

**ROC COMPARISON OF ACQUISITION PARAMETERS  
FOR TWO PET/CT SCANNERS BASED ON LESION  
DETECTABILITY IN A TORSO PHANTOM**

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

Kenneth Bernstein  
M.Sc., Louisiana State University, 1995  
Louisiana State University  
Baton Rouge, Louisiana

August 2005

## ACKNOWLEDGMENTS

I thank Dr. Kenneth Matthews for advising me through the graduate program in medical physics at LSU. He was generous enough to offer me financial support in my first semester on a research assistantship and has proven an excellent teacher in imaging science. In the three courses in medical imaging I have taken from him at LSU he has inspired me to work on a project in medical imaging.

I thank Dr. Ken Hogstrom for taking over the program and ensuring the highest quality instruction for both didactic and clinical training. He has added a great deal of value to my degree at LSU. I have been privileged to watch Dr. Gibbons raise the standard of care at Mary Bird Perkins Cancer Center. The wealth of knowledge in clinical training from Dr. Gibbons and Dr. Boyd has been fantastic and for this I will always be in their debt.

I also thank Dr. Sheldon Johnson for support during my clinical rotation at Mary Bird Perkins and for the use of the GE Discovery PET/CT scanner. I thank Dr. Steven Bujenovic for the use of the CTI Reveal scanners, for the very informative conversations, and for serving on my committee. I thank Dr. Warren Johnson for assistance with MATLAB programming and for serving on my committee.

I thank Dr. John Tyler for serving on my committee and for teaching an informative medical imaging class in the computer science department.

I thank Dr. Erno Sajo for serving on my committee, for mentoring me through a reading course, and teaching a very heavy course load to provide the students of my year with adequate training. He has always challenged me both in seminar and in the classroom. I owe my interest in radiation biology and genetics to his

stimulating lectures.

Most of all, I thank Dr. Blair Smith, whose debt I will always be in, for the tremendous amount of assistance he has provided me in editing this thesis.

I thank all the observers who participated in the study, some of whom have been already mentioned. The trained observers were: Dr. Bujenovic, Dr. McNeill, Dr. Dodson, Dr. Johnson, and Dr. Young. The untrained observers I need to acknowledge are: Yuri Ishihara, Rajesh Manoharan, Ines-Ana Jurkovic, John Jarret, Laurie Kelly, John Richert, Carol and Conrad Vorhoff, Keith Vorhoff, Roger Duhe, Dr. Joe Dugas, Dr. Smith and Dr. Matthews.

For all the support and assistance outside of the academic setting that made it possible for me to return to school, I thank my wife, Ashley, my children, Emma and Havah, my parents Audrey and Irving, and my sister, Serena.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b> . . . . .	<b>ii</b>
<b>LIST OF TABLES</b> . . . . .	<b>vi</b>
<b>LIST OF FIGURES</b> . . . . .	<b>vii</b>
<b>ABSTRACT</b> . . . . .	<b>ix</b>
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b> . . . . .	<b>1</b>
1.1 Overview . . . . .	1
1.2 Organization of the Thesis . . . . .	2
<b>2 QUESTIONS, MOTIVATION AND HYPOTHESES</b> . . . . .	<b>3</b>
2.1 Motivation for the Study of PET/CT Parameters Using ROC . . . . .	3
2.1.1 Medical Imaging Motivations . . . . .	4
2.1.2 Fundamental Technical Questions . . . . .	4
2.2 Specific Questions of This Thesis . . . . .	9
2.3 Initial Hypotheses . . . . .	10
<b>3 REVIEW OF PET/CT IMAGING AND ROC ANALYSIS</b> . . . . .	<b>13</b>
3.1 Overview of PET and CT Technology . . . . .	14
3.1.1 Some Characteristics of PET/CT Scanners . . . . .	14
3.1.2 Correction of Image Data for Random and Scattered Events . . . . .	16
3.1.3 Correction of Image Data for Attenuation . . . . .	17
3.1.4 Image Reconstruction . . . . .	18
3.2 Radionuclides Used in PET . . . . .	21
3.2.1 Standard Uptake Values (SUV) . . . . .	22
3.3 Overview of ROC Analysis . . . . .	23
3.3.1 The Need for ROC Analysis . . . . .	23
3.3.2 ROC Analysis . . . . .	26
3.3.3 Requirement for Diagnostic Truth in ROC Analysis . . . . .	31
<b>4 EXPERIMENTAL METHODS AND MATERIALS</b> . . . . .	<b>33</b>
4.1 Medical PET/CT Scanners . . . . .	33
4.1.1 The GE Discovery ST . . . . .	33
4.1.2 The CTI-Siemens Reveal HD . . . . .	35
4.1.3 Default Acquisition Protocols and Reconstruction Methods . . . . .	36
4.2 Laboratory Materials and Software . . . . .	37
4.2.1 Phantom and Sphere (Simulated Lesion) Set . . . . .	37
4.2.2 Preparation of the Phantom . . . . .	39

4.3	Preparation of the Images for the ROC Study . . . . .	41
4.3.1	Creating Images that Contain Lesions . . . . .	41
4.3.2	Design of Image Presentation and Confidence Survey . . . . .	42
4.4	Method of ROC Analysis . . . . .	45
4.4.1	Recording and Management of the ROC Database . . . . .	45
4.4.2	ROC Analysis Details . . . . .	45
4.4.3	Requirements of ROC Analysis . . . . .	47
<b>5</b>	<b>RESULTS OF ROC ANALYSIS . . . . .</b>	<b>49</b>
5.1	Overall ROC Analysis Results and Comparisons . . . . .	49
5.1.1	Overall ROC Results for Trained and Untrained Observers . . . . .	49
5.1.2	Results for All Observers Comparing Scanner Acquisition Modes . . . . .	50
5.1.3	Results for All Observers Comparing Reconstruction Algorithms . . . . .	53
5.1.4	ROC Results for All Observers Comparing Acquisition Times . . . . .	54
5.2	ROC Results for Trained Observers . . . . .	56
5.2.1	Differences Between Trained and Untrained Observers . . . . .	56
5.2.2	Comparisons of $A_z$ Values for Trained Versus Untrained Observers . . . . .	65
5.3	Diagnostic Accuracy, Sensitivity and Specificity . . . . .	67
<b>6</b>	<b>CONCLUSION . . . . .</b>	<b>72</b>
6.1	Validation of Hypotheses . . . . .	73
6.2	Proposed Future Related Projects . . . . .	75
	<b>REFERENCES . . . . .</b>	<b>78</b>
	<b>APPENDIX</b>	
	<b>A SAMPLE ROCKIT PROGRAM RUN . . . . .</b>	<b>84</b>
	<b>B SURVEY INSTRUCTIONS DOCUMENT . . . . .</b>	<b>93</b>
	<b>C MATLAB ROUTINES . . . . .</b>	<b>96</b>
	<b>D ROC INPUT DATA . . . . .</b>	<b>99</b>
	<b>VITA . . . . .</b>	<b>105</b>

## LIST OF TABLES

2.1	Physiological regions and conditions associated with $^{18}\text{F}$ -FDG uptake.	5
3.1	Radionuclides used in PET imaging. . . . .	22
3.2	Grouping of diagnostic decisions into two different partitions of two sets each . . . . .	28
4.1	Manufacturers' specifications for the Discovery ST and Reveal HD scanners. . . . .	34
5.1	Table of combined overall $A_z$ results from ROC curves. . . . .	50
5.2	Comparison of $A_z$ results for trained versus untrained observers. . . .	59
5.3	Summary of $A_z$ values and relative differences. . . . .	61
5.4	Sensitivity, specificity and accuracy for all observers, trained observers, and untrained observers. . . . .	69
5.5	The effect of acquisition times in the categories on the sensitivity, specificity and accuracy. . . . .	70
5.6	Comparison of reconstruction algorithms in the categories of sensitivity, specificity and accuracy. . . . .	71
6.1	Scanner performance parameters. . . . .	72
C.1	Example MATLAB commands for plotting ROC results. . . . .	96
C.2	Example MATLAB commands for lesion image fusion. . . . .	98
D.1	ROC survey results. . . . .	99

## LIST OF FIGURES

3.1	Image of lesion in left wall of a lung. . . . .	15
3.2	Illustration of a line of response (LOR) from PET scan. . . . .	19
3.3	The sinogram for the elliptical object shown in Figure 3.2. . . . .	19
3.4	An example of a poorly separated binormal (double Gaussian) distribution. . . . .	26
3.5	An example of a well-separated binormal (double Gaussian) distribution. . . . .	27
3.6	An example of the typical form for an ROC curve. . . . .	29
3.7	An idealized ROC curve obtained from fictional observers with zero discriminatory capacity. . . . .	30
3.8	The ROC curve for a perfect diagnostic test. . . . .	31
3.9	The ROC curve for an observer who reverses all positive and negative decisions. . . . .	32
4.1	The front view of the MBPCC Discovery ST PET/CT scanner. . . . .	35
4.2	Torso phantom preparation. . . . .	38
4.3	Photo of the micro-hollow spheres used to simulate lesions. . . . .	38
5.1	Composite ROC curves combining all scanner parameters. . . . .	51
5.2	ROC curves comparing the two PET/CT scanners, summarized over all observers; all other parameters are lumped together. . . . .	52
5.3	Comparison of ROC results for reconstruction algorithms. . . . .	54
5.4	Torso phantom images comparing three reconstruction algorithms. . . . .	55
5.5	Comparison of ROC results for different image acquisition times. . . . .	57
5.6	Summary of $A_z$ values grouped by comparison parameter for all observers, trained observers and untrained observers. . . . .	60

5.7	Chart of percentage difference in ROC curve $A_z$ values for trained versus untrained observers. . . . .	62
5.8	ROC curve for trained observers with different reconstruction algorithms.	63
5.9	Trained observer ROC results comparing image scanner acquisition modes. . . . .	63
5.10	Trained observer ROC results for different acquisition times. . . . .	65
5.11	Comparison of trained observer $A_z$ results alongside untrained observer $A_z$ results with mean and standard deviation across parameters. . . . .	66

## ABSTRACT

Positron emission tomography (PET) and computed tomography (CT) are well established and powerful tools for medical diagnostics but even integrated PET/CT scanner images still lack the necessary quality and resolution that would make medical diagnoses flawless. In this thesis experiments were performed to statistically determine the effect that image acquisition parameters have upon diagnostic accuracy. Images from different PET/CT scanners were assessed by comparing subject human diagnostic accuracy from a sample of both professional and student volunteers. The assessment results were compared to the objective NEMA-standards performance data provided by the manufacturers for each scanner. The data analysis method is the receiver operating characteristic (ROC) curve. We hypothesize that human performance in making accurate diagnoses from PET images correlates with the system performance. The data shows that human diagnostic performance correlates to spatial resolution and sensitivity of the PET imaging systems.

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

Positron emission tomography (PET) and computed tomography (CT) are well established and powerful tools for medical diagnostics but even integrated PET/CT scanner images still lack the necessary quality and resolution that would make medical diagnoses flawless. In this thesis experiments were performed to statistically determine the effect that PET image acquisition parameters have upon diagnostic accuracy.

This thesis presents a study to compare the diagnostic accuracy of the PET component of two clinical PET/CT scanners. The two available PET/CT scanners are systems installed respectively at Our Lady of the Lake Regional Medical Center (OLOLRMC) and Mary Bird Perkins Cancer Center (MBPCC), both located in Baton Rouge, Louisiana.

The comparison is based on the different physical characteristics and image acquisition parameters of the two systems, such as 2D versus 3D acquisition, length of acquisition, and reconstruction algorithm used. The study uses receiver operating characteristics (ROC) analysis to measure diagnostic accuracy. The diagnostic task was identification of lesions in PET images of an anthropomorphic torso phantom. Two sets of readers, medical professionals (radiologists and radiation oncologists) with extensive practice reading medical images and untrained volunteers (medical physics students, post doctoral researchers and faculty, as well as some additional non-professional volunteers) were used.

It is shown that human performance in making accurate diagnoses from PET images correlates with spatial resolution and system sensitivity.

## **1.2 Organization of the Thesis**

Chapter 2 presents hypotheses and motivations. Chapter 3 gives background information on PET/CT scanners and ROC analysis. In Chapter 4, the materials and methods used in this study are described. Chapter 5 presents the research results. The final chapter summarizes the project and provides suggestions for continued work on this topic.

## CHAPTER 2

### QUESTIONS, MOTIVATION AND HYPOTHESES

This chapter describes the motivations behind pursuing this thesis and presents identified goals for the research and the hypotheses that were adopted.

#### 2.1 Motivation for the Study of PET/CT Parameters Using ROC

The broad plan of this thesis is to investigate and assess PET/CT scanner image quality via correlation between design and acquisition choices and accuracy of human performance in making diagnoses from the images. The primary motivation is therefore fundamentally to help to guide medical professionals in making better decisions that will ultimately benefit their patients. While exploring these broad goals and motivations in depth, a number of further motivating questions and problems were outlined. First a set of motivations arising from medical diagnostic problems are listed. Secondly, a set of fundamental applied research (or technical) motivating issues are listed. This discussion should provide the reader with some appreciation for the relevance and importance of this research.

The initial consultations for the thesis proposal addressed the question of which of the two available PET/CT scanners at OLOLRMC and MBPCC provides the best diagnostic performance, the CTI-Siemens Reveal HD or the GE Discovery ST. In addition an interesting problem would be to examine the effects of different acquisition parameters on image quality, because this has important ramifications for clinical practice and diagnosis of diseases. A question can be asked, “what set of PET/CT scanner parameters ensures the best possible diagnosis?” It is apparent that this questions is too broad and the focus had to be narrowed. The

considerations for narrowing the focus are discussed in the rest of the chapter and various references [1]–[5].

### 2.1.1 Medical Imaging Motivations

The main utility of PET in diagnosing and staging cancer is in the modality's ability to view metabolic activity [6]. Cancer cells, which in general proliferate more quickly than normal cells, will have a higher glucose metabolism. By placing the radioactive positron emitting isotope  $^{18}\text{F}$  on a sugar molecule, the molecule fluorodeoxyglucose  $^{18}\text{F}$ -FDG is constructed. The presence of  $^{18}\text{F}$ -FDG makes cancer cells visible on a PET image [7]. The higher the concentration of  $^{18}\text{F}$ -FDG the brighter the appearance of the PET scan. Pathologies that show up as lesions in specific organs are a prime concern for forming accurate diagnoses from medical images. Some pathologies are very apparent on PET scans while others are impossible to image. The indication is the  $^{18}\text{F}$ -FDG uptake of the tissues affected. Both normal and pathologic cells have some FDG uptake. Table 2.1 shows the types of normal tissue uptake that will be present in most PET scans regardless of state of health and types of FDG uptake that are only present if a pathology.

Cancer diagnosis and perhaps therapy stand to benefit from delivering the highest diagnostic image quality; this thesis tries to address this in the limited context of patient care with the PET/CT scanners at OLOLRMC and MBPCC.

### 2.1.2 Fundamental Technical Questions

The fundamental technical questions address specific practical issues related to conducting the research project. This section outlines these issues. The list of questions has evolved over time. The current set of technical questions comprises

Table 2.1. Physiological regions and conditions associated with  $^{18}\text{F}$ -FDG uptake with the type of uptake (normal or pathological)\*.

$^{18}\text{F}$ -FDG Uptake Type	
Normal	Pathological
Myocardium	Metabolically active tumors
Brain.	Acute radiation injury
Urine	Subacute radiation injury
Muscle	Osteochondritis
Thymus	Bone infarctions
Saliva	Blood pooling (pancoast tumor)
Wall of large vessels	Blood pooling (SVC obstruction )
Bowel	Bronchiectasis
Joints	Lipoid pneumonia
Thyroid (including with thyroiditis)	Sarcoid
Female breast (including with lactation)	Granulomatous diseases
Brown fat (especially in younger adults)	Bone marrow (including with cytokines)
Uterus (especially with menses)	Collagen vascular disease.

\* Adapted from [8].

eight issues in two groups. The first group of questions required answering as part of the planning of the research project (see also [9]–[13]). As experimentalists, our procedure was to accumulate as much data as was reasonably possible in a small amount of time. The second group of questions deals with optimizing the available settings on the PET/CT scanners to maximize diagnostic performance (see also [14]–[17]).

- **Planning Questions**

1. Should one study patient images or images from an anthropomorphic phantom?
2. Should the PET and CT data be fused for the analysis and if so, how to do the registration?
3. What type of pathology should be considered [18]?
4. How should a pathology be represented and/or identified?

- **Optimization Questions**

1. What performance parameters should be considered when analyzing the results?
2. What acquisition parameters affect diagnostic performance?
3. What reconstruction algorithm provides the best diagnostic performance?
4. Given two acquisition modes on the Discovery ST, is either the 2D or the 3D mode better for diagnostic performance?

Also see [19] for questions related to optimization.

- **Discussion**

There are many obstacles to using patient data. For this study it would require institutional consent from LSU, OLOLRMC and MBPCC. On the other hand, if one uses phantom data there is no confidentiality requirement that would need approval. Also, patient images must be correlated by another means to determine that a patient is without pathology. By contrast, an ROC study of phantom images can be done in a simpler manner with definitive input of actual positives and actual negatives. Thus the ROC is easier to administer with phantom data. We know for certain whether the spheres that represent the pathology are present or not present. For the ROC study, ideally one can arrange relative numbers of positives and negatives in correct proportion for the prevalence of the disease being studied. This is very easy to achieve with phantom images but may not be with patient images.

To examine the effects of different acquisition parameters on image quality, ideally one needs multiple data sets for comparison. It would have been unreasonable to ask patients to have a second scan with a different acquisition time or in the case of the Discovery ST a second scan with the septa retracted for 3D acquisition. All of the parameters can be studied only by taking multiple acquisitions.

The main disadvantage to using phantom data is that the results at best only imply that they will be similar for the same type of patient data. Also, phantom images might be too simplistic. Patient images have the advantage of realism. Patient images exactly show all the real parts that are absent in the phantom or are artificially present in the phantom.

Whether using phantom or patient data it should be noted that any results in an ROC study only apply to a small subset of pathologies being studied [20, 21].

However plausible it is to extend the results to larger classes of pathologies, this cannot be a general extension of the statistical significance of an ROC study.

Extending the results to larger classes of pathologies would also not be reasonable for a small subset of real patient images. The thesis conclusions are technically valid for the simplified objects explored. The broader diagnostic significance of this thesis is therefore dependent on the assumption that the results can be extended to patient images with complex anatomy and physiology [22].

- **CT-PET Registration Characteristics**

PET and CT images can be fused in many ways [23]. These methods can be classified as “pre-acquisition” or prospective and “post-acquisition” or retrospective techniques. Retrospective fusion of PET and CT images must be done when the images are acquired from physically separate PET and CT scanners. With both classifications the images from the CT and PET are acquired separately, the difference being in the prospective technique the relative position of the body being imaged is known with a degree of precision. The images are then superimposed using various techniques, such as physical overlays, alignment of intensities, or fiducial markers. With prospective techniques, which are often called hardware fusion, the images can be directly overlaid because the two scanners are mechanically aligned in a clinical PET/CT system. The two scanners investigated in this project are dual PET/CT systems fused prospectively. The CT scanner is located a short table translation from the PET scanner. With patient immobilization devices such as VacLoc™ and Aquaplast RT™, the dedicated PET/CT scanners using prospective fusion do a superior job in image registration [23]. In addition, the short time interval between the CT and PET

scans decrease the chance for patient movement. By mapping directly the coordinate system on the CT scan to the coordinate system on the PET scan, many of the difficulties with retrospective alignment of images are minimized. Further motivation from medical imaging is obtained from a review of the characteristics of CT-PET fusion methods.

Over the past few years the physicians at MBPCC have identified several aspects of the impact of PET/CT on treatment planning [8].

- PET/CT allows more appropriate doses/volumes for extent of primary in certain head and neck, esophagus and lung with atelectasis;
- PET/CT allows more appropriate doses/volumes for nodal regions;
- PET/CT reduces “false positive” PET and occasionally “false positive” findings on CT;
- PET/CT improves determination of “indeterminate” CT nodes; and
- PET/CT improves sensitivity in cervical nodes.

It should be noted that CT-PET fusion methods were not part of the investigation in this thesis research. Expanding the research to include evaluation of fused PET/CT images, as well as patient images, is planned for future work.

## **2.2 Specific Questions of This Thesis**

The issues that were identified as worthy of investigation are listed here. These are referred to as the primary questions of this thesis.

**Question 1.** Using ROC analysis, can we distinguish a difference in diagnostic image quality between images taken with the Discovery ST in 2D mode and images taken with the Discovery ST in 3D mode?

**Question 2.** Using ROC analysis, can we distinguish a difference in diagnostic image quality between images taken with the Discovery ST in 2D mode and images taken with the Reveal HD for which only 3D mode is available?

**Question 3.** Using ROC analysis, can we distinguish a difference in diagnostic image quality between images taken with the Discovery ST in 3D mode and images taken with the Reveal HD?

**Question 4.** The PET images appear more visually appealing using an iterative reconstruction algorithm than filtered backprojection or other noniterative algorithms [24]. Does this correlate with an observer's ability to identify lesions?

**Question 5.** What is the correlation between system performance parameters and human observer performance in identifying lesions?

**Question 6.** Using ROC analysis, can we distinguish a difference in diagnostic image quality between images taken with 4-minute or 16-minute acquisition times?

### **2.3 Initial Hypotheses**

After consultation with medical physicists, radiologists and radiation oncologists at LSU, OLOLRMC and MBPCC, and with the foregoing motivational

problems in mind, six hypotheses were formed. The reader is referred to Chapter 3 for discussion of ROC analysis and a review of PET imaging methods.

**Hypothesis 1.** An ROC study will confirm that 2D acquisition of PET/CT data produces superior diagnostic image quality exhibited by a larger  $A_z$  value for the ROC curve than 3D acquisition.

*Reasoning:* The Discovery ST 2D mode of acquisition is used preferentially for whole-body imaging at MBPCC over the 3D mode. There are few clinical protocols in which the images at MBPCC are taken in 3D mode. An initial survey of patient images created with each mode from the Discovery ST by an OLOLRMC radiologist and an LSU physicist indicated that the 2D mode was generally preferable [25].

**Hypothesis 2.** An ROC study will show that the Discovery ST scanner operated in 2D mode shows inferior diagnostic image quality as compared to the Reveal HD scanner.

*Reasoning:* The Discovery ST 2D mode of acquisition has a larger spatial resolution and smaller sensitivity than the Reveal HD scanner as specified by the manufacturers.

**Hypothesis 3.** An ROC study shows that the Discovery ST scanner operated in 3D mode will show inferior diagnostic image quality as the Reveal HD scanner.

*Reasoning:* The Discovery ST 3D mode of acquisition has virtually the same sensitivity but inferior spatial resolution to the Reveal HD.

**Hypothesis 4.** An ROC study will show that the iterative reconstruction algorithm yields superior diagnostic image quality for both the 3D FORE and DIFT algorithms.

*Reasoning:* The images reconstructed with the iterative OSEM algorithm are more visually appealing and contain fewer reconstruction artifacts. This should make it easier to identify lesions in a PET image.

**Hypothesis 5.** Superior system performance parameters correlate to superior diagnostic image quality.

*Reasoning:* Crystal dimensions, spacing, ring separation, block size and detector electronics are all physical parameters that combine to determine the system performance parameters such as axial and transaxial resolution and sensitivity. Measurements of system performance parameters using NEMA guidelines are the standard used to evaluate PET/CT systems. Important system parameters are transverse resolution, axial resolution, sensitivity, crystal size, and noise equivalent count rate (NEC). Improvements in these parameters should improve diagnostic image quality.

**Hypothesis 6.** A longer acquisition improves the diagnostic image quality compared to shorter acquisitions.

*Reasoning:* The primary source of noise is in the Poisson statistics. That noise will decrease as  $1/\sqrt{N}$  where  $N$  is the number of counts.

## CHAPTER 3

### REVIEW OF PET/CT IMAGING AND ROC ANALYSIS

Any improvement in knowledge about how engineering standards either help or hinder medical diagnoses can be counted as potentially life-saving information. PET imaging provides information about the metabolic activity or function of the tissue being scanned. This is a powerful means by which a competent physician can glean knowledge about disease and various pathologies that structural imaging techniques such as x-ray imaging, CT, and magnetic resonance imaging (MRI) do not reveal. The latter techniques typically provide information about the density and composition of tissue and other structural features, but not functional information. PET scanning can reveal active tissue ‘hot spots’ that should not be present (signaling perhaps malignant tumors or other lesions); however, some hot-spots are indicative of normal physiology. PET can also reveal dead regions where one would expect activity (for example, in the heart this could signal dead myocardial tissue).

Combined with CT, a PET/CT system has tremendous diagnostic potential, but only if operated and interpreted correctly. A purpose of this thesis is to correlate the published performance parameters based on the NEMA standards for the two scanners to measured human observer performance of lesion detectability. Specific details of the two PET/CT scanners are give in Chapter 4. This chapter presents an overview of PET/CT imaging in §3.1, followed by an introduction to ROC analysis in §3.3. Section 3.2 briefly reviews PET radiopharmaceuticals and interpretation of PET images.

### 3.1 Overview of PET and CT Technology

The ability of PET scans to image tissue or cell function is due to positron emission from a radiopharmaceutical [23]. Section 3.2 describes the radiopharmaceutical  $^{18}\text{F}$ -FDG used in this study. A PET radionuclide is a positron emitter, or gives rise to a daughter decay product that emits a positron. Electrons capture the positrons during annihilation. The detected signal in a PET scan is the two 511 keV annihilation photons traveling at 180 degrees from one another [26].

The body's natural variability in blood flow and uptake of the administered materials for normal physiology is known in advance. PET scanning provides almost direct information about where the radionuclide-bearing compounds are preferentially absorbed or diverted in a patient's body. Variations of uptake in a PET scan from normal physiological uptake can be indicative of disease.

#### 3.1.1 Some Characteristics of PET/CT Scanners

- **Interpretation of PET Images**

The detectability of lesions is significantly improved with tools like maximum intensity projections and fusion of CT slices to obtain anatomical information. Often the lesion is the hottest area of pixels in an image, but not always. In general, the lesions in this study were not made significantly larger than the average fluctuation of pixel intensity in a relatively hot region of the image it is being fused with.

- **Positioning Uncertainties in PET**

PET scanners have inherently poor spatial resolution compared to CT and MRI; the positional uncertainty is very large compared to CT and MRI. For this

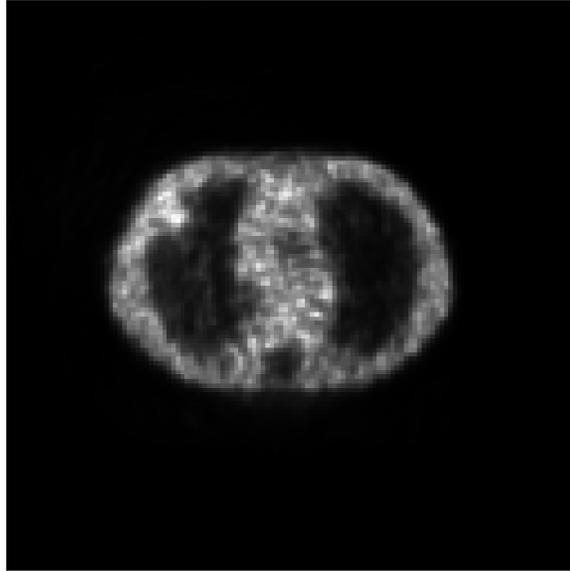


Figure 3.1. PET image of the anthropomorphic torso phantom with a simulated lesion in the right pulmonary wall.

reason, it is a poor modality to obtain anatomical or positional information. The main utility of PET is functional imaging. While efforts are being made to improve the quality (primarily spatial resolution) of PET scanners, there is a limit to the amount of improvement that can be achieved. Several issues govern the fundamental spatial resolution of PET scanners [26].

1. The positron range (average value  $\sim 0.2$  mm for  $^{18}\text{F}$ -FDG )
2. Non-collinearity
3. Stochastic energy deposition (stochastic fluctuations in energy deposition in detectors)
4. Stochastic scintillator photon production
5. Imperfect detector response (inefficiencies on top of fluctuations)

6. Detector readout (scatter and random events)
7. Signal processing (scatter, detector energy resolution, coincidence timing resolution)

For detector energy deposition lutetium oxyortho-silicate (LSO) detectors should provide significant improvement over bismuth germanate (BGO) detectors. The goal is to minimize the energy resolution to better remove scattered photons. A 511 keV photon in matter likely undergoes Compton scattering or photoelectric absorption. Compton scattered photons range in energy from 170 keV to 511 keV with recoil electron energies from 0 to 340 keV.

Range blurring and photon non-collinearity are uncertainties that are a result of positron physics. The path of a positron between emission and annihilation is highly irregular resulting in a distribution of path lengths prior to annihilation.

Photons arising from a positron annihilation event are colinear if the electron-positron annihilation pair has no residual momentum. Residual momentum of the electron-positron annihilation pair, which must be conserved, results in a deviation from colinearity. The uncertainty associated with this effect is minimized by using small separations between detectors. However, the detector separation must be large enough to encompass the majority of patients.

### **3.1.2 Correction of Image Data for Random and Scattered Events**

The three types of coincidence events in a PET scanner are true events, random events and scattered events [27, 28]. Scattered and random events provide lines of response that don't correspond to the true site of positron annihilation. Thus, scattered and random events create blurring and reduce contrast in the image

and should be removed. Removing these events can be done by various means, such as narrowing the coincidence time window and decreasing the width of the energy window. Intrinsic detector limitations allow only so much of these physical parameter manipulations. Additional corrections for scatter are made using models of scattered events based on an assumed distribution. Random events can be corrected using models based on detector count rates or measured using a delayed coincidence method. The problem with all algorithms for random and scatter corrections is that some true events are thrown away with the scattered and random events. These algorithms provide statistical corrections rather than deterministic identification of specific scattered or random events [29].

### 3.1.3 Correction of Image Data for Attenuation

For each annihilation event registered, other events are lost due to the attenuation of the 511 keV photons in the body being imaged. The probability of attenuation is given by a decaying exponential function of the distance the photon must travel. The probability of a photon being attenuated in a homogeneous medium in which the photon must travel a distance  $d$  is given by the equation:  $P = e^{-\mu d}$ . The attenuation coefficient,  $\mu$ , is dependent on energy. For a non-homogeneous medium  $\mu$  will vary with distance, and position, and the attenuation probability for a single photon is:

$$\text{Atten. Prob.} = e^{-\int_0^d \mu(E,l) dl} \quad (3.1)$$

and for two annihilation photons is

$$\text{Total. Prob.} = e^{-\int_0^D \mu(E,l) dl}, \quad (3.2)$$

where  $D$  is the patient diameter. The attenuation coefficient is obtained from a CT scan performed at 120 keV. Using measured values of attenuation for each energy and assumed densities of the tissues in the body being imaged, a map is generated to correct for photon attenuation. This allows for increased intensity to appear in the image in regions of activity where more of the true events are lost. Attenuation correction in a PET image makes interior sources easier to identify. All the images used in this study were attenuation corrected [23, 26].

### 3.1.4 Image Reconstruction

The data collected in the PET scanners will be of the form of a 2D or 3D sinogram. The sinogram is a plot of lines of response (LOR) of true events. For a 2D sinogram, one axis is the distance ( $s$ ) from the center to the LOR and the other axis is the inclination angle ( $\theta$ ) with  $\theta = 0$  defined as a LOR perpendicular to the  $x$ -axis.

Figure 3.2 illustrates the definition of a LOR; the figure also shows an elliptical object within the circular FOV of a scanner. Each LOR is the result of a positron annihilation event producing two 511-keV photons travelling 180 degrees apart. Detecting the pair of 511-keV photons registers an event along the LOR.

Figure 3.3 illustrates the sinogram for the uniform elliptical object shown in Figure 3.2. Reconstruction algorithms take the sinogram data  $g(s, \theta)$  and, through various computational techniques, reconstruct the object  $f(x, y)$ .

- **Basics of Filtered Backprojection**

The Radon transform is the mathematical description of projections by integrals over all directions around an object being imaged with a tomographic

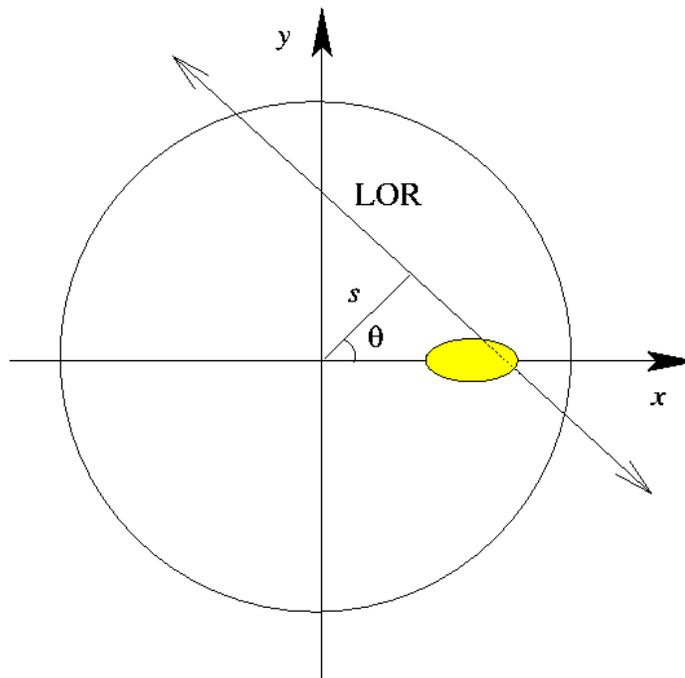


Figure 3.2. Illustration of a line of response (LOR) from PET scan, these form a sinogram used in image reconstruction algorithms.

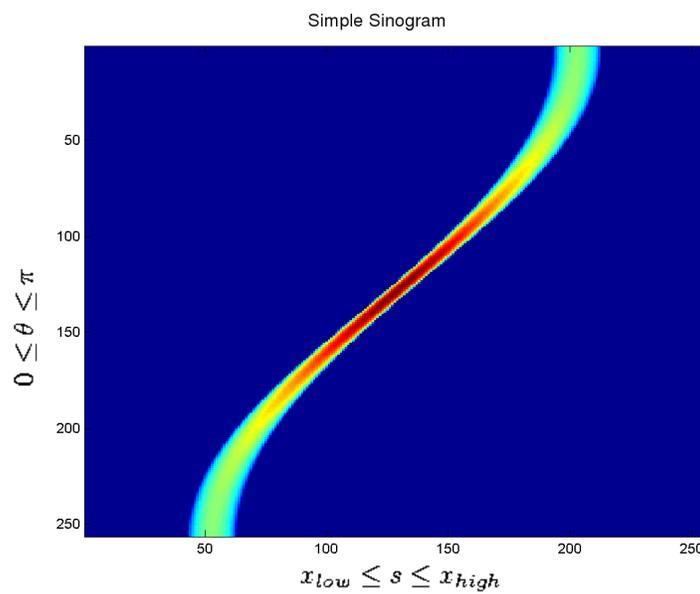


Figure 3.3. The sinogram for the elliptical object shown in Figure 3.2.

imaging device.

$$g(s, \vartheta) = \mathbb{R}f(x, y) = \iint_{-\infty}^{\infty} dx dy f(x, y) \delta(x \cos(\vartheta) + y \sin(\vartheta) - s) \quad (3.3)$$

The image of  $f(x, y)$  can be obtained by an inverse of this operation,

$$f(x, y) = \mathbb{R}^{-1}g(s, \theta) \quad (3.4)$$

This is done in practice by backprojecting the inverse Fourier transform of the Fourier transform of the sinogram multiplied by a ramp filter ( $|\xi|$ ) and an optional smoothing filter ( $H$ ),

$$\tilde{f}(x, y) = B[\mathfrak{F}^{-1}(|\xi|G(\xi, \theta))H] \quad (3.5)$$

$$G(\xi, \theta) = \mathfrak{F}(g(s, \theta)) \quad (3.6)$$

which is a method called filtered backprojection (FBP). The tilde is placed over the function  $f(x, y)$  to point out that what is reconstructed is a representation of the object, not the object itself, and no imaging system is perfect. The ramp filter is used to correct for the blurring of  $f(x, y)$  by the backprojection operator [30].

- **The DIFT Algorithm**

The DIFT (Direct Inversion Fourier Transform) algorithm uses the central slice theorem. In going from the sinogram data to the image, the data must be interpolated in Fourier space onto a rectangular grid [26]. Reconstruction is achieved by inverse Fourier transformation.

- **The OSEM Algorithm**

The ordered subset expectation-maximization (OSEM) algorithm is an iterative 2D or 3D algorithm. For 3D data acquisition, a 2D OSEM algorithm (used

in conjunction with another algorithm to convert 3D to 2D) is a relatively fast method for producing visually appealing PET images [26]. Images in this study specified as OSEM were not 3D acquired data Fourier rebinned and reconstructed with a 2D OSEM algorithm.

- **The 3D FORE Algorithm**

The 3D FORE algorithm is a method of Fourier rebinning. This algorithm converts a 3D sinogram into a set of 2D sinograms. It can then be reconstructed in various different reconstruction algorithms. In this study images that were specified as being reconstructed by 3D FORE were reconstructed by FBP after rebinning.

### **3.2 Radionuclides Used in PET**

There are several common positron-emitting radionuclides used in PET imaging. All of these elements have short half-lives. Therefore a PET imaging center requires a nearby or in-house supply of these isotopes. These isotopes must also be attached to larger molecules using methods of synthetic chemistry to make them useful for functional imaging. Table 3.1 lists common PET radionuclides.

Desirable features are longer half-life (especially if transport is needed), high specific activity (smaller volumes are needed), and smaller maximum positron energy (the smaller the energy, the less range blurring occurs). The only radionuclide used in this thesis was  $^{18}\text{F}$ . Southern Isotopes-Baton Rouge produced the  $^{18}\text{F}$  in a cyclotron. It was then attached to a glucose molecule by replacing a hydroxyl group from the glucose. The radiopharmaceutical produced is called fluorodeoxyglucose or  $^{18}\text{F}$ -FDG. The  $^{18}\text{F}$ -FDG mimics glucose in the body, allowing for the imaging of glucose metabolism. More detailed information of the chemistry

Table 3.1. Radionuclides used in PET imaging.\*

Radionuclides	Half-life (min)	Max. Positron Energy (MeV)	Mean Range in Water (mm)
Carbon-11	20.4	0.960	1.7
Nitrogen-13	10.0	0.491	1.2
Oxygen-15	2.0	1.732	2.7
Fluorine-18	109.8	0.633	1.4

\*Adapted from [31].

of producing  $^{18}\text{F}$ -FDG can be found in reference [32].

Advancements in the last ten years have improved the process of manufacturing  $^{18}\text{F}$ -FDG such that the time from bombardment in the cyclotron to synthesis is less than 40 minutes. This very fast production time for  $^{18}\text{F}$ -FDG has increased the utility of PET imaging significantly.

### 3.2.1 Standard Uptake Values (SUV)

Pathologic staging is broken up into several grades. The higher the grades the more poorly differentiated are the cells and the more cancerous. In general, higher-grade tumors proliferate more quickly and will have larger SUVs [33]. Cancer cells have higher metabolic glucose rates than normal tissue.  $^{18}\text{F}$ -FDG is a positron emitting simple sugar that, in general, will have higher concentrations in cancer cells than in normal tissue. Although the relationship is not linear, as the SUV increases so does the intensity. The SUV in normal tissue is by definition roughly

equal to one.

$$\text{SUV} = \frac{\text{Tracer uptake}}{\text{Administered activity}} \times 1/(\text{patient weight (kg)} \times 1000) \quad (3.7)$$

where tracer uptake is measured in MBq/mL and ‘administered activity’ is in units of MBq.

Physicians are concerned with the maximum SUV in a region of interest (ROI) as it has a better diagnostic value than an average SUV. Lesions are said to be “PET positive” or cancerous for SUVs above 4 in many clinical situations. There are many physiologic reasons for “PET positive” regions that are not cancerous. Therefore, this is not a task that simply can be replaced with computer-aided diagnosis [34, 35].

### 3.3 Overview of ROC Analysis

#### 3.3.1 The Need for ROC Analysis

ROC was first developed as a tool in signal detection theory to address problems with radar image observation [36]. ROC analysis is a branch of statistical decision theory. ROC analysis is used in business, economics, psychology, and many other non-medical fields [37]. Applications in medical imaging began in 1960 [38, 39]. The goal of ROC is to provide an objectively meaningful test for judging the performance and usefulness of a diagnostic imaging system [40]–[44]. This reduces down to a radiologist’s ability to correctly identify pathology based on the images created by the device. To have confidence in standards and equipment, a comparison must be done to show radiologists’ performance for correctly determining diseased and healthy states (see also [45, 46]).

The easiest test to administer and analyze is a straight average, historically

known as “diagnostic accuracy” [47]. This records how many images were correctly classified as diseased or healthy, normalized by the total number of images, giving an average performance. This number can be very high, and yet the performance of the imaging device simultaneously could be very bad. Take, for example, a rare disease in which the prevalence in the image set is 1%. If the radiologist, at a whim, classifies all images as healthy, the radiologist will have a 99% success rating. However, the diagnostic test has been of no use in identifying the diseased state [13, 48]–[50].

The diagnostic test should identify the frequency of false positive and false negative errors. For an initial screening test, a false positive may be far more desirable than false negatives, while on other tests the relative cost of each type of error could be equal. The relative cost of each type of error should be utilized when setting the decision boundary or confidence threshold, which is the criterion that determines (in the observer’s view) whether the image shows health or disease. With a computer model, the decision boundary can be set by parameters like relative intensity, contrast, or volume, giving a fixed classification criterion. The decision boundary can shift with human observers, often leading to a higher incidence of correct classifications with trained observers but potentially fluctuating from session to session even for the same observer. In image classification problems, a variable decision boundary based on each image is generally preferable, which can be achieved by human observers. This is a reason why trained human observers can out-perform computer-aided diagnoses in determining a unique decision boundary for every image [34, 35].

In ROC terminology, “sensitivity” is the true positive fraction (TPF), the fraction of patients actually having the disease who are correctly diagnosed as

“positive.” The ROC term true negative fraction (TNF) is “specificity”, the fraction of patients who actually are without the disease that are correctly diagnosed as “negative.” In the framework of the ROC analysis these two variables are independent [12, 51]. The implication of this is that the fractions of false positives and false negatives can also be measured separately. This is an essential assumption of ROC analysis, making ROC useful for highly prevalent diseases as well as for those of very low prevalence [52].

The relationships between false negative fraction (FNF) and sensitivity and between false positive fraction (FPF) and specificity are written as follows:

$$\text{FNF} = 1 - \text{TPF} = 1 - [\text{sensitivity}] \quad (3.8)$$

and

$$\text{FPF} = 1 - \text{TNF} = 1 - [\text{specificity}] \quad (3.9)$$

When evaluating the performance of a diagnostic system it is more revealing to compare two parameters, e.g., sensitivity and specificity or TPF and FPF, rather than the single parameter of diagnostic accuracy. However, simple comparison of two parameters is still not adequate in evaluating a system’s performance because the source of diagnostic misclassifications is not known: was the misclassification a result of observer error or of the imaging system’s inability to separate states of disease and health? An effective comparison must distinguish between these sources of error [53]–[55].

An individual human observer can move the decision boundary for each image or each image set. This corresponds to a set of pairs of comparisons that can be plotted on a unit square. The typical ROC curve places FPF on the horizontal axis and TPF on the vertical axis.

### 3.3.2 ROC Analysis

The ROC method takes continuous-distributed or discrete ordinal category diagnostic test results and fits them to a binormal distribution. The data typically are a confidence-rated ordinal category diagnostic test. The term binormal<sup>1</sup> refers to two normalized Gaussian distributions such as that in Figure 3.4. The left distribution shown in Figure 3.4 is a normalized Gaussian fitted distribution of actual positive cases. The right distribution is a normalized Gaussian fitted distribution of actual negative cases. The horizontal axis corresponds to a continuous classification of performance by a given diagnostic test. The better the diagnostic test, the larger the separation of the two Gaussians.

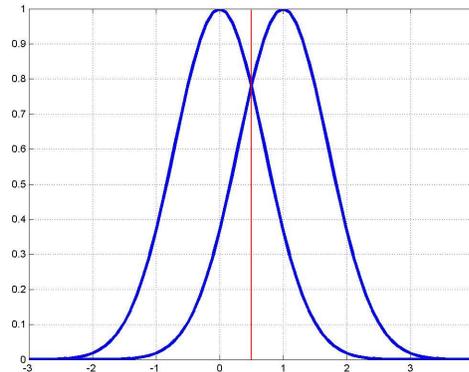


Figure 3.4. An example of a poorly separated binormal (double Gaussian) distribution.

Figure 3.5 shows an example of such a larger separation where the abscissa is an arbitrary test value. The decision boundary is the value that the test uses to

---

<sup>1</sup>In the mathematical literature the term “binormal distribution” usually refers to a 2D bivariate distribution, whereas in the ROC literature (and in this thesis) the word binormal is used for a 1D double Gaussian distribution (a ‘camel hump’ shaped PDF).

classify individual test results as positive or negative [56]. The decision boundary is set by the user's tolerance for an acceptable number of false positive decisions (i.e., actual negatives classified as positive) and an acceptable number of false negatives (i.e., actual positives classified as negative).

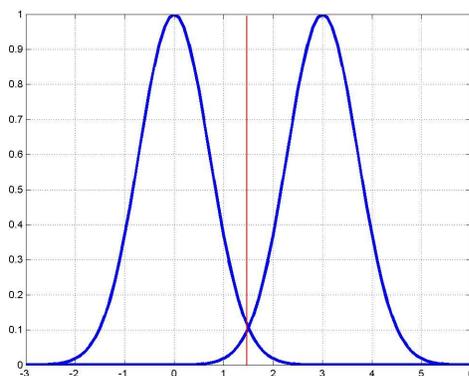
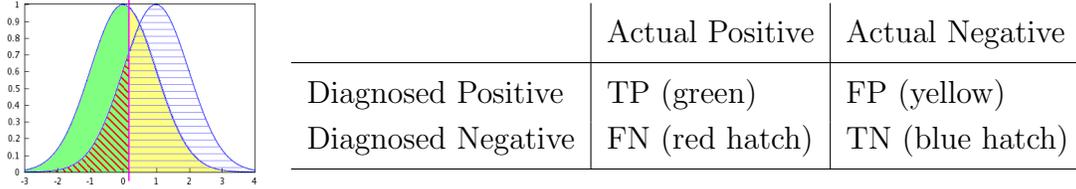


Figure 3.5. An example of a well-separated binormal (double Gaussian) distribution.

Each decision boundary separates the population into four groups shown in Table 3.2 and the accompanying figure insert. The purple line in the figure represents a virtual decision boundary about which observers might be presumed to have split their decisions when forced, whereas the two Gaussians represent the actual populations of images with and without lesions. The essential problem in statistical decision theory is finding the set of variables, the imaging device, the test or the gauge that best separates the two normal distributions of actual positive from actual negative. A poor test produces little or no separation of the two distributions. The best test is one that maximizes the separation, although external issues such as cost-effectiveness, convenience and patient discomfort may factor into determining if the test is useful clinically [57, 58].

Table 3.2. Grouping of diagnostic decisions into two partitions of two sets each. The diagonal table entries partition by truth value, whereas the row entries partition all diagnoses by observer decision.



A fundamental principle of ROC analysis is that movement through decision space can find the test metric(s) that produce the maximum separation and minimum overlap of the binormal distribution. The parameters that describe the binormal fit are  $a$  and  $b$ , where  $a$  is the separation and  $b$  is the relative widths of the two distributions. The ROC analysis attempts to maximize  $a$  and minimize  $b$  (see for example [13, 59, 60]). The binormal parameters specify a unique line on ‘normal-deviate’ axes in which each axis is the stretched-out TPF and FPF over the entire plane such that the origin maps to the center of the ROC unit square. Having the binormal parameters  $a$  and  $b$  is sufficient to generate any size set of pairs of TPF and FPF needed for plotting an ROC curve. Common ways to evaluate the goodness of a curve are the area under curve  $A_z$ , TPF at a given FPF, and FPF at a given TPF. ROC analysis is most useful in comparing differences in curves against a null hypothesis such as an area test, true positive fraction test, or bivariate test.

False positive decisions can prove expensive in both human and hospital resources and, therefore, require additional tests and analysis. Each observer or institution determines its acceptable rate of false positives. The movement of the decision boundary generates a set of indices that are plotted on a unit square with FPF on the horizontal axis and TPF on the vertical axis. This ROC curve represents the tradeoff of sensitivity and specificity attainable from the diagnostic

test. Analysis of the range of decision boundary(s) can identify the operating parameters of a diagnostic system that provide the best possible discriminating power [61, 62].

In practice, one sees a typical ROC curve similar to the curve shown in Figure 3.6. (This curve is intermediate to those shown in Figure 3.7 and Figure 3.8). Figure 3.7 illustrates a test with no discriminating power; there is no evidence from the data that an observer can tell a true positive from a false positive or a true negative from a false negative. The observer is basically guessing, giving the straight-line ROC curve shape. The area under this curve is  $A_z = 0.5$ . In decision space, this corresponds to two normal distributions that completely overlap. There is no clinical benefit of performing such a test.

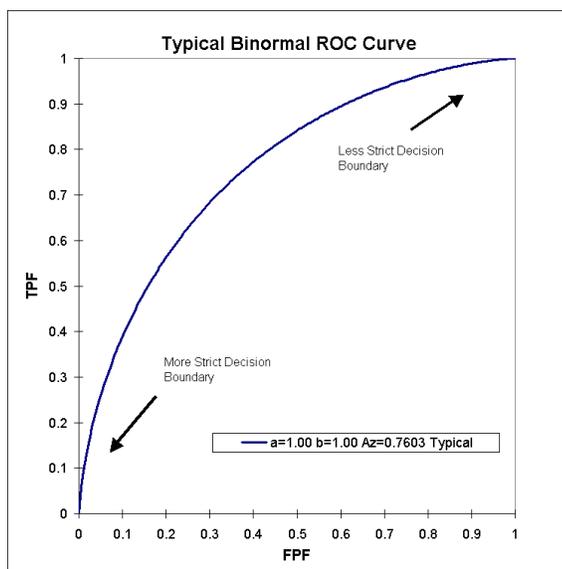


Figure 3.6. An example of the typical form for an ROC curve. The observers are able to occasionally distinguish between samples from the binormal distribution as representative of one or the other Gaussian, but not perfectly, yielding data points that are fitted by the line shown.

An ideal observer in a perfectly discernable test will give an ROC curve with  $A_z = 1.0$ . A perfect ROC curve has the idealized form shown in Figure 3.8. The binormal distribution of actual positives and actual negatives that produces this ROC curve has virtually no overlap.

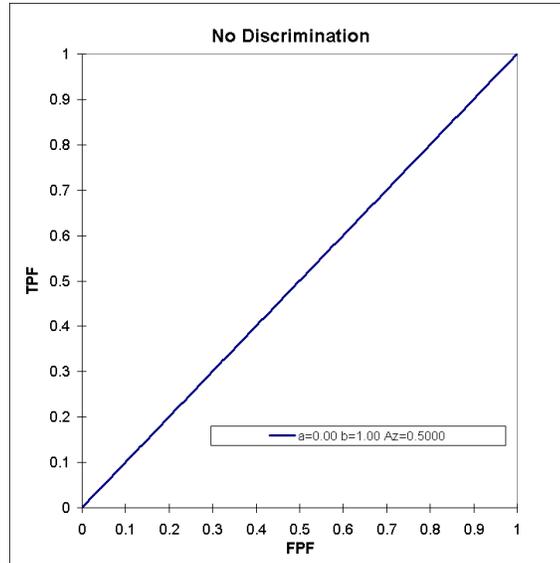


Figure 3.7. An idealized ROC curve obtained from fictional observers with zero discriminatory capacity. Unable to tell positive from negative cases, the observers give equal weight to each type of diagnosis. A blind observer or simple guessing might give one an approximation of this curve.

As a curiosity, a pathological observer produces an ROC curve with  $A_z < 0.5$ ; such a “dyslexic” observer is purposefully giving the wrong answer, inverting their diagnosis with a boolean NOT. The resulting ROC curve will have the form of Figure 3.9. In terms of the binormal distribution, the distribution of actual positives lies to the right of the actual negatives. The decision boundary, although correctly drawn, misclassifies all of the data.

\*

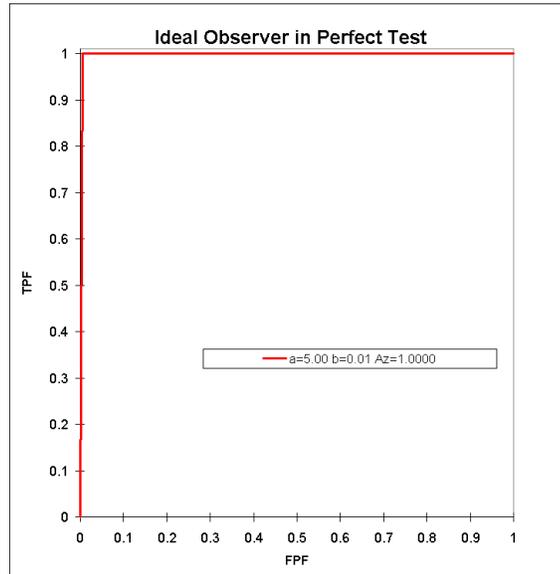


Figure 3.8. The ROC curve for a perfect diagnostic test. To obtain this ROC curve, all observers must be able to flawlessly distinguish between the two Gaussian distributions.

### 3.3.3 Requirement for Diagnostic Truth in ROC Analysis

In this thesis, truth is the presence or absence of the simulated lesions. The classification requirement of the observer comes down to confidence that the simulated lesions are present in the images. Diagnostic truth is a binary definitive answer. Future studies could have more than two states of diagnostic truth. For instance, the observers might be required to classify the image in four states according to size or intensity [63]. Another possibility for future work would be continuous classification of lesions based on size or location [64]–[66].

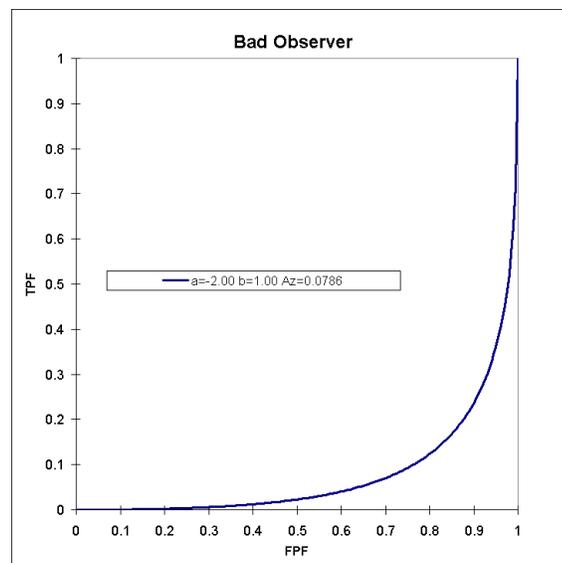


Figure 3.9. The ROC curve for an observer who reverses all positive and negative decisions. The observer reports the positives to be negatives and vice versa.

## CHAPTER 4

### EXPERIMENTAL METHODS AND MATERIALS

This study is confined to experimentation on two clinical PET/CT scanners. Medical imaging technology is perhaps one of the most advanced areas of application of noninvasive imaging of materials. PET/CT is one of the newest developments in medical imaging. Section 4.1 describes the two PET/CT systems that were used to obtain raw data for the ROC study. Section 4.2 details the phantom and spheres that were imaged. Section 4.3 describes data acquisition and preparation. Section 4.4 gives specific details of the ROC analysis used to test the hypotheses of §2.3.

#### 4.1 Medical PET/CT Scanners

Section §3.1.1 reviewed some of the general features of PET/CT scanners. In this section the two specific scanners used for the project are described. Table 4.1 summarizes the main specifications for these two PET/CT scanners used in this thesis research.

##### 4.1.1 The GE Discovery ST

The Mary Bird Perkins Cancer Center (MBPCC) owns a General Electric Discovery ST PET/CT scanner that serves the clinical professionals at MBPCC as well as research faculty and graduate students at LSU. The Discovery ST system was installed in Fall 2003; it has 2D and 3D PET acquisition modes and performs 4-slice helical CT scans. Figure 4.1 is a photo of the scanner.

The 2D PET acquisition mode is achieved by inserting septa between the

Table 4.1. Manufacturers' specifications for the Discovery ST and Reveal HD scanners.

Property	Scanner Properties	
	Reveal HD [16]	Discovery ST [14]
Energy Window Lower Level	350	375
Energy Window Upper Level	650	650
Ring Diameter (cm)	82.7	88.6
Number of Rings	32	24
Detectors per Ring	576	420
Total Number of Detectors	18432	10080
Axial FOV (cm)	15.5	15.7
Transaxial FOV (cm)	58.5	70
Scintillation Crystals	BGO	BGO
Crystal Dimensions (mm)	$4.05 \times 4.39 \times 30$	$6.3 \times 6.3 \times 30$
Number of PMTs	576*	280 <sup>†</sup>
Number of image planes per bed position	63	47
Coincidence Window (ns)	12	11.7

\*Four PMTs per block.

<sup>†</sup>One MA-PMT per block.

detector rings. The 0.8-mm thick and 54-mm long tungsten septa are retracted for 3D acquisition. The PET detector comprises an array of 10,080 individual BGO crystals put into 24 rings with 420 crystals per ring. Each crystal has an axial width of 6.3 mm, transaxial width of 6.3 mm and a length of 30 mm. The axial FOV is 15.7 cm and the transaxial FOV is 70 cm; the physical diameter of the PET detector ring is 88.6 cm. Each bed position produces 47 slices. The crystals are arranged in  $6 \times 6$  blocks coupled to a four-anode square photomultiplier. Additional information along with performance evaluations can be found in the literature [14, 15].

The Discovery ST has two available methods for PET image reconstruction in the 2D mode: filtered backprojection (FBP) and OSEM. In the 3D mode, the available algorithms are 3D reprojection or Fourier rebinning (FORE or 3D FORE)

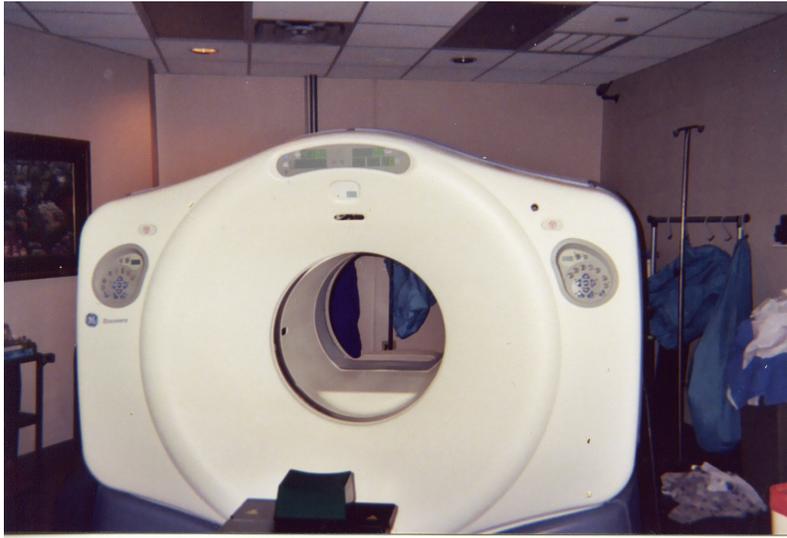


Figure 4.1. The front view of the MBPCC Discovery ST PET/CT scanner.

followed by either filtered backprojection or weighted least-squares OSEM iterative reconstruction.

#### 4.1.2 The CTI-Siemens Reveal HD

Our Lady of the Lake Regional Medical Center, has a Reveal HD PET/CT imager serving both clinical professionals and research personnel. This machine was installed in Spring 2002; it operates in 3D PET acquisition mode and acquires single-slice helical CT scans.

The Reveal HD has an ECAT™ PET gantry. ECAT is a model of a stand alone PET system developed independently from the CT scanner. The PET detector has only a 3D mode of acquisition. It is made of BGO crystals (4.05 mm axial  $\times$  4.39 mm transaxial  $\times$  30 mm radial) arranged in  $8 \times 8$  blocks. Each block has four PMTs. There are a total of 288 blocks, which defines the sinogram dimensions. The

ring diameter is 82.7 cm with an axial FOV of 15.5 cm and a transaxial FOV of 58.5 cm. There are 32 detector rings [16], making a total of 63 image planes per bed position.

The Reveal HD has the same three methods of PET image reconstruction that are present on the 3D Discovery ST. Section §4.1.3 gives more details on the default reconstruction protocols used clinically.

#### **4.1.3 Default Acquisition Protocols and Reconstruction Methods**

The default clinical acquisition protocols for whole-body imaging were used as the base acquisition parameters for the images in this thesis. These base parameters were altered to obtain the image variations investigated here, such as increasing the acquisition time from 4 minutes (default for Discovery ST) to 16 minutes. The raw sinogram data on the scanners is very large and not directly accessible to the user. It was therefore necessary to use the reconstruction algorithms bundled in the included scanner software. For the present research it was only possible to reconstruct the data within the bundled options for each scanner. Within the literature are examples of studies in which a null effect was shown on the effect of image processing on lesion detection for other modalities [67].

The default reconstruction for clinical use of both scanners is OSEM. OSEM in this study was performed with 4 iterations and 8 subsets. With iterative reconstruction, the activity distribution is initially estimated, then the measured projections are compared to these estimated projections. The estimation is then updated and the projection estimate recalculated. OSEM is an example of a maximum-likelihood expectation-maximization algorithm [26].

## 4.2 Laboratory Materials and Software

This section describes the experimental procedures and associated materials for acquiring the PET/CT images for performing the ROC analysis. The actual ROC analysis method is explained in §4.4.

### 4.2.1 Phantom and Sphere (Simulated Lesion) Set

Images for this thesis were acquired of a Data Spectrum Anthropomorphic Torso Phantom™ (Model ECT/TOR/P) with the optional Cardiac Insert™ (Model ECT/CAR/I). The main features of this torso phantom are a large body-shaped cylinder, separate chambers for each lung, a liver chamber, and a spine insert. The lung inserts are filled with Styrofoam beads and water to simulate lung tissue (density of  $\sim 0.3 \text{ g/cm}^3$ ). The size of the torso is consistent with average to large patients ( $38 \text{ cm} \times 26 \text{ cm}$ ). This phantom does not include contours to simulate breasts. The cardiac insert comprises two nested cylinders; the outer cylinder models the left ventricular wall while the inner cylinder represents the volume of the left ventricular chamber. The FDG concentration in each chamber of the phantom can be individually adjusted to simulate different physiologies and acquisition times after injection. Figure 4.2 shows a researcher filling the body-shaped outer chamber of the phantom. Details of the volumes of the individual chambers can be found on the manufacturer's website (<http://www.spect.com>).

The lesions were simulated with a set of spherical shells capable of holding small volumes of radioisotope. Four spheres were used to simulate lesions of different sizes. The spheres are the Data Spectrum Micro Hollow Sphere Set (4)™ (model SCT/MI-HS/ST). The inner diameters of the spheres are 3.95 mm, 4.95 mm,



Figure 4.2. Torso phantom preparation.

6.23 mm, and 7.86 mm; the outer diameters are 6.0 mm, 7.0 mm, 8.0 mm, and 10 mm, while the volumes are 0.031 mL, 0.063 mL, 0.125 mL, and 0.25 mL. The micro-hollow spheres are shown in Figure 4.3.



Figure 4.3. Photo of the micro-hollow spheres used to simulate lesions.

### 4.2.2 Preparation of the Phantom

The phantom is stored dry to inhibit bacterial growth, which might cloud the chambers and require additional cleanup effort. A few days prior to image acquisition the phantom and all of the inner chambers are filled with tap water. Forty-eight hours is sufficient for the water to reach thermal equilibrium and allow the air bubbles to separate and rise to the top. A small volume of air is left in each chamber to allow for thermal expansion and to prevent warping of the phantom.

In the hour or two prior to imaging, the phantom and inner chambers are filled completely with tap water and  $^{18}\text{F}$ -FDG. The radioactivity was assayed in a well chamber; the amount put into each phantom chamber was determined by assaying the syringe before and after dispensing the radioactivity into the chamber. A total activity of approximately 2 mCi was used each time the phantom was filled. The relative concentrations of  $^{18}\text{F}$ -FDG in each chamber were maintained at approximately the same amounts over all acquisition sets. The only exception was that activity was placed in the cardiac insert for some imaging sessions but not others.

The painstaking process of removing the air bubbles during preparation was the most difficult activity. Air bubbles contain no activity and will appear as dark spots in a PET image. We attempted to remove every air bubble in the phantom prior to imaging.

The activity in the phantom was chosen to represent approximately the activity per unit mass in a typical patient. A typical 70 kg patient will receive an injection of 10 mCi about one hour before imaging. The total volume of water in the phantom is 10.6 liters. The phantom's total mass is 10.6 kg plus the mass of the phantom itself. To obtain phantom images with a similar  $^{18}\text{F}$ -FDG activity per unit

mass, the activity was scaled down. Thus an activity in the phantom of 2 mCi ( $=10 \text{ mCi} \times 15 \text{ kg}/70 \text{ kg}$ ) was chosen to make the images similar to those attained clinically [68]. Each image acquisition was decay corrected to the time at which the phantom preparation began.

The process of preparing the phantom for imaging generally took between one and two hours. The preparation was done in the hot lab at MBPCC. All handling of radioactivity followed appropriate procedures including protection by a lead glass shield while assaying the syringe and filling the spheres.

The scans were acquired in the sequence of a scout x-ray image, the CT scan and then the PET scan. The phantom was placed on the imaging table on a protective pad. If the phantom were to leak, the protective pad would prevent the scanner from becoming contaminated with  $^{18}\text{F}$ -FDG. The phantom was centered in the FOV to minimize any effects of asymmetric and degraded resolution at the FOV edges. The scans were performed from the control room, which is safely shielded from the x-rays generated during the CT scan.

The scout scan is a conventional x-ray image of the object. It is used to select the location and number of bed positions that will be acquired in the PET and CT scans. The minimum size of a scan on either the Discovery ST or the Reveal HD is one bed position or about 15 cm. A typical scan of a patient may require 6 or more bed positions, depending on the volume to be imaged.

The CT scan produces an anatomical image; it also provides the attenuation correction factors for the PET images. Neither scanner permits the PET scan to take place without the CT scan. Other than for PET attenuation correction, the CT images were not used further in this study.

Images were obtained with 4 and 16 minute acquisition times on both scanners.

For the Discovery ST scanner, the septa were lowered and raised to obtain the 2D and 3D acquisitions. The spheres were put into the phantom and imaged; they were also imaged separately and added later into the phantom images by computer processing (see §4.3) to create additional image sets.

### **4.3 Preparation of the Images for the ROC Study**

#### **4.3.1 Creating Images that Contain Lesions**

Lesion-containing images of the phantom can be obtained directly by imaging the spheres physically inserted into the phantom. Alternatively, separate images of the spheres and phantom can be processed into combined images. Three methods for image creation were considered: the physical method and two image processing methods.

The method of physical insertion is the best method in many respects. With the spheres imaged in the lung in this way, the images most closely represent actual patient images. No additional image manipulation is needed to prepare the images for the ROC study. Real lesions may be asymmetric, while the simulated lesions are symmetrical which can detract from their realism. The drawback of this method is that many images must be acquired to create a wide range of data for the ROC study. Images in which the spheres were embedded physically in the lung chambers were not manipulated for the study.

Additive image manipulation combines separate images of the spheres and the phantom. The placement and intensity of the lesions is determined randomly. An additive operation on the pixel intensity is used to fuse the two images. A program written in MATLAB was used to implement this method. The advantage of this

method is that with only a few acquired images a wide range of data can be created for the ROC study.

XOR image manipulation is similar to the additive method, but it uses an XOR boolean operation to fuse the two images. If the intensity of a pixel in a sphere's image is greater than the intensity of the phantom's image for that same pixel, the XOR operation inserts the value from the sphere's image; otherwise the phantom image is not altered. Images were generated using all three methods, although only images from the first two categories were used finally in the ROC analysis. This is because there is no easily discernible difference between the images obtained from the additive method and the XOR method and the XOR method required additional processing steps.

#### **4.3.2 Design of Image Presentation and Confidence Survey**

The images consisted of roughly 50% actual positive cases and 50% actual negative cases. The positive cases consisted of two classes of images. The first class were images in which the spheres were directly inserted into the lungs and imaged without any artificial manipulation of the images. The second class of positive images was created by the additive method described in §4.3.1. Several operations were done to the simulated lesions before fusion. To simulate a wide range of pathologic tumor grades or standard uptake values (SUVs), the intensity of the lesions was randomly scaled. The lesions, without being scaled down, were far too intense for the torso phantom image and would have made the detection task trivial.

The second operation that was done to the lesions was a rotation followed by a translation. The placement of lesions should be different in each positive image to increase the difficulty in identifying the lesions. If the lesions were placed in the

same region or spot repeatedly, the task of detection would be trivial. The operations of rotation and translation were also performed using a random number generator. Although some areas are more likely to be cancerous than others, lesions can be found almost anywhere within the torso. This conclusion was reached after an initial conversation with Dr. Bujenovic at OLOLRMC.

Positive images from both methods of inserting lesions were compared by several graduate students. It appeared that it is not possible to distinguish the method of creation from simple inspection of most of the images. Although an algorithm was written to put in Gaussian blurring with the additive method to account for the spheres being imaged in air compared to water, it was not used. Due to the increased scatter in water, it was expected that the spheres would be less localized and more blurred, but the effect was not evident in the physical images and apparently less substantial than partial volume effect.

In the images, special attention was given to placing the lesions at least a few centimeters from the Teflon spine insert. The attenuation correction factors for Teflon are known to have errors on the order of 25% or larger [69]. Teflon is a poor substitute for bone at 511 keV. The high CT number is incorrectly treated as bone leading to a linear attenuation coefficient significantly lower than the actual value for 511 keV photons. This will lead to under correction in the vicinity of the Teflon. The same activity near the Teflon will appear less intense than if far away from the Teflon. For the other materials in the phantom, the attenuation correction accuracy is on the order of 3% [14].

The confidence survey was designed to maximize the value of each viewer's responses while minimizing the amount of time the viewer must spend on the image set [13, 70]–[73]. The complete set of images consisted of 120 images, 61 actual

positive images and 59 actual negative images. The observers were given unlimited time to complete the survey. A three-page training document was given to observers prior to the start of their survey. The training document was short, to minimize the time spent on it. The survey training document is given in Appendix B.

The training set consisted of twelve images: five actual negative images, five actual positive images, and two images of micro hollow spheres in air. A series of three images were shown that illustrated the additive method for creating positive images. The process of showing how the simulated lesions were inserted was meant to illustrate to the observers what they were looking for in the images and to illustrate that the lesions could be put anywhere and with any intensity. With the three-page introduction, a verbal explanation was given explaining the survey's motivation and disclaiming that the ROC study was not testing the observers but evaluating the imaging systems. Observers were told that the overall performance of the observer was not of interest, but rather the relative performance of the two systems, of the two acquisition modes, or of the reconstruction algorithms. Finally, the verbal and written instructions included brief explanations of PET imaging technology and the utility of PET imaging in viewing lesions with higher glucose metabolism than normal tissue.

The three anatomical regions present in the torso phantom images were transverse slices containing lung, heart, and liver. Not all three regions were present in every image. To demonstrate the appearance of each of these regions, the training set included one positive and one negative image for each of these three regions. The training set was presented both on paper and on the computer screen. All observers completing the survey used the same laptop computer, running a MATLAB routine for display.

The images were displayed as 120 separate windows making it possible for the observers to put an image aside, or minimize or maximize it individually. Each image is initially displayed at one quarter of the screen size. This size was chosen to make the images as clear as reasonably possible without making pixellation overly apparent.

#### **4.4 Method of ROC Analysis**

The principles of ROC analysis were provided in §3.3. This section gives specific details about the ROC analysis procedures used for this thesis.

##### **4.4.1 Recording and Management of the ROC Database**

A number of steps were taken to methodically record and analyze data for the ROC study. An outline of the procedure in simple stages is

1. Identifying observers
  - (a) Identifying trained observers (radiologists and oncologists)
  - (b) Identifying untrained observers (graduate students, post docs and professors)
2. Recording confidences in observers' abilities to identify lesions
3. Building database for analysis of results

##### **4.4.2 ROC Analysis Details**

This section adds to the general description of ROC methods given in §3.3 of Chapter 3. The ROC analysis was performed using ROCKIT software developed at

The University of Chicago. Beta Version 0.9B is available for free download from The University of Chicago website. The software is a 32-bit application and requires 2 MB of disk space and 2.1 MB of RAM.

The ROCKIT software takes continuous-distributed or discrete ordinal category diagnostic test results and fits them to a binormal distribution. The data for this work are a confidence-rated ordinal category diagnostic test. The data was input into ROCKIT in two ways; to obtain a maximum-likelihood estimate of the parameters for a conventional binormal model and to calculate the statistical significance of the difference between two ROC curves in an area test. The area test is a univariate z-score test of the difference between the  $A_z$  values that tests the null hypothesis. A  $P$ -value is generated that estimates the probability that the difference in  $A_z$  values from the two ROC curves is not due to random data fluctuations.

A sample ROCKIT program run is provided in Appendix A. It describes the typical commands that might be entered for a ROCKIT session, as well as sample output tables. The two options of entering the data are matrix format and list format. For this thesis all the data were entered using the matrix format. This required compilation of the responses in each category through Access database queries.

The ROCKIT software uses an iterative algorithm to calculate the binormal parameters  $a$  and  $b$  as well as parameter uncertainties and 95% confidence intervals. For this study the number of iterations needed for convergence to 3 significant digits was typically 4 iterations. Having the binormal parameters is sufficient to generate any size set of pairs of TPF and FPF needed for plotting an ROC curve.

Note: The ROC software will not allow idealized inputs and will fail to produce values for  $a$ ,  $b$ , and  $A_z$ . Idealized inputs correspond to a perfect test in which

everything is correctly classified (i.e., the ROC curve in Figure 3.8).

The ROCKIT authors created an Excel macro to produce an ROC curve from the binormal parameters from the ROCKIT session. The macro only requires two inputs, the binormal fitting parameters  $(a, b)$ . The statistical uncertainties obtained from the ROCKIT software are not reflected on the ROC curve.

#### 4.4.3 Requirements of ROC Analysis

For a valid ROC study, the diagnostic truth must be known, which means one must either use simulated pathologies or data taken from patients that have *a posteriori* been confirmed to be actually positive or actually negative with certainty (did or did not ever manifest the associated suspected illness). With patient data, diagnostic truth becomes a very important problem. What criterion should be used to determine diagnostic truth and to avoid the possibility of no clear classification of an image? These are two questions easily avoided by using simulated lesions in a phantom.

In this thesis, truth is dependent on the presence or absence of the simulated lesions. The classification requirement of the observer comes down to confidence that the simulated lesions are present in the images. Diagnostic truth is a binary definitive answer. Future studies could have more than two states of diagnostic truth. For instance, the observers might be required to classify the image in four states according to size or intensity [63]. Another possibility for future work would be continuous classification of lesions based on size or location [64]–[66].

For example, staging of some cancers is commonly done using the TNM system published by the American Joint Committee on Cancer (AJCC) [74]. The stage of a cancer is broken into three areas; ‘T’ for the primary tumor, ‘N’ for lymph node

involvement, and ‘M’ for metastatic tumors. The primary tumor classification, which is often related to size but can also be related to the extent it has penetrated tissue boundaries, is generally broken up into four states. Knowing the precise volume of the spheres, one could ask the observers to stage each lesion. This would have been much more time consuming for the observers and would have required a significant reduction in the number of images for the required time to remain the same for each observer.

In the comparison of the two PET/CT systems, it is important to have a large sample set with enough images so that differences in performance can be attributed to the imaging systems as opposed to sampling errors [75]–[81]. The required sample size for an ROC study can be case dependent, but general guidelines are given in the literature [57]. The type of study in this thesis is a multiple-reader, multiple-case (MRMC) ROC study [82]. There are three models for estimating the required sample size, outlined by Obuchowski *et al.* [75]. The basic principles of sampling are that with an unbiased sample, the larger the number of data points the smaller the uncertainty or the 95% confidence level can become. The more narrow the 95% confidence level, the better the ROC test can validate or rebuke a given hypothesis. The expected difference in  $A_z$  values dictates the sample size required. The smaller the expected difference in  $A_z$  values the larger the required sample size.

## CHAPTER 5

### RESULTS OF ROC ANALYSIS

In §5.1 the results of the ROC analysis are presented for all observers combined, including most of the generated ROC data curve fits. Section 5.2 goes into more detail about the observed differences between trained and untrained observer performance. Section 5.3 discusses the results for diagnostic accuracy, sensitivity and specificity.

#### 5.1 Overall ROC Analysis Results and Comparisons

A single parameter often used to summarize the performance of observers in a diagnostic test is the area,  $A_z$ , under the ROC curve. When comparing two systems, acquisition modes, sets of observers, reconstruction algorithms, or acquisition times, the statistical significance of different  $A_z$  values is sought. There are several ways to measure statistical significance. Obviously, translating statistical significance into clinical significance requires input and evaluation by physicians, administrators, and physicists.

The results for the binormal parameters,  $a$  and  $b$ , and  $A_z$  for all observers, which have been broken down by the categories of scanner, different acquisition mode, and reconstruction algorithm, are presented in Table 5.1. These results and the corresponding ROC curves are discussed in §5.1.1 through §5.1.4.

##### 5.1.1 Overall ROC Results for Trained and Untrained Observers

Figure 5.1 shows a composite ROC curve for all observers and all parameters lumped together. It also shows separate curves with all parameters lumped together

Table 5.1. Table of combined overall  $A_z$  results from ROC curves, tabulating the areas,  $A_z$ , under given ROC curves.

	$a \pm \sigma_a$	$b \pm \sigma_b$	$A_z \pm \sigma_{A_z}$
Overall	1.19±0.60	1.09±0.58	0.79±0.01
GE	1.05±0.07	1.08±0.07	0.76±0.01
CTI	1.57±0.13	1.06±0.11	0.86±0.02
2D GE	0.92±0.09	0.96±0.08	0.75±0.02
3D GE	1.24±0.11	1.23±0.11	0.78±0.02
ALL 3D	1.39±0.08	1.16±0.08	0.82±0.01
4 min	1.02±0.09	1.03±0.09	0.76±0.02
16 min	1.30±0.08	1.10±0.07	0.81±0.01
OSEM	1.05±0.07	0.93±0.07	0.78±0.02
3D FOR	1.24±0.11	1.23±0.11	0.78±0.02
DIFT	2.14±0.31	1.66±0.31	0.87±0.03

for the trained and untrained observers; these separate curves are discussed in detail in §5.2. The all-observer ROC curve is biased towards the curve for untrained observers because of the larger number ( $N = 12$ ) of untrained observers than trained observers ( $N = 5$ ). Observers are classified as trained if they routinely make diagnostic decisions based on medical images as part of their profession.

### 5.1.2 Results for All Observers Comparing Scanner Acquisition Modes

The combined observer performance is superior for the Reveal HD system compared to the 2D or 3D acquisition modes of the Discovery ST scanner in this study (Figure 5.2). Surprisingly, the  $A_z$  value ( $0.783 \pm 0.018$ ) for the 3D acquisition mode of the Discovery ST scanner is larger than  $A_z$  ( $0.749 \pm 0.020$ ) for the 2D Discovery ST acquisition mode. This contradicts our initial hypothesis 1 of Chapter 2. Recall that the reasoning for hypothesis 1 was

“The Discovery ST 2D mode of acquisition is used preferentially for

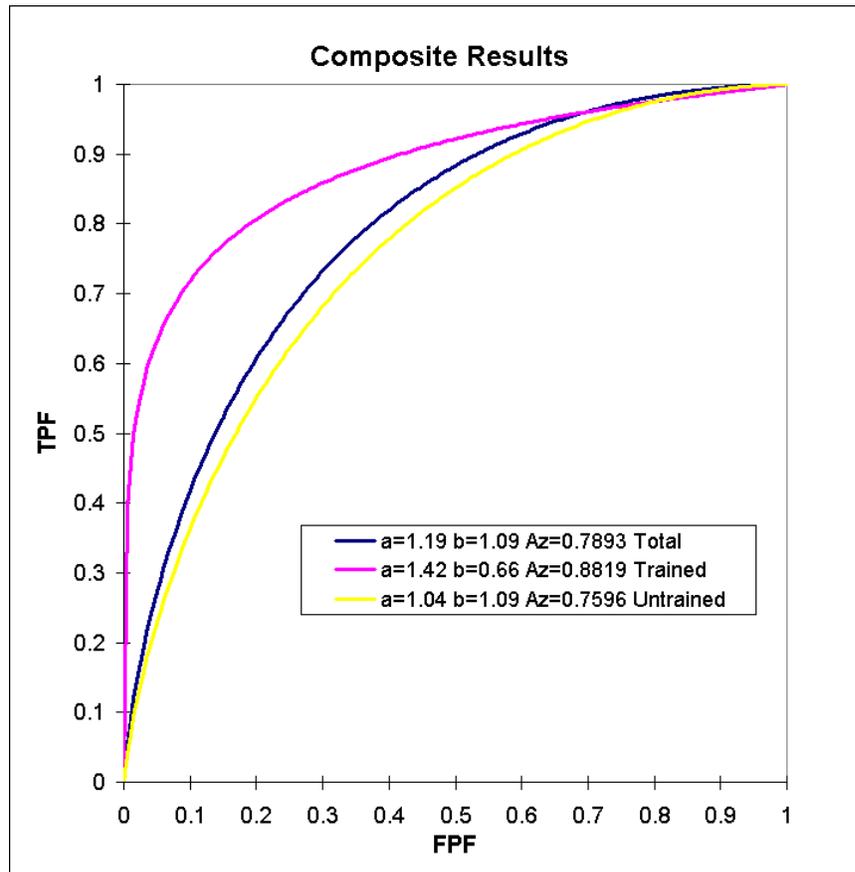


Figure 5.1. Composite ROC curves combining all scanner parameters for each observer class.

whole-body imaging at MBPCC over the 3D mode. There are few clinical protocols in which the images at MBPCC are taken in 3D mode. An initial survey of patient images created with each mode from the Discovery ST by an OLOLRMC radiologist and an LSU physicist indicated that the 2D mode was generally preferable.”

One must note that the default acquisition parameters and reconstruction algorithms were used for the initial survey.

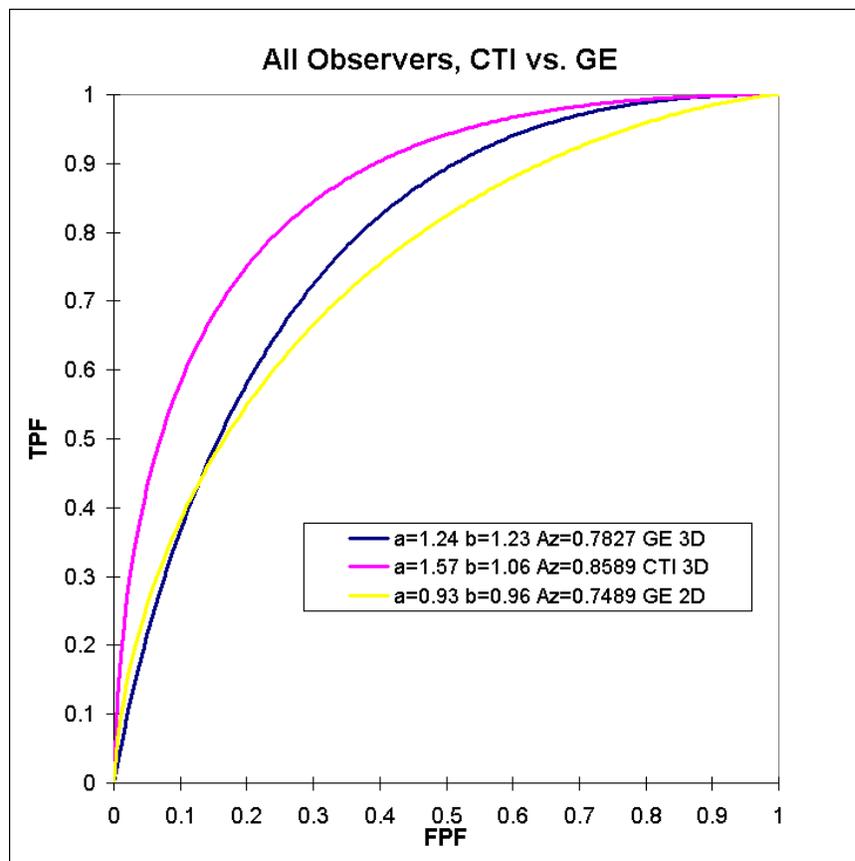


Figure 5.2. ROC curves comparing the two PET/CT scanners, summarized over all observers; all other parameters are lumped together.

### 5.1.3 Results for All Observers Comparing Reconstruction Algorithms

Some of the sinogram data were reconstructed with more than one algorithm, but most of the image data used in this study were reconstructed with only one algorithm. Data from the 3D acquisition mode were reconstructed by a 3D algorithm, and from the 2D acquisition mode data were reconstructed by a 2D algorithm. Alternatively, one could have reconstructed all data with the same reconstruction algorithm, such as FBP, to facilitate comparison. However, the 3D data would still have required some sort of pre-reconstruction processing or rebinning, so the comparison is still not exact. Therefore, we chose to compare the default clinical reconstruction algorithms to maintain similarity to actual patient imaging situations.

Figure 5.3 shows ROC curves comparing three reconstruction algorithms. The comparison suggests the superiority of the DIFT algorithm compared to the other two algorithms. Figure 5.4 shows images from the Reveal HD reconstructed by each method. The image at the top left is much clearer and is a result of the iterative reconstruction algorithm. The image at top right is reconstructed with the DIFT algorithm. The image on the bottom, exhibiting the most reconstruction artifacts, is reconstructed with FBP. The DIFT and FBP images exhibit more reconstruction artifacts than the OSEM image exhibits. The images using iterative reconstruction are visually preferable to those using DIFT or FBP, but this seems not to translate to superior lesion detection ability for observers.

Ideally, this analysis would have been done with the same images on the same system. The results of this ROC curve require further analysis, subgrouping the data into categories that can better isolate the parameter being considered.

Recall that the images are all scaled to a relative intensity. The single highest

activity voxel in the entire volume will have an intensity of  $2^{15} = 32768$ . The streakiness of the DIFT and FBP images largely results from a finite number of projections in the PET data. Clinically, the background region outside the object may be marked or windowed to minimize the appearance of the streakiness.

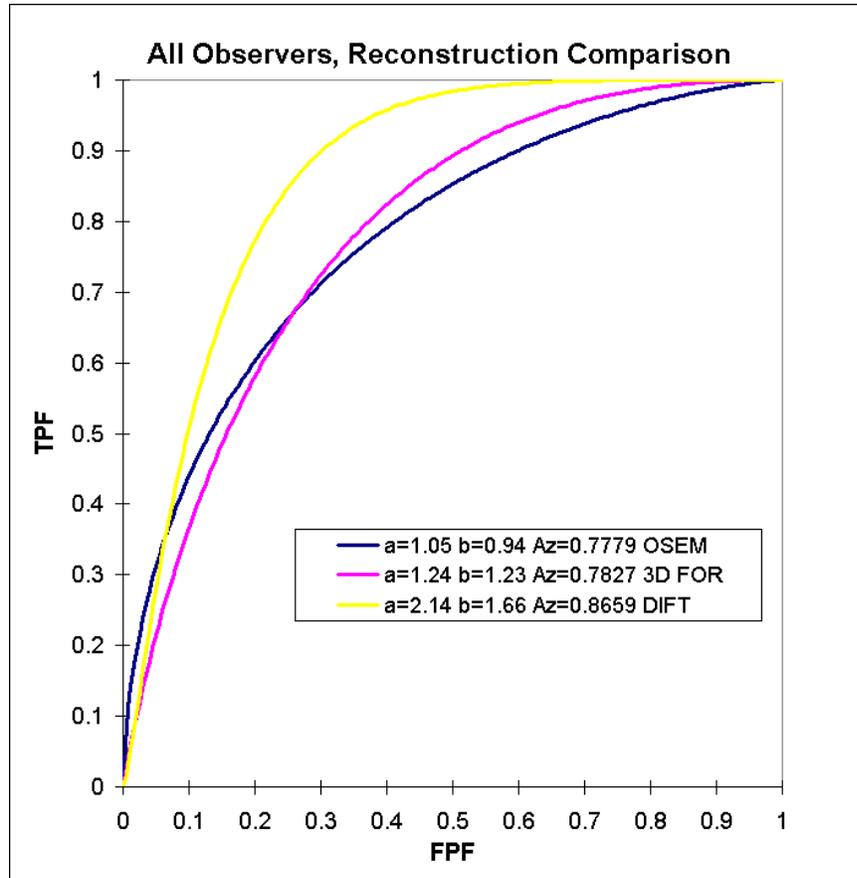


Figure 5.3. Comparison of ROC results for three different image reconstruction algorithms.

#### 5.1.4 ROC Results for All Observers Comparing Acquisition Times

Figure 5.5 is an ROC curve displaying lumped results of all scanners and all acquisition modes for 4-minute and 16-minute acquisitions. For each 4-minute

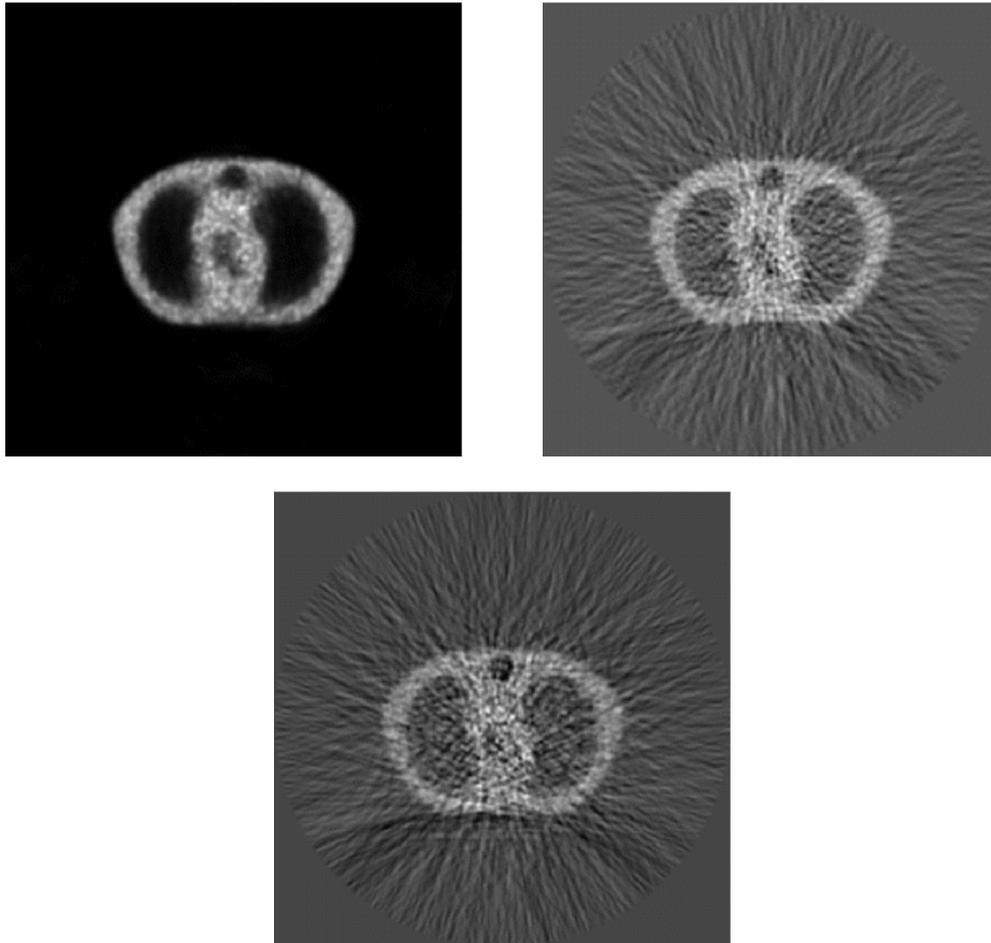


Figure 5.4. Torso phantom images comparing three reconstruction algorithms. Clockwise from top left are images reconstructed with OSEM, DIFT and FBP. The torso phantom was in the prone orientation for this acquisition set.

acquisition, a 16-minute scan was acquired also with all other scan parameters held fixed. The primary difference in the two images is that the number of counts in the 16-minute acquisition should be roughly four times as large. The standard error in pixel intensity due to counting statistics should be reduced by a factor of two.

One would expect the improved statistics to correspond to significantly better diagnostic image quality exhibited by much larger  $A_z$  values. The 16-minute acquisition's  $A_z$  ( $0.808 \pm 0.013$ ) is only slightly higher than the 4-minute acquisition's  $A_z$  ( $0.762 \pm 0.019$ ). The consequences of increased patient discomfort, increased patient motion, and decreased patient throughput probably greatly outweigh the diagnostic benefit of long acquisition times.

## 5.2 ROC Results for Trained Observers

### 5.2.1 Differences Between Trained and Untrained Observers

Figure 5.1 showed ROC curves with all parameters lumped together (trained observers as well as untrained observers). As noted, the all-observer ROC curve is biased towards the curve for untrained observers because of the larger number ( $N = 12$ ) of untrained observers than trained observers ( $N = 5$ ). Observers are classified as trained if they routinely make diagnostic decisions based on medical images as part of their profession.

With a low threshold for false positives (e.g., FPF=0.05), trained observers exhibit a much higher TP rate. The untrained observers only correctly identify 20% of the TPs at this level, while the trained observers correctly identify 65%. It is immediately apparent that the trained observers out-perform the untrained observers in the study of lesion detection in the torso phantom. The ROC curves in

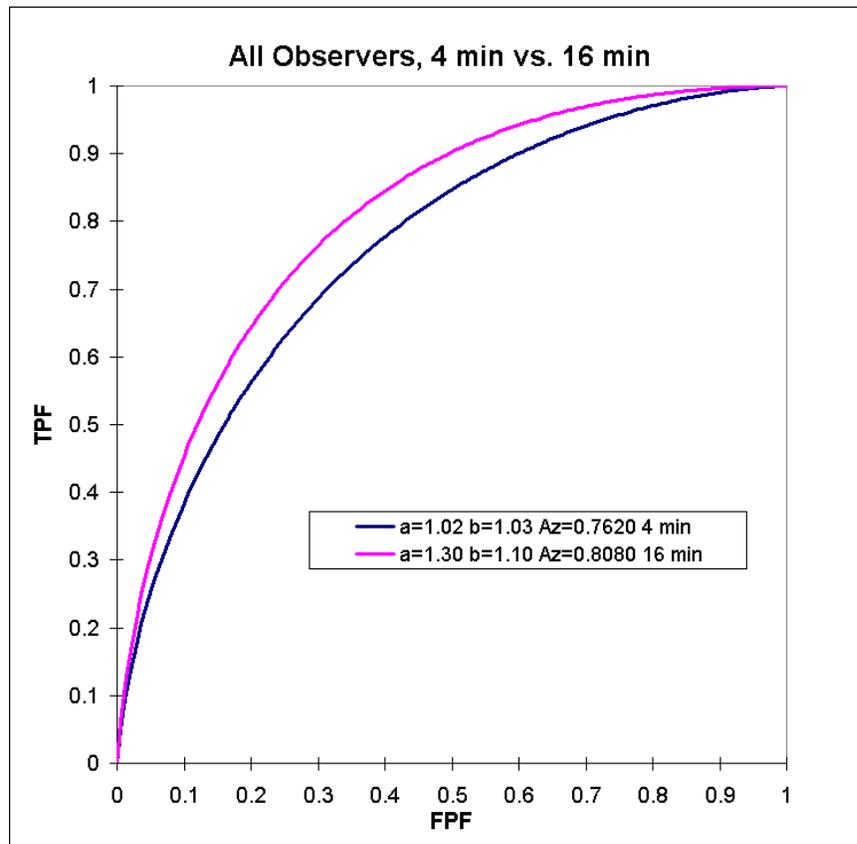


Figure 5.5. Comparison of ROC results for different image acquisition times. All other parameters are lumped together.

Figure 5.1 clearly indicate superior performance in this test is highly correlated to experience with reading PET images. The trend is further supported by examination of individual scores of trained observers. The dedicated PET nuclear medicine physician in this study outperformed the other trained observers.

The comparison of scanners is broken down for trained and untrained observers in Table 5.2. Figure 5.6 is a bar chart visualizing the results in Table 5.2.  $A_z$  values are charted as triplets of {Total, Trained, Untrained} in the bar chart. In all categories, the trained observers outperformed the untrained observers. In most categories, this difference ranged from 8.7% to 15.4%. The clinically-unused Discovery ST 3D acquisition mode produced a difference of 11.6% while the DIFT reconstruction mode produced a very large difference of 18.7%.

Table 5.2 summarizes the differences between trained and untrained observers for each comparison category. Figure 5.7 displays the percent differences visually. Some interesting trends are apparent in Table 5.2 and Figure 5.7. First, comparing acquisition length, the trained observers outperformed the untrained observers by a greater margin for the longer acquisition time than for the shorter acquisition time. This is probably a result of increased clarity of anatomical structures, which may preferentially decrease false positives for the trained observers compared to the untrained observers. The trained observers may be able to pickup additional anatomical clues of which the untrained observers are unaware. Second, when comparing the reconstruction algorithms, the effect of training is more apparent with the reconstruction methods that give apparently poor image quality. Our interpretation is that trained observers are able to read around the reconstruction artifacts for DIFT and 3D FORE, while the apparent benefit of the OSEM algorithm is that clearer images only increase the confidence of the untrained

Table 5.2. Comparison of  $A_z$  results for trained versus untrained observers.

Category	Group	$a \pm \sigma_a$	$b \pm \sigma_b$	$A_z \pm \sigma_{A_z}$
Overall	Trained	$1.41 \pm 0.12$	$0.65 \pm 0.14$	$0.88 \pm 0.02$
	Untrained	$1.04 \pm 0.07$	$1.08 \pm 0.07$	$0.76 \pm 0.01$
GE	Trained	$1.27 \pm 0.13$	$0.61 \pm 0.16$	$0.86 \pm 0.02$
	Untrained	$0.93 \pm 0.08$	$1.10 \pm 0.07$	$0.73 \pm 0.02$
CTI	Trained	$1.79 \pm 0.27$	$0.77 \pm 0.28$	$0.92 \pm 0.02$
	Untrained	$1.36 \pm 0.14$	$0.93 \pm 0.10$	$0.84 \pm 0.02$
2D GE	Trained	$1.16 \pm 0.17$	$0.55 \pm 0.20$	$0.85 \pm 0.03$
	Untrained	$0.80 \pm 0.10$	$0.95 \pm 0.09$	$0.72 \pm 0.02$
3D GE	Trained	$1.40 \pm 0.20$	$0.68 \pm 0.24$	$0.88 \pm 0.03$
	Untrained	$1.13 \pm 0.13$	$1.26 \pm 0.12$	$0.76 \pm 0.02$
ALL 3D	Trained	$1.57 \pm 0.16$	$0.72 \pm 0.18$	$0.90 \pm 0.02$
	Untrained	$1.24 \pm 0.09$	$1.14 \pm 0.08$	$0.79 \pm 0.02$
4 Min	Trained	$1.15 \pm 0.16$	$0.51 \pm 0.18$	$0.85 \pm 0.04$
	Untrained	$1.02 \pm 0.11$	$1.05 \pm 0.10$	$0.76 \pm 0.02$
16 Min	Trained	$1.59 \pm 0.18$	$0.71 \pm 0.19$	$0.90 \pm 0.02$
	Untrained	$1.07 \pm 0.09$	$1.10 \pm 0.08$	$0.76 \pm 0.02$
OSEM	Trained	$1.16 \pm 0.14$	$0.56 \pm 0.16$	$0.84 \pm 0.03$
	Untrained	$1.00 \pm 0.09$	$0.95 \pm 0.07$	$0.77 \pm 0.02$
3D FORE	Trained	$1.40 \pm 0.20$	$0.68 \pm 0.24$	$0.88 \pm 0.029$
	Untrained	$1.13 \pm 0.13$	$1.26 \pm 0.12$	$0.76 \pm 0.02$
DIFT	Trained	$2.97 \pm 0.95$	$1.55 \pm 0.99$	$0.95 \pm 0.037$
	Untrained	$1.26 \pm 0.28$	$1.39 \pm 0.25$	$0.77 \pm 0.04$

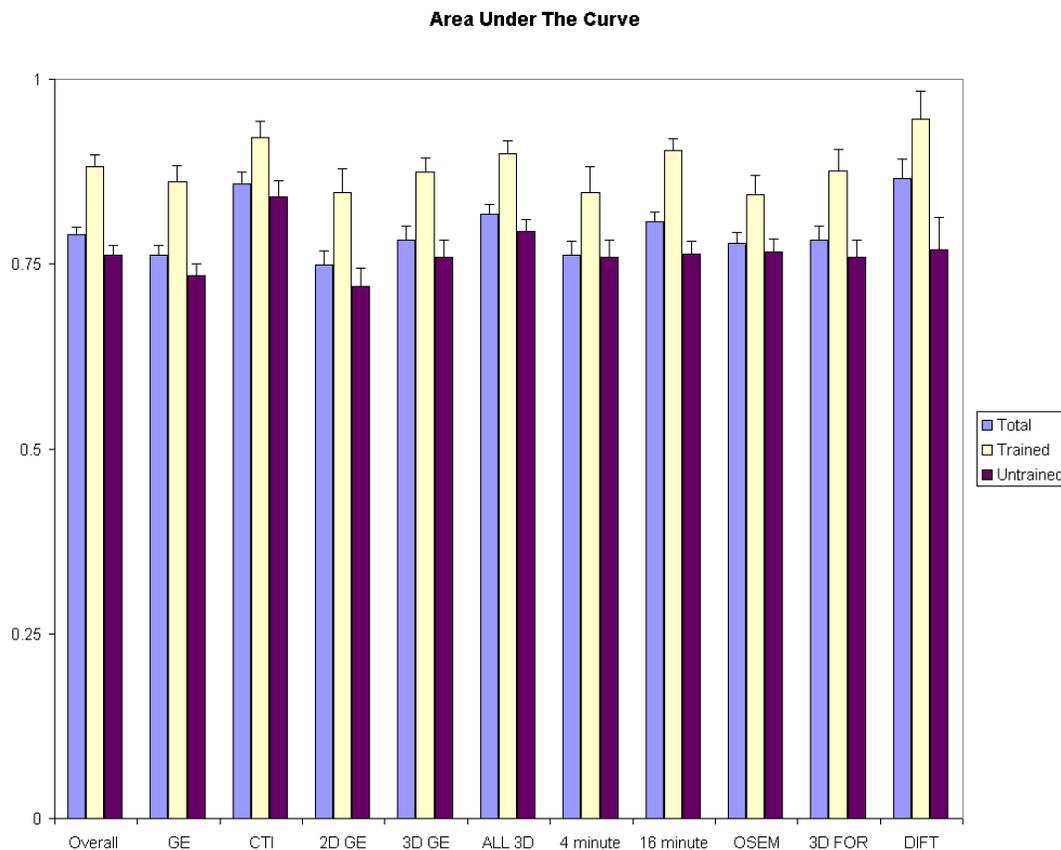


Figure 5.6. Summary of  $A_z$  values grouped by comparison parameter for all observers, trained observers and untrained observers.

observers as artifacts are removed.

While the untrained observer results are interesting, the results for the trained observers should be the guide for decisions based on ROC analysis. Figure 5.8 shows comparative performance for the trained observer for the different reconstruction algorithms. As noted with the performance for all observers, the DIFT algorithm produces a larger  $A_z$  value than the other algorithms. Although OSEM produces visually nicer images, it only outperforms DIFT at extremely low false positive rates.

Although the visual clarity of the OSEM algorithm images is superior to DIFT

Table 5.3. Summary of  $A_z$  values and relative differences.

Category	Trained	Untrained	% Difference*
Overall	<b>0.882</b>	0.762	13.6
GE	<b>0.861</b>	0.734	14.8
CTI	<b>0.921</b>	0.841	8.7
2D GE	<b>0.846</b>	0.720	14.9
3D GE	<b>0.875</b>	0.759	3.0
ALL 3D	<b>0.899</b>	0.794	11.7
4 min	<b>0.847</b>	0.759	10.4
16 min	<b>0.903</b>	0.764	15.4
OSEM	<b>0.844</b>	0.766	9.24
3D FOR	<b>0.875</b>	0.759	13.26
DIFT	<b>0.946</b>	0.769	18.7

$$*\% \text{ Diff.} = |A_z(\text{trained}) - A_z(\text{untrained})| / A_z(\text{trained}) \times 100.$$

and FBP, this does not translate into superior lesion detectability. This surprising result suggests the possible loss of information in this type of reconstruction.

The areas under the ROC curves for the trained observers were  $A_z = 0.844 \pm 0.027$  for OSEM,  $A_z = 0.876 \pm 0.029$  for 3D FORE, and  $A_z = 0.946 \pm 0.037$  for DIFT. The 3D FORE technique is a method of estimating 2D transaxial sinograms from the 3D data acquisition using Fourier rebinning which is then followed by FBP. The DIFT algorithm is a direct 3D inverse Fourier transform.

The results indicate that the ability of trained observers to identify lesions in these attenuated corrected PET images is inversely proportional to the sophistication and time of the reconstruction algorithm. This is in contrast to the untrained observers who have statistically identical performance across algorithms.

The results of trained observer performance for the different scanner acquisition modes are shown in Figure 5.9. The default operating mode for the Discovery ST scanner at MBPCC is 2D although 3D mode is available. The Reveal

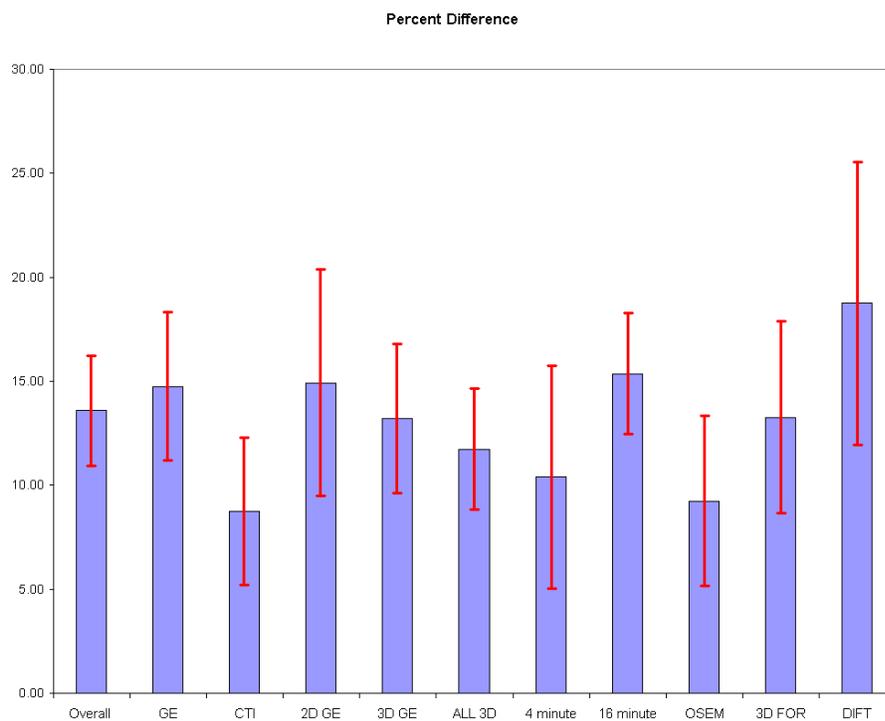


Figure 5.7. Chart of percentage difference in ROC curve  $A_z$  values for trained versus untrained observers.

HD scanner only has a 3D operating mode. Comparing the Reveal HD to the Discovery ST 2D mode, one sees that the Reveal HD outperforms the Discovery ST 2D mode. Interestingly, the nuclear medicine and x-ray technologists who work with both machines prefer working with the Discovery ST. This suggests that while the software and interface might be superior on the Discovery ST scanner, the diagnostic image quality is not.

The default reconstruction algorithms were used for this comparison. Thus, the difference reflects both 2D vs. 3D, but also different reconstruction algorithms; these two issues cannot be separated in this ROC study. For a refined study, all images should be reconstructed in the same basic way. A 2D filtered backprojection

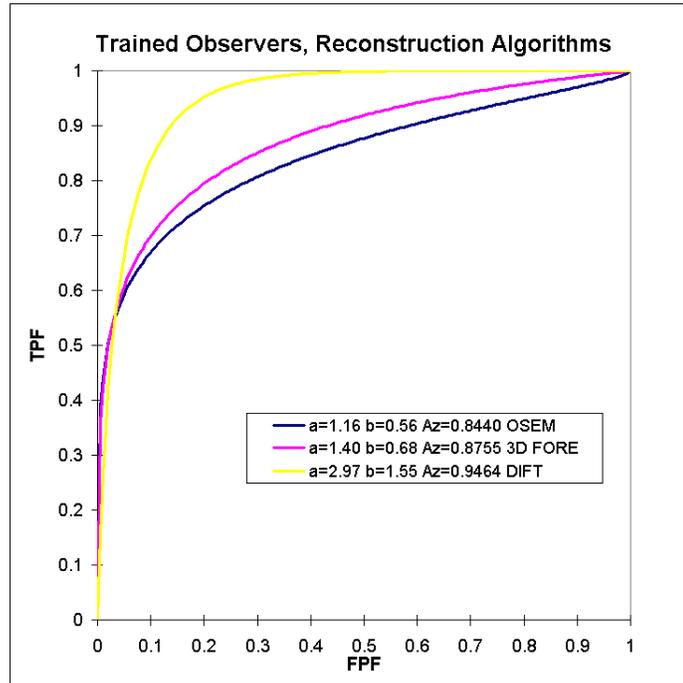


Figure 5.8. ROC curve for trained observers with different reconstruction algorithms.

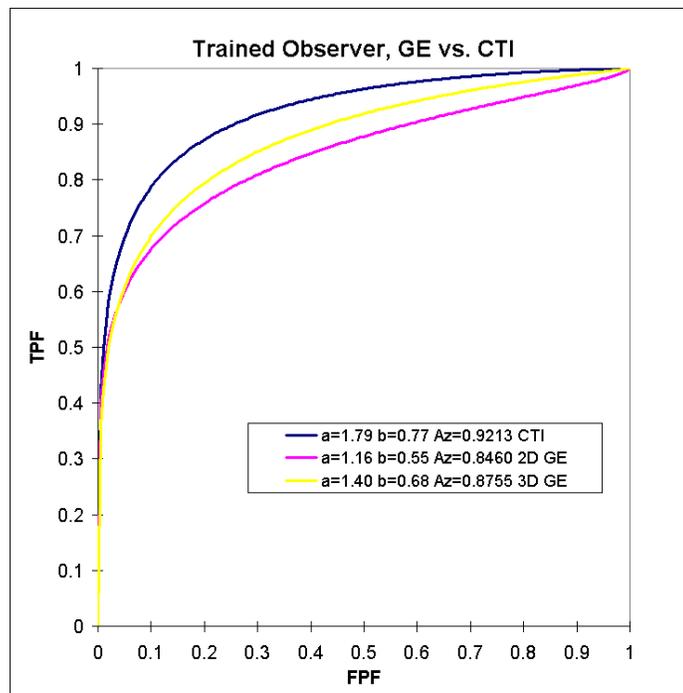


Figure 5.9. Trained observer ROC results comparing image scanner acquisition modes.

algorithm with ramp filter would be a logical choice, although one must still deal with the issue of resampling 3D data for use with a 2D algorithm.

Comparing 2D versus 3D acquisition modes of the Discovery ST there is a slightly higher  $A_z$  value ( $0.8755 \pm 0.029$ ) for the 3D mode than for the 2D mode ( $A_z = 0.8460 \pm 0.033$ ). This difference is less than a one sigma effect. The detectability of lesions in the two acquisition modes is statistically equal and no conclusion can be drawn about the superiority of one mode to the other.

The greater the sensitivity of the scanner, the less the effect of increased duration should be on increased detectability of lesions. For both scanners, the data indicate superior lesion detectability with increased scan duration for the trained observers. The  $A_z$  of  $0.9030 \pm 0.017$  for 16-minute acquisitions is significantly higher than the  $A_z$  of  $0.8427 \pm 0.035$  for the 4-minute acquisitions. The difference in the two  $A_z$  values for trained observers is larger than the difference for the overall or untrained observers. We conclude that a trained observer is able to ascertain additional information from higher-count scans more so than an untrained observer. These ROC curves, combined for all scanners and acquisition modes, are shown in Figure 5.10.

It is clear that with the levels of activity used, longer scan times do improve diagnostic image quality. The scans were acquired with count rates of 10–30 thousand counts per second (kcps) in the 2D mode and 50–230 kcps in the 3D mode. For a typical scan taken with the Discovery ST, count rates of 10 kcps total, 9 kcps trues, and 1 kcps random and scattered events, correspond to approximately 90% true events and 10% random and scattered events in the 2D mode. In the 3D mode the percentage of true events dropped to about 84% of all events with 16% random and scattered events; the rates were 50 kcps total, 42 kcps trues, and 8 kcps

random and scattered events. The relationship of true to scattered events is not a simple proportion in 2D or 3D mode [14]. Scans in this project were acquired with activities nowhere near the 50% random rate level upper threshold. For the Discovery ST, this rate is 686 kcps for the 2D mode and 921 kcps for the 3D mode [15].

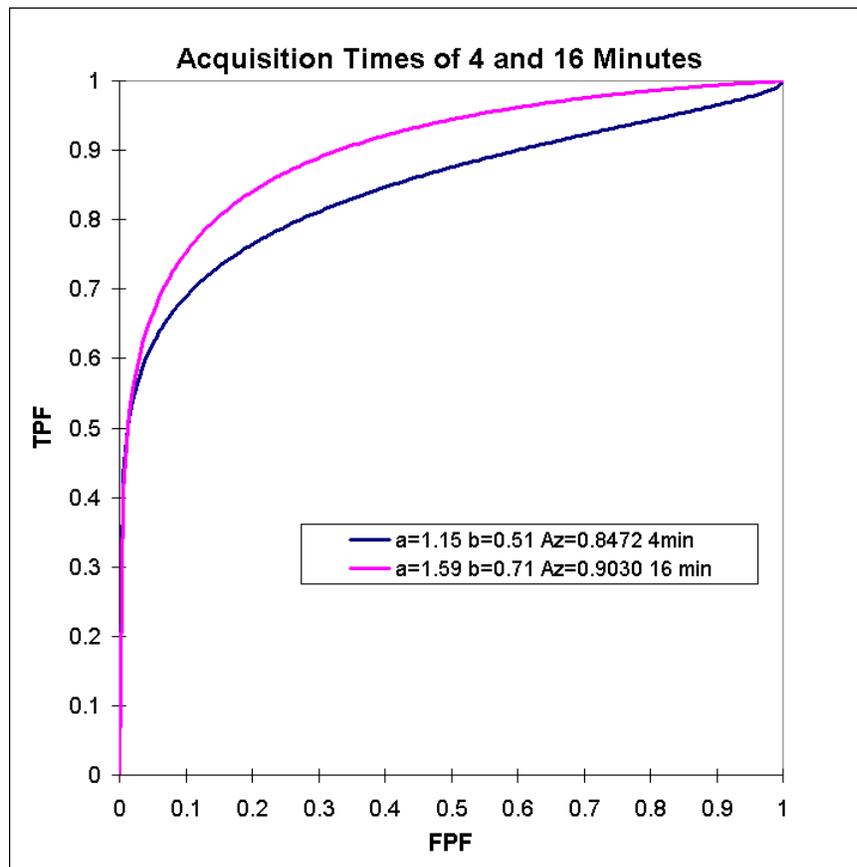


Figure 5.10. Trained observer ROC results for different acquisition times.

### 5.2.2 Comparisons of $A_z$ Values for Trained Versus Untrained Observers

Figure 5.11 reproduces the summary  $A_z$  values for trained and untrained readers from Table 5.2 and Figure 3.6. Figure 5.11 also indicates the average and

range of the standard deviation about the mean (gray rectangular boxes) for the trained and untrained observers. For the untrained observers there is very little fluctuation across parameters. This could have been a bias due to the method of image fusion or due to the heterogeneity of the untrained group. The poor performance of some untrained observers could be washing out any effects of the system parameters. With a larger number of observers, the data could have been split up into additional subgroups, perhaps better isolating the effects of the system parameters on lesion detectability.

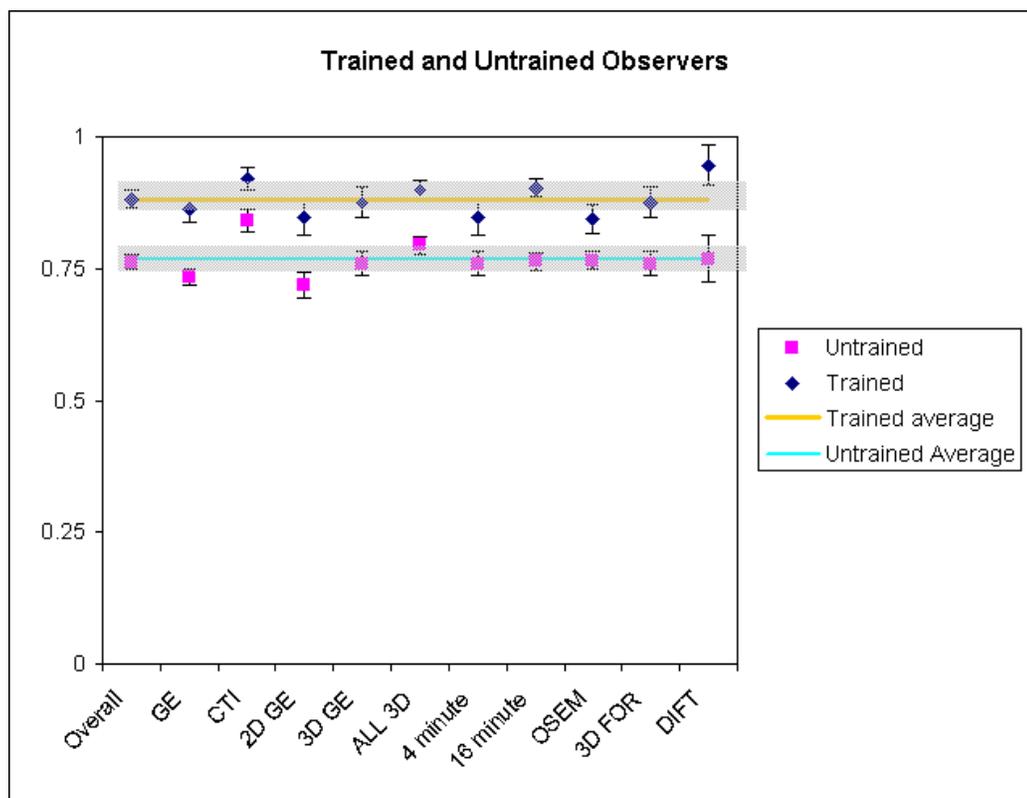


Figure 5.11. Comparison of trained observer  $A_z$  results alongside untrained observer  $A_z$  results with mean and standard deviation across parameters

The classic method to improve image quality is to increase acquisition length.

For the trained observers, the effect is as expected, but the increased time produces little benefit for the untrained observers. The obvious conclusion is that observers being unfamiliar with what they are looking at, is a bigger problem than the noise level in the image. For example, images containing the myocardium might be misclassified by the untrained observers due to lack of anatomical knowledge regardless of the noise level.

The motivation behind reconstructing images with iterative algorithms is to improve image quality and make more visual appealing. Iterative algorithms are more computationally expensive than the other algorithms considered. It was expected that the diagnostic image quality would be better for the OSEM algorithm. The untrained observers performed similarly irrespective of the reconstruction algorithm, while the trained observers performed worse with the iterative algorithm than for the other two algorithms. This unexpected result requires further study to verify and explain.

### **5.3 Diagnostic Accuracy, Sensitivity and Specificity**

The ROC survey was administered as a confidence-rating questionnaire for all but three of the trained observers. Those three observers were required to give a “yes or no” response to lesion presence with the same image set. The type of survey they were asked to respond to was a “two alternative forced choice” (2-AFC) experiment. Swets and Pickett showed that the area under the ROC curve will give the same value for both methods of administering the test [83]. One can conclude that using the 2-AFC test for three observers does not effect the results presented in previous sections.

Literature also suggests that the 2-AFC experiment takes less time to

administer than the confidence level test [84]. With the 2-AFC experiment the observer is only required to create a single decision boundary, while for a five-level confidence interval the user must have four decision boundaries to account for all five intervals. In this research, the length of time in evaluating the images was actually slightly longer for the observers doing the 2-AFC evaluation than for those doing the confidence rating evaluation.

The ROC analysis provides more information with the five-choice confidence responses than with the 2-AFC experiment. However, the basic quantities of sensitivity, specificity, and accuracy for classifying the value of a diagnostic test can be assessed equally for either method of administering the survey. To calculate these from the confidence survey, positive responses (rating categories one and two) and negative responses (rating categories three, four and five) are lumped together.

Table 5.4 shows the sensitivity, specificity and accuracy calculated for each of the parameters considered in comparing the two scanners. In the following tables the uncertainties are based on a standard deviation  $\sigma = 1/(2\sqrt{n})$ , where  $n$  is the number of images viewed [85], and the assumption that each view is independent is less plausible for the same image viewed by multiple observers. The errors are therefore probably slightly smaller than they would be if the viewed images were truly independent events.

Immediately apparent is the far superior ability of trained observers at correctly identifying actual positive images. The trained observers had better accuracy and sensitivity for all acquisition modes, compared to the untrained observers. However, specificity was not consistently better for the trained observers than the of untrained observers. The trained observers, unlike the untrained observers, were more likely to produce false positives than false negatives. The trained observers were willing to

Table 5.4. Sensitivity, specificity and accuracy for all observers, trained observers, and untrained observers, for all parameters combined and for each category of acquisition mode.\*

Category		All obsv.	Trained	Untrained
Overall:	Sensitivity	0.78±0.02	0.88±0.03	0.73±0.02
	Specificity	0.69±0.02	0.70±0.03	0.69±0.02
	Accuracy	0.73±0.01	0.79±0.02	0.71±0.01
All GE:	Sensitivity	0.79±0.02	0.87±0.04	0.76±0.02
	Specificity	0.63±0.02	0.65±0.04	0.62±0.02
	Accuracy	0.71±0.01	0.76±0.03	0.69±0.02
CTI:	Sensitivity	0.74±0.03	0.90±0.05	0.65±0.04
	Specificity	0.84±0.03	0.79±0.05	0.87±0.04
	Accuracy	0.79±0.02	0.85±0.04	0.76±0.03
GE 2D:	Sensitivity	0.82±0.03	0.86±0.05	0.80±0.03
	Specificity	0.55±0.03	0.62±0.05	0.52±0.03
	Accuracy	0.68±0.02	0.74±0.04	0.66±0.02
GE 3D:	Sensitivity	0.76±0.03	0.88±0.05	0.71±0.03
	Specificity	0.72±0.03	0.69±0.05	0.73±0.03
	Accuracy	0.74±0.02	0.79±0.04	0.72±0.02
All 3D:	Sensitivity	0.75±0.02	0.89±0.04	0.69±0.02
	Specificity	0.77±0.02	0.74±0.04	0.79±0.02
	Accuracy	0.76±0.01	0.82±0.03	0.74±0.02

\* All uncertainties based on a standard deviation  $\sigma = 1/(2\sqrt{n})$ , where  $n$  is the number of images viewed.

err towards a positive diagnosis, while the untrained observers were hesitant to give a positive result. The results for accuracy mirror those for the area under the ROC curve. The accuracy for the Reveal HD is higher than the Discovery ST in either 2D or 3D mode. Also, 3D mode for Discovery ST gives a larger accuracy than 2D mode.

Table 5.5 gives the sensitivity, specificity, and accuracy for the acquisition lengths of 4 minutes and 16 minutes. The 4 minute acquisition time exhibits a particularly low specificity for the trained observers. The conclusion is that long acquisitions decrease the number of false positives. This ideally leads to decreased

patient costs in terms of additional diagnostic tasks and unnecessary biopsies.

Other confounding effects such as patient motion and discomfort are not addressed in a phantom study such as that reported here.

Table 5.5. Comparison of the effect of acquisition times on the sensitivity, specificity and accuracy for all observer classes.\*

Category		All obsv.	Trained	Untrained
4 Minute:	Sensitivity	0.8±0.03	0.89±0.05	0.81±0.03
	Specificity	0.57±0.03	0.54±0.05	0.58±0.03
	Accuracy	0.70±0.02	0.72±0.04	0.70±0.02
16 Minute:	Sensitivity	0.74±0.02	0.88±0.04	0.68±0.03
	Specificity	0.76±0.02	0.78±0.04	0.75±0.02
	Accuracy	0.75±0.01	0.83±0.02	0.71±0.02

\* All uncertainties based on a standard deviation  $\sigma = 1/(2\sqrt{n})$ , where  $n$  is the number of images viewed.

Table 5.6 gives sensitivity, specificity, and accuracy for the different reconstruction algorithms. Interestingly, the sensitivity of trained observers with DIFT was 1.0; there were no misclassifications of actual positives with the reconstruction algorithm that coincidentally is the fastest. This is likely due to having a smaller image set for the DIFT (20) algorithm than for the other two algorithms, 3D FORE (40) and OSEM (60). There is a suggested bias in the small DIFT image set toward being too easy as compared to the entire set of images.

Table 5.6. Comparison of reconstruction algorithms in the categories of sensitivity, specificity and accuracy for all observer classes.\*

Category		All obsv.	Trained	Untrained
OSEM:	Sensitivity	0.78±0.02	0.84±0.04	0.76±0.03
	Specificity	0.65±0.02	0.67±0.04	0.64±0.03
	Accuracy	0.71±0.02	0.75±0.03	0.70±0.02
DIFT:	Sensitivity	0.78±0.05	1.00±0.07	0.60±0.07
	Specificity	0.82±0.05	0.82±0.07	0.82±0.07
	Accuracy	0.80±0.03	0.91±0.05	0.71±0.05
3D FORE:	Sensitivity	0.76±0.03	0.88±0.05	0.71±0.03
	Specificity	0.72±0.03	0.69±0.05	0.73±0.03
	Accuracy	0.74±0.02	0.79±0.04	0.72±0.02

\* All uncertainties based on a standard deviation  $\sigma = 1/(2\sqrt{n})$ , where  $n$  is the number of images viewed.

## CHAPTER 6

### CONCLUSION

We have found that the diagnostic image quality of a PET image is more closely related to physical design or actual performance standards than to initial inspection of visual image quality. In general, the Reveal HD performs better than the Discovery ST for the task of lesion detection in an anthropomorphic phantom. The Reveal HD has smaller crystals and better spatial resolution, aspects of performance that are closely related to the detection of small lesions. System sensitivity for 3D acquisition mode is comparable for the two systems. Table 6.1 lists the performance parameters for the two scanners considered in this thesis.

Table 6.1. Scanner performance parameters.\*

Parameter	Discovery ST 2D <sup>†</sup>	Discovery ST 3D <sup>†</sup>	Reveal HD <sup>‡</sup>
Spatial Resolution Axial	4.56 mm	5.68 mm	4.2 mm
Spatial Resolution Transaxial	6.28 mm	6.28 mm	4.5 mm
Sensitivity	8.6 kcps/kBq/cc	37 kcps/kBq/cc	38 kcps/kBq/cc

\* Spatial resolutions are measured @ FWHM.

<sup>†</sup> See reference [14].

<sup>‡</sup> See reference [16].

The technologists reported liking the Discovery ST software more than the Reveal HD software, but this does not influence diagnostic image quality. The conclusion is that investing in improved hardware is better than investing in “better” software, at least in terms of diagnostic image quality. It should be noted that the Discovery ST software is more stable and less prone to downtime due to

computer difficulty, as observed during the course of this project and noted by the technologists.

## 6.1 Validation of Hypotheses

The hypotheses of Chapter 2 were written prior to the acquisition of any image data or ROC analysis. Each hypothesis is reviewed in turn, with comments on the results that either verify or deny the hypothesis.

**Hypothesis 1.** An ROC study will confirm that 2D acquisition of PET/CT data produces superior diagnostic image quality exhibited by a larger  $A_z$  value for the ROC curve than 3D acquisition.

*Result:* Although the Discovery ST 2D mode of acquisition is used preferentially for whole-body imaging at MBPCC, it did not perform as well in this task as 3D mode. The 2D mode images were more visually appealing. This is partly due to the 2D OSEM algorithm disallowing negative pixel values, which decreases the visual artifacts in the reconstructed image. The results are contradictory to our hypothesis. The  $A_z$  values for all observer categories were higher for the 3D mode. A limitation to these results is that the images from the two acquisition modes were not reconstructed with the same algorithm.

**Hypothesis 2.** An ROC study will show that the Discovery ST scanner operated in 2D mode shows inferior diagnostic image quality compared to the Reveal HD scanner.

*Result:* The ROC analysis produced  $A_z = 0.921 \pm 0.022$  with the Reveal HD scanner for trained observers. The Discovery ST 2D mode of

acquisition, which has a larger spatial resolution and worse sensitivity than the Reveal HD scanner as specified by the manufacturers, performed worse in the ROC analysis with  $A_z = 0.846 \pm 0.033$  for the trained observers. The hypothesis was confirmed.

**Hypothesis 3.** An ROC study will show that the Discovery ST scanner operated in 3D mode will show inferior diagnostic image quality compared to the Reveal HD scanner.

*Result:* The hypothesis was confirmed. The trained observers performed better on the Reveal HD scanner. The Discovery ST 3D mode of acquisition has virtually the same sensitivity as, but inferior spatial resolution to, the Reveal HD. The Reveal HD trained observer  $A_z = 0.921 \pm .022$  is larger than the Discovery ST trained observer  $A_z = 0.876 \pm .029$ .

**Hypothesis 4.** An ROC study will show that the iterative reconstruction algorithm yields superior diagnostic image quality to both the 3D FORE and DIFT algorithms.

*Result:* Although the images reconstructed with the iterative OSEM algorithm are more visually appealing and contain fewer apparent reconstruction artifacts, this did not make it easier to identify lesions in a PET image. The hypothesis was incorrect.

**Hypothesis 5.** Superior system performance parameters correlate to superior diagnostic image quality.

*Result:* This hypothesis was proven correct. Smaller crystals, more

detectors, and both more and thinner slices are all physical parameters that combine to determine the system performance parameters of axial and transaxial resolution and sensitivity. Higher  $A_z$  values were obtained from the scanner with higher resolution and/or sensitivity.

**Hypothesis 6.** A longer acquisition improves the diagnostic image quality compared to shorter acquisitions.

*Result:* Increasing the number of counts by increasing the acquisition time does improve the diagnostic image quality for the acquisition lengths considered. The hypothesis is correct. The limitation of the phantom study is that it neglects potential issues of patient discomfort and motion due to longer acquisitions.

## 6.2 Proposed Future Related Projects

We have identified a variety of future studies or extensions of this project.

- Use patient data (One of the radiologists commented on the lack of realism in the phantom images). Patient data, as compared to phantom data, provides increased variety at the expense of increased complexity.
- The relevance of this work to radiation treatment planning would be better if the diagnostic task was extended to measurements of tumor volumes. This additional complexity would require a set of observers familiar with a common software platform that provides a measurement feature. The accuracy with which tumor boundaries are delineated has consequences on both cure rates and complications.

- Allow for CT images to be registered to the PET images. Using a PET/CT system to acquire image data, but only using the attenuation correction from the CT neglects the full utility of the PET/CT system. A diagnostic task using the combined image data must be identified. A common method of registration is probably necessary if combining data sets from different PET/CT scanners.
- Use a larger sample of radiologists. The trained reader population sample size collected in this thesis is too small. A larger set would increase the significance and ideally minimize certain systematic influences on the results due to individual responses.
- Similar to extending the study to delineate tumor boundaries, including the measurement of SUV would be interesting. This would allow for the possibility of incorporating classification or staging of lesions into the diagnostic task.
- A check for consistency of the observers should be added. For a portion of the readers, a second viewing would be an important check on the quality of the observers responses to the image set.

During the writing of this thesis an article was published discussing the current debate of who should read PET/CT images [86]. Conventionally, radiologists read CT scans and nuclear medicine physicians read PET scans. Registered PET/CT images may be a significant weapon in the diagnostic arsenal, but the question of who should read these images has become a turf war. Whether the reading should be a combination of both type of readers or whether one discipline can serve by obtaining additional training, there is not yet a standard for how registered PET/CT images should be read.

It would be a very interesting future study to investigate the performance of dedicated PET readers (nuclear medicine physicians) relative to dedicated CT readers (radiologists) for a combined diagnostic task on PET/CT. For the observers in this thesis, the one dedicated PET reader outperformed the other trained readers. Of course, a caveat to this observation is that the diagnostic task in this thesis was not a PET/CT task, but rather a task PET alone. The sample size of dedicated PET readers at MBPCC and OLOLRMC is too small to have addressed this issue in this project, even with a suitable PET/CT task.

## REFERENCES

- [1] C. B. Begg. Experimental design of medical imaging trials: Issues and options. *Investigative Radiology*, 24:934–936, 1988.
- [2] D. Gur, J. L. King, H. E. Rockette, C. A. Britton, F. L. Thaete, and R. J. Hoy. Practical issues of experimental ROC analysis: Selection of controls. *Investigative Radiology*, 25:583–586, 1990.
- [3] D. Gur, H. E. Rockette, W. F. Good, B. S. Slasky, L. A. Cooperstein, W. H. Straub, N. A. Obuchowski, and C. Metz. Effect of observer instruction on ROC study of chest images. *Investigative Radiology*, 25:230–234, 1990.
- [4] I. Hutt. *The Computer-Aided Detection of Abnormalities in Digital Mammograms*. PhD thesis, University of Manchester, Baton Rouge, Louisiana, 1996.
- [5] C. E. Metz. Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*, 24:234–245, 1989.
- [6] D. W. Townsend, J. P. J. Carney, J. T. Yap, and N. C. Hall. PET/CT today and tomorrow. *Journal of Nuclear Medicine*, 45(1 Suppl.):4S–14S, 2004.
- [7] V. Sharma, G. D. Luker, and D. Piwnica-Worms. Molecular imaging of gene expression and protein function in vivo with pet and spect. *Journal of Magnetic Resonance Imaging*, 16(4):336–351, 2002.
- [8] S. Johnson. Dose volume optimization in the IMRT era, Louisiana State University lecture notes. Unpublished lecture notes from a coures on “Clinical Principles of Radiation Therapy”, 2005.
- [9] C. B. Begg and R. A. Greenes. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, 39:207–215, 1983.
- [10] C. B. Begg. Biases in the assessment of diagnostic tests. *Statistics in Medicine*, 6:411–423, 1987.
- [11] C. B. Begg. Assessment of radiologic tests: Control of bias and other design considerations. *Radiology*, 167:565–569, 1988.
- [12] C. E. Metz. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8:283–298, 1978.
- [13] C. E. Metz. ROC methodology in radiologic imaging. *Investigative Radiology*, 21:720–733, 1986.
- [14] V. Bettinardi, M. Danna, A. Savi, M. Lecchi, I. Castiglioni, M. C. Gilardi, H. Bammer, G. Lucignani, and F. Fazio. Performance evaluation of the new whole-body PET/CT scanner: Discovery ST. *European Journal of Nuclear Medicine and Molecular Imaging*, 31(6):867–81, 2004.

- [15] O. Mawlawi, D. A. Podoloff, S. Kohlmyer, J. J. Williams, C. W. Stearns, R. F. Culp, and H. Macapinlac. Performance characteristics of a newly developed PET/CT scanner using NEMA standards in 2D and 3D modes. *The Journal of Nuclear Medicine*, 45(10):1734–1742, 2004.
- [16] R. Powers, J. Bland, C. Watson, and J. Hoffman. Cps innovations product specification ecat pet/ct. Technical Report No. 9111014-00, Revision G ECN:# 02-224, CPS Innovations™, April 6 2000.
- [17] T. G. Turkington. Introduction to PET instrumentation. *Journal of Nuclear Medicine Technology*, 29:1–8, 2001.
- [18] D. Lardinois, W. Weder, T. F. Hany, E. M. Kamel, S. Korom, B. Seifert, G. K. von Schulthess, and H. C Steinert. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *New England Journal of Medicine*, 348:2500–2507, 2003.
- [19] J. A. Swets. Measuring the accuracy of diagnostic systems. *Science*, 240:1285–1293, 1988.
- [20] D. D. Dorfman, K. S. Berbaum, and C. E. Metz. Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*, 27:723–731, 1992.
- [21] D. D. Dorfman and E. Alf. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. [binormal ROC curve-ordinal data]. *Journal of Mathematical Psychology*, 6:487–496, 1969.
- [22] K. S. Berbaum, D. D. Dorfman, and E. A. Franken Jr. Measuring observer performance by ROC analysis: Indications and complications. *Investigative Radiology*, 24:228–233, 1989.
- [23] D. W. Townsend, T. Beyer, and T. M. Blodgett. PET/CT Scanners: A Hardware Approach to Image Fusion. *Seminars in Nuclear Medicine*, XXXIII(3):193–204, 2003.
- [24] D. Jalihal and L. W. Nolte. Signal detection theory and reconstruction algorithms—performance for images in noise. *IEEE Transactions on Biomedical Engineering*, 41:501–504, 1994.
- [25] K. L. Matthews II. Evaluation of 3D mode of Discovery ST system. Unpublished Communication, 2005.
- [26] P. E. Valk. *Positron Emission Tomography: Basic science and clinical practice*. Springer, New York, 2003.
- [27] L-E. Adam, J. S. Karp, and G. Brix. Investigation of scattered radiation in 3D whole-body positron emission tomography using Monte Carlo simulations. *Phys. Med. Biol.*, 44:2879–2895, 1999.
- [28] S. Yamamoto, H. Horii, M. Hurutani, K. Matsumoto, and M. Senda. Investigation of single, random, and true counts from natural radioactivity in LSO-based clinical PET. *Annals of Nuclear Medicine*, 19:109–114, 2005.

- [29] H. Zaidi. Comparative evaluation of scatter correction techniques in 3D positron emission tomography. *European Journal Nuclear Medicine*, 27:1813–1826, 2000.
- [30] H. H. Barrett and W. Swindell. *Radiological imaging: the theory of image formation, detection and processing 1981*, volume 2. Academic Press, New York, 1981.
- [31] J. L. Humm, A. Rosenfeld, and A. Del Guerra. From PET detectors to PET scanners. *European Journal of Medical Molecular Imaging*, 30:1574–1597, 2003.
- [32] K. Hamacher, H. H. Coenen, and G. Stöcklin. Efficient stereospecific synthesis of no-carrier-added 2-[18f]-fluoro-2-deoxy-d-glucose using aminopolyether-supported nucleophilic substitution. *Journal of Nuclear Medicine*, 27:235–238, 1986.
- [33] M. Maisey, R. L. Wahl, and S. F. Barrington. *Atlas of clinical positron emission tomography*. Oxford University Press, New York, 1999.
- [34] S. Zabic and L. S. Bujenovic. A computer algorithm for gross tumor volume (GTV) determination in PET/CT images for radiation treatment planning (RTP). Unpublished manuscript.
- [35] W. Zhang, K. Doi, M. L. Giger, Y. Wu, R. M. Nishikawa, and R. A. Schmidt. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*, 21:517–524, 1993.
- [36] W. W. Peterson, T. G. Birdsall, and W. C. Fox. The theory of signal detectability. *Transactions of the IRE Professional Group in Information Theory PGIT*, 2–4:171–212, 1954.
- [37] L. S. Erdreich and E. T. Lee. Use of relative operating characteristic analysis in epidemiology: A method for dealing with subjective judgement. *American Journal of Epidemiology*, 114:649–662, 1981.
- [38] L. B. Lusted. Logical analysis in roentgen diagnosis. *Radiology*, 74:178–193, 1960.
- [39] L. B. Lusted. General problems in medical decision making, with comments on ROC analysis. *Seminars in Nuclear Medicine*, 8:299–306, 1978.
- [40] C. B. Begg. Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine*, 10:1887–1895, 1991.
- [41] R. M. Centor. Signal detectability: The use of ROC curves and their analyses. *Medical Decision Making*, 11:102–106, 1991.
- [42] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. Robert E. Krieger Publishing Co., Huntington, New York, 1974.
- [43] J. P. Egan. *Signal Detection Theory and ROC Analysis*. Academic Press, New York, 1975.

- [44] J. A. Swets. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1995.
- [45] B. J. McNeil and S. J. Adelstein. Determining the value of diagnostic and screening tests. *Journal of Nuclear Medicine*, 17:439–448, 1976.
- [46] J. C. Ogilvie and C. D. Creelman. Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, 5:377–391, 1968.
- [47] C. B. Begg. Statistical methods in medical diagnosis. *CRC Critical Reviews in Medical Informatics*, 1:1–22, 1986.
- [48] W. F. Good, D. Gur, W. H. Straub, and J. H. Feist. Comparing imaging systems by ROC studies: Detection versus interpretation. *Investigative Radiology*, 24:932–3, 1989.
- [49] B. J. McNeil, E. Keeler, and S. J. Adelstein. Primer on certain elements of medical decision making. *New England Journal of Medicine*, 293:211–215, 1975.
- [50] X. H. Zhou. Testing an underlying assumption on a ROC curve based on rating data. *Medical Decision Making*, 15:276–282, 1995.
- [51] D. J. Goodenough, K. Rossmann, and L. B. Lusted. Radiographic applications of receiver operating characteristic (ROC) curves. *Radiology*, 110:89–95, 1974.
- [52] D. A. Turner. An intuitive approach to receiver operating characteristic curve analysis. *Journal of Nuclear Medicine*, 19:213–220, 1978.
- [53] J. A. Hanley. Receiver operating characteristic (ROC) methodology: The state of the art. *Critical Reviews in Diagnostic Imaging*, 29:307–335, 1989.
- [54] J. A. Hanley and C. B. Begg. Response to ROC steady. *Medical Decision Making*, 7:244–246, 1987.
- [55] J. K. Hsiao, J. J. Bartko, and W. Z. Potter. Diagnosing diagnoses: Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry*, 46:664–667, 1989.
- [56] J. A. Hanley. The robustness of the “binormal” assumptions used in fitting ROC curves. *Medical Decision Making*, 8:197–203, 1988.
- [57] J. A. Hanley and B. J. Y. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36, 2004.
- [58] J. A. Hanley. The use of the ‘binormal’ model for parametric roc analysis of quantitative diagnostic tests. *Statistics in Medicine*, 15:1575–1585, 1996.
- [59] A. R. Henderson. Assessing test accuracy and its clinical consequences: a primer for receiver operating characteristic curve analysis. *Annals of Clinical Biochemistry*, 30:521–539, 1993.

- [60] R. M. Centor and J. S. Schwartz. An evaluation of methods for estimating the area under the receiver operating characteristic (ROC) curve. *Medical Decision Making*, 5:149–156, 1985.
- [61] A. N. Tosteson and C. B. Begg. A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8:204–215, 1988.
- [62] M. Treisman and A. Faulkner. The effect of signal probability on the slope of the receiver operating characteristic given by the rating procedure. *British J. Math. Statist+*, 37:199,215, 1984.
- [63] H. E. Rockette, D. Gur, L. A. Cooperstein, N. A. Obuchowski, J. L. King, C. R. Fuhrman, E. K. Tabor, and C. E. Metz. Effect of two rating formats in multi-disease ROC study of chest images. *Investigative Radiology*, 25:225–229, 1990.
- [64] H. E. Rockette, D. Gur, and C. E. Metz. The use of continuous and discrete confidence judgments in receiver operating characteristic studies of diagnostic imaging techniques. *Investigative Radiology*, 27:169–172, 1992.
- [65] H. C. Gifford, M. A. King, R. G. Wells, W. G. Hawkins, M. V. Narayanan, and P. H. Pretorius. LROC analysis of detector-response compensation in SPECT. *IEEE Transactions on Medical Imaging*, 19(5):463–473, 2000.
- [66] J. L. King, C. A. Britton, D. Gur, H. E. Rockette, and P. L. Davis. On the validity of the continuous and discrete confidence rating scales in receiver operating characteristic studies. *Investigative Radiology*, 28:962–963, 1993.
- [67] E. A. Krupinski, M. Evanoff, T. Ovitt, J. R. Standen, T. X. Chu, and J. Johnson. Influence of image processing on chest radiograph interpretation and decision changes. *Academic Radiology*, 5:79–85, 1998.
- [68] Y. Ishihara. Investigation of accuracy in quantitation of  $^{18}\text{F}$ -FDG concentration of PET/CT. Master's thesis, Louisiana State University, Baton Rouge, Louisiana, December 2004.
- [69] C. Burger, G. Goerres, S. Schoenes, A. Buck, A. Lonn, and G. von Schulthess. PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. *European Journal of Nuclear Medicine and Molecular Imaging*, 29(7):922–927, 2002.
- [70] J. A. Swets. ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14:109–121, 1979.
- [71] L. B. Lusted. Signal detectability and medical decision-making. *Science*, 171:1217–1219, 1971.
- [72] N. A. Macmillan and C. D. Creelman. *Detection Theory: A User's Guide*. Cambridge University Press, Cambridge, 1991.
- [73] C. E. Phelps. The methodologic foundations of studies of the appropriateness of medical care[see comments]. *New England Journal of Medicine*, 329(17):1241–1245, 1993.

- [74] AJCC, F. L. Greene, C. M. Balch, D. L. Page, D. G. Haller, I. D. Fleming, M. Morrow, and A. G. Fritz. *AJCC Cancer Staging Manual, Sixth Edition*. American Joint Committee on Cancer, Chicago, IL 60611, 6 edition, March 2002.
- [75] N. A. Obuchowski. Computing sample size for receiver operating characteristic studies. *Investigative Radiology*, 29:238–243, 1994.
- [76] G. A. Diamond. A receiver operating characteristic curve that is invariant relative to selection bias. *Medical Decision Making*, 7:247–249, 1987.
- [77] G. A. Diamond. What is the effect of sampling error on ROC analysis in the face of verification bias? *Medical Decision Making*, 12:155–6, 1992.
- [78] G. Champbell. Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*, 13:499–508, 1994.
- [79] C. E. Metz and J. H. Shen. Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of ROC analysis. *Medical Decision Making*, 12:60–75, 1992.
- [80] D. Mossman. Resampling techniques in the analysis of non-binormal ROC data. *Medical Decision Making*, 15:358–366, 1995.
- [81] B. Emir, S. Wieand, S. H. Jung, and Z. Ying. Comparison of diagnostic markers with repeated measurements: A non-parametric ROC curve approach. *Statistics in Medicine*, 19:511–523, 2000.
- [82] N. A. Obuchowski. Multireader, multicase receiver operating characteristic analysis: An empirical comparison of five methods. *Academic Radiology*, 11(9):980–995, 2004.
- [83] J. A. Swets and R. M. Pickett. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York, 1982.
- [84] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, fifth edition, 1966. (Reprint with corrections: Huntington, New York: Krieger, 1974).
- [85] P. R. Bevington and D. K. Robinson. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, New York, second edition, 1992.
- [86] G. Wiley. Disruptive Technology: The Conflict Over PET/CT. *Decisions in Imaging Economics*, 18(6):17–20, 2005.

## APPENDIX A

### SAMPLE ROCKIT PROGRAM RUN

The following listing shows the interactive command style user interface and responses from a user for a typical ROCKIT job. In this listing the software's prompted menu/UI lines, as well as a user's input, are prefixed by >, additional comments are added [*in square parentheses in slanted roman font*], which would not obviously really be seen in the sample ROCKIT session, they are included in the listing for explanatory purposes only. That is, if you were using ROCKIT you wouldn't actually type or see any of the comments in the square brackets in what follows.

Entry into the ROCKIT software

```
> [Prompt:] Do you want to use data from a previously created
           input file for the next run? (y/n, or q to quit)
```

```
> [Response:] n
```

```
[Reason: It is not necessary to create an input file prior
to using the ROC software.]
```

```
> [Prompt:] Do you want to create a file to store the data
           you will input? (y/n, or r to restart)
```

```
> [Response: n]
```

```
[Reason: It is not necessary to create an input file. Data
entry is simple.]
```

```
> [Alternate Response:] y
```

*[Reason: To be able to edit input or rerun program!]*

> [Prompt:] Do you want a (F)ull, (C)oncise or (N)o OUTPUT file?

> [Response:] n

*[Reason: The screen output will give the parameters needed  
for creating the ROC curve.]*

> [Prompt:] Do you want (F)ull or (C)oncise Screen output?

> [Response:] c

*[Reason: The parameters I need are the binormal fitted  
(a, b and  $A_z$ ) which are obtainable from the concise screen output.]*

> [Prompt:] Enter data description (up to 60 characters  
including blanks):

> [Response:] Physician complete response

*[Reason: This is just header information, which can be used  
for bookkeeping.]*

> [Prompt:] With Rating-Data Matrices, at most 2 datasets can  
be entered. Otherwise, up to 5 datasets may be input  
How many data sets do you wish to enter (1-5)?

> [Response:] 1

*[Reason: This example is for a single data set. The software  
tends to crash for larger numbers of data sets.]*

> [Prompt:] Enter Description for Condition 1: (12 chars max)

> [Response:] GE Discovery

*[Reason: This is just header information, which can be used  
for bookkeeping.]*

> [Prompt:] You are entering 1 datasets

1

GE Discovery

Are the data for condition 1 collected on a (D)iscrete  
rating or (C)ontinuous interval scale (D or C)?:

> [Response:] C

*[Reason: The test that was performed is a confidence level test with five  
discrete values as possible responses.]*

> [Prompt:] Is category (1) or (N) indicative of  
Actually-Positive cases (1 or N)?:

> [Response:] 1

*[Reason: Response 1 indicates the highest confidence that the observer  
has successfully identified a lesion.]*

> [Prompt:] Will you be entering the data in (M)atrix  
or (L)ist form?

> [Response:] M

*[Reason: This is the simplest way to enter the data. For  
actual positive cases tally up the number of each category  
responses and then for actual negative cases do the same.]*

> [Prompt:] How many categories does the discrete rating scale  
contain?: (3-20)

> [Response:] 5

*[Reason: This is the total number of possible responses that the observer  
can choose from.]*

> [Prompt:] For Condition 1: GE Discovery

Enter the Total Number of Actually-Negative Cases

(an integer):

> [Response:] 10

*[Reason: If there are 20 images and 10 are actually negative with a single onetime observer this would be a correct input.]*

> [Prompt:] Beginning with category 1 and separated by blanks,  
Enter (on one line, integers only) the number of responses to

Actually-Negative cases in each category:

1 2 3 4 5

> [Response:] 2 2 2 2 2

*[Reason: The program will give an error if the total does not equal the input from the previous prompt.]*

> [Alternate Response:] 2 3 3 3 3

[Readout:] ERROR: the total number of cases (14) is  
not equal to the number entered ( 10)

*[Comment: At this point the program will will return to the original prompt]:*

> [Enter the Total Number of Actually-Negative Cases

(an integer):]

> [Prompt:] Enter the Total Number of Actually-Positive Cases

(an integer):

> [Response:] 10

*[Reason: If there are 20 images and 10 are actually positive with a single onetime observer this would be a correct input.]*

> [Prompt:] Beginning with category 1 and separated by blanks,  
Enter (on one line, integers only) the number of responses to

Actually-Positive cases in each category:

1 2 3 4 5

> [Response:] 2 2 2 2 2

*[Reason: The program will give an error if the total does not  
equal the input from the previous prompt.]*

At this point the screen output is given.

Date - 18-May-05

Time - 16:25:50

ROCKIT (Windows95 version 0.9.1 BETA):

test

Maximum Likelihood Estimation of the Parameters

a Single Binormal ROC Curve

Name of Input File being used: 77777777777777777777

Condition 1: trues

Total number of actually-negative cases = 10.

Total number of actually-positive cases = 10.

Data effectively collected in 5 categories.

Category 1 represents the strongest evidence of positivity.

(e.g., that the disease is present)

Response Data:

Category 1 2 3 4 5

Actually-Negative Cases 2 2 2 2 2

Actually-Positive Cases 2 2 2 2 2

--- program is running; please wait ---

Operating Points Corresponding to the Input Data:

FPF: .000.200.400.600.800 1.000

TPF: .000.200.400.600.800 1.000

-----  
Initial Estimates of the Binormal ROC Parameters:  
-----

a = .0000

b = 1.0000

z(k) = .842.253 -.253 -.842

Procedure Converges after 3 Iterations

=====  
Final Estimates of the Binormal ROC Parameters  
=====

Binormal Parameters and Area Under the Estimated ROC :

a = 0.0000

b = 1.0000

Area (Az) = .5000

1: z(k) = -.842 -.253.253.842

Estimated Standard Errors and Correlation of these Values:

Std. Err. (a) = .4722

Std. Err. (b) = .4621

Corr(a,b) = 0.0000

Std. Err. (Az) = .1332

Symmetric 95% Confidence Intervals

For a : ( -.9255, .9255)

For b : (.0942, 1.9058)

Asymmetric 95% Confidence Interval

For Az: (.2564, .7436)

Estimated Binormal ROC curve, with Lower and Upper  
 Bounds of the Asymmetric Point-wise 95% Confidence  
 Interval for True-Positive Fraction at a Variety  
 of False-Positive Fractions:

FPF TPF (Lower Bound, Upper Bound)

.005.0050 ( 0.0000 ,.4737 )  
 .010.0100 ( 0.0000 ,.4900 )  
 .020.0200 ( 0.0000 ,.5095 )  
 .030.0300 (.0001 ,.5231 )  
 .040.0400 (.0002 ,.5340 )  
 .050.0500 (.0003 ,.5434 )  
 .060.0600 (.0006 ,.5518 )  
 .070.0700 (.0010 ,.5596 )  
 .080.0800 (.0014 ,.5669 )  
 .090.0900 (.0021 ,.5738 )  
 .100.1000 (.0028 ,.5804 )  
 .110.1100 (.0038 ,.5868 )  
 .120.1200 (.0049 ,.5930 )  
 .130.1300 (.0061 ,.5991 )  
 .140.1400 (.0076 ,.6051 )  
 .150.1500 (.0093 ,.6110 )  
 .200.2000 (.0207 ,.6396 )

.250.2501 (.0373 ,.6681 )  
 .300.3001 (.0589 ,.6971 )  
 .400.4002 (.1138 ,.7582 )  
 .500.5000 (.1773 ,.8227 )  
 .600.5998 (.2418 ,.8862 )  
 .700.6999 (.3029 ,.9411 )  
 .800.8000 (.3604 ,.9793 )  
 .900.9000 (.4196 ,.9972 )  
 .950.9500 (.4566 ,.9997 )

Estimates of Expected Operating Points on fitted ROC curve, with lower and upper bounds of asymmetric 95% confidence interval along the curve for those points:

Expected operating point	Lower bound	Upper bound
( FPF , TPF )	( FPF , TPF )	( FPF , TPF )
(.4000,.4000)	(.1313,.1313)	(.7302,.7302)
(.6000,.6000)	(.3160,.3160)	(.8378,.8378)
(.8000,.8000)	(.5436,.5436)	(.9422,.9422)
(.5000,.5000)	(.1930,.1930)	(.8070,.8070)

At this point the ROCKIT job is finished. This is the end of Appendix A. Check that the line spacing is still doublespace. The next appendix is about ROC curve errors, the one after that has some weird MATLAB worksheet stuff. I hope this is still double-spaced.

## APPENDIX B

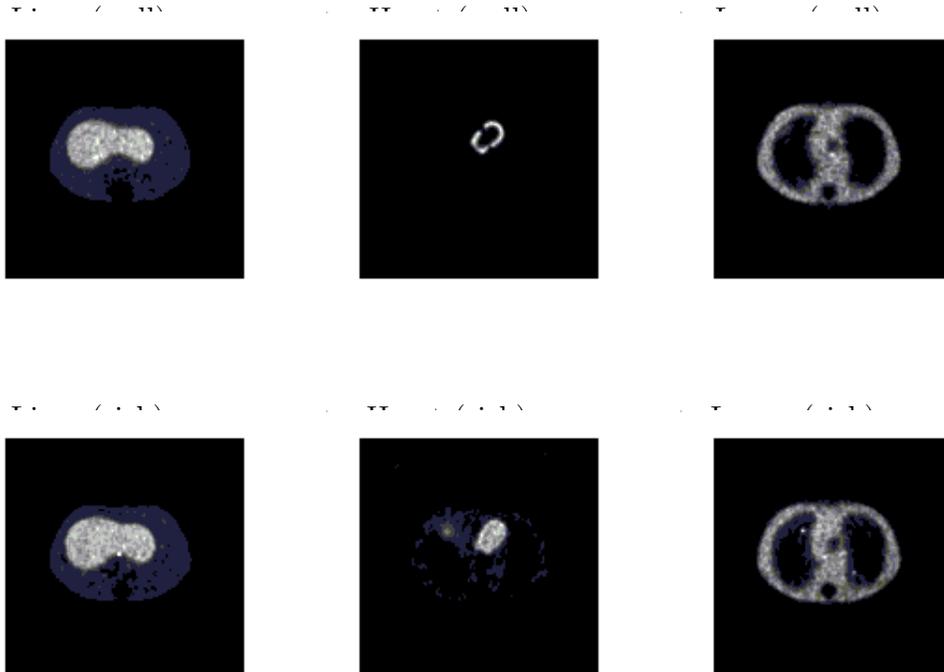
### SURVEY INSTRUCTIONS DOCUMENT

#### Instructions for evaluating images

The basic idea is to look at a series of images acquired from a PET/CT scanner and assess the likelihood of pathology based on the distribution of the radioisotope  $^{18}\text{F}$ . The images with pathology will differ from the normal or negative images with the presence of small-localized “hot spots”.

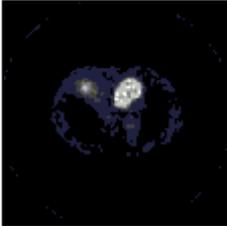
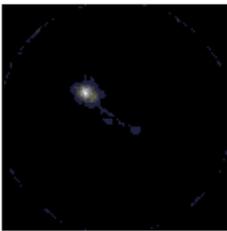
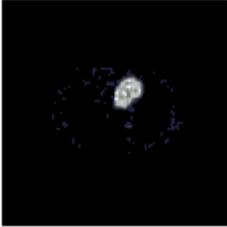
First Row: Normal appearance of PET slices obtained from phantom (Negatives)

Second Row: Appearance of PET slices obtained from phantom with pathology

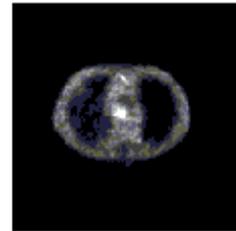
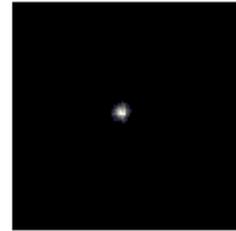
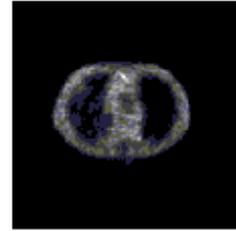


The basic process is to fuse the two pictures with random intensities of the tumor images.

Fusion of a tumor in a slice containing the heart.



or



**Statement for observer to assess:** The image exhibits strong evidence of pathology.

- 1) Strongly agree (I would say 1 or more tumor(s) are present)
- 2) Agree (There exists spots on the image that might be tumors)
- 3) Neutral (Cannot determine from the image presence or absence of tumors)
- 4) Disagree (Image looks reasonably clear of tumors)
- 5) Strongly disagree (I see no evidence of the presence of tumors)

## APPENDIX C

### MATLAB ROUTINES

MATLAB was used throughout the thesis research for analysis, image display, and image manipulation. For completeness, included in this appendix are the routine for display and a routine used to fuse the phantom images with lesion images.

Table C.1. Example MATLAB commands for plotting ROC results.

MATLAB Commands for Plotting ROC Curves
<pre>% path (path, 'C:\Documents and Settings\Ken\Desktop\ROC DICOM')  clear  %Start Num_file on 3 since . and .. are 1 and 2     num_file= dir;     nn=size(num_file);     n=nn(:,1)     AA=zeros(n-5,2); scrsz = get(0,'ScreenSize') j=0; for i = 3:n;     j=j+1;     fil_num = getfield(num_file(i),'name')</pre>

(table C.1 continued)

## MATLAB Commands for Plotting ROC Curves

```

nq=str2num(fil_num)
%[X,mapp,alph]=imread('temp.jpg');
Y=dicomread(fil_num);
info=dicominfo(fil_num);
figure('Position',[360 scrsz(4)/3 scrsz(3)/3 scrsz(4)/3]);
subplot(1,1,1);
imshow(Y, []);
title(['Case number #',int2str(nq)],'Color','b')
AA(i,1)=nq; AA(i,2)=j;
%ylabel(int2str(nq))
%text(1,-1/7,'\it Note symmetry.}','color','w')
%title=(['Case number #',(fil_num)],'color','r')
end
info

```

Table C.2. Example MATLAB commands for lesion image fusion.

MATLAB Commands for lesion fusion.
<pre>%X=imread(\GE\neg\2d_16min\1.bmp'); X=imread('1.bmp');  imshow(X, []); X2=imread('\GE\neg\2d_16min_lesion\2.bmp');  %BW = roipoly; subplot(2,2,1); imshow(X, []); subplot(2,2,2); imshow(X2, []); X3 = imlincomb(1,X2); subplot(2,2,3); imshow(X3, []);  X5=imadd(X3,X); subplot(2,2,4); imshow(X5, []); figure,imshow(X3, []);</pre>

## APPENDIX D

### ROC INPUT DATA

This appendix reproduces the raw data used for the ROC analysis. In Table D.1 the untrained and trained observer responses are tabulated in columns, with blocks of rows tabulating the corresponding categories of images. Columns 2 through 18 are labeled either “ $tn$ ” or “ $un$ ” respectively for trained and untrained observer data, the  $n$  is some identification number.

Table D.1. ROC survey results. The 19<sup>th</sup> column labeled “UTO” contains the totals for untrained observers, the 20<sup>th</sup> column labeled “TO” contains the totals for the trained observers (medical professionals).

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total
Overall Actual Negatives																				
5	5	9	14	53	20	46	15	23	0	13	0	5	14	39	1	2	8	111	156	267
4	8	22	24	0	0	0	17	14	1	13	5	13	15	9	14	14	6	137	38	175
3	25	20	11	0	0	0	9	8	36	21	19	21	10	0	18	12	23	214	19	233
2	15	7	8	0	0	0	12	10	15	10	24	10	3	2	16	19	14	147	18	165
1	8	3	4	8	41	15	8	6	9	4	13	2	9	1	2	4	0	63	74	137
Totals	61	61	61	61	61	61	61	61	61	61	61	51	51	51	51	51	51	672	305	977
Overall Actual Positives																				
5	0	0	0	11	2	12	1	1	0	3	0	0	3	6	1	0	2	16	26	42
4	0	3	1	0	0	0	5	0	1	2	3	3	9	4	5	2	0	37	1	38
3	9	12	4	0	0	0	8	4	16	11	11	11	9	7	13	9	8	124	8	132
2	31	24	5	0	0	0	15	11	15	30	36	26	13	6	24	25	18	263	16	279
1	19	20	49	48	57	47	30	43	27	13	9	9	15	26	6	13	21	208	244	452
Totals	59	59	59	59	59	59	59	59	59	59	59	49	49	49	49	49	49	648	295	943

(table D.2 continued)

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total
GE Discovery Actual Negatives (41)																				
5	1	5	7	34	14	28	7	15	0	8	0	5	10	29	1	1	6	73	98	171
4	4	12	14	0	0	0	9	7	1	4	3	9	10	9	10	9	3	83	21	104
3	16	14	8	0	0	0	7	7	20	15	14	16	9	0	12	9	18	150	15	165
2	12	7	8	0	0	0	10	8	12	10	15	9	3	2	16	18	14	128	16	144
1	8	3	4	7	27	13	8	4	8	4	9	2	9	1	2	4	0	58	55	113
Totals	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	492	205	697
GE Discovery Actual Positives (39)																				
5	0	0	0	9	1	9	0	1	0	1	0	0	3	5	1	0	2	12	20	32
4	0	1	0	0	0	0	3	0	0	1	2	2	6	4	4	2	0	25	0	25
3	6	5	2	0	0	0	5	3	10	7	3	7	9	5	8	6	6	77	5	82
2	16	17	5	0	0	0	10	7	11	19	29	21	12	6	20	23	16	200	12	212
1	17	16	32	30	38	30	21	28	18	11	5	9	9	19	6	8	15	154	158	312
Totals	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	468	195	663
CTI Reveal Actual Negatives (20)																				
5	4	4	7	19	6	18	8	8	0	5	0	0	4	10	0	1	2	38	58	96
4	4	10	10	0	0	0	8	7	0	9	2	4	5	0	4	5	3	54	17	71
3	9	6	3	0	0	0	2	1	16	6	5	5	1	0	6	3	5	64	4	68
2	3	0	0	0	0	0	2	2	3	0	9	1	0	0	0	1	0	19	2	21
1	0	0	0	1	14	2	0	2	1	0	4	0	0	0	0	0	0	5	19	24
Totals	20	20	20	20	20	20	20	20	20	20	20	10	10	10	10	10	10	180	100	280
CTI Reveal Actual Positives (20)																				
5	0	0	0	2	1	3	1	0	0	2	0	0	0	1	0	0	0	4	6	10
4	0	2	1	0	0	0	2	0	1	1	1	1	3	0	1	0	0	12	1	13
3	3	7	2	0	0	0	3	1	6	4	8	4	0	2	5	3	2	47	3	50
2	15	7	0	0	0	0	5	4	4	11	7	5	1	0	4	2	2	63	4	67
1	2	4	17	18	19	17	9	15	9	2	4	0	6	7	0	5	6	54	86	140
Totals	20	20	20	20	20	20	20	20	20	20	20	10	10	10	10	10	10	180	100	280

(table D.2 continued)

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total
GE Discovery 2D Acquisition Actual Negatives (21)																				
5	0	1	3	19	6	12	4	8	0	1	0	1	1	13	0	0	0	21	48	69
4	0	5	4	0	0	0	2	3	0	1	0	3	6	6	3	2	3	31	7	38
3	9	10	6	0	0	0	3	4	8	8	7	10	7	0	6	5	7	80	10	90
2	8	4	6	0	0	0	6	4	8	9	10	6	2	2	12	11	11	89	10	99
1	4	1	2	2	15	9	6	2	5	2	4	1	5	0	0	3	0	31	30	61
Totals	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	252	105	357
GE Discovery 2D Acquisition Actual Positives (19)																				
5	0	0	0	5	0	6	0	0	0	0	0	0	1	4	0	0	1	6	11	17
4	0	0	0	0	0	0	1	0	0	0	0	0	4	1	1	1	0	8	0	8
3	1	2	1	0	0	0	1	1	6	5	1	3	4	2	2	2	2	31	2	33
2	6	7	3	0	0	0	5	3	5	9	16	11	6	5	11	11	9	101	6	107
1	12	10	15	14	19	13	12	15	8	5	2	5	4	7	5	5	7	82	76	158
Totals	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	228	95	323
GE Discovery 3D Acquisition Actual Negatives (20)																				
5	1	4	4	15	8	16	3	7	0	7	0	4	9	16	1	1	6	52	50	102
4	4	7	10	0	0	0	7	4	1	3	3	6	4	3	7	7	0	52	14	66
3	7	4	2	0	0	0	4	3	12	7	7	6	2	0	6	4	11	70	5	75
2	4	3	2	0	0	0	4	4	4	1	5	3	1	0	4	7	3	39	6	45
1	4	2	2	5	12	4	2	2	3	2	5	1	4	1	2	1	0	27	25	52
Totals	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	240	100	340
GE Discovery 3D Acquisition Actual Positives (20)																				
5	0	0	0	4	1	3	0	1	0	1	0	0	2	1	1	0	1	6	9	15
4	0	1	0	0	0	0	2	0	0	1	2	2	2	3	3	1	0	17	0	17
3	5	3	1	0	0	0	4	2	4	2	2	4	5	3	6	4	4	46	3	49
2	10	10	2	0	0	0	5	4	6	10	13	10	6	1	9	12	7	99	6	105
1	5	6	17	16	19	17	9	13	10	6	3	4	5	12	1	3	8	72	82	154
Totals	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	240	100	340

(table D.2 continued)

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total
All 3D Acquisition Actual Negatives (20)																				
5	5	8	11	34	14	34	11	15	0	12	0	4	13	26	1	2	8	90	108	198
4	8	17	20	0	0	0	15	11	1	12	5	10	9	3	11	12	3	106	31	137
3	16	10	5	0	0	0	6	4	28	13	12	11	3	0	12	7	16	134	9	143
2	7	3	2	0	0	0	6	6	7	1	14	4	1	0	4	8	3	58	8	66
1	4	2	2	6	26	6	2	4	4	2	9	1	4	1	2	1	0	32	44	76
Totals	40	40	40	40	40	40	40	40	40	40	40	30	30	30	30	30	30	420	200	620
All 3D Acquisition Actual Positives (20)																				
5	0	0	0	6	2	6	1	1	0	3	0	0	2	2	1	0	1	10	15	25
4	0	3	1	0	0	0	4	0	1	2	3	3	5	3	4	1	0	29	1	30
3	8	10	3	0	0	0	7	3	10	6	10	8	5	5	11	7	6	93	6	99
2	25	17	2	0	0	0	10	8	10	21	20	15	7	1	13	14	9	162	10	172
1	7	10	34	34	38	34	18	28	19	8	7	4	11	19	1	8	14	126	168	294
Totals	40	40	40	40	40	40	40	40	40	40	40	30	30	30	30	30	30	420	200	620
4 Minute Acquisition Times Actual Negatives																				
5	1	0	3	17	4	4	2	4	0	3	0	1	4	14	0	0	3	28	32	60
4	1	6	10	0	0	0	6	5	0	2	1	5	3	5	7	5	3	44	15	59
3	7	7	2	0	0	0	3	5	8	7	7	7	7	0	3	3	9	68	7	75
2	7	6	4	0	0	0	3	6	6	6	9	6	2	1	8	11	5	70	10	80
1	4	1	1	3	16	16	6	0	6	2	3	1	4	0	2	1	0	30	36	66
Totals	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	240	100	340
4 Minute Acquisition Times Actual Positives																				
5	0	0	0	5	1	3	0	0	0	1	0	0	0	2	1	0	0	4	9	13
4	0	0	0	0	0	0	1	0	0	0	1	1	3	3	0	1	0	10	0	10
3	2	1	0	0	0	0	2	2	6	4	1	1	5	2	2	2	4	32	2	34
2	6	7	2	0	0	0	4	4	4	7	15	11	6	3	12	12	9	96	6	102
1	12	12	18	15	19	17	13	14	10	8	3	7	6	10	5	5	7	98	83	181
Totals	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	240	100	340

(table D.2 continued)

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total
16 Minute Acquisition Times Actual Negatives																				
5	4	8	8	17	10	30	9	11	0	9	0	3	9	12	1	2	5	62	76	138
4	7	11	10	0	0	0	9	6	1	10	4	5	6	-2	4	7	0	62	16	78
3	9	3	3	0	0	0	3	-1	20	6	5	4	-4	0	9	4	7	66	2	68
2	0	-3	-2	0	0	0	3	0	1	-5	5	-2	-1	-1	-4	-3	-2	-12	-2	-14
1	0	1	1	3	10	-10	-4	4	-2	0	6	0	0	1	0	0	0	2	8	10
Totals	20	20	20	20	20	20	20	20	20	20	20	10	10	10	10	10	10	180	100	280
16 Minute Acquisition Times Actual Positives																				
5	0	0	0	1	1	3	1	1	0	2	0	0	2	0	0	0	1	6	6	12
4	0	3	1	0	0	0	3	0	1	2	2	2	2	0	4	0	0	19	1	20
3	6	9	3	0	0	0	5	1	4	2	9	7	0	3	9	5	2	61	4	65
2	19	10	0	0	0	0	6	4	6	14	5	4	1	-2	1	2	0	66	4	70
1	-5	-2	16	19	19	17	5	14	9	0	4	-3	5	9	-4	3	7	28	85	113
Totals	20	20	20	20	20	20	20	20	20	20	20	10	10	10	10	10	10	180	100	280
OSEM Actual Negatives																				
5	0	1	3	29	6	22	7	11	0	1	0	1	5	23	0	1	2	41	71	112
4	3	10	11	0	0	0	6	7	0	6	2	7	11	6	7	7	6	71	18	89
3	14	15	9	0	0	0	4	5	17	13	9	15	8	0	12	8	12	127	14	141
2	10	4	6	0	0	0	8	5	8	9	16	7	2	2	12	12	11	101	11	112
1	4	1	2	2	25	9	6	3	6	2	4	1	5	0	0	3	0	32	41	73
Totals	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	372	155	527
OSEM Actual Positives																				
5	0	0	0	7	1	9	1	0	0	0	0	0	1	5	0	0	1	8	17	25
4	0	0	1	0	0	0	1	0	1	0	0	1	7	1	2	1	0	14	1	15
3	3	5	3	0	0	0	3	2	8	8	4	7	4	4	7	5	4	62	5	67
2	12	10	3	0	0	0	6	5	6	14	21	16	7	5	15	13	11	136	8	144
1	14	14	22	22	28	20	18	22	14	7	4	5	10	14	5	10	13	128	114	242
Totals	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	348	145	493

(table D.2 continued)

ID	u1	u2	t1	t2	t3	t4	u3	t5	u4	u5	u6	u7	u8	u9	u10	u11	u12	UTO	TO	Total	
DIFT Actual Negatives																					
5	4	4	7	9	6	8	5	5	0	5	0	0	0	0	0	0	0	18	35	53	
4	1	5	3	0	0	0	4	3	0	4	0	0	0	0	0	0	0	14	6	20	
3	4	1	0	0	0	0	1	0	7	1	3	0	0	0	0	0	0	17	0	17	
2	1	0	0	0	0	0	0	1	3	0	3	0	0	0	0	0	0	7	1	8	
1	0	0	0	1	4	2	0	1	0	0	4	0	0	0	0	0	0	4	8	12	
Totals	10	10	10	10	10	10	10	10	10	10	10	0	0	0	0	0	0	60	50	110	
DIFT Actual Positives																					
5	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	2	
4	0	2	0	0	0	0	2	0	0	1	1	0	0	0	0	0	0	6	0	6	
3	1	4	0	0	0	0	1	0	4	1	5	0	0	0	0	0	0	16	0	16	
2	9	4	0	0	0	0	4	2	3	6	2	0	0	0	0	0	0	28	2	30	
1	0	0	10	10	10	10	3	8	3	0	2	0	0	0	0	0	0	8	48	56	
Totals	10	10	10	10	10	10	10	10	10	10	10	0	0	0	0	0	0	60	50	110	
3D FOR Actual Negatives																					
5	0	3	-2	-21	-2	0	-3	-5	0	3	0	2	4	-11	1	1	3	3	-30	-27	
4	3	-4	-4	0	0	0	-1	-4	1	0	2	-2	-5	-8	-3	0	-6	-23	-8	-31	
3	-9	-13	-6	0	0	0	-2	-6	-4	-8	-7	-11	-12	0	-3	-4	-5	-78	-12	-90	
2	-11	-7	-8	0	0	0	-5	-6	-10	-14	-14	-9	-3	-3	-16	-15	-13	-120	-14	-134	
1	-4	0	-1	0	-19	-21	-10	0	-8	-2	-2	-1	-5	1	0	-3	0	-34	-41	-75	
Totals	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-21	-252	-105	-357	
3D FOR Actual Positives																					
5	0	0	0	-6	0	-6	0	1	0	0	0	0	1	-5	0	0	0	-4	-11	-15	
4	0	1	0	0	0	0	0	0	0	1	1	1	-5	-1	2	-1	0	-1	0	-1	
3	2	0	0	0	0	0	1	-1	-8	-7	0	0	-4	-1	2	0	-2	-17	-1	-18	
2	-2	-4	-3	0	0	0	-4	-3	-3	-6	-18	-12	-6	-7	-14	-11	-11	-98	-6	-104	
1	-19	-16	-16	-13	-19	-13	-16	-16	-8	-7	-2	-8	-5	-5	-9	-7	-6	-108	-77	-185	
Totals	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-19	-228	-95	-323	

## VITA

Kenneth Bernstein was born in New York City on November 1, 1967, in the same Greenwich Village neighborhood as former heavyweight boxing champion Gene Tunney. After fulltime employment as a licensed financial planner, an Alaskan fish processing plant worker, waiter, snowmaker and home renovator, he completed his undergraduate work at the University of Minnesota in 1992. He came to Baton Rouge in 1992 to work on the Super Conducting Super Collider. When the government cut the funding for the project, he began research in theoretical particle physics. He received his Master of Science degree in 1995 from Louisiana State University. In 1998 he began a fulltime faculty-teaching career with Baton Rouge Community College where he took a program with 20 students enrolled in physics and physical science to over 200 students by the time he left four years later. He reenrolled at LSU in Spring 2003 in the medical physics program. On August 11, 2005, he received his Master of Science in medical physics before moving to Des Moines, Iowa, to work as a radiation oncology medical physicist.