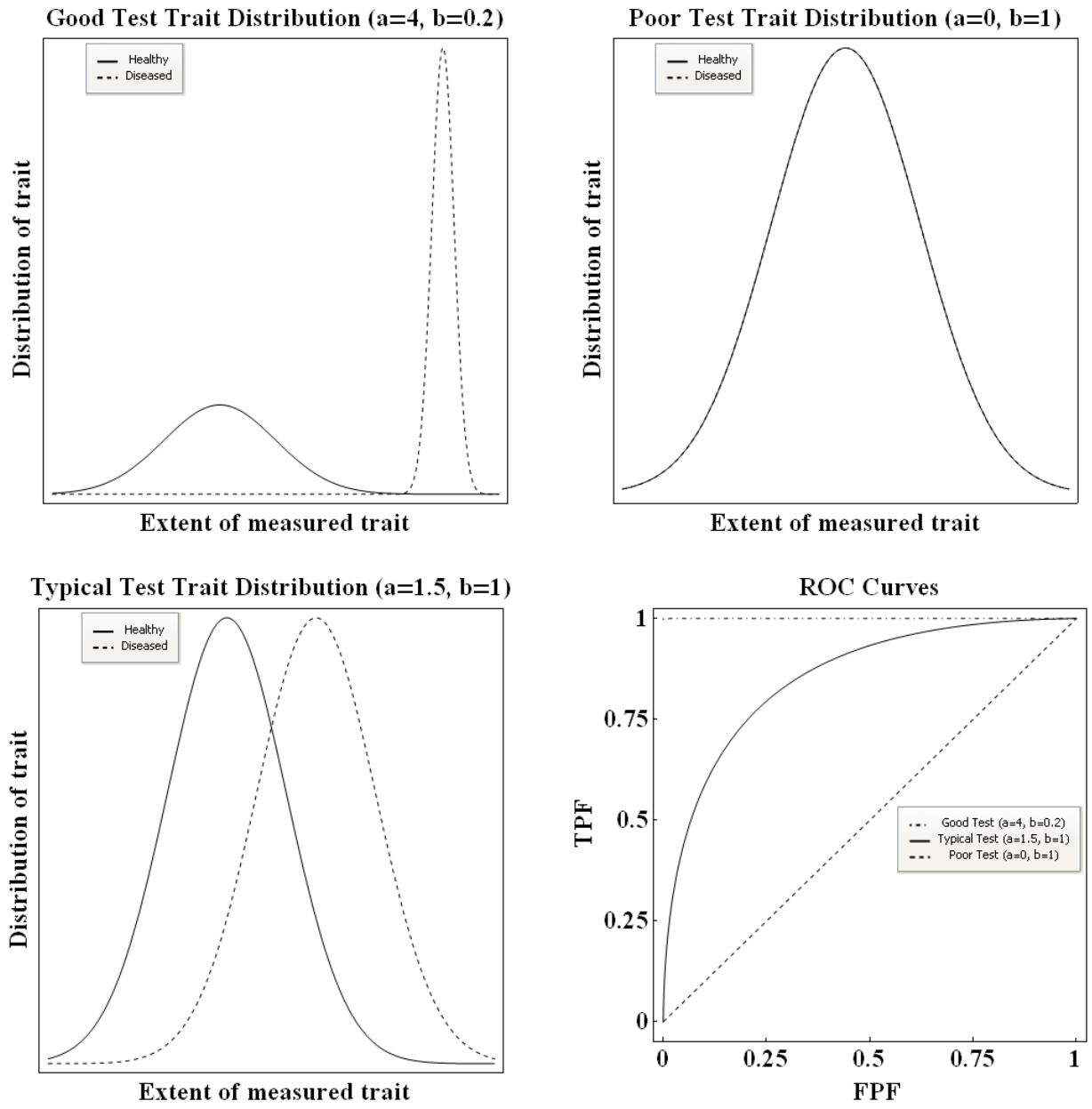

Figure 3: Distributions of healthy and diseased populations for good (top left), poor (top right), and typical (bottom left) tests. The corresponding ROC curves are plotted at bottom right. For the poor test, the healthy and diseased populations overlap completely.

CHAPTER 4:
METHODS AND MATERIALS

4.1 AIM 1: DEVELOP THE SUV CORRECTION SCHEME

In this section, the implementation of the recovery coefficient method of correcting for PVE is described. The recovery coefficient method was explained in Section 3.1.2; our implementation uses observer-drawn ROIs for case-by-case correction (4).

4.1.1 Implementation of the Correction Scheme Algorithm in Software

The correction scheme proceeded in several steps to determine the correction factor for a lesion. First, an ROI was drawn on the correlated anatomical CT data that matched the lesion boundary; the ROI was drawn using the CT data because of its higher resolution relative to the PET data. The PET ROI was created by copying the CT ROI to the corresponding PET image. Because PET voxel size did not match CT voxel size, a new PET boundary was made to determine the mean PET value. Voxels were assigned to the PET ROI if at least 50% of the PET voxel was overlapped by voxels from the CT drawn ROI. The mean PET value within the PET ROI was calculated. A simulated lesion image, defined by the CT drawn boundary, was created at the CT voxel size, given a uniform SUV, and blurred by the assumed PSF of the PET data. Blurring was done to mimic the effects of the imaging system. To blur the image of the lesion, the simulated lesion was convolved with the PSF. The PSF was a Gaussian distribution with dimensions of 6.5 x 6.5 x 6.0 mm^3 FWHM (in lateral, anterior-posterior, and craniocaudal directions, respectively), derived from results from measured PSF data for the same model PET/CT scanner used in this work (20). This blurred image was re-sampled to the PET voxel size to mimic the effect of voxelation. The known SUV of the simulated lesion was divided by the mean of the blurred and pixelated simulated lesion within the PET ROI to obtain the

correction factor. The measured mean PET value was then multiplied by the correction factor to create the corrected mean.

4.1.2 Evaluation of the Correction Scheme with Simulation Data

Simulated data were created both to evaluate the behavior of the SUV correction scheme used in this investigation and to help guide how corrected values and bounds should be created and displayed to the observers in the observer study.

To evaluate the correction scheme, the output of the scheme was investigated for a variety of adjustable scheme parameters. These parameters included lesion radius, scanner PSFs in the transverse and axial directions, and lesion placement relative to the voxel grid. For each parameter set, a simulated lesion image was created at the specified lesion size and positioned on a 0.25 mm x 0.25 mm x 0.25 mm voxel grid. The boundary of the lesion was drawn using CT sized voxels (0.98 mm x 0.98 mm x 3.27 mm). The lesion was blurred by the specified PSF. Finally, the lesion image was resampled to the larger PET voxel size (4.69 mm x 4.69 mm x 3.27 mm). This simulated lesion image was then used as input to the correction scheme.

The effects of several other variables were also investigated. One was the correction factor simulation PSF, which may not match the actual scanner PSF. Another variable was error in the CT drawn ROI. A new ROI, which slightly mismatched the ROI of the simulated PET lesion, was drawn for the correction factor simulation, which provided a means to evaluate the effect of boundary accuracy.

4.2 AIM 2: DEVELOP THE MATERIALS AND METHODS FOR THE OBSERVER STUDY

The second task of this study was to acquire and process PET/CT data for the observer study. Phantom images were used, instead of human images. With the use of a phantom, the exact values of radiotracer uptake and lesion size were known and controllable. Phantom images were

acquired for various lesion sizes and lesion radiotracer activity concentrations, as well as for two acquisition modes and two reconstruction algorithms.

4.2.1 The GE Discovery ST PET/CT Scanner

The PET/CT scanner used in this study was a GE Medical Systems Discovery ST PET/CT scanner. System specifications are summarized in Table 1. Retractable septa exist between the rings of detectors to allow acquisitions in either 2D or 3D mode. The scanner provides several reconstruction algorithms, including the OSEM and FBP algorithms used in this project. The scanner acquires the image data sequentially, acquiring helical CT data first then PET data. The CT data is used for both attenuation correction of the PET data during tomographic reconstruction and for co-registered display of anatomical images with the PET images. The spatial resolution of the PET imaging component, as measured for this model PET/CT scanner by Bettinard et. al. (20), is summarized in Table 2.

Table 1: GE Discovery ST manufacturer's specifications (15)

Attribute	Value
Detector material	Bismuth Germanate
Energy window (keV)	375-650
Crystal dimensions (mm^3)	6.3 x 6.3 x 30
Number of photomultiplier tubes	280
Coincidence window (ns)	11.7
Number of detector rings	24
Number of detectors per ring	420
Number of image planes	47
Detector ring diameter (cm)	88.6
Transverse FOV (cm)	70 (PET), 55 (CT)
Axial FOV (cm)	15.7

Table 2: GE Discovery ST resolution at different positions within the scanner (20)

Position/acquisition plane	FWHM (mm) (2D mode)	FWHM (mm) (3D mode)
1 cm radius/transverse	6.28	6.29
1 cm radius/axial	4.56	5.68
10 cm radius/transverse radial	6.94	7.01
10 cm radius/transverse tangential	6.82	6.64
10 cm radius/transverse axial	6.11	6.05

4.2.2 Phantom

The phantom comprised several components (Figure 4). The anthropomorphic torso phantom and the simulated lesions were the components imaged. The scatter phantom was used to mimic out of field scatter.

Figure 4: Pictured from left to right: The anthropomorphic torso phantom, a sample of micro hollow spheres used to simulate lesions, and the scatter phantom.

The torso phantom was the Data Spectrum Anthropomorphic Torso Phantom™ (Model ECT/TOR/P)[1] with the Cardiac Insert™ (Model ECT/CAR/I)[1]. The torso phantom simulates patients of average size (38 cm x 26 cm). The torso phantom includes two lung compartments filled with Styrofoam® beads. These compartments, when filled with water, simulate lung tissue with a density of ~0.3 g/mL. The other structures modeled in the torso phantom are the spine, which is a Teflon® cylinder, and the liver. The heart is simulated with two chambers: one for the myocardium and the other for the left ventricle: In nuclear medicine imaging, only the left ventricle is typically visualized, so it is the only chamber that is modeled in the phantom.

Clinically, radiotracer is not confined to the portion of the patient within the imaging FOV. To ensure that the acquired images reasonably represented clinical conditions for scatter, the NEMA PET Scatter Phantom™ (Model PET/NEMA-SCT/P)[1] was used. The scatter phantom was positioned distal to the anthropomorphic torso phantom and outside the imaging FOV. The

[1] Data Spectrum Corporation, 437 Dimmocks Mill Rd, Suite 17, Hillsborough NC 27278-2300

scatter phantom is a right circular polyethylene cylinder 70 cm in length and 20.3 cm in diameter. The scatter phantom has a fillable line source insert.

Fillable hollow acrylic spheres were used to simulate lesions. The spheres were from the Hollow Sphere Set (6)™ (Model ECT/HS/SET6)[1] and the Micro Hollow Sphere Set (4)™ (Model ECT/MI-HS/SET4)[1]. Images acquired of radioactivity filled spheres used sphere sizes of 0.125, 0.25, 0.5, 1.0, and 2.0 mL (fillable volumes). Images acquired of non-radioactive spheres used sphere sizes of 0.031, 0.063, 0.125, 0.25, 0.5, 1.0, and 2.0 mL (fillable volumes). Radioactivity-filled "hot" spheres represented lesions with PET radiotracer uptake, or PET active lesions, while "cold" spheres represented lesions with no PET radiotracer uptake. These "cold" lesions had the same uptake as their surrounding medium, the simulated lung.

4.2.3 Phantom Preparation

For this study, images were displayed in units of tissue to background ratio (TBR) instead of SUV. TBR is unitless: It is the radiotracer concentration in a given compartment of the phantom divided by the radiotracer concentration in the background portion of the phantom. The background portion of the phantom was meant to simulate the total body mean; total body mean SUV is one, assuming a patient density equal to that of water. The background portion of the phantom surrounded the lungs, heart, liver, and spine. Because each image was normalized to the activity concentration within the background portion of the phantom, the values for SUV and TBR were thus equivalent in magnitude. The actual TBRs used in the different compartments of the phantom were zero for the heart ventricle and lungs, 8.94 - 10.6 for the myocardium, 2.52 - 2.79 for the liver, 1 for the background, 0.968 - 1.14 for the scatter phantom, and 0 to 7.56 for the simulated lesions. Healthy lungs in patients normally have an SUV of 0.6 to 0.8. However, placing radioactivity in the lungs would have complicated the image acquisition and greatly increased the likelihood of contamination of personnel and equipment. The total activity injected

into the phantom, at time of assay, ranged from 5.16 mCi to 7.44 mCi for the various acquisition sessions.

After filling, the phantom was positioned on the PET/CT scanner table. The scatter phantom was placed according to NEMA recommendations for evaluating image quality (21). For acquisitions that included a simulated lesion, a single sphere was placed in one of the lung chambers. Multiple acquisitions with various lesion sizes were made during each session by replacing the spheres and repositioning the phantom. For each acquisition session, a different actual lesion TBR was used.

4.2.4 Acquisition Protocols and Reconstruction Methods

Following typical clinical protocols on the Discovery ST scanner used for this project, 4 minutes per bed position were used for 2D acquisitions and 3 minutes per bed position were used for 3D acquisitions. The 2D acquisitions were imaged prior to 3D acquisitions so that a clinically realistic count rate would be seen for the 2D image sets and then the activity could decay sufficiently to provide a reasonable count rate for the subsequent 3D acquisitions. The 2D sets were imaged at 1-2.5 hours after assay while the 3D sets were acquired at 2.5-4 hours after assay. Once a sinogram had been acquired, it was reconstructed both with FBP and with OSEM. The 3D OSEM data sets were reconstructed at 256x256 pixels per slice, but were re-binned for the observer study to 128x128 pixels per slice to match the other data sets. Figure 5 illustrates two fused PET/CT images acquired for this study. Table 3 lists the reconstruction parameters used in this study.

4.2.5 Image Processing

The PET images were manually rebinned and cropped to match the CT pixel size and FOV. The CT images were shifted by half of a PET pixel in both the lateral and anterior-posterior

Table 3: Reconstruction parameters used in this study

Reconstruction type	Iterations/Subsets	Post/ Loop filter (mm FWHM)	Randoms Correction	Transaxial Filter
2D OSEM	2/30	6.00/4.29	Delayed event subtraction	N/A
3D OSEM	5/32	2.57/2.34	Correction by singles	N/A
2D FBP	N/A	N/A	Correction by singles	Hanning, 9.3 mm cutoff
3D FBP	N/A	N/A	Correction by singles	Hanning, 9.3 mm cutoff

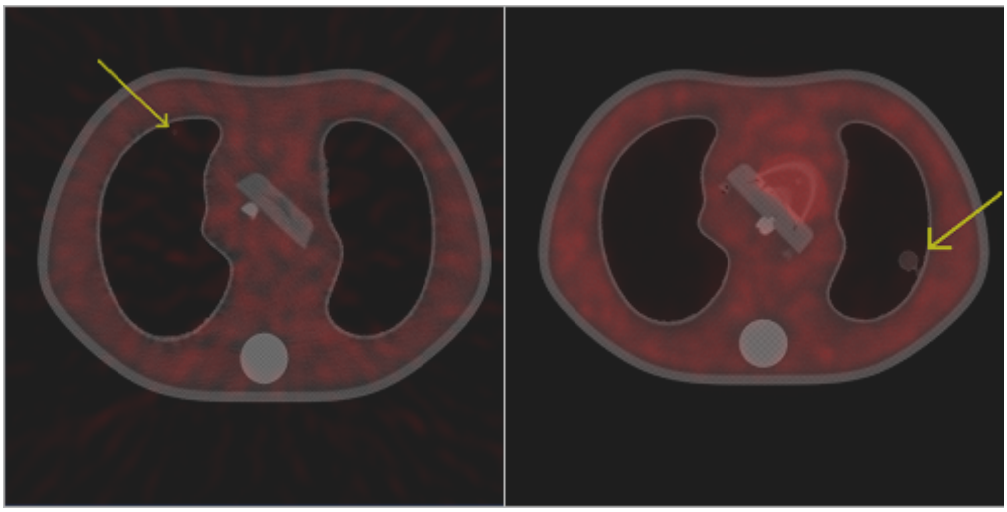

Figure 5: PET/CT images demonstrating lesion visibility vs. lesion size, TBR, acquisition mode, and reconstruction algorithm: (left) small lesion with high TBR acquired in 3D mode reconstructed with FBP; (right) large lesion with low TBR acquired in 2D mode reconstructed with OSEM. The lesions are marked by yellow arrows.

directions to align the images, because the images, as they are stored in the DICOM files, are not aligned.

The hollow spheres have stems attached to them, as seen in Figure 4. The appearance of these stems in the CT data could give away the position of the smallest lesions when they might otherwise be unseen, as they blended in well with the surrounding simulated lung tissue. To remove the stems from the CT images, the pixel values in the region of the stem were replaced by simulated phantom lung pixels. To do this, an ROI was drawn around the stem. All points in the ROI were replaced by a number randomly picked from a Gaussian distribution that was

similar to the histogram of the CT values in the phantom lung. The modified ROI, along with the original CT values adjacent to the ROI, was then convolved with a Gaussian distribution with a 2 mm FWHM to blend the modified ROI into the original CT data.

The PET data was stored with values in units of activity concentration, Bq/mL. Although OSEM applies a non-negativity constraint during reconstruction, FBP does not. The nature of FBP can create negative values which are spurious, as there is no such thing as a negative activity concentration. Thus all negative values in the PET data were replaced with a value of zero.

To convert the PET images from Bq/mL to TBR, several steps were required. First, the average activity concentration of the PET data set, A_t, was found by summing the total activity in the PET data set and dividing that value by the total fillable internal volume of the torso phantom (11,050 mL). Only image slices containing the torso phantom were used to find this average. The entire data set was then normalized to this average value. To finally convert the pixel data to TBR, the images were scaled to produce an activity concentration of one inside the background portion of the torso phantom. The background portion of the phantom was very large compared to both the scanner resolution and the PET image voxel size and thus would not be greatly affected by PVE, so the scaling could be based on the background portion of the phantom. The scaling factor used to accomplish this was derived assuming that the ratio of the measured activity concentration in the background portion of the phantom, A_b, to A_t in the PET image would be equal to the ratio of the actual background $A_{b,\,true}$ to total $A_{t,\,true}$ activities injected into the phantom. That is,

$$\frac{A_b}{A_t} = \frac{A_{b,true}}{A_{t,true}}. \tag{5}$$

Thus, for each compartment of the phantom, the scaling factor was

(6) $$A_i * \frac{A_{t,true}}{A_t * A_{b,true}} = TBR_i,$$

where i represents the compartment, e.g., liver or myocardium.

4.2.6 Image Display Software

The image display software (Figure 6) that was used to present images to observers was written in house in IDL[2] and had a number of features. The PET and CT data and ROIs could be

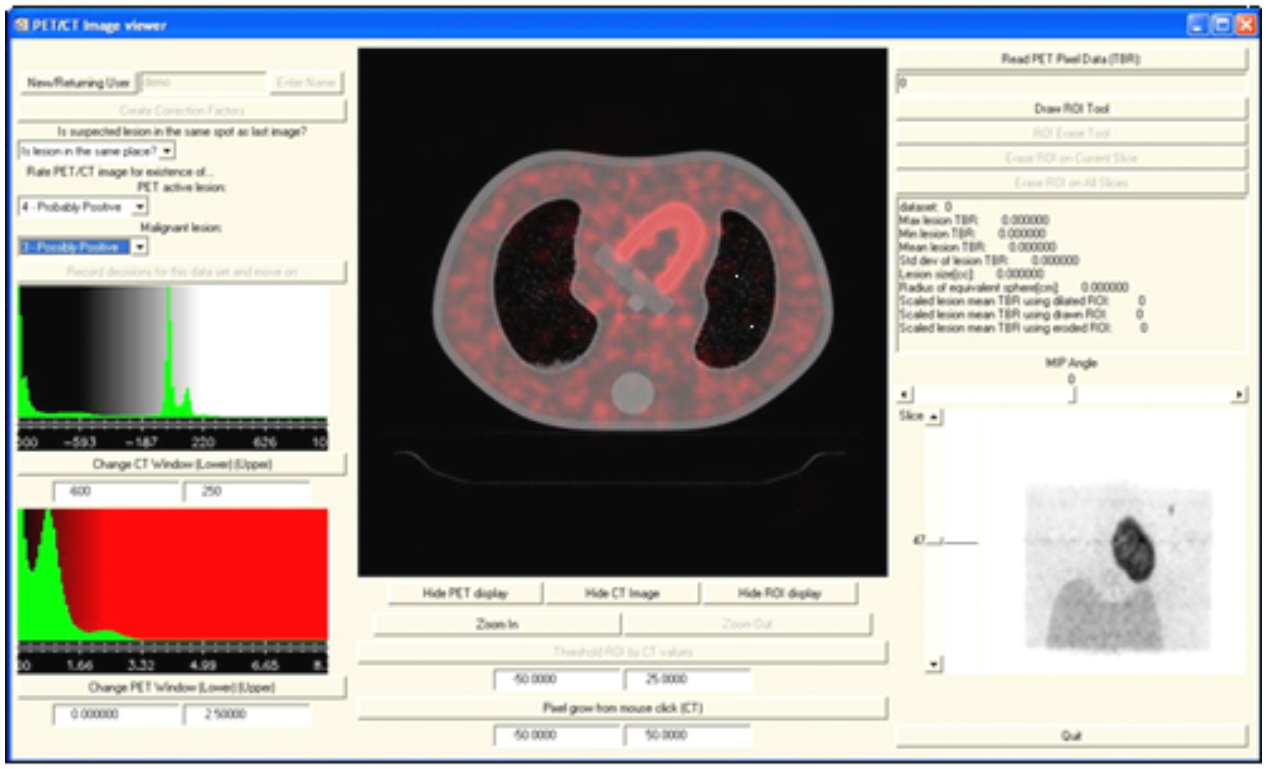

Figure 6: Screen shot of the image presentation software. The user can view and zoom PET, CT and fused PET/CT images, can alter the color table and display windows, and enter rating responses.

displayed individually or fused for simultaneous viewing. Only transverse slices were displayed. Observers were given the option to draw a single ROI per slice. A number of ROI manipulation options were available – the user was allowed to draw an ROI using the mouse, erase an ROI using the mouse, erase the ROI completely or per slice, set the ROI based on CT threshold values, or grow the ROI from a selected point based on CT values. Observers were allowed to

[2] ITT Corporation, 4990 Pearl East Circle, Boulder, CO 80301

change the window and level of both the PET and CT data, to zoom in and out, to move between slices in the data set, to view PET TBRs either point-by-point or within an ROI, and to review PET statistics within an ROI.

4.2.7 Final Image Data Sets

There were 104 image sets total, of which 78 images had PET active lesions (lesions that contained any amount of activity) while 26 images had no lesion or a lesion with no activity. Thirty-eight images had malignant lesions (lesions that had an activity concentration at least 2.5 times that of the background) while 66 had no lesion or a benign lesion. Of the 78 PET active lesions, 40 were not malignant.

4.3 AIM 3: CONDUCT THE OBSERVER STUDY

The third task of this project was to conduct the observer study with appropriate experienced observers (radiation oncologists, nuclear medicine physicians and medical physicists) using the software described in Section 4.2.6 to sequentially display the processed PET/CT data, to record image ratings (for the existence of a PET active lesion and for malignant lesion classification), and both to obtain the ROIs needed for the SUV correction scheme and to present the corrected data to the observer. The data acquired from the observer study were used to evaluate the hypotheses of this work.

4.3.1 Lesion Detection and Malignant Lesion Classification Studies

Observer performance was assessed for two tasks. The lesion detection study (LDS) gauged observers' abilities to identify the presence of a PET active lesion in a PET image or in a PET/CT image. The malignant lesion classification study (MLCS) judged observers' abilities to classify a lesion as benign or malignant based on TBR values.

4.3.2 Image Rating

For each image set, the PET image alone was first displayed to the observer. The observer rated the PET image for the existence of a PET active lesion; this rating was part of the LDS. The rating was based on a five point scale, one meaning "a PET active lesion definitely does not exist" and five meaning "a PET active lesion definitely exists". Image sets with no lesions or lesions with no PET activity were considered to be negative as PET active lesions. The PET image was also rated for the existence of a malignant lesion for MLCS. The MLCS rating was also based on a five point scale, one meaning "the lesion is definitely not malignant" and five meaning "the lesion is definitely malignant". In the case of non-small cell lung cancer, lesions with SUV greater than 2.5 may be considered malignant (22).

After rating the PET image alone, the correlated CT data was displayed with the PET data. The PET/CT data was rated for both the LDS and MLCS using the same scale as for the PET image alone. Observers were instructed to use the CT data only for lesion localization; ratings for PET active lesions and malignancy were based on the PET data.

The observer was then asked to draw ROIs on all slices of the PET/CT data for which the lesion was visible. These ROIs were used as input to the correction scheme. When drawing the ROI, the observer was instructed to exclude the acrylic shell of the simulated lesion; the shell did not include radioactivity, so it should not be included in the SUV correction scheme's calculations (real lesions would not have a shell).

For each case, the correction scheme calculation was repeated for dilated and eroded versions of the observer's ROIs. The ROIs were dilated or eroded on a slice by slice basis creating ROIs that were larger or smaller, respectively, by 1 mm in the radial direction. This showed the observer how a change in ROI size altered the corrected TBRs. A 1 mm change in boundary was selected because the CT pixel size was approximately 1 mm^2. These additional correction values

were displayed to the observers along with the value based on the original ROI. The observer rated the data for malignancy using the corrected values; the LDS was not used here because the SUV correction scheme does not alter the detection of lesions.

The process of rating the PET data, the PET/CT data, and then the corrected PET/CT data was repeated for all 104 image sets in this study. The order of presentation of the PET/CT image sets was randomized for each observer.

To populate the points for each ROC curve, the ratings were grouped at four different thresholds. For the threshold with the highest FPF and TPF, all ratings between two and five were considered positive ratings while the lowest rating, one, was considered a negative rating. TPF and FPF were calculated using these considerations. For the next threshold with a slightly lower FPF and TPF, all ratings between three and five were considered positive ratings while the two lowest ratings, one and two, were considered negative ratings. This process was repeated similarly for the next two decision thresholds.

4.3.3 Observer Performance Assessment

The observers' response data were categorized by acquisition mode (2D or 3D), reconstruction algorithm (FPB or OSEM), observer type (nuclear medicine physicians, radiation oncologists, medical physicists, numerical), and study type (LDS or MLCS). The response data were then entered into ROCKIT 0.9B[3]. ROCKIT was used to obtain a maximum-likelihood estimate of the parameters, $\boldsymbol{a}$ and $\boldsymbol{b}$, for the conventional binormal model of the ROC curve and for the area under the curve A_z, as well as the standard deviation for all of these values. A_z is one metric that can be used to compare different ROC curves.

In addition to A_z, the performance markers of sensitivity, specificity, and accuracy were used to judge the observer data. Sensitivity is the number of true positives (TP) divided by the sum of

[3] University of Chicago (webpage url: http://www-adiology.uchicago.edu/krl/KRL_ROC/software_index6.htm)

the TP and false negatives (FN). Specificity is the number of true negatives (TN) divided by the sum of the TN and the false positives (FP). Accuracy is the sum of the TP and TN divided by the total number of all cases, or the sum of the TP, TN, FP, and FN. To calculate these values for each observer, the two most positive responses and the three most negative responses from the ROC study were lumped together. When referring to statistical significances between pairs of values, "no significance" means that the values lie within each others' 1-σ standard deviation, "marginal significance" means that the values overlap at the level of two standard deviations, and "significance" means that the values are further than two standard deviations apart. Equations (8) through (14) give the equations that determine the standard deviations (σ) for the sensitivity, specificity, and accuracy (23).

(8) $$\sigma_{TP} = \sqrt{\frac{TP*FN}{TP+FN}}$$

(9) $$\sigma_{FN} = \sigma_{TP}$$

(10) $$\sigma_{TN} = \sqrt{\frac{FP*TN}{FP+TN}}$$

(11) $$\sigma_{FP} = \sigma_{TN}$$

(12) $$\sigma_{sensitivity} = sensitivity * \sqrt{(\frac{\sigma_{TP}}{TP})^2 + (\frac{\sqrt{\sigma_{TP}^2+\sigma_{FN}^2}}{TP+FN})^2}$$

(13) $$\sigma_{specificity} = specificity * \sqrt{(\frac{\sigma_{TN}}{TN})^2 + (\frac{\sqrt{\sigma_{FP}^2+\sigma_{TN}^2}}{FP+TN})^2}$$

(14) $$\sigma_{accuracy} = \sqrt{(\frac{TP+FN}{TP+FN+TN+FP}\sigma_{sensitivity})^2 + (\frac{TN+FP}{TP+FN+TN+FP}\sigma_{specificity})^2}$$

4.3.4 Assessing the Quantitative Accuracy of the Correction Scheme

Once all observers were finished with the study, a numerical observer was used to further evaluate the efficacy of the SUV correction scheme, as part of the first hypothesis of this study.

The numerical observer was used to quantitatively evaluate the impact of the SUV correction scheme. The numerical observer used the ROIs drawn by the experienced observers. To evaluate each data set, the numerical observer computed an average TBR from all human observer drawn ROIs. This was repeated for the eroded and dilated versions of the human observer drawn ROIs to evaluate the effect of a change in lesion boundary definition. ROIs with a volume greater than 2.5 mL were not included in the average. The largest spherical lesion used in this study was 2.0 mL in size so any ROI larger than 2.5 mL was created by an observer that was not careful in the use of the pixel growing method used to create the ROI. Fortunately, this was a rare occurrence and only occurred with one set of data with a few of the observers where the largest lesion was placed close to the bottom corner of the simulated lung. For these cases, the ROI grew outside of the lesion and into the background compartment of the phantom.

The sensitivity, specificity, and accuracy of the numerical observer were calculated in the same manner as for the human observers. The numerical observer used a fixed TBR threshold of 2.5 to classify images as malignant. For an average TBR of 2.5 or larger, the image was classified as malignant; for an average TBR less than 2.5, the image was classified as not malignant.

CHAPTER 5:
RESULTS

The layout of this chapter follows the order of the research problems enumerated in Chapter 2. The results of the scheme characterization are in Section 5.1, the impacts of the correction scheme are given in Section 5.2, and the effects of acquisition mode and reconstruction algorithm selection on lesion detection and classification in PET images, PET/CT images, and PET/CT images with PVE correction are given in Section 5.3.

5.1 CHARACTERIZATION OF THE CORRECTION SCHEME

The SUV correction scheme was tested with simulation data. This simulation data evaluated factors such as tumor size, positioning of the tumor with respect to the PET voxel matrix, mismatch of the PSF of the scanner to the PSF used by the correction scheme, mispositioning of ROI boundaries, and PET voxel size. Figure 7 illustrates TBR recovery with and without PVE

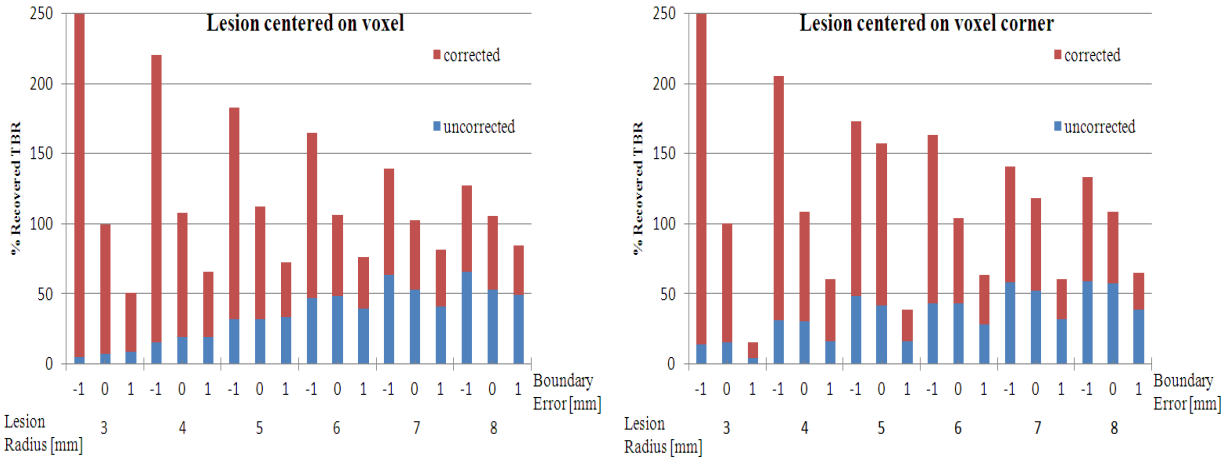

Figure 7: Percent recovered mean TBR as a function of lesion radius using simulated data when the correction PSF matched the simulated scanner PSF, for lesions centered over the middle of a PET voxel (left) vs. centered over a PET voxel corner (right).

correction as a function of lesion size, ROI size mismatch, and lesion location relative to the voxel grid. The lesion is centered over a voxel in Figure 7 (left) and over a voxel corner in Figure 7 (right). The results shown in Figure 7 used a PET voxel size of 4.69 mm x 4.69 mm x

3.27 mm and a CT voxel size of 0.98 mm x 0.98 mm x 3.27 mm. The correction scheme's PSF was a Gaussian function with an axial width of 6.5 mm FWHM and a longitudinal width of 6.0 mm FWHM. Figure 7 plots the results for the correct ROI size and for a ±1 mm mismatch in ROI size. The ±1 mm mismatch shows the bounds on the corrected TBR given such an error in ROI size.

Error in ROI size had a noticeable effect on the corrected TBRs for small lesions, especially if the ROI was too small compared to the lesion. Corrected TBRs for the larger lesions were less affected by ROI mismatch. The simulation results in Figure 7 show that the simulated PSF used for the correction factor simulation produced 99-112% TBR recovery when the lesion was centered over a voxel and 101-157% TBR recovery when the lesion was centered on a voxel corner.

In Fig. 7, the correction scheme PSF exactly matched the simulated scanner PSF. Figure 8 shows the effect of mismatch between the correction scheme PSF and the assumed scanner PSF as a function of ROI size. The simulation for Figure 8 used a PET voxel size of 4.69 mm x 4.69 mm x 3.27 mm and a CT voxel size of 0.98 mm x 0.98 mm x 3.27 mm. The simulated lesions were located over the center of a PET voxel. For the Discovery ST scanner used to acquire the PET/CT images of the torso phantom, the actual PSF varied not only with position in the scanner, but also with the acquisition mode used to create each image (Table 2). The PSFs for the acquisition modes were similar except in the axial direction at the center of the field of view (FOV), where 3D mode was poorer than 2D mode. The resolution worsened for both modes when moving away from the center of the FOV. Throughout this work, a PSF of 6.5 mm FWHM in the axial direction and 6.0 mm FWHM in the longitudinal direction was used in the correction scheme at all times.

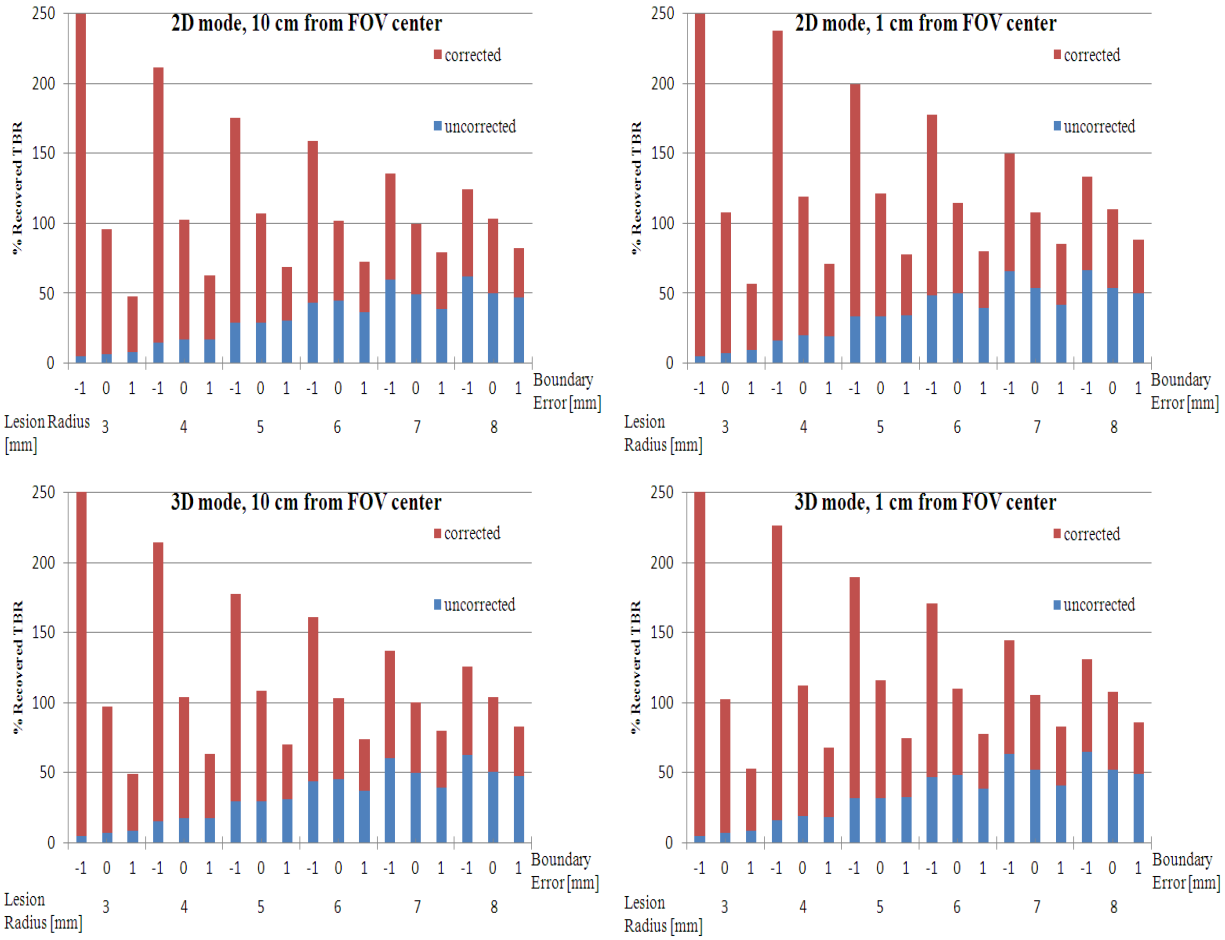

Figure 8: Percent recovered mean TBR as a function of lesion radius using simulated data when the correction PSF did not match the scanner's local PSF, for lesions located over the center of a PET voxel. Scanner PSF varied with position in the FOV, as summarized in Table 2.

The simulation results in Fig. 8 show that the mismatched PSF of the correction scheme produced 96-107% TBR recovery for 2D mode and 96-109% TBR recovery for 3D mode when the lesion was 10 cm from the center of the FOV and the ROI matched the actual lesion boundary. For lesions 1 cm from the center of the FOV, the correction scheme over-reported TBR by 8-21% for 2D mode and 2-16% for 3D mode. Thus, PSF mismatch has an effect, although it is relatively minor when the mismatch is small. One should note that the lesions were typically on the order of 10 cm from the FOV center for the PET/CT phantom images acquired

for the observer study. Figure 8 shows similar trends for TBR recovery versus ROI size mismatch as those observed with the results shown in Figure 7.

5.2 IMPACT OF CORRECTION SCHEME ON OBSERVER PERFORMANCE

The impact of PVE correction was measured for human observers and a numerical observer. Here, ratings data for all acquisition mode and reconstruction algorithm combinations were pooled. Note that the physicians were clinically experienced in reading PET/CT images while the medical physicists were not. The numerical observer categorized each PET/CT image, with or without PVE correction, by average TBR only without the possibility of subjective judgment calls. The implementation of the numerical observer was described in section 4.3.3.

5.2.1 Effect of the SUV Correction Scheme on Human Observers

Table 4 presents the A_z results for the MLCS both without and with the SUV correction scheme. Performance, as measured by A_z values, did not improve or worsen with the addition of a corrected TBR. With the corrected TBR, A_z values for the medical physicists were statistically no different than the A_z values for the physicians. TPF vs. FPF data and fitted ROC curves from the MLCS using corrected and uncorrected PET/CT data are displayed in Figure 9. Qualitatively, for all observers, the ROC curves for the MLCS without PVE correction and for the MLCS with correction appear to overlap (Figure 9). Once given a corrected TBR, observers gave more positive ratings, both correctly and incorrectly, which shifted data points up and to the right. Sensitivity was increased at the most difficult decision thresholds for most ROC curves and data. For example, the medical physicists had 98.8% sensitivity on the hardest decision

Table 4: A_z using PET/CT data

MLCS	MLCS, corrected TBR	MLCS	MLCS, corrected TBR
Nuclear medicine physicians		All physicians	
0.80 ± 0.04	0.84 ± 0.03	0.83 ± 0.03	0.83 ± 0.02
Radiation oncologists		Medical physicists	
0.87 ± 0.03	0.82 ± 0.03	0.90 ± 0.05	0.86 ± 0.03

threshold using the corrected PET data, with a corresponding specificity of 45.4%. Values for the other observer groups were similar.

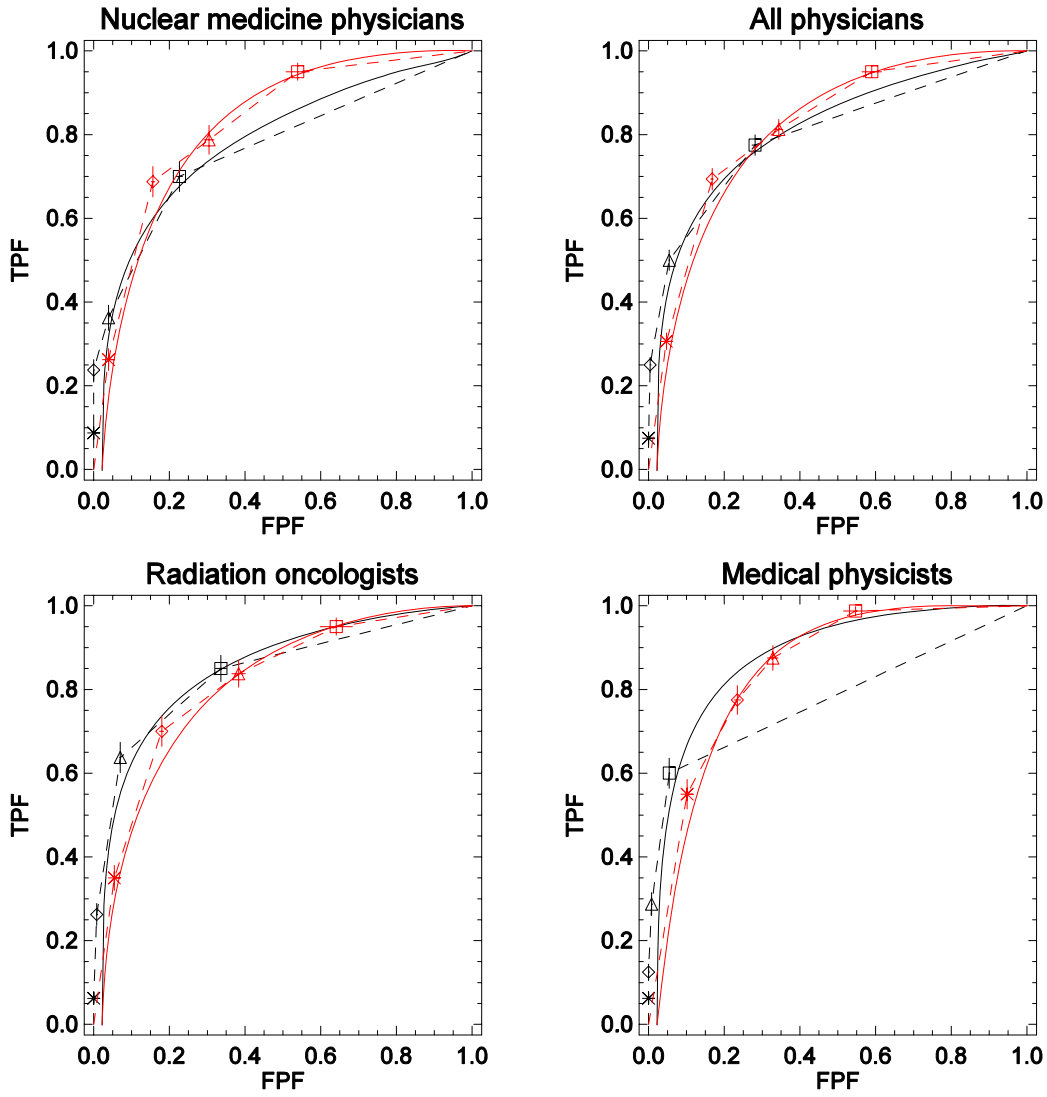

Figure 9: Data for the MLCS using uncorrected (black symbols) and corrected (red symbols) PET/CT data. Lines connecting symbols (uncorrected data: dashed black line; corrected data: dashed red line) are for visual aid only as the data are not continuous. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square symbols, respectively. Solid black lines and red lines are fitted ROC curves representing studies without and with PVE correction, respectively.

Table 5 summarizes the sensitivity, specificity and accuracy results for the MLCS both without and with the SUV correction scheme. For all observers, the corrected TBRs significantly (the values are further than two standard deviations apart) increased sensitivity by 44.4% for

physicians and 65% for medical physicists. Corrected TBRs produced a marginal increase in accuracy for physicians and significantly increased accuracy for medical physicists. Corrected TBRs caused a significant decrease in specificity, between 15 and 17 percent for all groups of observers. The physicians' pooled MLCS data exhibited marginally higher sensitivity and accuracy with uncorrected TBRs than the medical physicists. The medical physicists appeared to score better on sensitivity than the physicians when using the corrected TBRs, but without statistical significance.

Table 5: Percent sensitivity, specificity, and accuracy using PET/CT data

MLCS			MLCS, corrected TBR		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Nuclear medicine physicians					
23.8 ± 5.0	100 ± 0.0	70.7 ± 1.9	68.8 ± 7.2	84.4 ± 0.9	78.4 ± 2.8
Radiation oncologists					
26.2 ± 5.2	99.3 ± 0.0	71.2 ± 2.0	70.0 ± 7.2	82.1 ± 1.1	77.4 ± 2.9
All physicians					
25.0 ± 3.6	96.1 ± 0.0	70.9 ± 1.4	69.4 ± 5.1	83.3 ± 0.7	77.9 ± 2.0
Medical physicists					
12.5 ± 3.8	100 ± 0.0	66.3 ± 1.4	77.5 ± 6.9	76.6 ± 1.7	76.9 ± 2.9

5.2.2 Classification Using a Numerical Observer

Using uncorrected TBRs, the numerical observer was unable to correctly identify any actually positive lesions (Table 6) because the uncorrected TBRs were never above 2.5. Figure 10 illustrates this issue by plotting uncorrected and corrected TBRs versus lesion volumes and actual TBRs. Figure 10 also illustrates how errors in ROI size result in errors in corrected SUV. When presented with a corrected TBR, the numerical observer was more successful at identifying actually positive lesions than when presented with an uncorrected TBR, as illustrated

Table 6: Percent sensitivity, specificity, and accuracy for the numerical observer using a threshold of 2.5 TBR, using the observer drawn ROIs

MLCS			MLCS, corrected TBR		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
0.0*	100 ± 0.0	60.0*	53.1 ± 7.8	82.3 ± 6.0	70.6 ± 4.8

*Lack of true positives caused division by zero in calculation of variation

by the sensitivity results in Table 6. When using corrected TBRs, specificity significantly decreased by 17.7%, and accuracy increased by 10.6%; the change in accuracy is probably not significant, if one assumes a comparable level of error as for the calculated accuracy with uncorrected TBRs. (Note that this significance cannot be assessed directly because the lack of any true positives by the numerical observer prevents the calculation of variance for accuracy with uncorrected TBRs.)

The numerical observer, when using the dilated version of the human observers' ROIs, had a similar accuracy, sensitivity, and specificity it did when using the ROIs as drawn, as seen in Table 7. When using the eroded ROIs, the numerical observer scored the same on accuracy and sensitivity as when using the drawn ROIs, but had a decreased specificity.

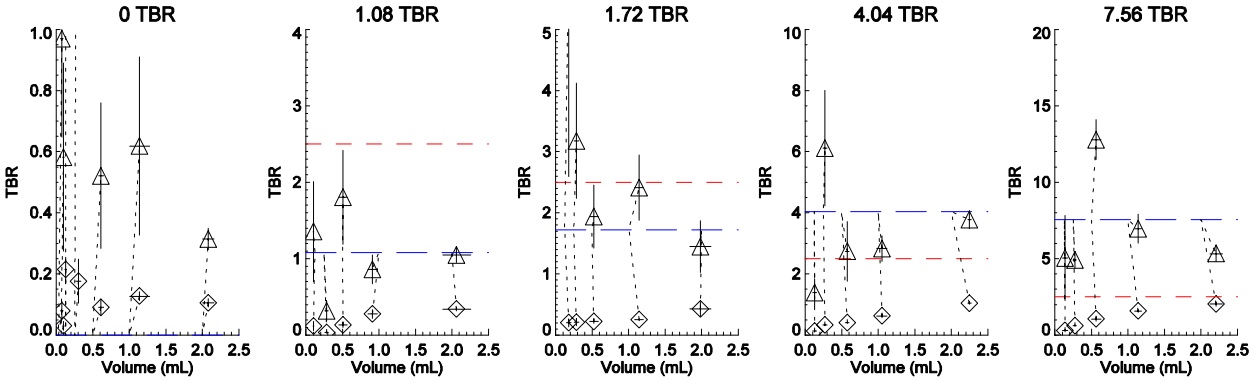

Figure 10: Each graph plots the average volumes (mL) of the ROIs drawn by the observers versus the average TBRs in these ROIs. Diamonds represent uncorrected TBRs and triangles represent corrected TBRs. Error bars for these points represent the standard deviation of the average TBR (vertical bars) and ROI volume (horizontal bars). Dotted lines connect the plotting symbols to the corresponding true lesion volumes and TBRs (blue long-dashed lines). The red short-dashed horizontal line represents a TBR of 2.5.

The percent recovered TBR for uncorrected and corrected TBRs were tabulated for all actual sizes and actual TBRs, for the observer drawn ROIs (Table 8) and the dilated and eroded ROIs (Table 9). The percent recovered TBR was closer to 100% for corrected TBRs than uncorrected TBRs for 74 out of 76 of the drawn ROIs, 75 out of 76 of the dilated ROIs, and 68 out of 76 of the eroded ROIs. On average, the uncorrected percent recovered TBR was only 15.1% while the

average corrected TBR was 100%, 125%, and 210% for the dilated, observer drawn, and eroded ROIs, respectively. This implies that the observers generally drew their ROIs slightly too small.

Table 7: Percent sensitivity, specificity, and accuracy for the numerical observer using a threshold of 2.5 TBR and using dilated and eroded ROIs

Dilated ROIs			Eroded ROIs		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
53.1 ± 7.8	81.3 ± 6.1	70.0 ± 4.8	59.4 ± 8.0	68.8 ± 6.6	65.0 ± 5.1

Table 8: Average recovered uncorrected and corrected TBRs, as a percentage of the actual TBR, for observer-drawn ROIs for all actual lesion sizes and TBRs.

		0.125 mL	0.25 mL	0.5 mL	1.0 mL	2.0 mL	Mean
1.08 TBR	Uncorrected	15.3	3.6	12.6	23.7	32.1	17.5
	Corrected	183	33.9	165	158	109	130
1.72 TBR	Uncorrected	7.5	13.9	12.4	14.1	25.3	14.7
	Corrected	205	281	111	146	107	170
4.04 TBR	Uncorrected	3.4	9.8	9.8	16.2	29.0	13.6
	Corrected	51.2	199	59.4	81.5	113	101
7.56 TBR	Uncorrected	4.9	8.0	13.1	21.1	25.3	14.5
	Corrected	137	62.9	142	95.8	67.7	101
Mean	Uncorrected	7.8	8.8	12.0	18.8	27.9	15.1
	Corrected	144	144	119	120	99.1	125

Table 9: Average recovered corrected TBRs, expressed as a percentage of the actual TBR, for dilated and eroded ROIs.

		0.125 mL	0.25 mL	0.5 mL	1.0 mL	2.0 mL	Mean
1.08 TBR	Dilated	146	28.4	154	142	101	114
	Eroded	283	42.5	233	200	119	176
1.72 TBR	Dilated	123	202	98.8	138	94.9	131
	Eroded	459	486	143	184	119	278
4.04 TBR	Dilated	32.2	133	51.9	70.4	100	77.5
	Eroded	112	404	80.7	103	122	164
7.56 TBR	Dilated	71.0	48.5	121	84.1	63.6	77.7
	Eroded	651	82.0	167	128	71.9	220
Mean	Dilated	93.1	103	106	109	89.9	100
	Eroded	376	254	156	154	108	210

The correction scheme produced overestimates of TBR in some situations, but underestimates of TBR in others (Table 10). Overall, underestimation of the corrected TBR was more likely to occur than overestimation. Under- or overestimation of corrected TBR had no

obvious trends with either lesion size or actual TBR. Underestimation by the corrected TBR occurred more often than overestimation when the observer drawn ROI was too large compared to the actual lesion size, which agreed with the results from the characterization simulations.

Table 10: Correlation between ROI size and correction accuracy versus actual lesion size and TBR. Percent occurrence is given per each actual lesion size or actual TBR.

Lesion size [mL]	Drawn ROI too small		Drawn ROI too large	
	Underestimated TBR	Overestimated TBR	Underestimated TBR	Overestimated TBR
0.125	34.2 %	13.7 %	34.2 %	17.8 %
0.25	19.1 %	13.8 %	40.4 %	26.6 %
0.5	13.0 %	17.4 %	35.9 %	33.7 %
1.0	15.1 %	15.1 %	35.5 %	34.4 %
2.0	19.8 %	9.9 %	51.6 %	18.7 %
Actual TBR	Underestimated TBR	Overestimated TBR	Underestimated TBR	Overestimated TBR
1.08	24.8 %	24.8 %	32.4 %	29.5 %
1.72	14.4 %	14.4 %	23.4 %	41.4 %
4.04	17.5 %	17.5 %	52.6 %	15.8 %
7.56	22.1 %	22.1 %	49.6 %	20.4 %

However, when ROI size was smaller than the lesion size, the simulation predicted that overestimation was more likely to occur. This finding with the experimental results may be due to the correction scheme PSF being less broad than the PET scanner PSF. The narrow correction scheme over-corrects for the effect of the too-small ROI.

5.3 IMPACT OF ACQUISITION MODE AND RECONSTRUCTION ALGORITHM ON OBSERVER PERFORMANCE

The impact of acquisition mode and reconstruction algorithm selection was measured for human observers and the numerical observer in terms of lesion detection and malignant lesion classification. PET data, PET/CT data, and PET/CT data with PVE correction were evaluated. As mentioned previously, the physicians were clinically experienced in reading PET/CT images while the medical physicists were not. The numerical observer assessed each PET/CT image, with or without PVE correction, by average TBR only, without the possibility of subjective judgment calls.

5.3.1 Effect of Acquisition Mode and Reconstruction Algorithm Selection Using PET Data Alone

The nuclear medicine physicians who took part in this study only use a 3D mode scanner and reconstruct PET images with OSEM in their daily clinical practice. Using their responses on the PET data, A_z for the 3D OSEM and 3D FBP algorithms were not significantly different for the LDS ($p = 0.52$) but were marginally different for the MLCS ($p = 0.07$). The radiation oncologists who took part in this study almost exclusively use the 2D mode on the Discovery ST scanner and reconstruct PET images with OSEM. Using their responses on the PET data, the differences in A_z between FBP and OSEM modes with 2D acquisition were not significantly different for either the LDS ($p = 0.85$) or MLCS ($p = 0.68$). Table 11 summarizes A_z results for these two groups of observers using PET data only.

Table 11: A_z for physicians using PET data only

	Nuclear medicine physicians		Radiation oncologists	
	LDS	MLCS	LDS	MLCS
2D FBP	0.67 ± 0.08	0.75 ± 0.08	0.60 ± 0.08	0.65 ± 0.09
2D OSEM	0.64 ± 0.09	0.77 ± 0.08	0.58 ± 0.09	0.60 ± 0.11
3D FBP	0.68 ± 0.08	0.61 ± 0.09	0.44 ± 0.08	0.70 ± 0.09
3D OSEM	0.74 ± 0.07	0.83 ± 0.07	Degenerate†	0.72 ± 0.10

†Marginal data

For the LDS, the nuclear medicine physicians using 3D OSEM data had a higher A_z than the radiation oncologists using 2D OSEM data, although not significantly higher ($p = 0.17$). For the MLCS, the nuclear medicine physicians' 3D OSEM ROC curve had a higher A_z than the radiation oncologists' 2D OSEM ROC curve with marginal significance ($p = 0.07$)

However, ROC data pooled for all physicians showed a significant difference between 3D FBP and 3D OSEM reconstructions for the LDS ($p = 0.03$), with 3D OSEM outperforming 3D FBP. For the MLCS, 3D OSEM and 3D FBP were not significantly different ($p = 0.18$). Table 12 summarizes these pooled A_z results for the physicians and also for the medical physicists using PET data only. Performance for 3D OSEM and 3D FBP were not significantly different (p

= 0.18) however. Much of the medical physicists' data failed to produce valid ROC curves and A_z values because the physicists rarely responded with ratings other than the two extreme values (definitely negative or definitely positive). TPF vs. FPF for several decision thresholds for the LDS using PET data only is displayed in Figure 11; the corresponding MLCS data is displayed in Figure 12. For all of the plots, notice the lack of points in the higher false positive range, especially for the MLCS, and the vertical grouping between different thresholds. Some of the observers noted that they were having trouble finding the lesions in the PET only data. Despite the lack of activity in the surrounding simulated lung tissue, distinguishing lesions from background was apparently difficult. The difficulty of the data set appears to skew the ROC results with many false negatives, few false positives, and consequently low sensitivity. Vertical grouping of the ROC data points prevented the ROCKIT software from generating fitted ROC curves in some cases (Table 12). Improving the data set would require adding more images

Table 12: A_z for all observers using PET data only

	All physicians		Medical physicists	
	LDS	MLCS	LDS	MLCS
2D FBP	0.63 ± 0.06	0.70 ± 0.06	ROCKIT failed	Degenerate[†]
2D OSEM	0.61 ± 0.06	0.68 ± 0.07	Degenerate[†]	0.88 ± 0.11
3D FBP	0.56 ± 0.06	0.66 ± 0.07	Degenerate[†]	Degenerate[†]
3D OSEM	0.75 ± 0.07	0.78 ± 0.06	0.65 ± 0.09	0.77 ± 0.15

[†]Marginal data

spanning a larger range of TBRs and tumor sizes; this increase in the data set size would result in a substantially larger time commitment on the part of the observers to complete the study. This was not deemed practical within the context of the current project.

Table 13 summarizes the results for sensitivity, specificity, and accuracy for the two groups of physicians for the MLCS and LDS using PET data. Again, to calculate these values for each observer, the two most positive responses and the three most negative responses from the observer study were lumped together. When referring to statistical significances between these

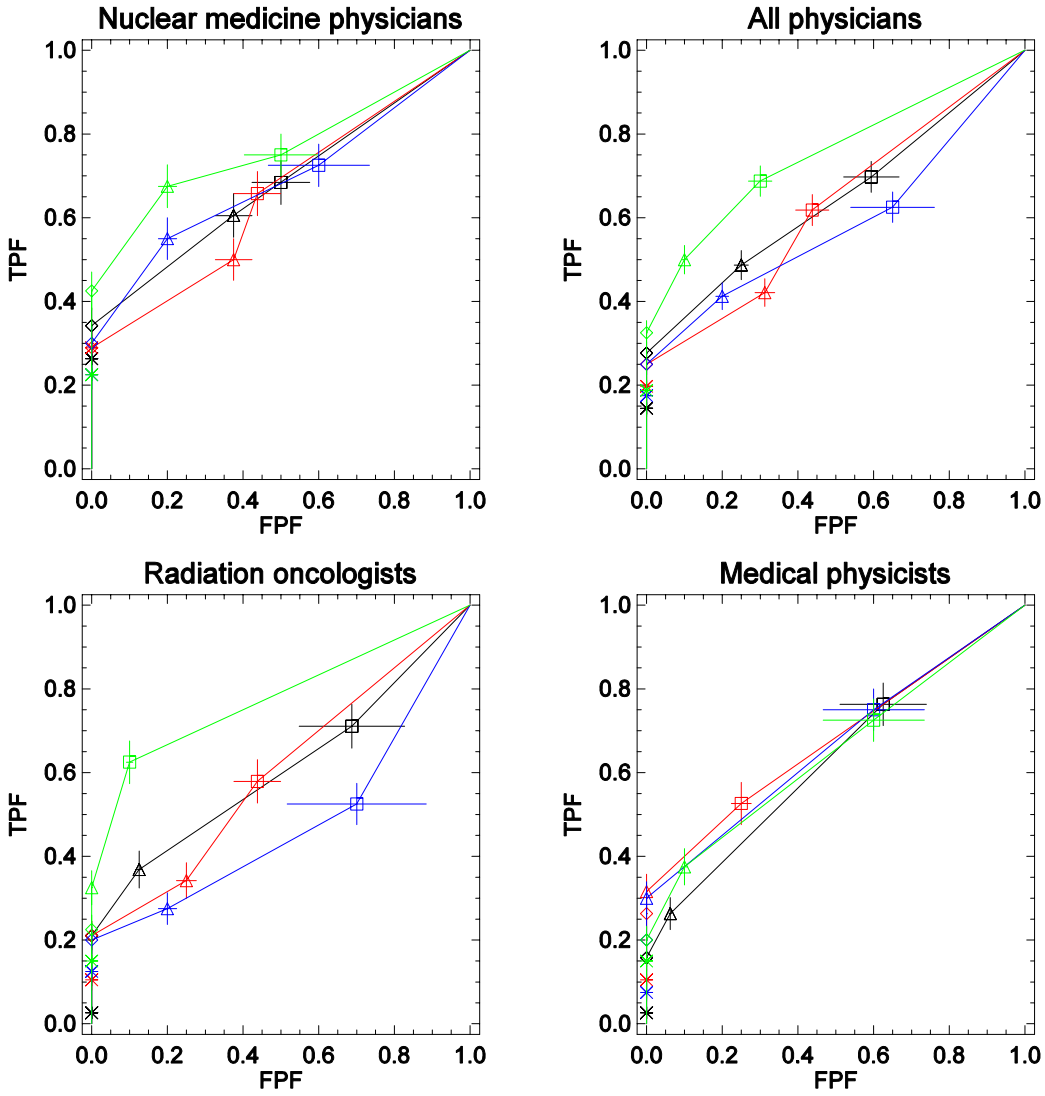

Figure 11: Data for the LDS using PET data only. Black lines represent 2D FBP, red lines represent 2D OSEM, blue lines represent 3D FBP, and green lines represent 3D OSEM. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively; the lines connecting the symbols are for visual aid only.

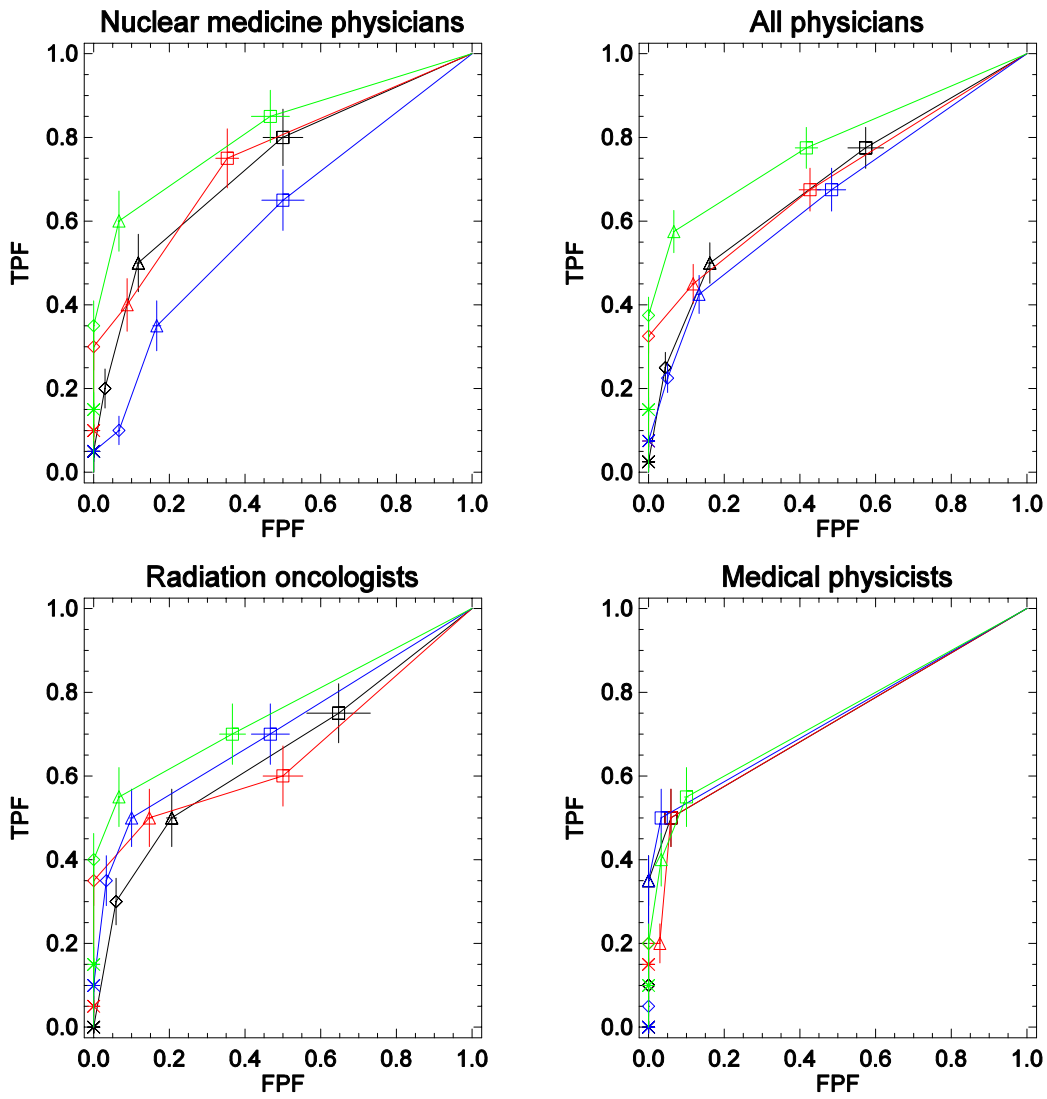

Figure 12: Data for the MLCS using PET data only. Black lines represent 2D FBP, red lines represent 2D OSEM, blue lines represent 3D FBP, and green lines represent 3D OSEM. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively; the lines connecting the symbols are for visual aid only.

Table 13: Percent sensitivity, specificity, and accuracy for nuclear medicine physicians and radiation oncologists using PET data only

Nuclear medicine physicians						
	LDS			MLCS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	34.2 ± 8.5	100 ± 0.0	53.7 ± 6.0	20.0 ± 9.3	97.1 ± 4.9	68.5 ± 4.6
2D OSEM	28.9 ± 7.9	100 ± 0.0	50.0 ± 5.6	30.0 ± 11.1	100 ± 0.0	74.1 ± 4.1
3D FBP	30.0 ± 7.9	100 ± 0.0	44.0 ± 6.3	10.0 ± 6.8	93.3 ± 7.5	60.0 ± 5.3
3D OSEM	42.5 ± 9.1	100 ± 0.0	54.0 ± 7.3	35.0 ± 12.0	100 ± 0.0	74.0 ± 4.8
Radiation oncologists						
	LDS			MLCS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	21.1 ± 6.9	100 ± 0.0	44.4 ± 4.9	30.0 ± 11.1	94.1 ± 6.7	70.4 ± 5.9
2D OSEM	21.1 ± 6.9	100 ± 0.0	44.4 ± 4.9	35.0 ± 12.0	100 ± 0.0	75.9 ± 4.4
3D FBP	20.0 ± 6.6	100 ± 0.0	36.0 ± 5.2	35.0 ± 12.0	96.7 ± 5.6	72.0 ± 5.9
3D OSEM	22.5 ± 6.9	100 ± 0.0	38.0 ± 5.5	40.0 ± 12.6	100 ± 0.0	76.0 ± 5.0

values, no significance means that the pair of values lie within each others' standard deviation, marginal significance means that the values overlap between one and two standard deviations, and significance means that the values are further than two standard deviations apart. The nuclear medicine physicians performed marginally better using 3D OSEM than 3D FBP for sensitivity and accuracy for the MLCS when using only PET data. The radiation oncologists fared slightly better using 2D OSEM than 2D FBP on the MLCS with respect to sensitivity, specificity, and accuracy using although not significantly. The nuclear medicine physicians had a marginally higher sensitivity for the LDS using 3D OSEM than the radiation oncologists had using 2D OSEM in the LDS. Table 14 summarizes the LDS and MLCS results when using PET data only for the physicians as a group and for the medical physicists, for which no mode fared significantly better than any other. A specificity of 100% with zero variation, which is seen in these tables is caused by the lack of false positives in this difficult image data set, as mentioned previously.

Table 14: Percent sensitivity, specificity, and accuracy for physicians and medical physicists using PET data only

All physicians						
	LDS			MLCS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	27.6 ± 5.5	100 ± 0.0	49.1 ± 3.9	25.0 ± 7.3	95.6 ± 4.2	69.4 ± 3.8
2D OSEM	25.0 ± 5.3	100 ± 0.0	47.2 ± 3.7	32.5 ± 8.2	100 ± 0.0	75.0 ± 3.0
3D FBP	25.0 ± 5.1	100 ± 0.0	40.0 ± 4.1	22.5 ± 6.9	95.0 ± 4.7	68.0 ± 4.0
3D OSEM	32.5 ± 5.8	100 ± 0.0	46.0 ± 4.6	37.5 ± 8.7	100 ± 0.0	75.0 ± 3.5
Medical physicists						
	LDS			MLCS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	15.8 ± 6.1	100 ± 0.0	40.7 ± 4.3	10.0 ± 6.8	100 ± 0.0	66.7 ± 2.5
2D OSEM	26.3 ± 7.6	100 ± 0.0	48.1 ± 5.4	20.0 ± 9.3	100 ± 0.0	70.4 ± 3.4
3D FBP	20.0 ± 6.6	100 ± 0.0	36.0 ± 5.3	5.0 ± 4.9	100 ± 0.0	62.0 ± 2.0
3D OSEM	20.0 ± 6.6	100 ± 0.0	36.0 ± 5.3	20.0 ± 9.3	100 ± 0.0	68.0 ± 3.7

5.3.2 Effect of Acquisition Mode and Reconstruction Algorithm Selection Using PET/CT Data

The A_z values for the LDS using PET/CT data (Table 15) were generally higher than the A_z values for the LDS using PET data alone (Tables 11 and 12) for the nuclear medicine physicians. However, much of the data for radiation oncologists and medical physicists failed to produce a valid ROC curve due to poor grouping of observer ratings; other than the nuclear medicine

Table 15: A_z using PET/CT data for the LDS

	Nuclear medicine physicians	Radiation oncologists	All physicians	Medical physicists
2D FBP	0.80 ± 0.07	ROCKIT failed	0.80 ± 0.05	Degenerate†
2D OSEM	0.71 ± 0.09	Degenerate†	0.78 ± 0.07	0.66 ± 0.08
3D FBP	0.65 ± 0.08	Degenerate†	0.70 ± 0.06	Degenerate†
3D OSEM	Degenerate†	Degenerate†	Degenerate†	Degenerate†

†Marginal data

physicians, observers tended to rate lesion detection as either definitely positive or definitely negative with few ratings in between. TPF vs. FPF for several decision thresholds for the LDS using PET/CT data is displayed in Figure 13. Again, lack of false positives caused vertical grouping of ROC data points. Because of this grouping, ROCKIT failed to produce fitted ROC curves for some cases. With the addition of CT data, no ROC data or fitted ROC curve appeared

to outperform any other, with the exception of the 2D FBP curve for the nuclear medicine physicians.

In terms of sensitivity, specificity, and accuracy, no modality outperformed any other in the LDS for any observer type when using PET/CT data. Table 16 summarizes sensitivity, specificity, and accuracy results for the LDS study using PET/CT data.

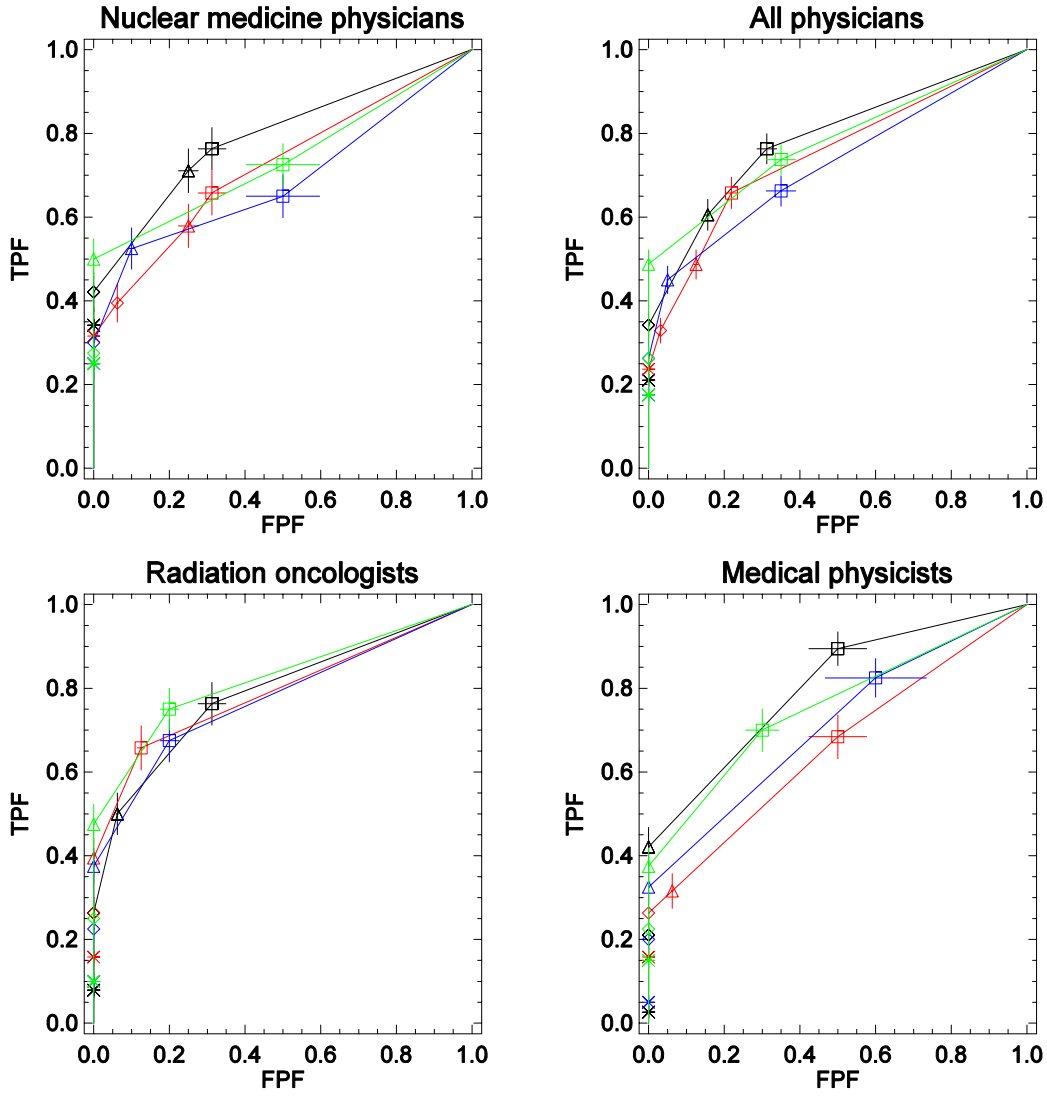

Figure 13: Results for the LDS using PET/CT data. Black lines represent 2D FBP, red lines represent 2D OSEM, blue lines represent 3D FBP, and green lines represent 3D OSEM. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively; the lines connecting the symbols are for visual aid only.

Table 16: Percent sensitivity, specificity, and accuracy for all observers using PET/CT data for the LDS

	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	Nuclear medicine physicians			All physicians		
2D FBP	42.1 ± 9.3	100 ± 0.0	59.3 ± 6.6	34.2 ± 6.0	100 ± 0.0	53.7 ± 4.3
2D OSEM	39.5 ± 9.1	93.8 ± 10.0	55.6 ± 7.1	32.9 ± 5.9	96.9 ± 5.2	51.9 ± 4.5
3D FBP	30.0 ± 7.9	100 ± 0.0	44.0 ± 6.3	26.3 ± 5.2	100 ± 0.0	41.0 ± 4.2
3D OSEM	27.5 ± 7.6	100 ± 0.0	42.0 ± 6.1	26.3 ± 5.2	100 ± 0.0	41.0 ± 4.2
	Radiation oncologists			Medical physicists		
2D FBP	26.3 ± 7.6	100 ± 0.0	48.1 ± 5.4	21.1 ± 6.9	100 ± 0.0	44.4 ± 4.9
2D OSEM	26.3 ± 7.6	100 ± 0.0	48.1 ± 5.4	26.3 ± 7.6	100 ± 0.0	48.1 ± 5.4
3D FBP	22.5 ± 6.9	100 ± 0.0	38.0 ± 5.5	20.0 ± 6.6	100 ± 0.0	36.0 ± 5.3
3D OSEM	25.0 ± 7.3	100 ± 0.0	40.0 ± 5.8	22.5 ± 6.9	100 ± 0.0	38.0 ± 5.5

A_z values for the MLCS were generally higher with the addition of correlated CT data. No A_z value was significantly higher than any other, as shown in Table 17. Again, some of the ROC data was degenerate due to observers never giving false positive decisions, as seen by the vertical grouping of the data in Figure 14. Table 18 shows that sensitivity, specificity, and accuracy were not significantly different between any acquisition mode or reconstruction method. Qualitatively, the ROC data for each combination of acquisition mode and reconstruction algorithm appear to overlay one another as shown in Figure 14.

Table 17: A_z values for the MLCS using PET/CT data

	Nuclear Medicine Physicians	Radiation Oncologists	All Physicians	Medical Physicists
2D FBP	0.78 ± 0.08	0.87 ± 0.06	0.82 ± 0.05	0.91 ± 0.06
2D OSEM	0.86 ± 0.05	0.76 ± 0.09	0.77 ± 0.06	Degenerate†
3D FBP	Degenerate†	0.91 ± 0.05	0.85 ± 0.06	Degenerate†
3D OSEM	Degenerate†	0.94 ± 0.04	0.88 ± 0.05	Degenerate*

*Perfect performance implied
†Marginal data

5.3.3 Effect of Acquisition Mode and Reconstruction Algorithm Selection Using PET/CT Data with PVE Correction

When the PVE correction scheme was used, no acquisition mode and reconstruction algorithm combination or observer type outperformed any other in terms of A_z (Table 19). Nor did any

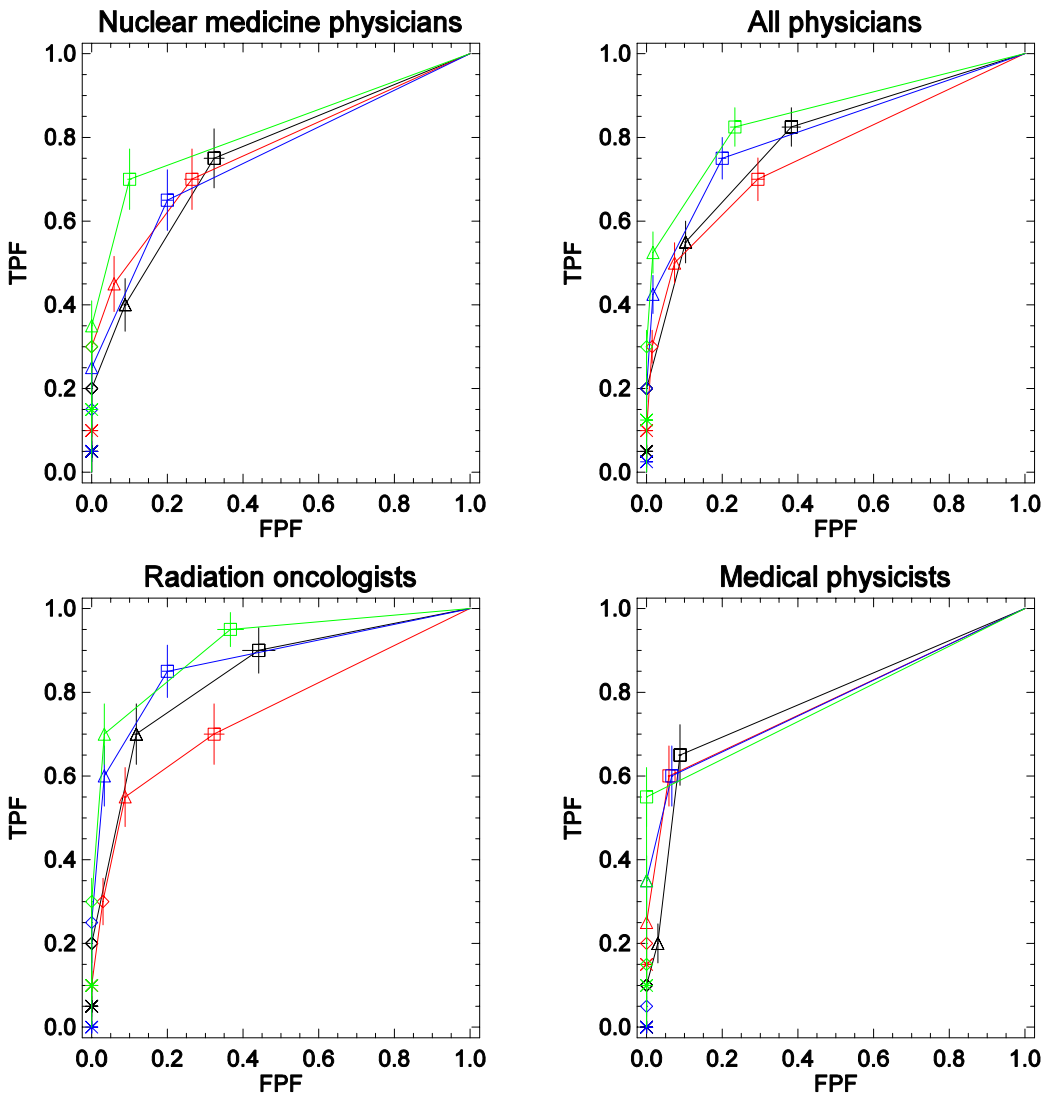

Figure 14: Data for the MLCS using PET/CT data. Black lines represent 2D FBP, red lines represent 2D OSEM, blue lines represent 3D FBP, and green lines represent 3D OSEM. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively; the lines connecting the symbols are for visual aid only.

Table 18: Percent sensitivity, specificity, and accuracy for all observers using PET/CT data for the MLCS

	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	Nuclear medicine physicians			All physicians		
2D FBP	20.0 ± 9.3	100 ± 0.0	70.4 ± 3.4	20.0 ± 6.6	100 ± 0.0	70.4 ± 2.4
2D OSEM	30.0 ± 11.1	100 ± 0.0	74.1 ± 4.1	30.0 ± 7.9	98.5 ± 2.5	73.1 ± 3.3
3D FBP	15.0 ± 8.2	100 ± 0.0	66.0 ± 3.3	20.0 ± 6.6	100 ± 0.0	68.0 ± 2.6
3D OSEM	30.0 ± 11.1	100 ± 0.0	72.0 ± 4.5	30.0 ± 7.9	100 ± 0.0	72.0 ± 3.1
	Radiation oncologists			Medical physicists		
2D FBP	20.0 ± 9.3	100 ± 0.0	70.4 ± 3.4	10.0 ± 6.8	100 ± 0.0	66.7 ± 2.5
2D OSEM	30.0 ± 11.1	97.1 ± 4.9	72.2 ±5.2	20.0 ± 9.3	100 ± 0.0	70.4 ± 3.4
3D FBP	25.0 ± 10.3	100 ± 0.0	70.0 ± 4.1	5.0 ± 4.9	100 ± 0.0	62.0 ± 2.0
3D OSEM	30.0 ± 11.1	100 ± 0.0	72.0 ± 4.5	15.0 ± 8.2	100 ± 0.0	66.0 ± 3.3

combination of acquisition mode and reconstruction algorithm or observer type outperform any other in terms of sensitivity, specificity, or accuracy with any significance (Table 20). Qualitatively, the ROC data for each combination of acquisition mode and reconstruction algorithm appear to overlay one another, as shown in Figure 15. Given PET/CT data with PVE correction, there was no inter-observer variation or effect of acquisition mode and reconstruction algorithm on the diagnostic performance.

Table 19: A_z values for the MLCS using PET/CT data with corrected PET data

	Nuclear medicine physicians	Radiation oncologists	All physicians	Medical physicists
2D FBP	0.86 ± 0.05	0.83 ± 0.06	0.84 ± 0.04	0.84 ± 0.06
2D OSEM	0.79 ± 0.07	0.86 ± 0.05	0.82 ± 0.04	0.86 ± 0.05
3D FBP	0.86 ± 0.06	0.80 ± 0.07	0.83 ± 0.04	0.86 ± 0.05
3D OSEM	0.83 ± 0.06	0.82 ± 0.06	0.82 ± 0.04	0.89 ± 0.05

5.3.4 Effect of Acquisition Mode and Reconstruction Algorithm Selection on the Numerical Observer

The numerical observer performed exactly the same for all combinations of acquisition mode and reconstruction algorithm when working without PVE correction (Table 21). Furthermore, when the numerical observer used PVE correction, its diagnostic performance, in terms of sensitivity, specificity, or accuracy did not vary between combinations of acquisition mode and reconstruction algorithm. When using dilated or eroded ROIs, diagnostic performance still was

not significantly different between combinations of acquisition mode and reconstruction algorithm (Table 22).

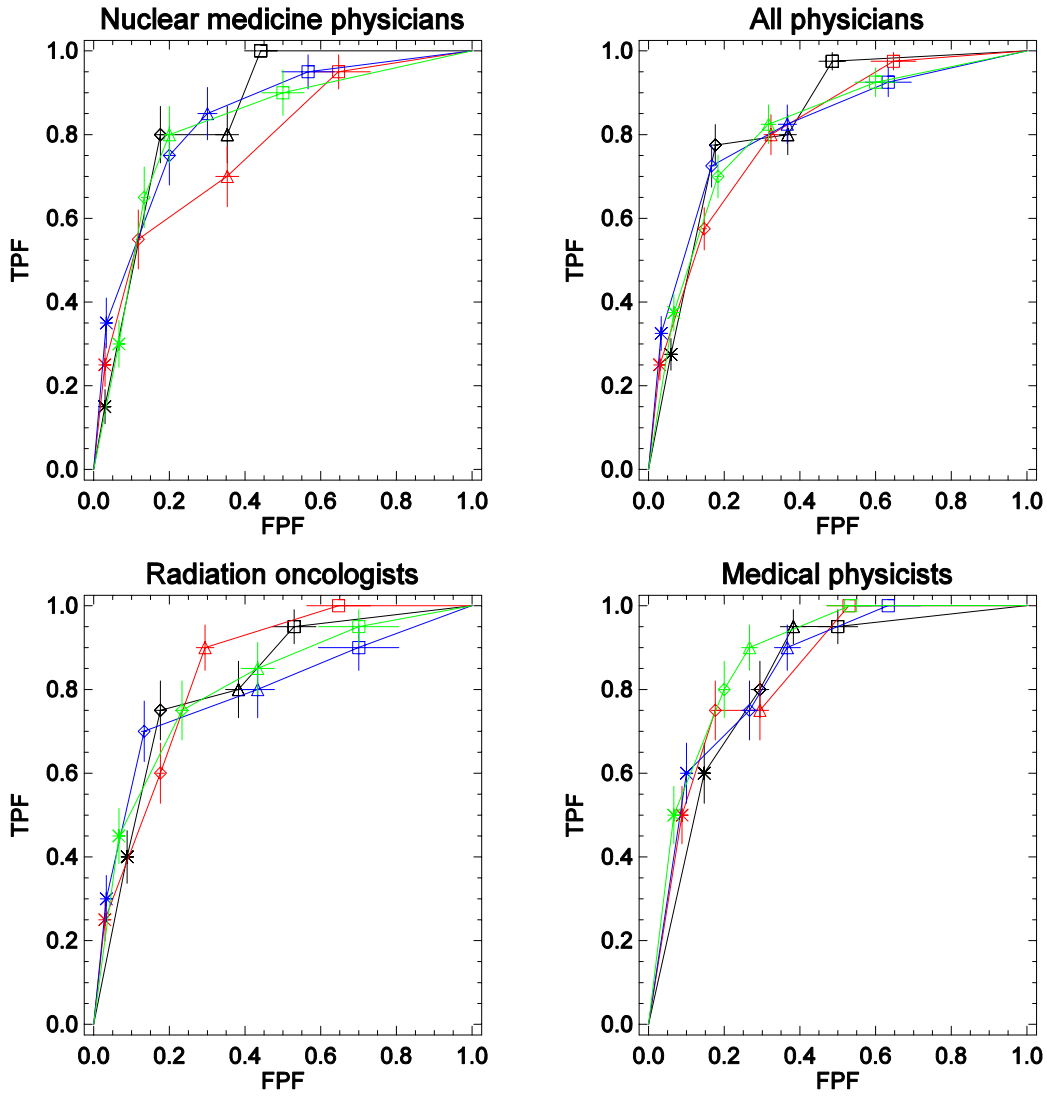

Figure 15: Data for the MLCS using PET/CT data with corrected PET data. Black lines represent 2D FBP, red lines represent 2D OSEM, blue lines represent 3D FBP, and green lines represent 3D OSEM. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively; the lines connecting the symbols are for visual aid only.

While diagnostic performance did not vary between combinations of acquisition mode and reconstruction algorithm, quantitative accuracy did vary. The only two cases where the uncorrected TBR was closer to the actual TBR than the corrected TBR using an observer drawn

Table 20: Percent sensitivity, specificity, and accuracy using PET/CT data with corrected PET data for the MLCS

	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	Nuclear medicine physicians			All physicians		
2D FBP	80.0 ± 13.5	82.4 ± 10.0	81.5 ± 8.1	77.5 ± 9.8	82.4 ± 7.1	80.6 ± 5.8
2D OSEM	55.0 ± 14.1	88.2 ± 8.8	75.9 ± 7.6	57.5 ± 10.1	85.3 ± 6.7	75.0 ± 5.6
3D FBP	75.0 ± 14.1	80.0 ± 11.0	78.0 ± 8.7	72.5 ± 10.1	83.3 ± 7.4	79.0 ± 6.0
3D OSEM	65.0 ± 14.5	86.7 ± 9.8	78.0 ± 8.3	70.0 ± 10.2	81.7 ± 7.6	77.0 ± 6.1
	Radiation oncologists			Medical physicists		
2D FBP	75.0 ± 14.1	82.4 ± 10.0	79.6 ± 8.2	80.0 ± 13.5	70.6 ± 11.0	74.1 ± 8.6
2D OSEM	60.0 ± 14.4	82.4 ± 10.0	74.1 ± 8.3	75.0 ± 14.1	82.4 ± 10.0	79.6 ± 8.2
3D FBP	70.0 ± 14.4	86.7 ± 9.8	80.0 ± 8.2	75.0 ± 14.1	73.3 ± 11.6	74.0 ± 9.0
3D OSEM	75.0 ± 14.1	76.7 ± 11.4	76.0 ± 8.9	80.0 ± 13.5	80.0 ± 11.0	80.0 ± 8.5

Table 21: Percent sensitivity, specificity, and accuracy for numerical observer using a threshold of 2.5 TBR

	MLCS			MLCS, corrected TBR		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	0.0*	100 ± 0.0	60.0*	50.0 ± 15.3	75.0 ± 12.9	65.0 ± 9.9
2D OSEM	0.0*	100 ± 0.0	60.0*	50.0 ± 15.4	87.5 ± 10.7	72.5 ± 8.9
3D FBP	0.0*	100 ± 0.0	60.0*	56.3 ± 15.8	79.2 ± 12.4	70.0 ± 9.8
3D OSEM	0.0*	100 ± 0.0	60.0*	56.3 ± 15.8	87.5 ± 10.7	75.0 ± 9.0

*Lack of true positives caused division by zero in calculation of variation

Table 22: Percent sensitivity, specificity, and accuracy for numerical observer in the MLCS using a threshold of 2.5 TBR and using dilated and eroded ROIs

	Dilated ROIs			Eroded ROIs		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
2D FBP	50.0 ± 15.3	75.0 ± 12.9	65.0 ± 9.9	62.5 ± 16.1	70.8 ± 13.1	67.5 ± 10.2
2D OSEM	43.8 ± 14.6	91.7 ± 9.2	72.5 ± 8.0	50.0 ± 15.3	75.0 ± 12.9	65.0 ± 9.9
3D FBP	50.0 ± 15.3	83.3 ± 11.8	70.0 ± 9.3	62.5 ± 16.2	79.2 ± 12.4	72.5 ± 9.9
3D OSEM	56.3 ± 15.8	87.5 ± 10.7	75.0 ± 9.0	56.3 ± 15.8	79.2 ± 12.4	70.0 ± 9.8

ROI were when the images in question were reconstructed with FBP. These cases are highlighted in bold in Table 23. Using dilated ROIs, there was only one case where the uncorrected TBR was closer to the actual TBR than the corrected TBR. Of the eight eroded ROIs where the uncorrected TBR was closer to the actual TBR than the corrected TBR, six were reconstructed with FBP and two were reconstructed with OSEM. These cases are highlighted in bold in Table 24.

Corrected TBRs were closer, on average, to actual TBRs and varied less for images reconstructed with OSEM than those reconstructed with FBP (Table 25 and Figure 16). This was true for both observer drawn and dilated ROIs, but not for eroded ROIs.

5.4 IMPACT OF THE ADDITION OF CORRELATED CT DATA TO PET DATA ON OBSERVER PERFORMANCE

The addition of correlated CT data to the PET data enhanced diagnostic performance for all observer types, as measured by A_z values, for both the LDS (Table 26) and MLCS (Table 27). Also, qualitatively, the ROC curves for all observers when PET/CT data were used lay above the curves obtained when PET data alone were used (Figure 17 for the LDS and Figure 18 for the MLCS). This enhancement seen in diagnostic performance was in accord with other studies (24) (25) (26) (27) (28).

However, when the impact of viewing PET data simultaneously with CT data was evaluated in terms of sensitivity, specificity, and accuracy, no statistically significant improvement was seen for any observer type in either the LDS or MLCS (Table 28 for the LDS and Table 29 for the MLCS). The sensitivities, specificities, and accuracies evaluated correspond to the diamond shape seen in the plots in Figures 17 and 18. If these values were evaluated at another decision threshold, e.g. the decision threshold with the highest FPF (the square shapes in Figures 17 and 18), diagnostic performance was enhanced with the addition of correlated CT data. In most of the plots in Figures 17 and 18, the red square for the PET/CT data lay closer to the top left corner of the ROC square (meaning increased TPF at all FPFs) than the black square for the PET only data.

Table 23: Percent recovered mean uncorrected and corrected TBRs using mean value from all observer-drawn ROIs. Boldface numbers highlight where uncorrected TBRs were closer to 100% than the corrected TBRs.

		1.08 TBR		1.72 TBR		4.04 TBR		7.56 TBR		Mean	
	[mL]	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected
2D FBP	0.125	*	*	1.1	9.0	3.2	50.9	5.8	66.5	3.4	42.1
	0.25	2.1	25.4	23.6	415	6.9	86.8	8.8	75.7	10.4	151
	0.5	19.8	284	16.2	112	10.7	58.3	9.0	131	13.9	147
	1	**36.8**	**310**	15.7	152	15.2	104	18.9	71.9	21.7	160
	2	31.7	132	44.9	211	27.7	92.4	23.4	67.1	31.9	126
2D OSEM	0.125	*	*	6.0	62.3	5.1	68.8	2.0	22.1	4.5	51.1
	0.25	7.0	72.7	13.4	375	7.7	102	8.4	69.8	9.1	155
	0.5	13.0	167	14.1	105	8.6	33.8	11.3	90.0	11.8	99.1
	1	20.1	202	12.7	113	13.5	75.4	20.0	79.5	16.6	118
	2	27.2	131	24.1	109	33.1	104	29.8	83.6	28.6	107
3D FBP	0.125	4.0	45.6	**19.3**	**647**	2.5	61.7	6.3	317	7.3	268
	0.25	2.1	16.4	8.0	126	15.0	343	7.7	57.6	8.2	136
	0.5	9.0	111	13.1	153	9.8	71.3	16.8	176	12.2	128
	1	20.2	62.9	16.1	187	16.2	71.3	20.4	104	18.2	107
	2	33.4	83.5	15.2	53.3	24.5	128	20.8	51.6	23.5	79.2
3D OSEM	0.125	26.6	320	3.8	99.5	2.9	23.4	5.3	143	9.7	147
	0.25	3.14	21.3	10.8	205	9.4	263	7.3	48.7	7.7	135
	0.5	8.5	98.4	6.2	72.8	10.3	74.3	15.3	171	10.1	104
	1	17.8	54.5	12.0	132	19.9	75.3	25.1	127	18.7	97.2
	2	36.3	87.3	17.0	55.3	30.6	125	27.0	68.5	27.7	84
Mean		17.5	130	14.7	170	13.6	101	14.5	101.2	15.1	125

* No ROI drawn by human observers for this set

Table 24: Percent recovered corrected mean TBRs using mean value from all dilated and eroded observer drawn ROIs. Boldface numbers highlight where uncorrected TBRs were closer to 100% than the corrected TBRs.

		1.08 TBR		1.72 TBR		4.04 TBR		7.56 TBR		Mean	
	[mL]	Dilat-ed	Erod-ed	Dilat-ed	Erod-ed	Dilat-ed	Erod-ed	Dilat-ed	Erod-ed	Dilat-ed	Erod-ed
2D FBP	0.125	*	*	6.1	17.6	32.0	131	52.8	97.6	30.3	82.1
	0.25	19.7	25.0	304	**845**	66.9	92.2	58.7	106	112	269
	0.5	267	418	95.8	137	51.1	94.1	107	172	246	205
	1	**279**	**398**	146	183	90.8	136	63.0	87.4	145	201
	2	124	144	192	**233**	86.8	103	63.7	71.4	117	138
2D OSEM	0.125	*	*	48.4	126	45.8	130	18.5	29.5	37.6	95.2
	0.25	63.8	89.7	257	**761**	82.1	104	56.3	86.9	115	260
	0.5	157	240	83.3	135	31.4	46.3	75.3	119	86.8	135
	1	181	257	104	152	63.8	99.2	69.6	96.8	105	151
	2	121	145	95.0	121	96.6	118	78.9	85.3	97.9	117
3D FBP	0.125	38.5	67.5	376	**1474**	35.2	150	121	**160**	143	463
	0.25	12.8	23.9	95.8	131	218	**792**	42.7	71.5	92.3	255
	0.5	102	142	147	206	61.7	91.1	156	189	117	157
	1	57.1	78.8	178	230	62.3	84.8	91.8	147	97.3	135
	2	76.7	90.7	45.9	60.4	111	135	48.2	55.3	70.5	85.4
3D OSEM	0.125	254	**498**	60.4	220	15.9	37.9	91.8	871	106	407
	0.25	17.3	31.5	152	207	165	627	36.3	63.3	92.7	232
	0.5	90.9	133	67.3	93.2	63.4	91.4	147	188	92.2	126
	1	49.9	67.8	123	171	64.8	92.5	112	182	87.4	128
	2	82.2	94.6	47.4	63.3	106	132	63.6	75.4	77.3	91.3
Mean		114	176	131	278	77.5	164	77.7	220	100	210

* No ROI drawn by human observers for this set

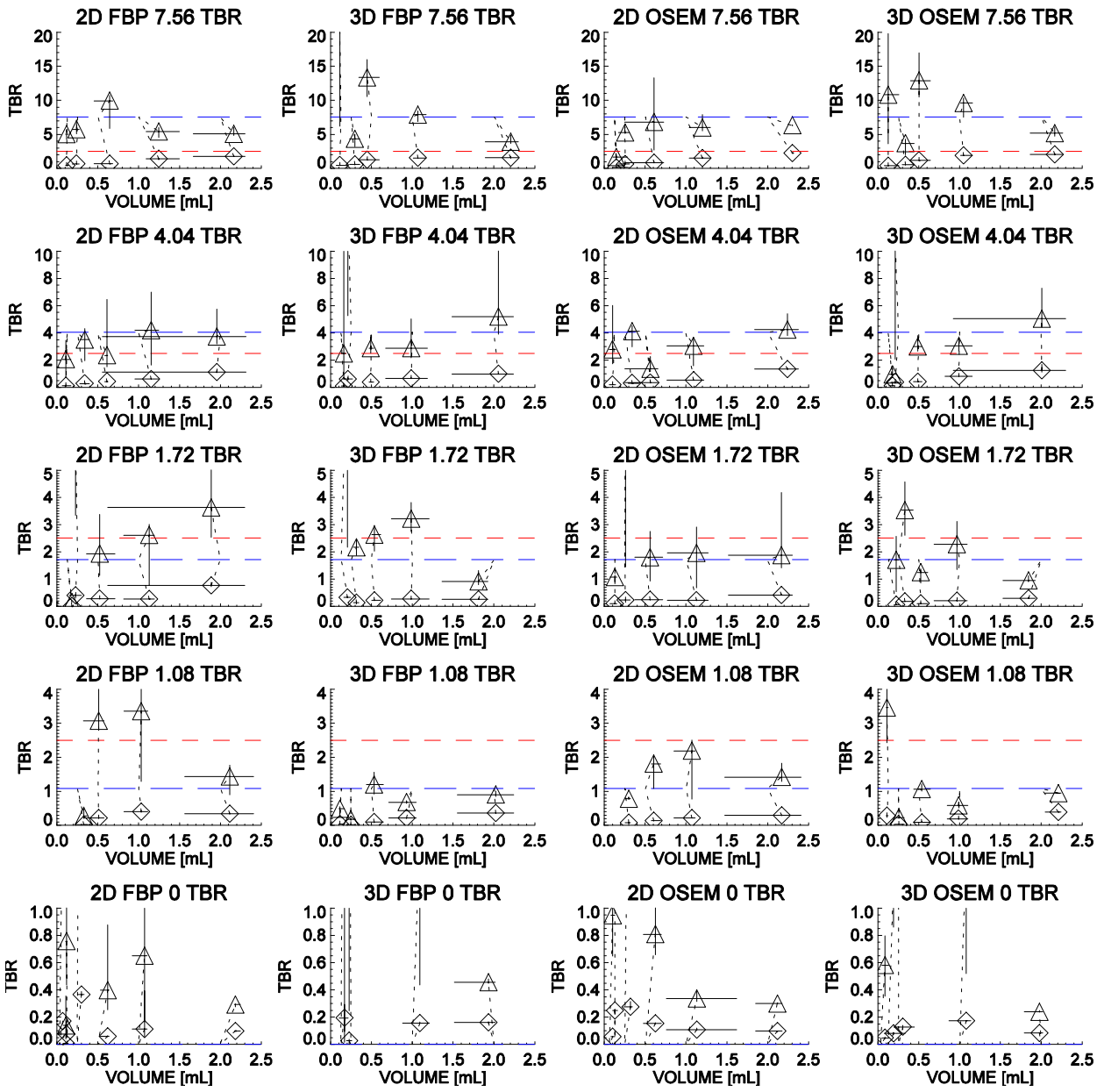

Figure 16: Each graph plots the average volumes (mL) of the ROIs drawn by the observers versus the average TBRs in these ROIs. Diamonds represent uncorrected TBRs and triangles represent corrected TBRs. Error bars for these points represent the standard deviation of the average TBR (vertical bars) and ROI volume (horizontal bars). Dotted lines connect the plotting symbols to the corresponding true lesion volumes and TBRs (blue long-dashed lines). The red short-dashed horizontal line represents a TBR of 2.5.

Table 25: Percent recovered TBR, averaged over all actual TBRs and lesion sizes

	2D FBP	3D FBP	2D OSEM	3D OSEM
ROIs, as drawn	129 ± 106	144 ± 146	109 ± 77.0	113 ± 77.6
Dilated ROIs	111 ± 88.4	104 ± 88.3	91.0 ± 56.5	90.0 ± 57.1
Eroded ROIs	184 ± 192	291 ± 456	155 ± 157	197 ± 218

Table 26: A_z values using pooled data for the LDS

PET	PET/CT	PET	PET/CT
Nuclear medicine physicians		All physicians	
0.68 ± 0.04	0.72 ± 0.04	0.63 ± 0.03	0.77 ± 0.03
Radiation oncologists		Medical physicists	
0.59 ± 0.5	0.81 ± 0.04	0.67 ± 0.04	0.75 ± 0.04

Table 27: A_z values using pooled data for the MLCS

PET	PET/CT	PET	PET/CT
Nuclear medicine physicians		All physicians	
0.74 ± 0.04	0.80 ± 0.04	0.71 ± 0.03	0.83 ± 0.03
Radiation oncologists		Medical physicists	
0.67 ± 0.05	0.87 ± 0.03	0.82 ± 0.09	0.90 ± 0.05

Table 28: Percent sensitivity, specificity, and accuracy using PET/CT data for the LDS

PET			PET/CT		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Nuclear medicine physicians					
34.0 ± 4.2	100 ± 0.0	50.5 ± 3.2	34.6 ± 4.2	97.1 ± 0.0	50.5 ± 3.2
Radiation oncologists					
21.2 ± 3.4	100 ± 0.0	40.9 ± 2.6	25.0 ± 3.7	100 ± 0.0	43.8 ± 2.8
All physicians					
27.6 ± 2.7	100 ± 0.0	45.7 ± 2.0	29.8 ± 2.8	99.0 ± 0.0	47.1 ± 2.1
Medical physicists					
20.5 ± 3.4	100 ± 0.0	40.4 ± 2.5	22.4 ± 3.5	100 ± 0.0	41.8 ± 2.6

Table 29: Percent sensitivity, specificity, and accuracy using PET/CT data for the MLCS

PET			PET/CT		
Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Nuclear medicine physicians					
23.8 ± 5.0	97.7 ± 0.0	69.2 ± 1.9	23.8 ± 5.0	100 ± 0.0	70.7 ± 1.9
Radiation oncologists					
35.0 ± 6.0	97.7 ± 0.0	73.6 ± 2.3	26.2 ± 5.2	99.3 ± 0.0	71.2 ± 2.0
All physicians					
29.4 ± 3.9	97.7 ± 0.0	71.4 ± 1.5	25.0 ± 3.6	96.1 ± 0.0	70.9 ± 1.4
Medical physicists					
13.8 ± 3.9	100 ± 0.0	66.8 ± 1.5	12.5 ± 3.8	100 ± 0.0	66.3 ± 1.4

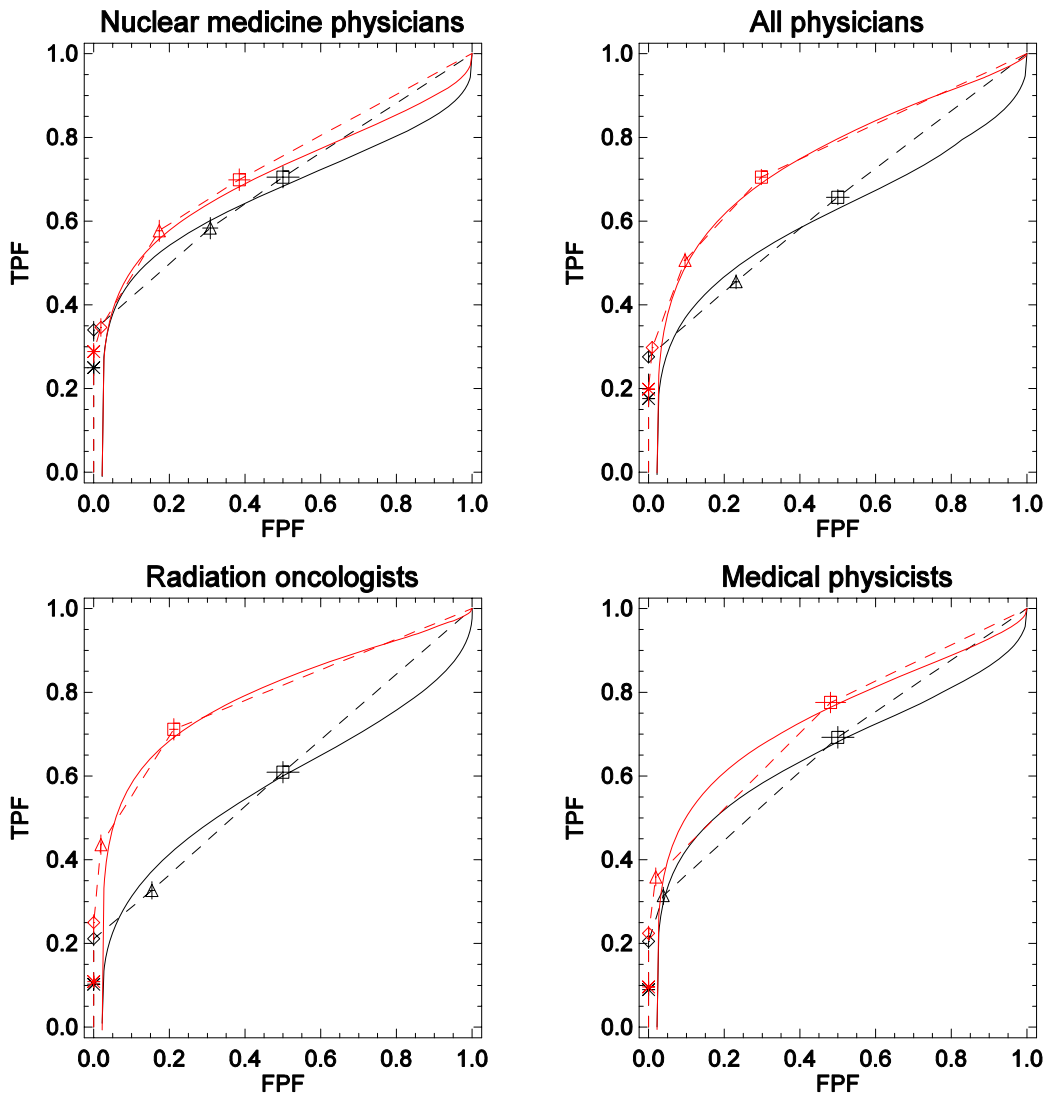

Figure 17: Data for the LDS using PET data alone (black symbols) and PET/CT data (red symbols). Lines connecting symbols (PET data: dashed black line; PET/CT data: dashed red line) are for visual aid only as the data are not continuous. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively. Black lines and red lines are fitted ROC curves representing studies without and with correlated CT images, respectively.

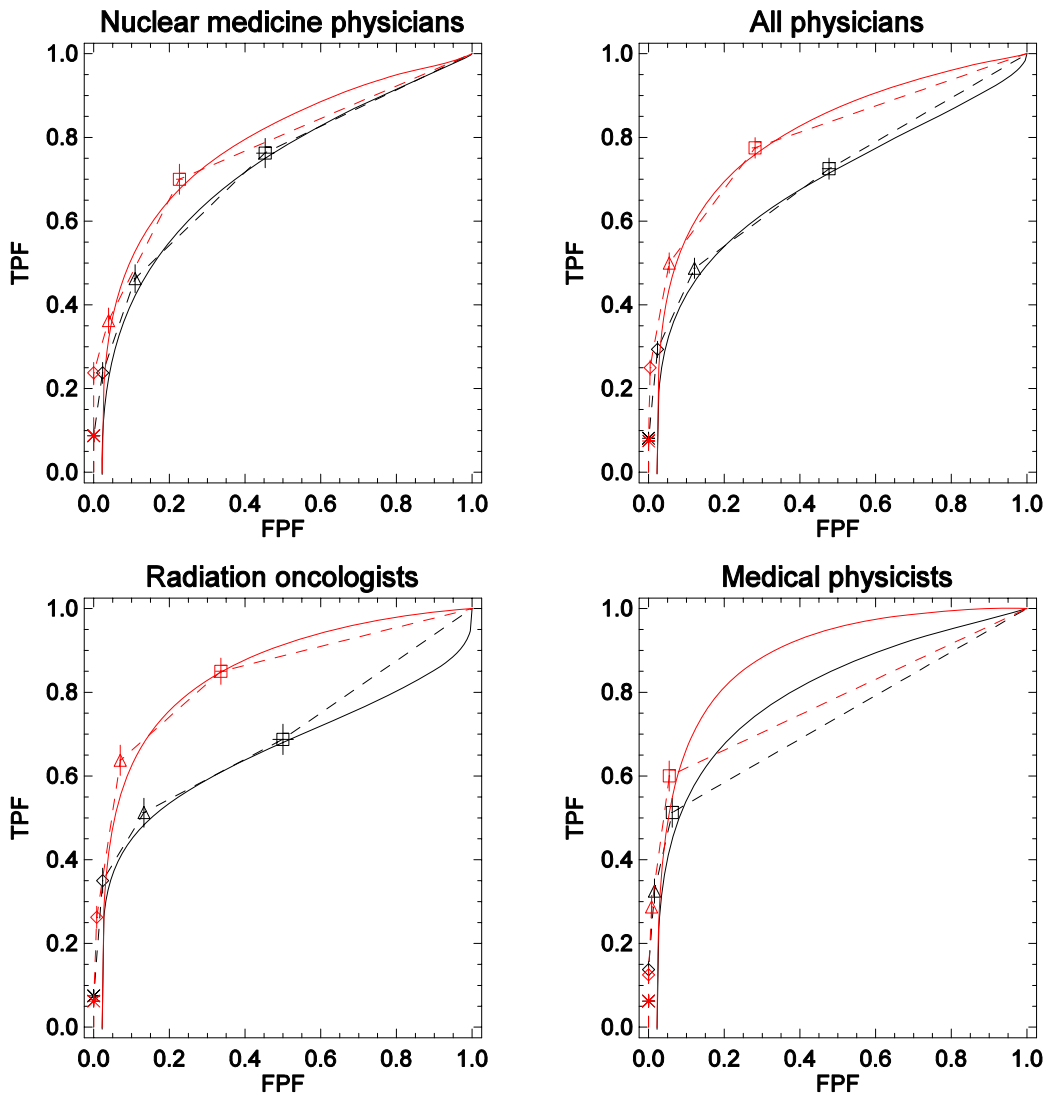

Figure 18: Data for the MLCS using PET data alone (black symbols) and PET/CT data (red symbols). Lines connecting symbols (PET data: dashed black line; PET/CT data: dashed red line) are for visual aid only as the data are not continuous. First, second, third, and fourth decision thresholds correspond to the asterisk, diamond, triangle, and square shapes, respectively. Black lines and red lines are fitted ROC curves representing studies without and with correlated CT images, respectively.

CHAPTER 6:
CONCLUSION

PET images of single pulmonary nodules should be corrected for PVE. PVE correction results in better diagnostic performance, reduced inter-observer variability, and increased quantitative accuracy of PET data.

In this study, the PVE correction resulted in many more actually malignant cases being labeled as malignant at the cost of a slight increase in the number of actually benign cases being labeled as malignant. In other words, observers, when using PVE correction, were much better at finding disease and a little worse at ruling out disease than when they were not using PVE correction.

In this study, the medical physicists had much less experience evaluating PET/CT images than the physicians. Thus, as expected, the physicians outperformed the medical physicists when using uncorrected TBRs. However, when using corrected TBRs, the medical physicists performed similarly to the physicians. This suggests that PVE correction makes evaluating PVE a less subjective task than when not using PVE correction, thus reducing inter-observer variation.

PVE correction is more important than selection of reconstruction algorithm and acquisition mode. In this study 3D OSEM allowed for the best diagnostic performance when PET data was displayed alone. However when PET and CT data were displayed with and without PVE correction, no selection of reconstruction algorithm and acquisition mode outperformed any other. Thus the selection of acquisition mode and reconstruction algorithm should be made based on factors other than diagnostic performance. For instance, if the main factor in selection of acquisition mode is the length of time required to acquire a PET image, 3D mode should be used because of its increased sensitivity. This finding also implies that the capability for 2D versus 3D acquisitions may not be a critical issue when purchasing new equipment if PVE

correction is used. Selection of reconstruction algorithm could depend on a physician's preference based on his or her familiarity with the selected algorithm or based on an expectation that the PET images alone may be used for some other diagnostic purpose.

6.1 FUTURE DIRECTION

Extensions of this work include:

1. Expanding this work with more observers and more image data to further clarify the results of this study
2. Evaluation of the influence of PVE correction on diagnostic performance regarding human PET images for a variety of possible medical conditions
3. Finding the optimal method of PVE correction in terms of diagnostic performance regarding human PET images for a variety of possible medical conditions

The statistical significance of many of the conclusions found in this study regarding acquisition mode and reconstruction algorithm were lacking because of the combination of low populations of observers and images. Also, because of a lack of true and false positives, fitted curves could not be found for some of the ROC data. Expanding this work with more observers and image data would further clarify the results found in this study. Logistically, expanding this work could be difficult. Evaluation of the 104 image sets took around six hours per observer; reducing standard deviations by one half would require roughly four times as many data sets or observers.

Evaluating the influence of PVE correction on diagnostic performance of experienced observers regarding human PET images would be a natural extension of this work. Results found in this study were promising enough to suggest that PVE correction should be used with human PET images. However, the efforts to collect the relevant image data are time consuming. These efforts include obtaining information on the true condition of the patient. It

may be impossible to determine the true tumor uptake unless some other independent method is employed, such as lesion biopsy. Also, determining the true state of disease often requires following the patient for several years. Although a data set for PET and CT images was recently published (29), the PET and CT images had been acquired on separate scanners; this data set preceded the common use of PET/CT. An equivalent PET/CT data set would take an amount of time to accumulate that is prohibitive in terms of a Masters research project.

The second extension of this work is suggested because any given method of PVE correction may not be suitable for all possible medical conditions. For instance, the method of PVE correction used in this study is suitable only for lesions with homogenous radiotracer uptake that are surrounded by a structure with homogenous radiotracer uptake. This method would not be suitable for necrotic lesions with heterogeneous radiotracer uptake, where a pixel-by-pixel correction would be more suitable, or for lesions surrounded by two different structures, where the geometric transfer matrix method would be more suitable.

REFERENCES

1. J. Bradley, W. Thorstad, S. Mutic, T. Miller, F. Dehdashti et. al., "Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer", International Journal of Radiation Oncology, Biology, Physics, **59**, 78–86, 2004

2. F. Ciernik, E. Dizendorf, B. Baumert, B. Reiner, C. Burder et. al., "Radiation treatment planning with an integrated positron emission and computed tomography PET/CT : A. feasibility study", International Journal of Radiation Oncology, Biology, Physics, **57**, 853–863, 2003

3. J. Vanuytsel, J. Vansteenkiste, S. Stroobants, P. De Leyn, W. De Wever et. al., "The impact of 18 F-fluoro-2-deoxy-D-glucose positron emission tomography FDG-PET lymph node staging on the radiation treatment volumes in patients with non-small cell lung cancer", Radiotherapy and Oncology, **55**, 317–324, 2000

4. M. Soret, S. Bacharach, I. Buvat, “Partial-volume effect in PET tumor imaging”, Journal of Nuclear Medicine, **48**, 932-945, 2007

5. C. Lartizien**,** P. Kinahan, R. Swensson, C. Comtat, M. Lin et al., “Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging”, Journal of Nuclear Medicine, **44**, 276-290, 2003

6. H. Son, M. Yun, T. Jeon, D. Kim, H. Jung et al., “ROC analysis of ordered subset expectation maximization and filtered back projection technique for FDG-PET in lung cancer”, IEEE Transactions on Nuclear Science, **50**, 37-41, 2003

7. R. Boellaard, A. Van Lingen, A. Lammertsma, "Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: Quantitative characteristics and effects on kinetic modeling", Journal of Nuclear Medicine, **42**, 808-817, 2003

8. V. Bettinardi, P. Mancosu, M. Danna, G. Giovacchini, C. Landoni, M. Gilardi et al., "Two-dimensional vs three-dimensional imaging in whole body oncologic PET/CT: a Discovery-STE phantom and patient study", Quarterly Journal of Nuclear Medicine and Molecular Imaging, **51**, 214-223, 2007.

9. S. Kim, M. Allen-Auerbach, J. Goldin, B. Fueger, M. Dahlbom et al., “Accuracy of PET/CT in characterization of solitary pulmonary lesions”, Journal of Nuclear Medicine, **48**, 214-220, 2007

10. J. Bushberg, J. Seibert, E. Leidholdt, J. Boone, "The Essential Physics of Medical Imaging, Second Edition", Philadelphia, PA, Lippincott Williams & Wilkins, 2002

11. R. Boellaard, K. Nanda, O. Hoekstra, A. Lammertsma, "Effects of Noise, Image Resolution, and ROI Definition on the Accuracy of Standard Uptake Values: A Simulation Study", Journal of Nuclear Medicine, **45**, 1519-1527, 2004

12. B. Teo, Y. Seo, S. Bacharach, J. Carrasquillo, S. Libutti et al., “Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data”, Journal of Nuclear Medicine, **48**, 802-810, 2007

13. M. Hickeson, M. Yun, A. Matthies, H. Zhuang, L. Adam et. al., "Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET", European Journal of Nuclear Medicine, **29**, 1639-1647, 2002

14. V. Frouin, C. Comtat, A. Reillhac, M. Gregoire, "Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness", Journal of Nuclear Medicine, **43**, 1715 - 1726, 2002

15. G. Bolard, J. Prior, L. Modolo, A. Delaloye, M. Kosinski et al., “Performance comparison of two commercial BGO-based PET/CT scanners using NEMA NU 2-2001”, Medical Physics, **34**, 2708-2717, 2007

16. O. Mawlawi, D. Podoloff, S. Kohlmyer, J. Williams, C. Stearns et al., “Performance characteristics of a newly developed PET/CT scanner using NEMA standards in 2D and 3D modes”, Journal of Nuclear Medicine, **45**, 1734-1742, 2004

17. B. McNeil, S. Adelstein, "Determining the Value of Diagnostic and Screening tests", Journal of Nuclear Medicine, **17**, 439-448, 1976

18. T. Fawcett, "An Introduction to ROC analysis", Pattern Recognition Letters, **27**, 861-874, 2006

19. C. Metz, "Basic Principles of ROC Analysis", Seminars in Nuclear Medicine, **8**, 283-298, 1978

20. V. Bettinard, M. Danna, A. Savi, M. Lecchi, I. Castiglioni et al., “Performance evaluation of the new whole-body PET/CT”, European Journal of Nuclear Medicine and Molecular Imaging, **48**, 867–881, 2004

21. "Performance Measurements of Scintillation Cameras", National Electrical Manufacturers Association, 2001

22. U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig et al., “Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy for patients with non-small cell lung cancer”, Journal of Nuclear Medicine, **48**, 1342-1348, 2005

23. P. Bevington, D. Robinson, “Data reduction and error analysis for the physical sciences”, McGraw-Hill, New York, second edition, 1992

24. F. Mottaghy, C. Sunderkotter, R. Schubert, P. Wohlfart, N. Blumstein et. al., "Direct comparison of [F-18]FDG PET/CT with PET alone and with side-by-side PET and CT in patients with malignant melanoma", European Journal of Nuclear Medicine and Molecular Imaging, **34**, 1355 - 1364, 2007

25. T. Hany, H. Steinert, G. Goerres, A. Buck, G. von Schulthess, "PET diagnostic accuracy: Improvement with in-line PET-CT system: Initial results", Radiology, **225**, 575-581, 2002

26. W. De Wever, S. Ceyssens, L. Mortelmans, S. Stroobants, G. Marchal et. al., "Additional value of PET-CT in the staging of lung cancer: comparison with CT alone, PET alone and visual correlation of PET and CT", European Radiology, **17**, 23-32, 2007

27. E. Even-Sapir, U. Metser, G. Flusser, L. Zuriel, Y. Kollender et. al., "Assessment of malignant skeletal disease: Initial experience with F-18-fluoride PET/CT and comparison between F-18-flouride PET and F-18-fluoride PET/CT", Journal of Nuclear Medicine, **45**, 272-278, 2004

28. G. Goerres, G. Schulthess, H. Steinert, "Why most PET of lung and head-and-neck be cancer will be PET/CT", Journal of Nuclear Medicine, **45**, 66S-71S, 2004

29. J. Fletcher, S. Kymes, M. Gould, N. Alazraki, R. Coleman et. al., “A comparison of the diagnostic accuracy of F-18-FDG PET and CT in the characterization of solitary pulmonary nodules”, Journal of Nuclear Medicine, **49**, 179-185, 2008

VITA

Andrew Morrow was born in New Orleans, Louisiana, on March 20, 1983. Since then, he has earned a bachelor's degree in science, major physics, magna cum laude, married his high school sweetheart, and fathered a daughter. He enjoys jogging, video games, using commas, and, with the help of his daughter, poking ant piles with sticks.