Virtual Clinical Trials in PET Imaging for Improved Diagnosis of and Evaluation of Therapies for Cancer

Kristen A. Wangerin

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Dr. Paul E. Kinahan, Chair

Dr. Ravindra Manjeshwar

Dr. Matthew O'Donnell

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Virtual Clinical Trials in PET Imaging for Improved Diagnosis of and Evaluation of Therapies for Cancer

Kristen A. Wangerin

Chair of the Supervisory Committee:
Dr. Paul E. Kinahan
Bioengineering

Positron emission tomography (PET) imaging is a diagnostic tool used to both quantify and verify the extent of disease, such as in cancer staging, and to monitor treatment response or measure disease progression. However, missed detection of tumors can lead to incorrect cancer staging and treatment selection, and uncertainty in quantifying tumor change can result in continuation of ineffective therapy. The objective of this work was to develop a virtual clinical trial methodology in PET imaging to enable evaluation of the impact of parameters on the final image analysis metrics. Numerous tools and techniques were combined, including radiotracer kinetic modeling, data generation simulations, image reconstruction algorithms, and human and model observer analyses. Improved understanding of parameters can be used to inform the

design of prospective clinical trials to improve detection of tumors and measurement of response to therapy and ultimately improve cancer patient management.

Lesion detectability using a new penalized likelihood (PL) reconstruction algorithm with a relative difference prior (RDP) was compared to that using OSEM, the standard clinical reconstruction algorithm. Due to the characteristics of the RDP, there was concern that low-contrast lesions were at risk of being smoothed into the background. Lesions in the liver and the lung were evaluated, and equivalent or improved detectability was demonstrated using the new algorithm.

Optimum imaging time post-radiotracer injection for detection was investigated. Previous studies have shown that the tumor uptake increases for many hours past the standard imaging time of one-hour post injection, while uptake in the normal tissue decreases. Noise, however, also increases with time. It was hypothesized that there should be a time when the tumor signal-to-noise, and thus detectability, would be maximized. Lesion detectability was found to increase for several hours, indicating that delayed PET imaging may reveal low-conspicuity lesions that would have otherwise gone undetected.

Finally, the uncertainties of static SUV and dynamic Ki metrics to measure change in tumor radiotracer uptake in response to therapy were characterized. For non-high-grade breast cancer tumors, it was found that Ki, while having higher variability, outperformed SUV in an ROC analysis of pre- and post-therapy parameter uncertainty distributions. Kinetic analysis, which accounts for metabolic state of the radiotracer, may better detect or enable earlier assessment of response to therapy, especially for low-uptake tumors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I, firstly, thank my parents who have always been there for me!

I thank my one and only sister, who I can and did call when things got overwhelming, and who knew I would persevere through whatever had and would come up.

I thank my boyfriend for his unending patience. Whether I had to work evenings or weekends, was tired or frustrated, or just needed to complain about something, he would listen and provide those needed words of encouragement.

I thank my nearby friends for their company when I needed that social time to decompress, for providing cheer when times got rough, and for their understanding when I had to decline extracurricular events in favor of research or writing. I thank my far away friends for the many phone conversations or in-person adventures that helped keep things in perspective.

I thank our collaborators at the SCCA, my co-workers at GE Global Research (and GE Healthcare), and all of my co-authors. Without these amazing research collaborations, this work would not have been possible.

I thank the members of the Imaging Research Laboratory for endless productive and insightful discussions, practice presentation feedback, baked goods, and unmatched camaraderie.

Finally, I thank my advisor for his amazing guidance and support through this journey!

DEDICATION

I dedicate my thesis to the loving memories of my grandmother, Lois Mills, and friends Al Richter and Stas Gray. Their incredible passion for and philosophy of life lives on in the memories and lives of those who knew them.

Chapter 1. INTRODUCTION

1.1 PET IMAGING

Positron emission tomography (PET) imaging is a type of nuclear medical imaging, which uses radioactive material to investigate a targeted cellular process. Because PET imaging investigates biological activity within the body, it is also known as functional or molecular imaging. It is complementary to x-ray based imaging, such as computed tomography (CT), which reveals anatomic structure based on density. As a result, the two imaging modalities are often combined to produce a superimposed image of both anatomic and functional information.

In PET imaging, a radioactive nuclide is attached to a biological tracer molecule, such as fluorine-18 attached to glucose to produce 2-fluoro-2-deoxy-D-glucose (FDG). Metabolic processes involving the molecule in vivo are observed without disturbing the natural biology. PET imaging can detect on the order of picomolar concentrations of the tracer, and this extremely high sensitivity has advantages over other types of functional imaging. Functional magnetic resonance imaging (fMRI), for comparison, has only micromolar sensitivity. This low sensitivity prevents the direct imaging of molecular receptors such as those sensitive to glucose and, instead, requires exogenous tracers such as a Gadolinium-based contrast agent [1]. While optical imaging has similar sensitivity limits as PET has, optical imaging has limited depth penetration [2].

1.2 APPLICATIONS

Two principal fields for application of PET imaging are neurology and oncology, where functional information is critical in detecting and assessing extent of disease. Though a detailed discussion is beyond the scope of this summary, some highlights of clinical applications and current research efforts are presented here.

1.2.1 *Clinical*

In neurology, PET is particularly valuable in imaging Alzheimer's disease, which manifests itself as decreased metabolism of both glucose and oxygen in the brain. Seizures and other central nervous system disorders can also be investigated.

In oncology, PET is not typically used for initial screening, but instead, when it is known that a patient has cancer, PET is used to determine if and to where the cancer may have metastasized. The number and location of metastases are important for proper disease staging and treatment planning. PET is also used to assess response to therapy. If the cancer is not responding to treatment, other treatment options can be considered. While other imaging modalities such as CT and MRI can also determine tumor location and size, PET imaging provides metabolic activity. Change in metabolic activity has been shown to indicate tumor response sooner than change in size [3]. However, large uncertainties in quantitation currently limit the use of PET for determining response to therapy [4]. Lastly, PET imaging is used to screen for cancer recurrence.

1.2.2 *Research*

Research and development efforts aimed at improving the performance of PET imaging in these tasks span a complex space of new tracers and applications to post-reconstruction image processing. Other areas include investigation of patient protocol and improvements in hardware, systems, and software.

Glucose metabolism is the most common biological process imaged with PET, but targeted radiotracers have been developed to image hypoxia, DNA synthesis, beta-amyloid protein in Alzheimer's disease, and estrogen receptor status in breast cancer [5,6]. In addition to the information biological tracers provide, there is ongoing research investigating the combination of PET imaging with other biological markers, either from pathology [7] or blood-based analyses [8].

The two main applications for PET are detection and quantitation. Detection of small features is partially limited by the spatial resolution of the system, and the potential to resolve finer image details has motivated the development of higher resolution scanners, particularly application-specific scanners for brain and breast imaging. Current whole-body systems have a resolution limit for detecting tumors between 4 and 10 mm in diameter [4], and low FDG uptake and partial volume effects can make these tumors even harder to detect [9]. Whereas a whole-body PET/CT scanner has a sensitivity of 75% for tumors from 1 to 10 mm in diameter, a dedicated breast PET scanner has a sensitivity of 93% (specificity for this study was not reported) [10]. Higher resolution scanners will improve tumor detection and enhance visualization of fine structures in both oncology and neurology applications. For improved use of PET imaging for measurement of response to therapy, quantitation accuracy is being improved through scanner calibrations and corrections and image reconstruction algorithms [11,12].

1.3 OVERVIEW OF DISSERTATION

1.3.1 *Motivation*

Despite consistent improvements in cancer outcomes [13], cancer remains a difficult disease to detect and treat; cancers are often detected at a late stage, and treatments are often ineffective. Missed detection of tumors can lead to incorrect cancer staging and treatment, and uncertainty in quantifying tumor change can result in continuation of ineffective therapy. Ideally, it would be straightforward to realize potential improvements of new technologies, therapies, and techniques developed to improve imaging tasks. Due to differences between patients and uncertainties in the imaging process, this realization is not straightforward. Evaluations are often made through prospective clinical trials, where patients are enrolled to measure some difference or impact between two study arms. Based on the magnitude of the difference to be observed, a certain number of patients are needed to measure that difference with some degree of certainty. While differences between patients are a fixed uncertainty, other uncertainties that confound the measurement and comparison of differences can be improved. By improving the imaging process, the chance of a successful clinical trial in evaluating or testing the differences between study arms can be improved.

PET imaging has been shown to improve patient management through more accurate disease staging and earlier assessment of response to therapy [14,15]. PET imaging can also assist in evaluation of new therapies undergoing clinical trials [16]. Due to its value in clinical trials and personalized medicine, much effort is aimed at increasing the quantitative accuracy of PET, and in particular the reliability of standard uptake values (SUVs) as a quantitative biomarker [12]. The sources of SUV instability are vast, including such disparate factors as calibrations, physics modeling, and if the patient ate breakfast instead of fasting. The constant evolution of PET as a quantitative measurement presents a challenge for researchers, whose interpretation of data necessarily relies on knowledge of the biases and variability of PET image metrics, which are often poorly characterized.

While the gold standard to evaluate any change in the imaging process is to perform a prospective study, clinical trials are costly, time-consuming, and must be designed carefully in order to be effective. Alternative methods include evaluation using retrospective patient data, a phantom study, or imaging simulation. The evaluation of some parameters is difficult to determine using patient data alone, as patient data lacks ground truth, and phantom studies and simulations can lack clinical realism. Therefore, there is a need to understand the factors limiting the accuracy of detection and quantitation in PET imaging, such as uncertainties in the imaging process, and there is a need for a method to do so.

1.3.2 *Proposed Solution*

The proposed solution in this work was to improve the ability of PET to guide patient management through the development and application of a virtual clinical trial (VCT) methodology. The VCT methodology was then applied to investigate three clinically driven hypotheses seeking to improve tumor detection and measurement of tumor response to therapy.

VCTs are a flexible and versatile tool that can be applied to many imaging scenarios and help reduce cost, time, and radiation dose associated with prospective clinical trials. The VCT methodology links a complete chain of simulation, modeling, and image analysis tools, and is therefore an important resource to characterize the uncertainties in the imaging process and the resulting impact of parameters on image quality metrics. However, in the evaluation of new imaging techniques or procedures, it is impractical to perform the large number of phantom studies or patient scans required to thoroughly do so. In VCTs, the imaging simulations are based on existing patient data with a known ground truth. Because ground truth is known, the contribution to and relationship of each study parameter on the final outcome can be independently evaluated in a way that could not be answered using patient data alone. Otherwise, the impact of any variable on the final study output can be obscured based on the unknown uncertainty distributions. This improved understanding can then be used to optimize PET imaging in a task-specific manner, as well as guide the design of prospective clinical trials, where new technologies and therapies are tested for evaluation before potentially being transitioned into clinical practice.

1.3.3 *Structure of Dissertation*

Chapter 2 summarizes all necessary background information. Chapter 3 describes how the numerous tools and techniques are applied to the four PET imaging studies of tumor detection and measurement of tumor response to therapy. Chapters 4 through 7 contain the four studies performed. Chapter 8 summarizes the outcomes of this thesis and describes limitations and future work. Chapter 9 presents final conclusions.

Chapter 2. BACKGROUND

2.1 PHYSICS

2.1.1 *Positron Emission*

PET imaging is based on position emission. A positron, denoted by β+, is the anti-matter equivalent of an electron, having the opposite charge but identical mass, rest energy, and charge magnitude. In positron decay, a proton is converted into a neutron with the release of a positron and a neutrino. The radioactive isotope most often used in PET imaging is Fluorine-18, which undergoes the decay reaction

$$ {}_{9}\mathrm{F}^{18} \rightarrow {}_{8}\mathrm{O}^{18} + \beta^{+} + \nu_{e}. \tag{2.1}$$

The energy released in this decay is divided between the energy of the neutron and the kinetic energy of the emitted positron. The mean kinetic energy of the positron produced in ^{18}F decay is 250 keV, and the maximum energy is 635 keV, which occurs when the neutrino energy is zero. These kinetic energies result in mean and maximum travel distances of 0.64 mm and 2.3 mm in water [17].

When the positron comes into contact with an electron, the two annihilate and produce two gamma-ray photons in the reaction

$$\beta^{+} + e^{-} \rightarrow \gamma + \gamma, \tag{2.2}$$

which must satisfy the momentum and energy laws of conservation. The photons are emitted in opposite directions, each with energy of 511 keV that is equal to their rest mass. Because neither the positron nor electron is completely at rest when they annihilate (their energies are on the order of 10 eV), the photon trajectories are slightly acollinear. The maximum possible deviation from 180° is ±0.25°, which adds 1.8 mm of resolution uncertainty to an 80 cm diameter PET system [18]. A narrower diameter system with shorter lines of response has less position uncertainty.

A schematic of positron emission in PET imaging is summarized in Figure 2.1 [19]. The positron emitting radionuclide is attached to a biological tracer molecule to form a radiotracer. The radiotracer is injected into the bloodstream of a patient and, as it is circulated throughout the body, it is preferentially taken up into cells based on the targeted biological mechanism. The patient is positioned within a scanner comprised of many rings of detector modules that detect the coincident photons and convert the energy into electronic signals. It is known that the origin of the decay is located somewhere along the line connecting the coincidence events. For example, if there were only one point of radioactivity in the field of view, the intersection of all the lines would reveal the location of that point. By detecting millions of events, enough information is obtained to reconstruct any radioactivity distribution.

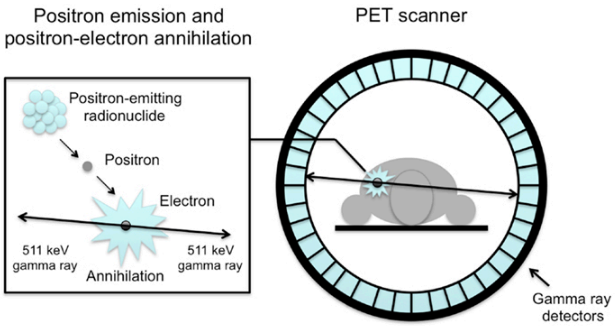


Figure 2.1. Schematic of position emission and detection of a coincident event within a ring of detectors [19].

2.1.2 *Attenuation*

One or both of the photons can interact with electrons within the patient, however, before reaching a detector. The combined effect of these interactions is called attenuation and is calculated using

$$N = N_0 e^{-\mu x}, \qquad (2.3)$$

where N is the original number of particles; N_o is the fraction of particles remaining after traveling a certain distance, x, in the absorbing material; and μ is the linear attenuation coefficient or probability of interacting per unit distance traveled. The linear attenuation coefficient includes the effects of three energy loss mechanisms: the photoelectric effect, Compton scattering, and pair production.

The photoelectric effect dominates at low energies of around 50 keV or less. In this process, the photon is absorbed, and an energetic photoelectron is ejected from one of the bound shells of the atom with energy equal to that of the original photon minus that of the binding energy of the electron. The filling of the inner shell vacancy with an electron from an outer, higher energy shell can produce an x-ray photon [20].

At higher energies of around 100 keV or greater, Compton scattering becomes important. As the energy of the photons from positron annihilation is 511 keV, Compton scattering is the dominant source of attenuation in PET imaging. In Compton scattering, the incoming photon interacts with an electron in the absorbing material and is deflected through an angle, θ, with respect to its incident direction. The photon transfers a portion of its energy to the electron, assumed to be initially at rest, causing this electron to recoil. The energy of the deflected photon is dependent upon the scattering angle [20]. The scattering of one or both of the coincident photons can significantly affect the ability to localize the source of the radioactive decay, and there are statistical correction algorithms to reduce the impact of scatter [21].

In pair production, the photon is absorbed and reemitted as two particles: a positron and electron. This process requires the energy of the incoming photon to be greater than 1.022 MeV, however, so it is not a factor in PET imaging [20].

2.2 BIOCHEMISTRY

The most common clinically used tracer is a glucose-analog, ^{18}F-FDG [22]. The chemical description of FDG is 2-deoxy-2-(^{18}F)fluoro-D-glucose, which indicates that the hydroxyl group at the 2' position in the glucose is replaced with a fluorine-18 atom. FDG, like glucose, enters the cells via facilitated diffusion through sodium-independent glucose transporters. These transporters, including GLUT-1 and to a lesser extent GLUT-3 and GLUT-12, are often overexpressed in tumor cells, resulting in preferential uptake of glucose into tumor cells in comparison to normal cells [23].

Once in the cell, glucose is most often metabolized through the glycolytic pathway, which provides energy for the cell. As shown in Figure 2.2, both glucose and FDG are first phosphorylated to glucose-6-phosphate (G6P) and FDG-6-P, respectively, by hexokinase. This reaction is reversible with glucose-6-phosphatase, which catalyzes the dephosphorylation of G6P or FDG-6-P back to glucose or FDG, but the concentration of this enzyme in most cells is quite small and the phosphorylation step is considered irreversible. Therefore, G6P and FDG can only leave the cell by being further metabolized through the glycolytic pathway. After phosphorylation, G6P is converted to fructose-6-phosphate by phosphoglucose (or phosphohexose) isomerase via the 2' hydroxyl group. However, FDG-6-P is missing this 2' hydroxyl and so is not a substrate for phosphoglucose and cannot be broken down through the glycolytic pathway until undergoing radioactive decay. Further, FDG-6-P is a highly polar molecule and is therefore unable to diffuse out of the cell [24]. As a result of these trapping mechanisms, the concentration of FDG increases in the cell. Further, it has been shown that FDG is preferentially taken up into tumor cells in comparison to normal cells, to support the continued growth of the tumor [25].

There are alternative pathways for metabolism of glucose, including the pentose phosphate pathway for production of NADPH (a reducing agent) and ribulose-5-phosphate (nucleotide precursors of DNA for growth and synthesis), but there are few actual studies of metabolic pathway of FDG. Small amounts of FDG metabolites have been found in brain in liver tissue in prolonged studies, indicating FDG can be slowly metabolized before radioactive decay [26], but neither this pathway nor dephosphorylation is considered to be of quantitative importance in the typical imaging time of 45 to 60 minutes [27].

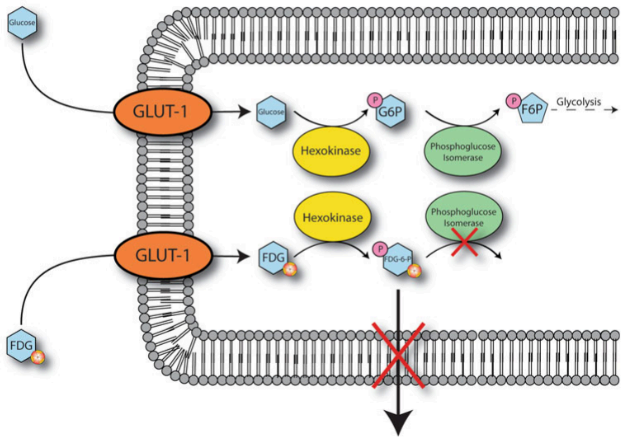


Figure 2.2. Comparison of metabolism for FDG compared to glucose [23].

2.3 IMAGING PROCESS

There are many steps in performing PET imaging studies as shown in Figure 2.3. The first step is generating the data from a given object inside of a scanner. Once the data is acquired, an image reconstruction algorithm is used to generate the image. Finally, the image is used to perform a given task such as detection. Each of these components are described in more detail in the following sections.

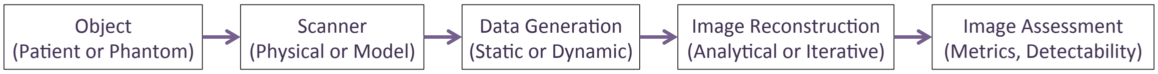
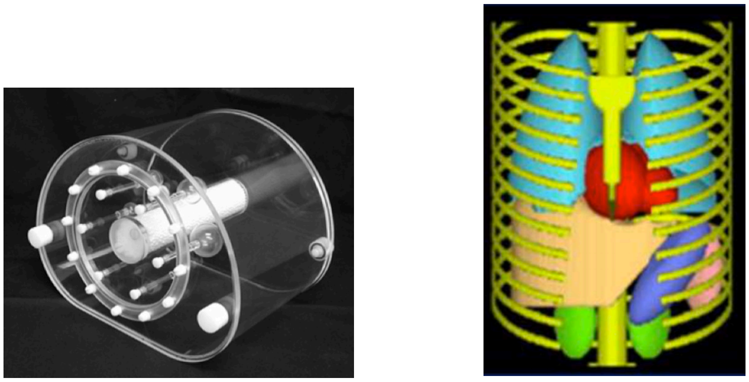


Figure 2.3. Steps in the imaging process.

2.4 OBJECT

PET imaging requires an object with a radiotracer distribution such as a human patient, a small animal, or a phantom. One benefit of an actual patient study is that the imaging is most realistic to the final clinical scenario. However, there are limitations to performing a patient study, including radiation dose, time and cost of scanning, and the absence of ground truth. In the case of a clinical trial, it can be difficult to recruit and enroll patients. Therefore, phantoms are often used, where a phantom is a specially designed object to characterize some aspect or aspects of the imaging process. Phantom studies are particularly useful for generating many data realizations, which, for example, can be used to characterize noise or test-retest repeatability.

One commonly used phantom is the NEMA image quality phantom, shown in Figure 2.4(a), which is used to assess different aspects of system performance, including background variability and contrast recovery [28]. This NEMA phantom can be physically scanned or simulated using a digital representation of it. Another example is the digital MCAT anthropomorphic torso phantom [29], shown in Figure 2.4(b), where specific organs and tissues are assigned different radiotracer uptake values and attenuation properties to be representative of an actual patient.



(a) (b)

Figure 2.4. Examples of imaging phantoms: (a) NEMA image quality phantom [28] and (b) MCAT torso phantom [29].

2.5 SCANNER AND DATA GENERATION

Data can be acquired using a physical scanner or simulated using a virtual scanner. While acquired data is the gold standard, benefits to simulating scanner performance include preliminary investigation of new scanner components or geometries.

2.5.1 *Data Acquisition*

Data acquisition can largely be divided into static and dynamic acquisitions. Static acquisitions are the clinical standard and provide a spatial distribution of the radiotracer uptake at a single time point. Patient data are acquired at roughly 60 minutes post injection over a two to five minute time window. The axial field of view of the scanner ranges from 15 to 25 cm, so a whole body scan requires repeat acquisitions at different patient bed positions.

Dynamic acquisitions provide considerably more information about the biology and physiology of the patient by providing the spatial distribution of the radiotracer uptake as a function of time. For a dynamic patient study, data acquisition typically begins immediately after the radiotracer injection and continues for 45 minutes or longer. The data is binned into different time frames. The time frame duration is initially very short while the radiotracer distribution is changing rapidly, and they are gradually lengthened as the distribution changes more slowly. An example of a 60-minute dynamic acquisition sequence is 15×5-sec, 5×15-sec, 5×30-sec, 4×1-min, 4×3-min, and 9×5-min time frames [30]. It is more difficult to incorporate the dynamic nature of radiotracer uptake into a simulation study, but it can be done by performing each time point separately using tracer uptake information based on dynamic patient studies.

2.5.2 *Data Simulation*

There are different methods for simulating data, with the trade-off of computation time and complexity. One software package for simulating the scanner and generating data is SimSET [31]. SimSET uses Monte Carlo techniques to track the interactions of individual photons through the object. SimSET then models the physics that occur in the detectors and outputs the information on those interactions.

Another software package is ASIM [32], which uses analytical techniques based on line integrals through the object in the imaging field of view to obtain the emission data. An input phantom is combined with the model scanner definition for comprehensive simulation of most physical and instrumentation factors impacting the data collection. For example, Figure 2.5 shows the MCAT phantom implementation in ASIM with a spherical 10 mm diameter breast cancer lesion. ASIM has the functionality to include many of the system physics impact the simulation of emission data, including attenuation, randoms, scatter, and detector efficiencies. The advantage of an analytic simulation is the speed compared to photon tracking simulations, as scanner effects are calculated rather than statistically estimated using Monte Carlo techniques.

Therefore, ASIM is particularly well suited to the rapid generation of multiple noise realizations with realistic noise and resolution properties. However, the lack of randomized photon tracking is also the main drawback of ASIM. The statistical element of the data acquisition must instead be added to the initial noise-free projection by sampling from a Poisson distribution based on the number of detected counts.

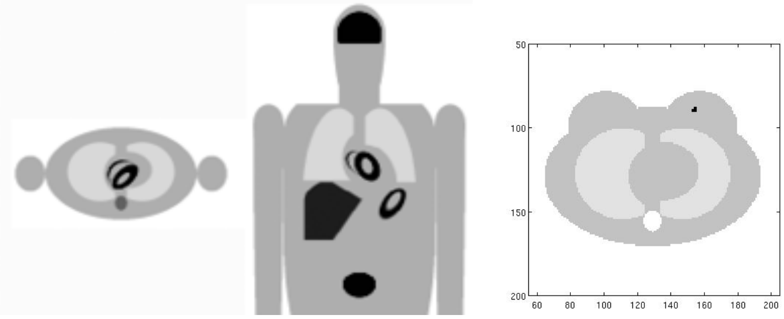


Figure 2.5. Adapted MCAT phantom for input into ASIM (left) and inserted breast cancer lesion (right).

2.5.3 *Hybrid Data Generation*

One benefit to phantom-based data is that the ground truth of the feature sizes, shapes, locations, and signals are known. However, the drawback is that, for either acquired or simulated phantom studies, the data will never completely represent or replace real patient data. One solution is to merge patient data with a ground truth lesion. A digital phantom lesion is forward projected into data space while taking into account the physics and scanner effects. The synthetic lesion data is then added to the patient data. Similarly, a physical sphere can be imaged separately from a patient, and the data later combined. Common locations for inserting lesions are into the liver and the lung, due to their differing properties with respect to activity and noise texture in the reconstructed image. An example application of this method is the comparison of different reconstruction algorithms and parameters.

2.5.4 *Sinograms*

PET data is represented in what is called a sinogram. The sinogram is a collection of projections, p, where each projection is made up of line integrals through an object as a function of angle. Each row is a unique angle, ϕ, and each column is the radial distance, s, from the origin [33]. Figure 2.6 shows an example projection for one angle and how the lesion is represented in the sinogram.

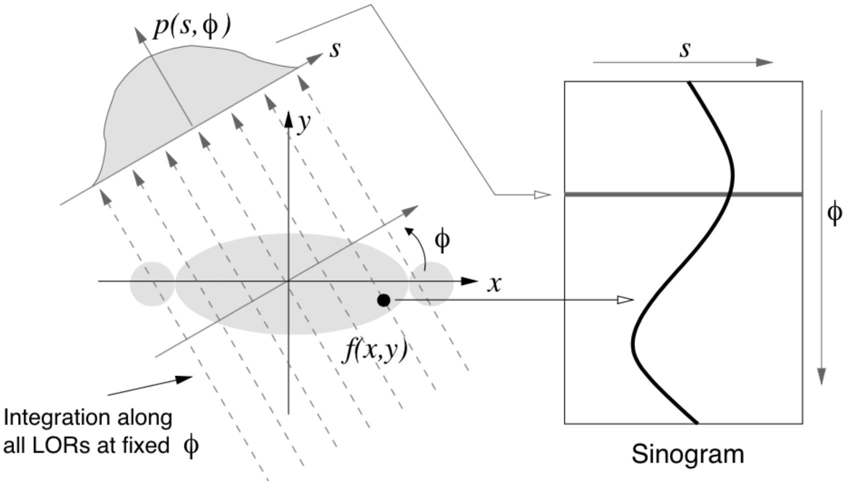


Figure 2.6. Example image projection and how the sinogram looks like for that projection, including the representation of the lesion [33].

2.6 IMAGE RECONSTRUCTION

Once the data are acquired or simulated, the next step is the image reconstruction. Image reconstruction is an inverse problem defined by the imaging equation,

$$\bar{y} = A\,x + n, \tag{2.4}$$

which describes the relationship of the image, x, with that of that mean acquired data, $\bar{y}$. A is the system model, which includes all of the physics effects, and n is the error in the data observations [34].

Reconstruction algorithms can be analytical or iterative. The most common analytical method is filtered backprojection (FBP), which reconstructs an image using the direct inversion of the Radon transform that describes the relationship between the projections and line integrals through the object [35]. Analytical methods, however, allow for only simple statistical modeling before the mathematical problem becomes quite complex. In order to more fully model non-idealities in the imaging system, iterative methods are used. Iterative methods link the estimated image and measured data through the system model. Although there are benefits to iterative reconstruction, these algorithms do require more computing power.

2.6.1 *Iterative Reconstruction*

All iterative methods contain five components: a model for the image, a system model, the data measurement, an objective function, and an algorithm to maximize the objective function [36]. The image is typically modeled as voxels, but they can also be modeled as pixels or blobs. The system model describes the effects that relate the object or image to the data, including patient attenuation and scanner geometry [37]. The data is typically modeled as a Poisson or Gaussian process. The objective function defines what is considered the best image estimate based on the acquired data and defined model. The algorithm then determines how that image estimate is achieved.

The Poisson noise model is often a good description of the data acquired in PET imaging, where

$$L(Y = y|x) = \prod_{i=1}^{M} \frac{\bar{y}_i{}^{y_i} e^{-\bar{y}_i}}{y_i!}, \tag{2.5}$$

describes the likelihood of obtaining the random variable, Y, given the observed data, y, and the object or image, x. The likelihood is based on multiplication of M probabilities, where each one of these probabilities is based on the individual measurement and mean of a data point, y_i and $\bar{y}$, respectively [61].

The objective function can be defined as simply as the maximum of the data likelihood function. The natural logarithm of the likelihood function often simplifies the problem and so the log-likelihood is used [61]. Equation 2.5 becomes

$$L(Y = y|x) = \sum_{i=1}^{M} y_i \log(\bar{y}_i) - \bar{y}_i - \log(y_i!). \tag{2.6}$$

The objective function can then be optimized using maximum likelihood or maximum *a posteriori* (MAP) estimation.

2.6.2 *Maximum Likelihood*

Maximum-likelihood (ML) optimization is a method to estimate a desired set of parameters using a statistical model and observed data. In tomography, the parameters to be estimated are the image voxels, the statistical model is the Poisson noise model, and the measured data are the acquired projections. ML maximizes the log-likelihood function to produce an estimate of the image [38]. The estimated mean image, $\hat{x}$, is

$$\hat{x}_{ML} = arg\,max_{x \geq 0} \log(f(Y = y|x). \tag{2.7}$$

The main drawback to ML is that the solutions are unstable because of the low number of counts per image voxel. Therefore, the solution of the imaging equation does not exist or is not unique, and many iterations are required for the image to reach convergence. The image noise also increases with iteration number [39].

2.6.3 *Maximum a Posteriori*

To solve the problem of ill-posedness, *Maximum a Posteriori* (MAP) optimization adds a regularization term to the likelihood function, where prior information is incorporated through a penalty function. The MAP estimate is therefore the posterior probability that is conditioned on both the log-likelihood function and the log of the assumed prior distribution,

$$\hat{x}_{MAP} = arg\,max_{x\geq 0}\ \log(f(Y=y|x) + \log\left(f(x)\right), \tag{2.8}$$

where $f(x)$ is the *a priori* distribution [61]. The benefit of MAP image reconstruction is that the image reconstruction problems can be convergent, and the parameters can be more predictably tuned to achieve desired properties, such as low image noise. Drawbacks of MAP image reconstruction include more intensive computational demands and undesirable image texture [40]. MAP and penalized maximum likelihood (PL) terminologies are often used interchangeably.

2.6.4 *Expectation Maximization*

Expectation maximization (EM) is a common algorithm for performing the optimization. The acquired data from the object represents a subset of the complete data space. The EM algorithm maximizes the conditional expectation of the complete data, given the image estimate. The comparison of the conditional expectation of the complete data to the acquired data is used to update the estimated image. The well-known EM update equation is

$$x_k^{n+1} = x_k^n \frac{1}{\sum_{j=1}^{N} a_{j,k}} \sum_{j=1}^{N} a_{j,k} \frac{y_j}{\sum_{k'=1}^{M} a_{j,k'}\, x_{k'}^{n}}. \tag{2.9}$$

where *x* the image estimate, *y* is the data, *a* is the system matrix, j are the pixels, and k are the projections [41].

In MLEM iterative image reconstruction, the first step is to assume an image estimate, typically assumed to be all ones. Each iteration then comprises four steps. The estimated image is forward projected through the system matrix to generate the corresponding estimated data. A ratio is taken of the measured data to the estimated data. This ratio is backprojected and then normalized with the system matrix. Finally, this term is multiplied to the current image estimate to form the new image estimate. The new image estimate becomes the current image estimate in the next iteration.

2.6.5 *Ordered Subsets Expectation Maximization*

Ordered subsets expectation maximization (OSEM) is a block iterative implementation of MLEM [42], where the data are divided into subsets. Each subset should contain an adequate number of well-distributed projections. The image estimate is updated using each subset of data, and the use of all of the data comprises one iteration. The benefit of OSEM is that it dramatically speeds image convergence compared to the MLEM algorithm by a factor approximately equal to the number of subsets. However, OSEM does not generally converge to the maximum likelihood solution [43].

2.7 IMAGE ASSESSMENT

Image quality is assessed by how well the desired information can be extracted from the reconstructed image. There are many influencing factors that result in distinct noise and signal properties of an image. Numerous metrics can be used to analyze the properties of the image, and techniques have been developed to quantify the ability of a specific task to be performed. For example, classification tasks include signal detection or discrimination, such as the signal being present or not present, and estimation tasks involve one or more parameters that describe the object, such as tumor size, location, or activity. Image assessment techniques employed in the research studies of this work are summarized here.

2.7.1 *Image Metrics*

Image metrics have been defined to assess both noise and signal properties of a reconstructed image. These enable systematic evaluation or comparison of objects, scanners, image reconstruction algorithms, and post-processing.

Metrics to characterize noise include image roughness, background variability, ensemble noise, and the standard deviation image [71]. Some of these metrics can be calculated on a single image, and others are calculated over multiple noise realizations. Image roughness (IR), which defines the pixel-to-pixel variability in an image, for example, is first calculated over an ROI in a single image as

$$\mathrm{IR}_{r,k} = \frac{\sqrt{\frac{1}{I-1}\sum_{i\in ROI_k}\left(f_{ir}-m_{r,k}\right)^2}}{m_{r,k}}, \tag{2.10}$$

where r is an image realization, k represents each ROI, i represents the pixels in the ROI, I is the toal number of pixels in the ROI, f is the reconstructed pixel value, and m is the mean value of the pixels in the ROI of a given image. An average IR can be calculated for all ROIs in the image as

$$\mathrm{IR}_r = \frac{1}{K}\sum_{k=1}^{K}\mathrm{IR}_{r,k}. \tag{2.11}$$

Finally, the mean and standard deviation of IR can be taken across different image realizations, r.

The image signal is characterized by the mean or maximum activity of pixels in a region of interest. One metric is the contract recovery coefficient (CRC), which is defined by the ratio of the measured activity divided by the true activity. Therefore, this metric can only be used when ground truth is known. Another common metric is the standardized uptake value (SUV), which is defined by the activity normalized by the injected dose of radiotracer and radiotracer distribution throughout the body, or

$$\mathrm{SUV} = \frac{A\ (\mathrm{kBq/mL})}{\frac{D\ (\mathrm{MBq})}{W\ (\mathrm{kg})}}, \tag{2.12}$$

where A is the radioactivity activity concentration measured by the PET scanner within a ROI, D is the decay-corrected amount of injected radiolabeled FDG, and W is the approximate distribution volume of the tracer. Patient weight is often used as the distribution volume of the tracer, but lean body mass has been shown to be a potentially more accurate measure, as it accounts for lower uptake of FDG by adipose tissue [12]. In a phantom simulation, when the true activity is known, the activity divided by the true sphere activity is the contrast recovery coefficient (CRC).

Unfortunately, many factors such as scanner calibration, reconstruction parameters, tumor heterogeneity, and ROI definition strongly influence the SUV estimate [44], and the uncertainty can be quite high with a test-retest variability of 10 to 12% [45]. The SUV can be based on different ROI shapes as well as maximum or mean values within the entire ROI or in a smaller "peak" region around where the activity is highest [46]. The maximum SUV is most often reported, but it is not necessarily the best choice; it can be highly impacted by image noise and voxel size, leading to more bias and uncertainty in the estimate [4]. The mean SUV also has drawbacks in that it is more sensitive to the ROI definition and is not as reproducible [12].

2.7.2 *Human Observers*

Human observers are the gold standard in image quality analysis, as radiologists interpret patient images. The impacts of modifications to the imaging progress on the ultimate decision-making in the clinic are important to understand. To characterize the detection performance from two or more imaging systems, a detection study can be conducted, where a human observer performs a specific task. One example is a yes-no detection task, where one image is presented to the observer, and the observer must decide if a signal is present or absent. Another example is a two-alternative forced-choice (2AFC) detection task, where two images are presented to the observer, and the observer must decide which image contains the signal [47]. The proportion of correct responses gives the percent correct (PC). The PC can be converted to a detectability index, which is an estimate of the human signal-to-noise ratio, using

$$d_A = 2\ \mathrm{erf}^{-1}(2\ PC - 1), \tag{2.13}$$

assuming that the underlying probability distributions are binormal [48]. The PC an be compared between two systems. The location of the signal can be known or unknown. The detection task can be made more complicated and realistic by having the observer identify the location of one or more tumors in the image.

Instead of a binary decision of the lesion being present or absent, the observer can rate his or her confidence about the presence of the lesion using a scale from 0 to 10, for example. This produces two vectors with the test scores for the lesion present and absent images, respectively. A receiver operating characteristic curve analysis can be performed using this test results, which is described in more detail in section 2.7.4.

If a nonrandom signal is imposed on a nonrandom background, the detection task is signal-is-known exactly (SKE). If any of the parameters are only known statistically, such as the tumor size or signal magnitude, the task is signal known statistically (SKS). Similarly, the background characteristics can also be background known exactly (BKE) or statistically (BKS) [49].

2.7.3 *Computer Model Observers*

Human observer studies are costly and time intensive, requiring many observers and images for statistical accuracy. Therefore, mathematical model observers are often used to mimic and predict human observers. A model observer computes scalar test statistics, λ_0 and λ_1, for image vectors, f_0 and f_1, where 0 and 1 indicate lesion-absent and lesion-present, respectively. Each test statistic is compared to a threshold value to classify the image as lesion present or absent. When the test statistic is greater than the threshold, a positive (lesion present) decision is made.

Model observer types include ideal, ideal linear, and anthropomorphic. The non-linear Bayesian ideal observer has full knowledge of the image statistics and produces the maximum possible area under the curve. The test statistic for the ideal observer is the likelihood ratio, or the ratio of the probability of making a particular decision in the case of the null hypothesis (signal absent) compared to the alternative hypothesis (signal present). The ideal observer, however, is difficult to calculate in realistic imaging scenarios because the likelihood probabilities are complicated or unknown, and it has little predictive value of human performance [50]. As an alternative, the Hotelling observer linearizes the model observer task. With full knowledge of the noise and signal statistics, the Hotelling observer is the ideal linear observer. It uses the means and covariance matrices of the signal-absent and signal-present image classes to form the discriminant function, which is used to calculate the test statistics. However, the covariance matrix is difficult to find, and once found, it is even more difficult to invert. The difficulty to invert the covariance matrix is the main drawback for this model [51]. A nonprewhitening observer is a linear observer that does not incorporate the covariance matrix. As a result, this observer may not perform well in situation where there are significant noise correlations [50]. Finally, the anthropomorphic observer, such as the channelized Hotelling (CHO) observer, incorporates aspects of the human visual system. This observer has many advantages over the ideal or Hotelling observers, including that it is computationally simpler and is designed to more closely predict human observer performance [52].

Model observers have been studied for over two decades [49,53], and the focus in this work is not to develop a new observer but to apply existing methods to new imaging scenarios. The two model observers that are applied in this work are the nonprewhitening (NPW) and channelized Hotelling (CHO) model observers. NPW was chosen because of its simplicity of implementation, and CHO was chosen because of its flexibility in implementation and good agreement with humans over a wide range of imaging scenarios.

For the NPW observer model, the image template is a matched filter, or the difference of the mean image vectors,

$$\mathrm{w}_{NPW} = \overline{f}_1 - \overline{f}_0. \tag{2.14}$$

This template is then applied to each image separately as

$$\lambda_0 = w^t f_0 \ \text{ and } \ \lambda_1 = w^t f_1, \tag{2.15}$$

to obtain the lesion absent and lesion present test statistics, respectively [54].

For the CHO observer model, the image template is

$$\mathrm{w}_{CHO} = K^{-1} U\left(\overline{f}_1 - \overline{f}_0\right), \tag{2.16}$$

which includes a prewhitening step that is the inverse of the covariance matrix, K, as well as frequency-based channels that act on the difference of the mean image vectors. Common channel types are rectangular, Laguerre-Gauss, Gabor, difference of Gaussian, and difference of mesa [49]. The test statistic is found by applying the template to the feature vector, which is the output of the channels applied to each image. The feature vector is calculated as

$$\nu_0 = U f_0 \ \text{ and } \ \nu_1 = U f_1, \tag{2.17}$$

and the test statistic is calculated as

$$\lambda_0 = w^t \nu_0 \ \text{ and } \ \lambda_1 = w^t \nu_1. \tag{2.18}$$

A signal-to-noise (SNR) metric can be calculated using

$$\mathrm{SNR} = d' = \frac{\langle \lambda^+ \rangle - \langle \lambda^- \rangle}{\sqrt{\frac{1}{2}\left(\sigma_{\lambda^+}^2 + \sigma_{\lambda^-}^2\right)}}, \tag{2.19}$$

where $\langle \lambda^{+} \rangle$ and $\langle \lambda^{-} \rangle$ are the mean of the test statistics with results for the lesion present and lesion absent images, and $\sigma_{\lambda^{+}}^{2}$ and $\sigma_{\lambda^{-}}^{2}$ are the variances of the test statistics [55]. The variance in the SNR metric is calculated using

$$\sigma_{d'}^{2} = \frac{2}{\sigma_{\lambda^{+}}^{2}+\sigma_{\lambda^{-}}^{2}}\left[\frac{\sigma_{\lambda^{+}}^{2}}{N_1}+\frac{\sigma_{\lambda^{-}}^{2}}{N_2}+\frac{\left(\mathrm{d}'/2\right)^{2}}{\sigma_{\lambda^{+}}^{2}+\sigma_{\lambda^{-}}^{2}}\left(\frac{\sigma_{\lambda^{+}}^{4}}{N_1-1}+\frac{\sigma_{\lambda^{-}}^{4}}{N_2-1}\right)\right], \tag{2.20}$$

where N is the number of samples [70]. The SNR is one metric to assess observer performance, and another is a receiver operating characteristic curve analysis.

2.7.4 *Receiver Operating Characteristic Curve*

Receiver operating characteristic (ROC) curve analysis provides a reproducible, statistically quantitative measure to characterize performance of a specific task, allowing the comparison of systems and image quality [56]. The ROC curve is a plot of the true positive versus false positive fraction, or specificity versus 1-sensitivity, as the decision threshold is swept from -∞ to +∞ over the lesion absent and lesion present distributions. An example is shown in Figure 2.7. System performance can be evaluated by comparing the true positive and false positive rates for particular operating points on the ROC curve. Performance across systems can be compared using the area under the ROC curve (AUC or AUROC). Random chance has an AUC of 0.5, and a perfect system has an AUC of 1.

The AUC is related to the SNR by

$$\mathrm{AUC} = \Phi\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right), \tag{2.21}$$

assuming that the test statistics are normally distributed and that the variance is the same for both image classes [57].

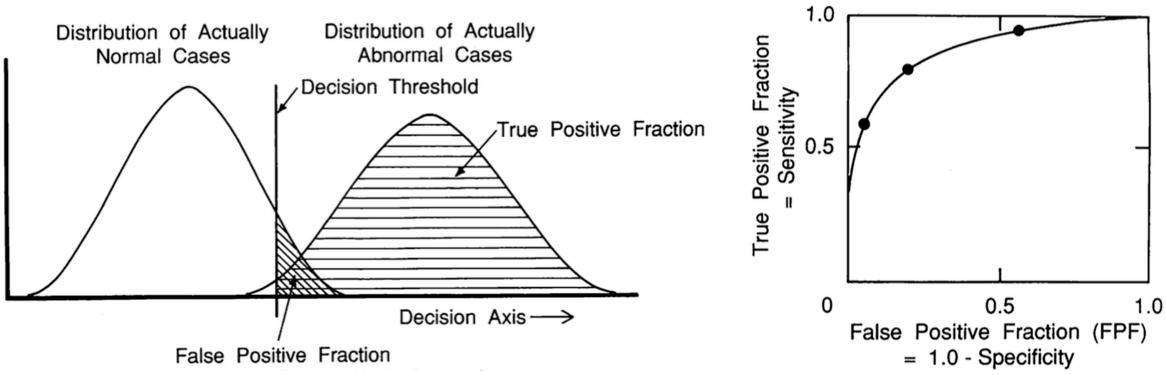


Figure 2.7. A plot of the actually normal (lesion absent) and actually abnormal (lesion present) distributions (left) and the ROC curve obtained by sweeping a decision threshold across the decision axis (right) [56]. The three points on the ROC curve represent three operating points with different sensitivity and specificity trade-offs.

2.8 TRACER KINETIC MODELING

Dynamic PET imaging measures radiotracer uptake as a function of time, and tracer kinetic modeling is used to estimate relevant physiological information about the radiotracer. For example, different metabolic forms or binding states of the radiotracer can be distinguished. There are different types of kinetic models, and a detailed discussion is beyond the scope of this work. Compartmental modeling is the most common method, as it provides a balance between model complexity and the ability to estimate biological parameters [58].

In compartmental modeling, each compartment represents a tissue or metabolic form of the radiotracer within the body. The compartments are connected, and the rates at which the contents are exchanged from one compartment to another is estimated. A few simplifying assumptions are made. One assumption is that each compartment is well mixed, meaning that it is homogenous and that there are no spatial concentration gradients. Another assumption is that the underlying physiological processes are in steady state, meaning that the rate constants of the system are not changing. Lastly it is important and assumed that the tracer itself does not perturb the physiological system under study [168].

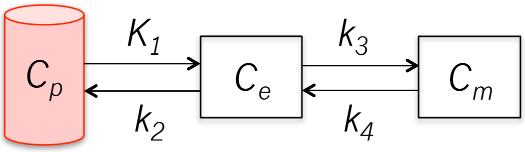


Figure 2.8. Two-tissue compartment kinetic model, where C_p is the plasma compartment and C_e and C_m are the exchangeable and metabolized (or bound or trapped) tissue compartments. $K_1, k_2, k_3,$ and k_4 are the rate constants describing the flow of the compartmental contents between compartments.

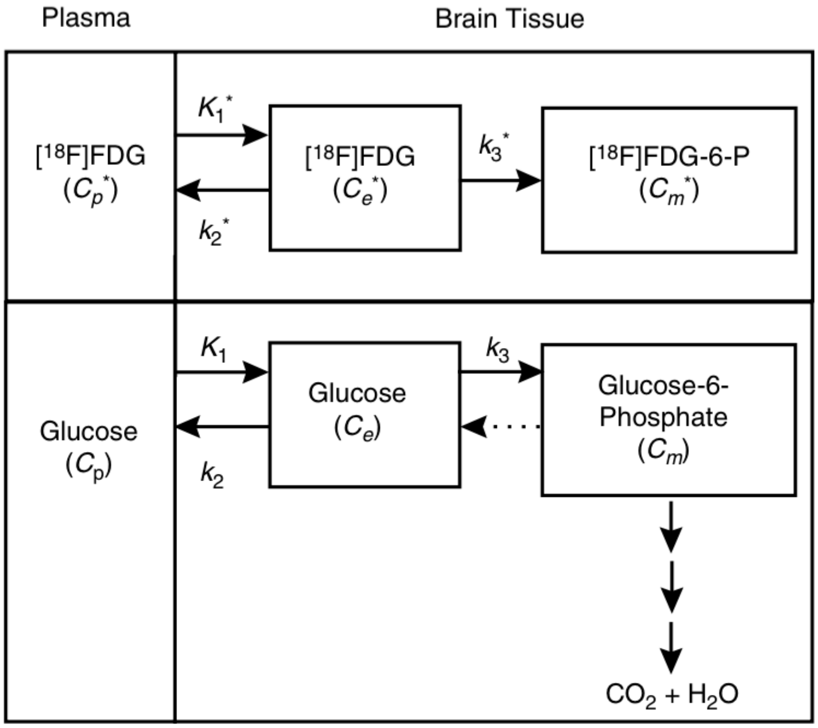


Figure 2.9. Two-tissue compartmental kinetic model of FDG compared to that of glucose [59].

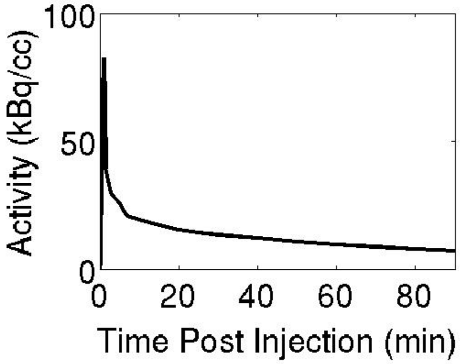


Figure 2.10. Averaged population arterial input function.

For the modeling of FDG, a two-tissue compartment model is assumed, a general example of which is shown in Figure 2.8. A more detailed kinetic model for both glucose and FDG are shown and compared in Figure 2.9. Both compounds begin in the plasma compartment, C_p. Both enter the tissue exchangeable compartment, C_e, with a rate of K_1 and return to the plasma with a rate of k_2. Both can be phosphorylated with a rate of k_3, moving to the metabolized compartment, C_m [59]. As described in Section 2.2, whereas glucose can be dephosphorylated with a rate of k_4, FDG is trapped in the metabolized compartment.

The arterial input function (AIF) is a measurement of the concentration of the tracer in the plasma, which is represented by C_p in the compartment model. The AIF data are ideally obtained from measurements of arterial blood samples from the patient as a function of time. The AIF can also be estimated from the dynamic images by drawing a region of interest over the left ventricle of the heart [166]. Alternatively, a population-based input function, averaged from a number of previously acquired AIFs, can be used and appropriately scaled for an individual patient. An example population AIF that has been averaged over many patients is shown in Figure 2.10. The peak represents the bolus injection of the tracer, and the tail represents the slow washout of the tracer.

The relationship of the rate constants, or kinetic parameters, and the compartments can be described mathematically using an ordinary differential equation for each compartment and one equation describing the relationship between the compartments. For a two-tissue compartment model, the change in tracer concentration with time in the exchangeable and metabolized compartments is expressed as

$$\frac{\partial C_e}{\partial t} = K_1 C_p(t) - k_2 C_e(t) - k_3 C_e(t) + k_4 C_m(t) \tag{2.22}$$

$$\frac{\partial C_m}{\partial t} = k_3 C_e(t) - k_4 C_m(t) \tag{2.23}$$

and the plasma, exchangeable, and metabolized compartments are related as

$$C_t(t) = V_b C_p + (1 - V_b)(C_e + C_m), \tag{2.24}$$

where V_b is the fractional volume of blood in the tissue.

The solution to these equations describes the concentration of the tracer in the exchangeable and metabolized compartments as a function of time as

$$C_e(t) = C_p K_1 \mathrm{e}^{-(k_2+k_3)t} \tag{2.25}$$

$$C_m(t) = C_p \frac{K_1 k_3}{k_2+k_3}\left(1 - \mathrm{e}^{-(k_2+k_3)t}\right). \tag{2.26}$$

The constant $\frac{K_1 k_3}{k_2+k_3}$ is defined as K_i and describes the net metabolic rate of tracer uptake into the metabolized compartment [58].

In PET imaging, SUV measurements are acquired from a region of interest in the dynamic PET reconstructed images. The SUVs reflect the combined radiotracer concentration in both the exchangeable and metabolized compartments, as the signal from radioactive decay of FDG and FDG-6-P cannot be discriminated. An optimization algorithm is used along with the SUVs, AIF, and compartmental model to estimate the tracer kinetic parameters in the region of interest.

Chapter 3. OVERVIEW OF PERFORMED STUDIES

The VCT methodology was applied to four studies. These studies investigated different clinically relevant questions, but all possessed a similar framework requiring common tools and skills as summarized in Table 3.1. A list of all acronyms in this work is provided in Table 3.2.

The first and second studies characterized tumor detection using a new penalized likelihood reconstruction algorithm in comparison to the standard clinical algorithm, OSEM. The retrospective patient cohort included 81 whole-body ^{18}F-FDG PET imaging datasets. Synthetic, ground truth lesions were added to the original patient data. After image reconstruction, human and model observer studies were performed followed by an ROC analysis to evaluate the difference between algorithms.

The third study evaluated the impact of patient imaging time post-radiotracer injection on tumor detection, with the goal of finding the optimum imaging time. In particular, delayed imaging after the current standard imaging time of 60 minutes was investigated. The fourth study quantified the ability of two metrics, SUV and Ki, to measure tumor change in response to therapy. Both of these studies utilized a dynamic ^{18}F-FDG breast cancer patient dataset, consisting of 24 non-high-grade, ER+ tumors, and relied heavily on tracer kinetic modeling and data simulations. In the optimum imaging time study, a model observer was applied followed by an ROC detectability analysis. In the response to therapy study, an ROC analysis was performed to quantify the ability of SUV versus Ki metrics to measure change in tumor tracer uptake between pre- and post-therapy scans.

Table 3.1. Tools applied in each of the VCT PET imaging studies.

	Penalized Likelihood Detectability (2 studies)	Optimum Imaging Time	Response to Therapy
Patient Data	X	X	X
Kinetic Modeling		X	X
Data Simulation		X	X
Synthetic Lesions	X		
Image Reconstruction	X	X	X
Human Observers	X		
Model Observers	X	X	
ROC Analysis	X	X	X

Table 3.2. List of acronyms.

AFC	Alternative forced choice
AIF	Arterial input function
ASIM	Analytical PET simulation software
AUAIF	Area under the arterial input function
AUC	Area under the curve
AUROC	Area under the receiving operating characteristic
BKE	Background known exactly
BKS	Background known statistically
BSREM	Block sequential regularized expectation maximization
C	Contrast
CHO	Channelized Hotelling
CNR	Contrast to noise ratio
CRC	Contrast recovery coefficient
CT	Computed tomography
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
ER	Estrogen receptor
FBP	Filtered back projection
FDG	Fluorodeoxyglucose
FDG-6-P	Fluorodeoxyglucose-6-phosphate
FLT	Fluorothymidine
fMRI	Functional magnetic resonance imaging
FWHM	Full width at half maximum

G6P	Glucose-6-phosphate
GE	General Electric
GUI	Graphical user interface
HER2	Human epidermal growth factor receptor
IR	Image roughness
IRB	Institutional review board
LABC	Locally advanced breast cancer
MAP	Maximum *a posteriori*
MATLAB	Matrix-based numerical computing environment
MCAT	Mathematical cardiac torso
MLEM	Maximum likelihood estimation maximization
MRI	Magnetic resonance imaging
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NEMA	Positron emission tomography
NPW	Non-prewhitening
NPWMF	Non-prewhitening matched filter
OSEM	Ordered subsets expectation maximization
PC	Percent correct
PET	Positron emission tomography
PL	Penalized [maximum] likelihood
PMOD	Pharmacokinetic modeling software
RDP	Relative difference prior
PR	Progesterone receptor
ROC	Receiver operating characteristic
ROI	Region of interest
S/B	Signal to background
SimSET	Simulation system for emission tomography
SKE	Signal known exactly
SKS	Signal known statistically
TAC	Time activity curve
TOF	Time of flight
VCT	Virtual clinical trial

Chapter 4. HUMAN AND MODEL OBSERVER EVALUATION OF LESION DETECTABILITY IN PET USING A PENALIZED LIKELIHOOD IMAGE RECONSTRUCTION ALGORITHM

4.1 ABSTRACT

Ordered Subset Expectation Maximization (OSEM) is currently the most widely used image reconstruction algorithm for clinical PET. However, OSEM does not necessarily provide optimal image quality, and a number of alternative algorithms have been explored. We have recently shown that a penalized likelihood image reconstruction algorithm using the relative difference penalty, block sequential regularized expectation maximization (BSREM), achieves more accurate lesion quantitation than OSEM, and importantly, maintains acceptable visual image quality in clinical whole-body PET. The goal of this work was to evaluate lesion detectability with RDP-based PL versus OSEM. We performed a two-alternative forced choice study using 81 patient datasets with lesions of varying contrast inserted into the liver and lung. At matched imaging noise, PL and OSEM showed equivalent detectability in the lungs, and PL outperformed OSEM in the liver. These results suggest that PL provides not only improved quantitation and clinically acceptable visual image quality as previously shown but also improved lesion detectability compared to OSEM. We then modeled this detectability study, applying both nonprewhitening (NPW) and channelized Hotelling (CHO) model observers to the reconstructed images. The CHO model observer showed good agreement with the human observers, suggesting that we can apply this model to future studies with varying simulation and reconstruction parameters.

4.2 INTRODUCTION

OSEM [60] is currently the standard image reconstruction algorithm for clinical PET. Noise is controlled in OSEM by performing a limited number of iterations and post-filtering, and, despite its popularity, the OSEM algorithm does not necessarily result in the best possible image for a desired task. In contrast, penalized likelihood (PL) image reconstruction algorithms [61] use penalty functions and the tuning of parameters to control image properties such as the trade-off between noise and resolution. The PL algorithms have a potential for improved images when the penalty functions and parameters are optimized.

We have implemented a PL algorithm based on the edge preserving relative difference penalty [62,63] and have recently shown that it achieves superior lesion quantitation compared to OSEM [63] while maintaining clinically acceptable visual image quality and improved lesion conspicuity [64,65,66]. However, edge-preserving priors only preserve edges beyond a certain limit, and low contrast lesions at the threshold of detectability are at risk of being smoothed into the background. The question remains whether the PL algorithm provides better lesion detectability than OSEM.

In order to evaluate and quantify the impact of the PL algorithm on lesion detectability, we first conducted a human observer study, which is ideally how all images are evaluated. As an alternative to human observer studies, which are time intensive, model observers have often been applied [67]. Therefore, we also performed NPW and CHO model observer studies and compared to the human observer results. A model observer that matches the human observer could be used to predict detectability trends in future studies.

4.3 METHODOLOGY

4.3.1 *Image Generation*

We obtained 81 list-mode patient datasets acquired on a GE Discovery PET/CT 690 scanner (Waukesha, WI, USA) at the University Hospital of Zurich, Switzerland. From each list-mode dataset, two non-time-of-flight projection datasets were generated by binning alternating events into the two sinograms and histogramming within non-overlapping time windows, providing independent and identically distributed noise realizations for the lesion-absent and lesion-present images.

Synthetic spherical lesions of 15 mm were inserted into both the liver and the lung regions of one of the noise realizations of the patient projection data prior to image reconstruction. As illustrated in Figure 4.1, scanner (resolution, geometric efficiency and detector efficiency variations) and patient effects (attenuation and scatter) were modeled. For both the liver and lungs, lesion source-to-background ratios (S/B) of 1.5:1, 1.75:1, and 2.0:1 were chosen such that the lesions were near the limit of detectability; the contrast is defined with respect to the local backgrounds.

The datasets were reconstructed using OSEM and PL algorithms. The OSEM algorithm was run with 32 subsets and up to 10 iterations. The OSEM reconstructed images were post-filtered by a 5 mm FWHM Gaussian in-plane filter. The PL algorithm used the relative difference penalty (RDP) [62] and was implemented using the block sequential regularized expectation maximization (BSREM) algorithm [68,69] as in [63]. The RDP parameter γ was set to 2, and the PL algorithm was run with penalty strength parameter β of 250, 350, and 450, which we found are clinically realistic (see Equations (1) and (2) of [63] for mathematical expressions, including γ and β). The OSEM iterations were down-selected to those whose image roughness most closely matched that from the three PL beta settings, which ended up being the first three iterations.

Considering all of the parameters (liver/lung, lesion absent/present, OSEM/PL, 3 reconstruction parameters with different resolution and noise trade-offs, and 3 lesion contrasts), we generated 5832 images. Pairing lesion absent/present images, there were 2916 trials for each observer.

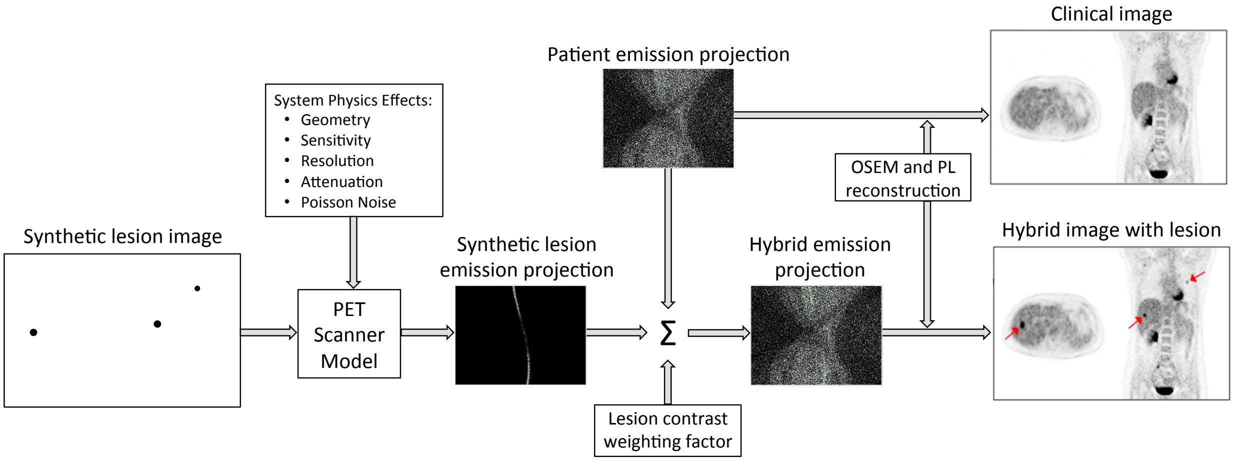


Figure 4.1. Method of inserting synthetic lesions into raw patient data: OSEM reconstructed patient image (right, top) and image with an inserted lesion with a S/B = 5.0 (right, bottom).

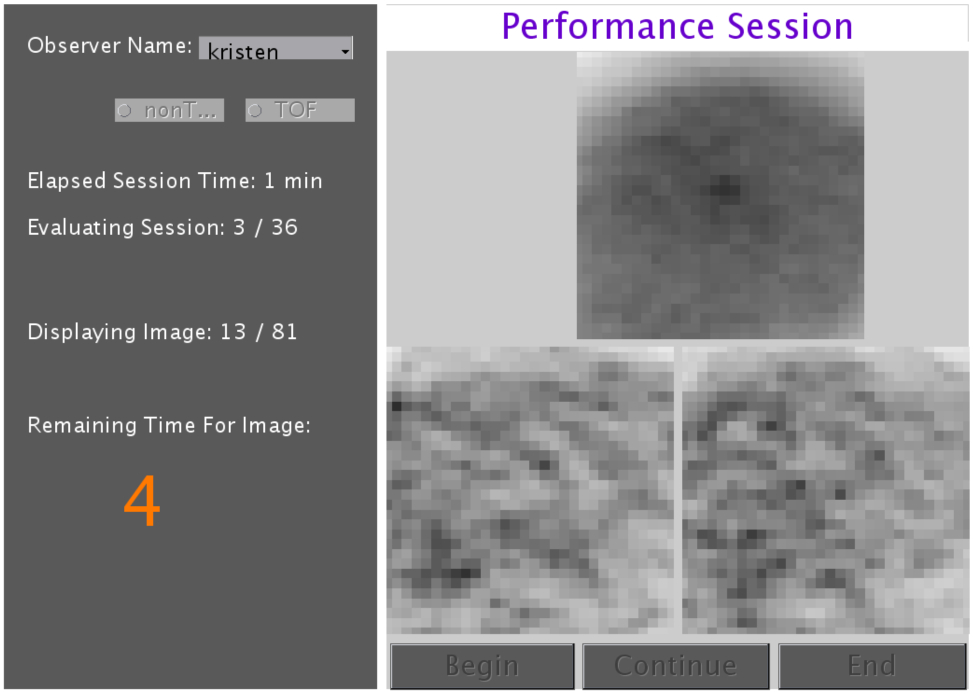
4.3.2 *Human Observer Study*

The detectability of the lesions was evaluated first using human observers in a signal known statistically and background known statistically (SKS-BKS) task. Although the signal to local background was kept constant among images, the signal magnitude varied across patients and, therefore, was only statistically known. Similarly, the background of every patient was different.

Four medical imaging scientists acted as the human observers in a two-alternative forced choice study. Each observer performed the study using the same workstation to ensure consistency of viewing conditions (monitor settings, ambient lighting, etc.). Observers completed the study in

36 sessions with 81 paired lesion absent/present images per session. A session consisted of all liver or lung images, randomly determined, and the display left/right of the paired images was also random. For each of the lung and liver sets, images over all parameters (contrast, reconstruction algorithm and reconstruction parameters) were pooled together to prevent bias as a function of time. Lung and liver images over all patients and scenarios were averaged, and the mean image was shown above the lesion absent and present images as a template for lesion size and location.

Figure 4.2 shows an example display screen with liver images. Observers were given a maximum decision time of 15 seconds per image, after which an image would be randomly chosen.



Performance Session

Observer Name: kristen

○ nonT... ○ TOF

Elapsed Session Time: 1 min

Evaluating Session: 3 / 36

Displaying Image: 13 / 81

Remaining Time For Image:

4

Figure 4.2. Display panel for human observer studies with example liver images. A lesion size and location template is shown above paired lesion-absent and present images.

Before each session, the observers were trained with five randomly selected image pairs so that the lesion characteristics (size, shape, contrast, and location) would be learned. Feedback was given immediately after each selection as to which image was the lesion-present image. These training images were included randomly with the other images for performance estimation of correctly identifying lesion-present images.

After completion of all studies, the percent correct (PC) for each observer was calculated by dividing the number of correct responses for each contrast level by 81, the patient sample number. PC was then converted to a detectability index.

4.3.3 *Model Observer Study*

We then applied 2D nonprewhitening (NPW) and channelized Hotelling (CHO) model observers to the reconstructed images. The "training" of each model observer was the creation of a template that was derived from a region of interest (ROI) across all images, and the lesion was centered in this 30 x 30 pixel ROI. The NPW model template was a matched filter, defined by the difference of the mean signal present (f_1) and signal absent (f_0) image vectors,

$$w_{NPW} = \overline{f_1} - \overline{f_0}. \tag{4.27}$$

The CHO model used four non-overlapping square profile channels [54]. The channels, U, were represented in the 4 x 30^2 matrix, where the columns described each channel's impulse response centered on the tumor location. They were defined by the frequency intervals $[\,f_c\, Q^{n-1},\ f_c\, Q^n)$, where $f_c = 1/32$, $Q = 2$, and $n = 1{:}4$. The channels were applied individually to the 81 lesion-present and lesion-absent images as Uf_0 and Uf_1 to extract the feature 4 x 81 vectors, v_0 and v_1, respectively. The CHO template was described by

$$w_{CHO} = K^{-1}U\left(\overline{f_1} - \overline{f_0}\right), \tag{4.28}$$

where K was the 4 x 4 unconditional covariance matrix of the channel output.

The test statistics, λ_0 and λ_1 for lesion-absent and lesion-present images, were calculated for each image using

$$\lambda_0 = w^t f_0 \text{ and } \lambda_1 = w^t f_1, \tag{4.29}$$

for NPW, and

$$\lambda_0 = w^t v_0 \text{ and } \lambda_1 = w^t v_1. \tag{4.30}$$

for CHO. The more similar the image was to the template, the higher the test statistic. The signal-to-noise ratio, often called the detectability metric, d', was calculated using

$$d' = SNR = \frac{\langle\lambda_1\rangle - \langle\lambda_0\rangle}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_0^2)}}, \tag{4.31}$$

where σ_0^2 and σ_1^2 are the variances of the test statistics for lesion-absent and lesion-present images, respectively. The standard error of d' was calculated using propagation of error of the SNR expression [70]. The detectability was plotted as a function of mean image roughness in the liver, defined as the coefficient of variation of the pixel values [71].

4.4 RESULTS

4.4.1 *Reconstructed Images*

Liver and lung ROIs from the OSEM and PL reconstructed images are shown in Figure 4.3 for one patient. The leftmost column is the lesion-absent image. The middle columns are lesion-present images with increasing OSEM iteration or PL beta value, where the effects of increased noise or increased smoothing can be seen. While the lesions in this patient are somewhat visible, the very low contrast makes them visually difficult to detect in many patients. The rightmost column shows the averaged ROI across all 81 patients, where the lesions are clearly visible. The reconstructed signal to background for the liver and lung lesions as a function of iteration or beta is shown in Figure 4.4. The contrast recovery trend is similar for all contrast levels for the same algorithm.

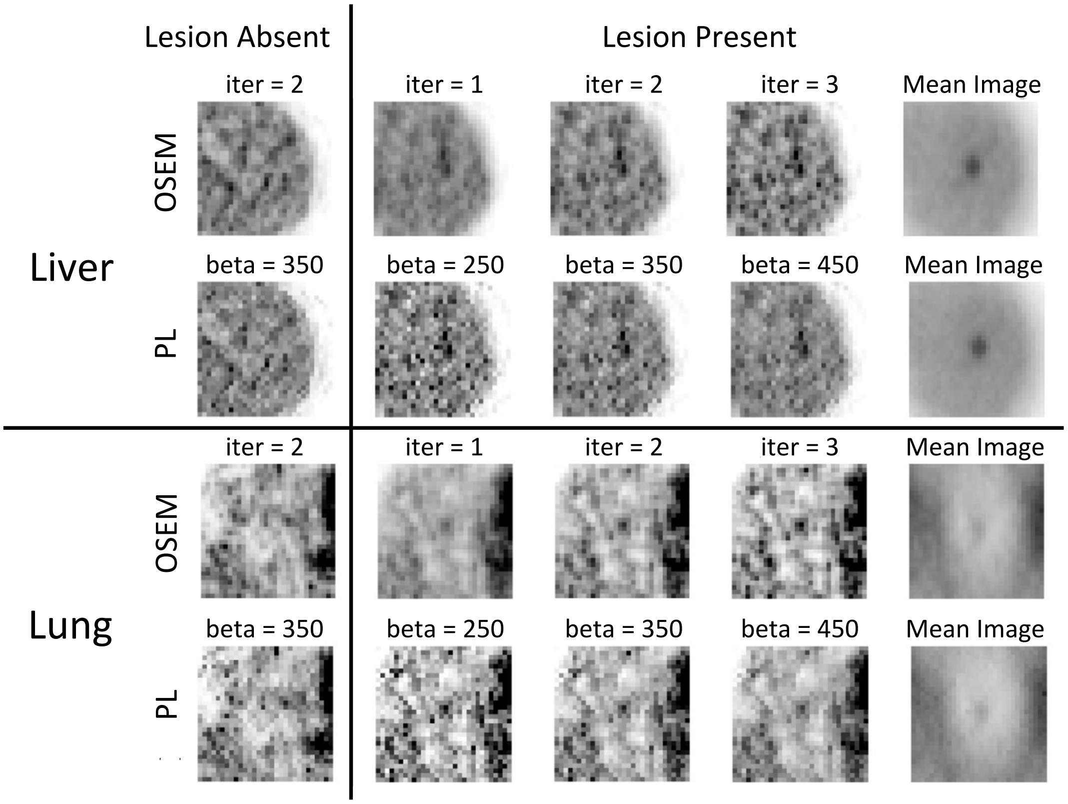


Figure 4.3. For liver and lung regions of interest, OSEM and PL reconstructions are shown for lesion-absent and lesion-present cases. The lesion present case is for S/B = 2 and three OSEM iterations or PL beta values. The mean lesion image is across all 81 patient datasets. The color scale is fixed within liver and lung images.

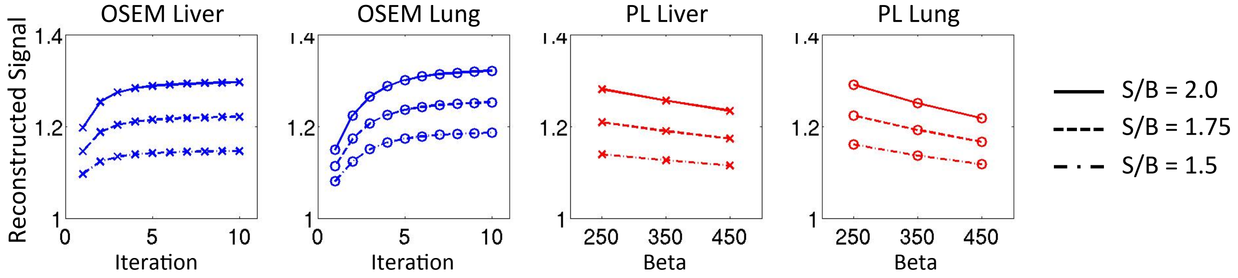


Figure 4.4. Reconstructed liver and lung lesion S/B as a function of iteration and beta.

4.4.2 *Human Observer Results*

Human observer results are shown in Figure 4.5, where the individual data points represent each observer and the solid line is a best linear fit through all data points for each lesion location and reconstruction algorithm. Detectability with PL versus OSEM is higher in the liver and comparable in the lung. The detectability generally decreases slightly with increasing image roughness. Finally, liver lesions were more detectable than the lung lesions for the same signal to local background ratio, likely due to the larger difference in signal to background in the liver.

4.4.3 *Model Observer Results*

The detectability indices for the NPW and CHO model observers are shown in Figure 4.6. PL and OSEM are compared for the liver and lung lesions as a function of image roughness in the liver. For NPW, the detectability is very similar between algorithms and increases with increasing image roughness for both locations. For CHO, the detection performance for PL compared to OSEM is similar in the lung and improved in the liver for matched image roughness, and the detectability trend decreases slightly with increasing image roughness.

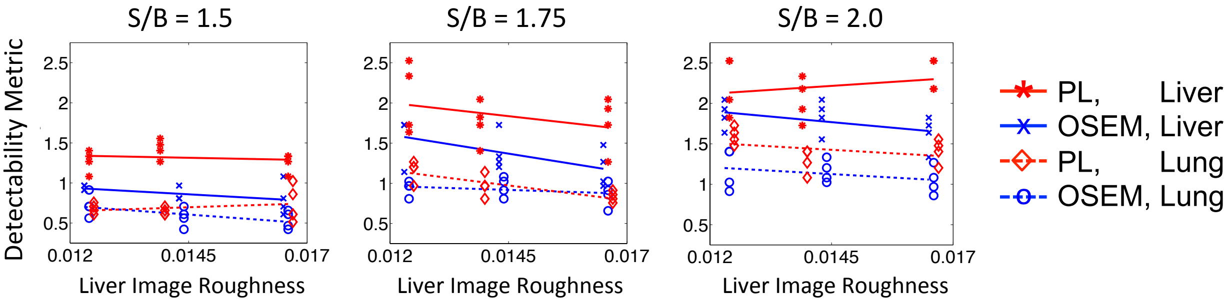


Figure 4.5. Human observer results, shown as a data point for each observer and a best fit line through all 12 points for each algorithm and location pair (top).

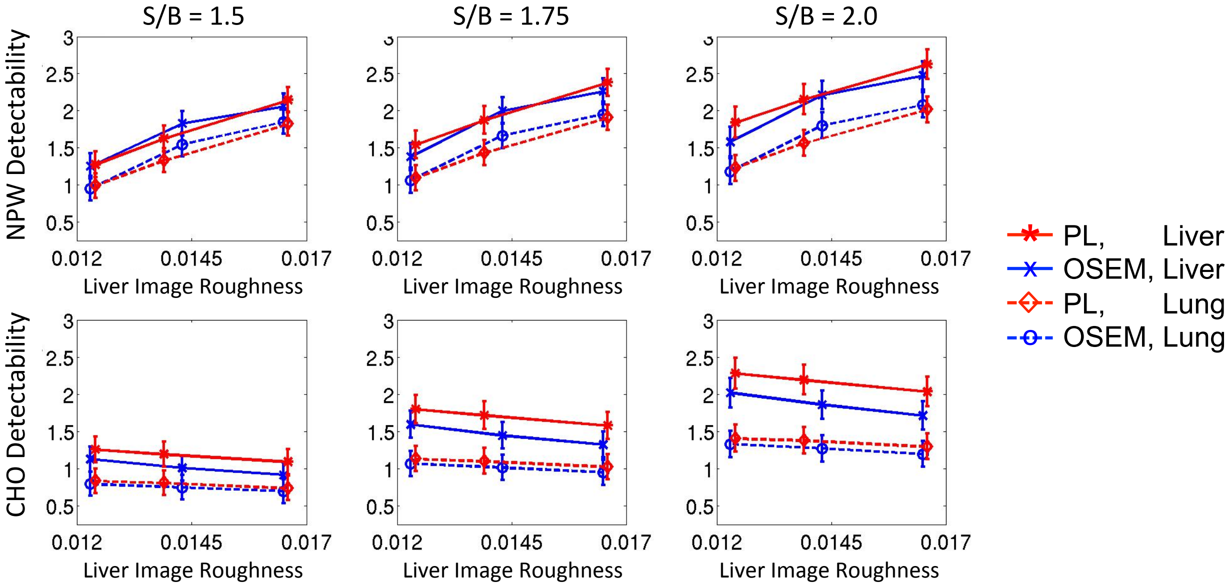


Figure 4.6. NPW and CHO model observer detectability as a function of image roughness in the liver.

4.4.4 *Comparison Between Human and Model Observers*

The relationship between human and model observers is shown in Figure 4.7. The detectability for PL and OSEM and all contrasts and reconstruction parameters are shown on a signal plot. The linear relationships for PL show good agreement between the human observer and CHO model observer. Human and CHO detectabilities are compared as a function of image roughness in Figure 4.8, again showing matched detectability trend and magnitude. Finally, CHO and human observer detectability as a function of reconstructed signal to background is plotted in Figure 4.9, showing that detectability improves linearly at all noise levels as a function of contrast. Note we did not apply internal noise in this study.

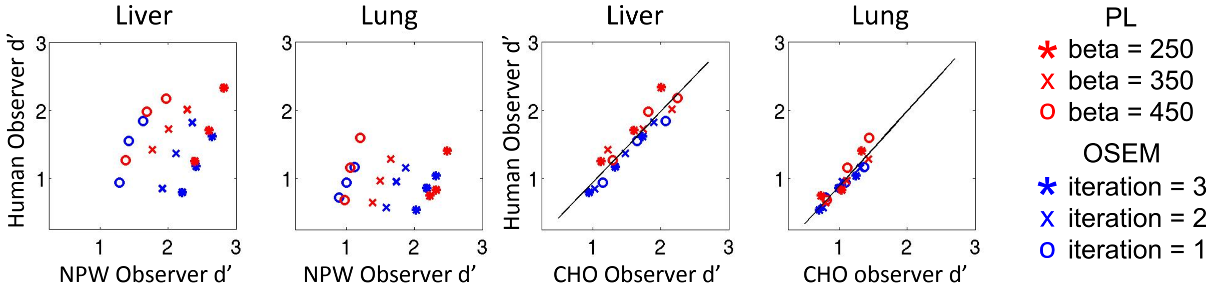


Figure 4.7. Human observer detectability versus NPW and CHO model observer detectability.

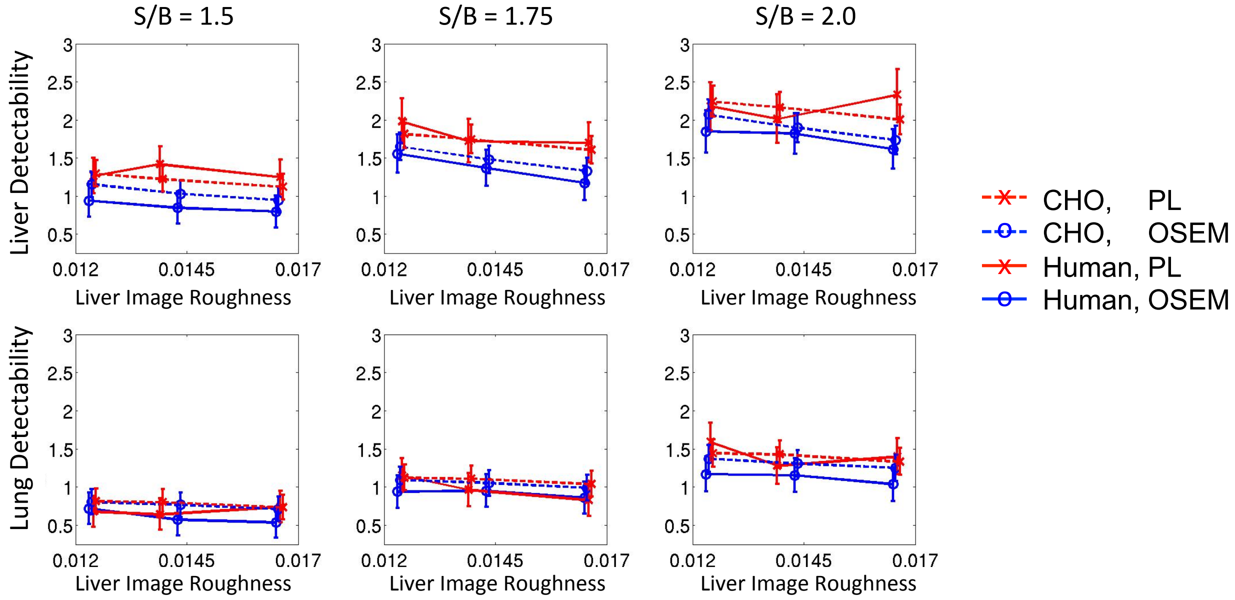


Figure 4.8. Comparison of NPW and CHO model observer detectability to the human observer detectability.

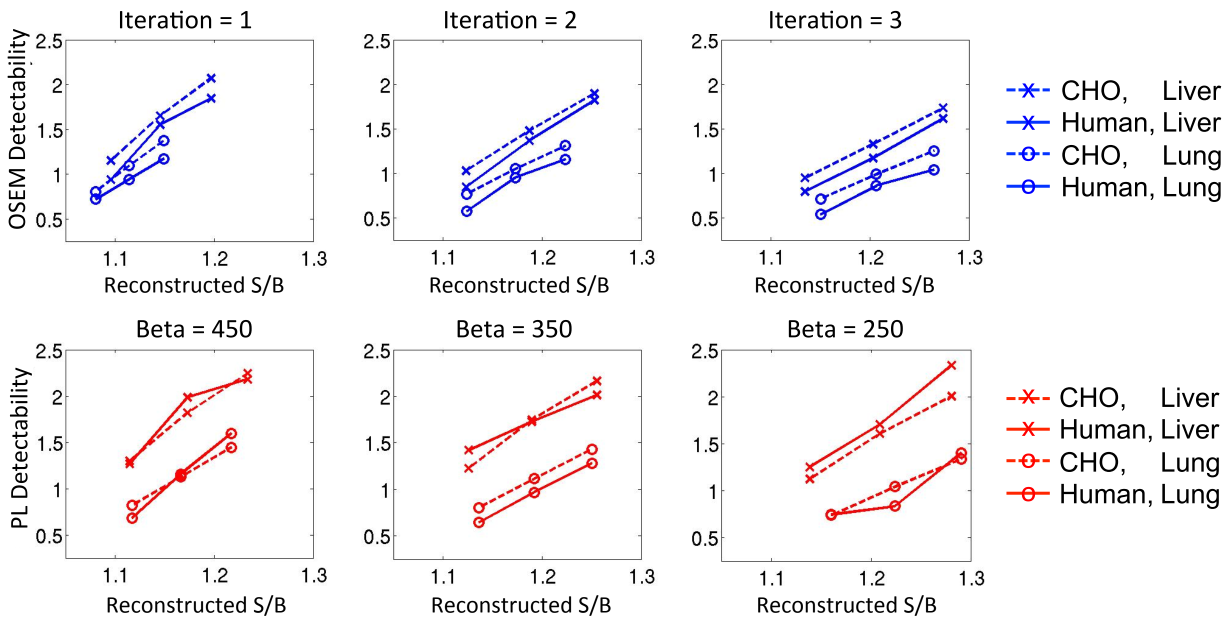


Figure 4.9. Comparison between the CHO and human observers as a function of reconstructed S/B.

4.5 DISCUSSION

We demonstrated in a previous study [63] that PL improves lesion standardized uptake value (SUV) quantitation accuracy compared to OSEM at matched image noise, particularly in cold background regions such as the lung. The quantitative improvement is partly due to the edge-preserving property of RDP. Since that quantitation study used relatively high contrast lesions for evaluation, we questioned whether the RDP would smooth out low contrast lesions, subsequently worsening lesion detectability. The results in this chapter indicate that, at clinically relevant count density, PL does not decrease lesion detectability compared to OSEM and may even improve detectability in the liver.

The trends of the NPW and CHO observers are observed to be opposite from one another. For NPW, lesion detectability increased as image noise increased, that is, as the OSEM iteration number increased or the PL regularization parameter beta decreased. This trend is consistent with that seen in the literature [72,73]. For CHO, lesion detectability decreased as image noise increased, which is also consistent with many studies [73,74,75,76,77]. In contrast, we note that there were two studies, one applying CHO and the other CNPW, that indicated an increase in detectability with increased image noise [78,79]. These differing trends may be due to the different noise correlations that arise with different reconstruction algorithms and parameters. For example, with increasing OSEM iteration or lower PL beta, the image noise is whiter and has fewer correlations; therefore, the NPW algorithm, which does not prewhiten, performs better.

However, our goal here was not to specifically compare model observers but to model the human observer results. We found CHO and human observers agree well, and it is well known in the community that CHO more consistently predicts human performance for a variety of correlations [67,80]. As these reconstruction algorithms have many parameters, we can apply this tool to predict detectability trends in future studies, saving the cost of human observer studies.

4.6 CONCLUSIONS

We compared lesion detectability of OSEM and the PL algorithm with RDP implemented by PL. Realistic lesion-present datasets were generated by adding simulated low-contrast liver and lung lesions to clinical datasets. A 2AFC human observer study showed that PL outperformed OSEM at matched image noise, particularly in the liver. We then aimed to model these results and found good agreement using the CHO model observer. We conclude that PL achieves improved lesion detectability as well as improved quantitation and visual image quality, and we can apply the CHO model observer in future studies with these patient data to further evaluate imaging parameters.

Chapter 5. EVALUATION OF LESION DETECTABILITY IN PET WHEN USING A CONVERGENT PENALIZED LIKELIHOOD IMAGE RECONSTRUCTION ALGORITHM

5.1 ABSTRACT

We have previously developed a convergent penalized likelihood (PL) image reconstruction algorithm using the relative difference prior (RDP) and showed that it achieves accurate lesion quantitation compared to ordered subsets expectation maximization (OSEM), currently the most widely used image reconstruction algorithm for clinical PET, without compromising subjective visual image quality. However, we were concerned about the behavior of the RDP for small, low contrast lesions. In this work, we evaluated the detectability of low-contrast liver and lung lesions using our RDP-based PL image reconstruction algorithm compared to OSEM. We performed a two-alternative forced choice model observer study using a channelized Hoteling (CHO) model that was previously validated against human observers. At matched image roughness, lesion detectability using PL was equivalent or improved compared to using OSEM. Lesion detectability using PL showed a stronger dependence on lesion size, which was likely due to nature of the RDP that applies more smoothing in both low contrast regions and low background activity regions such as the lungs. Lesion detectability also improved using TOF reconstruction, with greater benefit for the liver compared to the lung and with increasing benefit for decreasing lesion size and contrast. PL detectability was statistically significantly higher than OSEM for 20 mm liver lesions when contrast ≥ 0.5 ($p < 0.05$), and TOF PL detectability was statistically significantly higher than TOF OSEM for 15 mm liver lesions with contrast ≥ 0.5 and 20 mm lesions with contrast ≥ 0.25. For all other cases, there was no statistically significant

difference between PL and OSEM ($p > 0.05$). These results suggest the PL algorithm with RDP provides equivalent or improved lesion detectability compared to OSEM for the range of studied lesion properties.

5.2 INTRODUCTION

Positron emission tomography (PET) imaging is clinically used in the detection and quantitation of tracer-avid lesions or tumors in cancer. Ordered subsets expectation maximization (OSEM) [81] is currently a standard image reconstruction algorithm for clinical PET imaging. The OSEM algorithm is not convergent, and noise increases rapidly with iteration. Noise is controlled in OSEM by stopping after a few iterations and post-filtering the image. Despite its popularity, the behavior of the algorithm is unpredictable and the joint performance of detection and quantitation using OSEM is not well understood. In contrast, penalized likelihood (PL) image reconstruction algorithms [82] show predictable behaviors, and parameters can be tuned to control the trade-off between image noise and resolution [83,84,85,86,87]. These algorithms, therefore, have the potential to improve image quality such as lesion detectability [88,89] and region of interest (ROI) quantitation accuracy [90] when the penalty functions and parameters have been optimized. Although PL algorithms have been in the research community for over three decades [91], they have not been adopted clinically due to the different and unfamiliar image appearance compared to OSEM as well as increased computation time. For example, non-quadratic edge-preserving penalties often produce undesirable patchy or 'mosaic'-like images [92,93].

In previous work, we developed a PL algorithm. [94] based on the relative difference prior (RDP) [95] that penalized relative differences using the block sequential regularized expectation maximization (BSREM) optimizer [96,97]. The RDP applies activity-dependent smoothing and suppresses image noise in low activity background regions. The degree of edge preservation is also controlled. Studies comparing RDP-based PL and OSEM have demonstrated that RDP-based PL achieves better lesion quantitation accuracy with higher signal-to-noise ratio (SNR) [94,98, 99] while maintaining subjective image quality [99,100,101,102,103] compared to OSEM. Studies have also shown that RDP-based PL produces higher lesion standardized uptake values (SUVs) [94,98,99,103,104,105] than OSEM, suggesting a potential improvement in lesion detectability; indeed, recent clinical studies have reported that the PL algorithm with RDP produced higher lesion detectability scores than OSEM in visual image quality evaluation studies [99,101].

However, the existing studies use relatively high contrast lesions and one may question whether the edge-preserving RDP would smooth out low contrast lesions, subsequently worsening lesion detectability. A study to evaluate the lesion detectability using low contrast lesions with varying sizes and contrasts for our RDP-based PL algorithm [94] compared to OSEM has not yet been performed except for our preliminary study [106]. In this work, we performed a two-alternative forced choice detectability study [107,108] using a CHO model observer [109,106]. We compared the detectability results as a function of lesion location, size, and contrast, and algorithm parameters for both for time-of-flight (TOF) and non-TOF data. For the remainder of this paper, the RDP-based PL algorithm is referred to as simply PL.

5.3 METHODOLOGY

5.3.1 *Image generation*

We obtained 81 TOF list-mode patient datasets using ^{18}F-FDG acquired on a Discovery PET/CT 690 scanner (General Electric, Waukesha, WI, USA) at the University Hospital of Zürich, Switzerland. The TOF timing resolution of the system was 544 ps [110]. From each list-mode dataset, two projection datasets were generated for both TOF and non-TOF imaging scenarios by binning alternating events into the two sinograms and histogramming with non-overlapping time windows. Dividing the data provided independent and identically distributed (i.i.d.) noise realizations for the lesion-absent and lesion-present images.

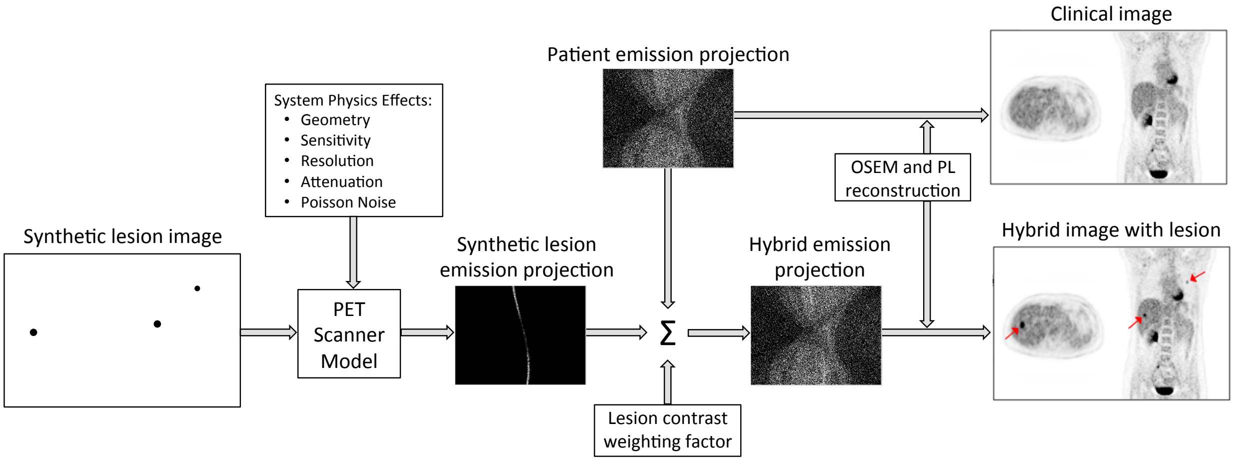


Figure 5.1. Synthetic lesions are inserted into the raw patient data after accounting for all PET system physics effects. Examples of a lesion-absent (top right) and lesion-present (bottom right) image reconstruction are shown for a lesion contrast of 4.0.

Synthetic spherical lesions with diameter of 10, 15, and 20 mm were generated in both the liver and lung regions using the procedure shown in Figure 5.1. The synthetic lesions were then forward projected through the scanner model, where scanner effects (resolution, geometric efficiency and detector efficiency variations) and patient effects (attenuation) were considered. The synthetic lesion emission projection data was added to patient emission projection data using different lesion contrast weighting factors. The lesion contrasts for the liver and lung lesions studied in this paper are shown in Table 5.1, and the lesion contrast is defined as [(lesion activity)–(local background activity)]/(local background activity). The contrasts were chosen so that the lesions were near the limit of detectability. Lesion recovery is lower for lesions in low-activity background regions, such as the lungs, than for those in high-activity background regions such as livers. This is due to slow convergence rates of OSEM in the lungs and more smoothing in the lungs applied by the penalty function, RDP, used in PL. As a result, the lesion detectability is also lower in the lungs, and we used higher contrasts for lung lesions than for liver lesions.

The two different sinogram noise realizations were used for the lesion-absent and lesion-present cases. Both sinograms were reconstructed using both OSEM and PL algorithms for TOF and non-TOF cases and with point spread function (PSF) modeling [111]. The OSEM algorithm was run with 24 subsets and 3 iterations and axially and transaxially post-filtered. Details about the post-filters are given in Section 3. PL was implemented using the block sequential regularized expectation maximization (BSREM) algorithm [96,97] as in Ahn et al [86].

The PL objective function is

$$\Phi(x) = \sum_i y_i \log([Px]_i + r_i) - ([Px]_i + r_i) - \beta R(x), \tag{5.32}$$

where x is the vector representing the activity image, y_i are emission sinogram data, P is the forward projection operator that includes attenuation, normalization and point spread function resolution modeling, r_i are estimated background contributions of scatter and randoms, $R(x)$ is the regularization or penalty function, and β is the regularization or penalty parameter, which controls the global strength of regularization.

The regularization function is the RDP and is defined by

$$\beta\, \mathrm{R}(x) = \beta \sum_{j,k:\text{neighbors}} \alpha_{jk} \frac{(x_j - x_k)^2}{(x_j + x_k) + \gamma |x_j - x_k|}. \tag{5.33}$$

The RDP is calculated between each pixel j and its k neighbor pixels. There are two smoothing terms. In the numerator, the smoothing is applied by penalizing the square of neighboring pixel differences, and in the denominator smoothing is applied based on the absolute amount of activity. For example, low-activity regions will have a larger penalty and be smoothed more. The ratio of the difference in activity to the absolute activity gives this penalty its relative characteristic. Finally, gamma controls edge-preservation, and this term is also based on the difference between pixels. In this work, the weights α_{jk} as used in Ahn et al. [94] were chosen. The RDP parameter, γ, was 2, and the penalty strength parameter β was 250, 350, and 450. The range of the β parameters was found clinically relevant in Ahn et al. [94], and the edge-preservation parameter $\gamma = 2$ was previously chosen as a reasonable trade-off between visual image quality and lesion quantitation accuracy [94,122]. No post-filtering was applied to PL images. We found the image roughnesses from the first three OSEM iterations with post-filtering to most closely match the image roughnesses from the three PL betas, which allowed us to compare the lesion detectability between algorithms based on matched image noise.

Considering all of the parameters (81 patients, nonTOF/TOF, liver/lung, lesion absent/present, OSEM/PL, 3 lesion sizes, 3 reconstruction parameters with different resolution and noise trade-offs, and 3 lesion contrasts), 46,656 images were generated. An example of lesion absent and lesion present image patches are shown in Figure 5.2.

Table 5.1. Lesion contrasts for the liver and lung lesion locations and three spherical lesion diameters.

	10 mm			15 mm			20 mm		
Liver	0.75	1.0	1.25	0.25	0.5	0.75	0.25	0.5	0.75
Lung	1.25	1.75	2.25	0.75	1.0	1.25	0.5	0.75	1.0

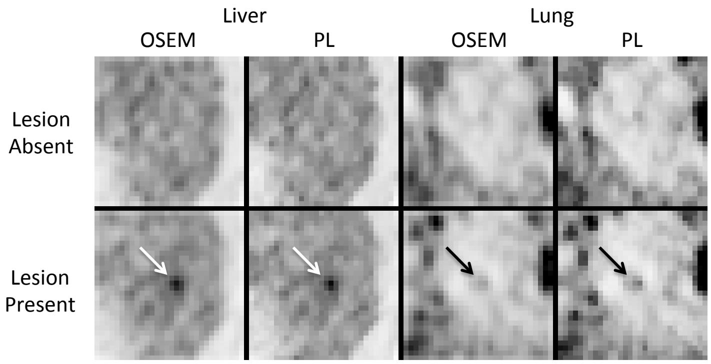


Figure 5.2. Lesion absent and lesion present image patches reconstructed using OSEM and PL algorithms. The liver and lung lesion examples are 20 mm with a contrast of 1 and 2.25, respectively. OSEM images are shown for 2 iterations with post-filtering, and PL images are shown for beta=350.

5.3.2 *Model Observer Study*

The detectability was evaluated in a signal known statistically and background known statistically (SKS-BKS) task. For each lesion size and contrast, although the true lesion signal to local background was kept constant among images, the reconstructed lesion signal magnitude as well as the true lesion signal magnitude varied across patients. Therefore, the signal magnitude was only statistically known. Similarly, the background of every patient was different and only statistically known.

We applied a 2D channelized Hotelling (CHO) model observer using four non-overlapping square profile channels [112] to the reconstructed images using the center image slice through each lesion. The channels, U, were represented in the 4 x 30 x 30 matrix, where the columns described each channel's impulse response centered on the tumor location. They were defined by the frequency intervals $[f_c\, Q^{n-1},\ f_c\, Q^n)$, where $f_c = 1/32$, $Q = 2$, and $n = 1{:}4$, shown in Figure 5.3. The channels were applied individually to the 81 lesion-absent and lesion-present images as Uf to extract the 4 x 81 feature vectors, v_0 and v_1.

The lesion template was derived from a region of interest (ROI) across all images for a given parameter set, described by

$$w = K^{-1}U\left(\overline{f_1} - \overline{f_0}\right), \tag{5.34}$$

where K was the 4 x 4 unconditional covariance matrix of the channel output and $\overline{f_0}$ and $\overline{f_1}$ were the mean of the lesion-absent images and of the lesion-present images, respectively. The lesion was centered in this 30 x 30 pixel ROI. The test statistics, λ_0 and λ_1 for lesion-absent and lesion-present images, were calculated for each image using

$$\lambda_0 = w^t v_0 \text{ and } \lambda_1 = w^t v_1. \tag{5.35}$$

The more similar the image was to the template, the higher the output test statistic was. The signal-to-noise ratio, often called the detectability index, d', was calculated using

$$d' = SNR = \frac{\overline{\lambda_1} - \overline{\lambda_0}}{\sqrt{\frac{1}{2}\left(\sigma_1^2 + \sigma_0^2\right)}}, \tag{5.36}$$

where σ_0^2 and σ_1^2 were the variances of the test statistics for lesion-absent and lesion-present images, respectively. The standard error of d' was calculated using propagation of error of the SNR expression [113]. The detectability as measured by SNR was compared as a function contrast as well as mean image roughness in the liver, defined as the coefficient of variation of the pixel values [114].

Our goal was not to develop a model observer but, rather, to model the human observer results. It is well known in the community that CHO more consistently predicts human performance for a variety of correlations [115,116]. The CHO observer has become a popular model observer within the nuclear medicine community due to its flexible and predictive characteristics [117]. Further, others have used the CHO model observer for both non-TOF/TOF and without and with PSF modeling [118,119]. In our previous study [106], we evaluated this model observer and found good agreement with humans. Although it is not uncommon for model observer results to outperform human results [120], this difference was small in our previous study, potentially due to the template estimation (observer training) [121], ROI, or image noise properties. While internal noise can be added to reduce the model observer performance, we did not find this step necessary in our study.

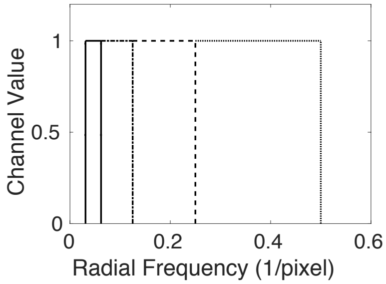


Figure 5.3. The four constant-Q channel model applied in the CHO model observer.

5.3.3 *Statistical Analysis*

The significance of the differences in lesion detectability between images reconstructed using the OSEM and PL algorithms were evaluated using a resubstitution bootstrapping method [121]. To generate 1000 samples, each of which consisted of 81 patient image pairs, for each combination of parameters (nonTOF/TOF, liver/lung, OSEM/PL, lesion size, lesion contrast and reconstruction parameter such as iteration number for OSEM and beta parameter for PL), images were randomly sampled from the 81 patient images; for each sampling, some patients were selected one or more times and some patients were not selected at all. Because images were reconstructed from the same sinogram noise realization for both algorithms, the same random sampling was used for both algorithms to maintain that correlation. Each random sample of lesion detectability for OSEM was subtracted from that for PL, and the results were histogrammed to form a difference distribution. If such a difference distribution centered on zero, for example, it would indicate no difference in lesion detectabilities between the two algorithms. Confidence intervals were calculated using the standard deviation of the difference bootstrap distribution. The p-value for each case was determined by calculating two times the fraction of the difference distribution tail that crossed zero to represent a two-tailed test.

5.4 RESULTS

We applied our CHO model observer to OSEM and PL reconstructed images with varying simulation parameters, and studied the impact of lesion location, lesion size, lesion contrast, TOF, PSF, transaxial and axial post-filtering, and OSEM iteration number or PL beta value on detectability.

OSEM and PL SNR are compared as a function of contrast in Figure 5.4 for lesion location, lesion size, and non-TOF and TOF image reconstruction. These results are for 2 iterations of OSEM with a [1/6, 4/6, 1/6] three-point axial post-filter and 6 mm transaxial post-filter and for PL with a beta value of 350; these reconstruction parameters are clinically reasonable ones [94]. Detectabilities were higher in the liver than the lung for matched contrast. Detectability increased with increasing contrast and lesion size, as well as for TOF compared to non-TOF, as was expected. Detectability using PL showed more of a dependence on lesion size; for example, for a contrast of 0.75 in the liver, the detectability was 186% higher for the 20 mm lesion compared to the 10 mm for PL and 164% higher for OSEM for the TOF case. The percent improvement in TOF versus non-TOF SNR increased with decreasing lesion size and contrast, as shown in Figure 5.5. TOF also provided more benefit in the liver than the lung, increasing detectability by an average of 22.0% compared to 11.8%, respectively, when averaging over algorithm, lesion size, and contrast. Modeling PSF also improved lesion detectability (data not shown), as other studies have found [118,119].

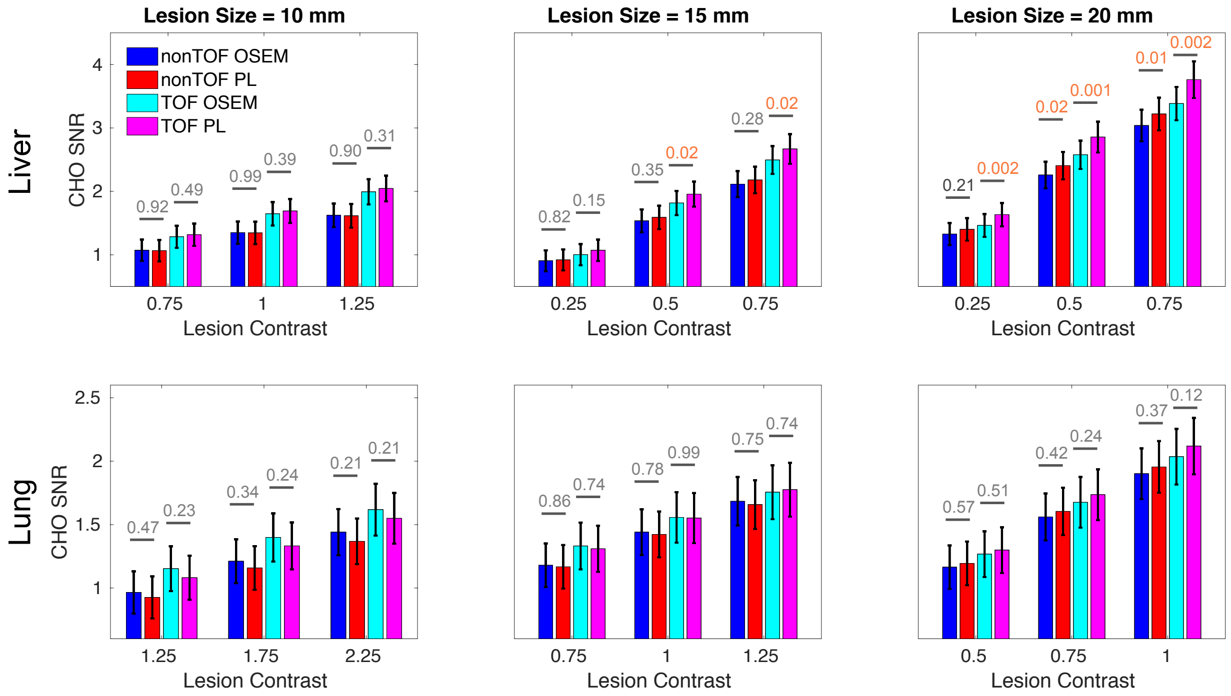


Figure 5.4. CHO SNR as a function of contrast for liver and lung lesion locations for the three lesion sizes and nonTOF and TOF reconstructions. The p-values indicate the statistical significance of the difference in CHO SNR between non-TOF OSEM and PL and between TOF OSEM and PL. Statistically significant p-values (< 0.05) are highlighted in orange.

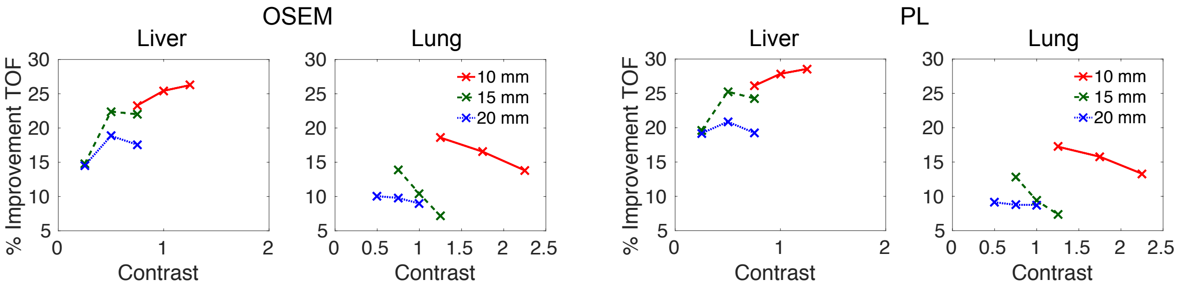


Figure 5.5. Percent improvement of TOF over non-TOF image reconstruction as a function of contrast for both reconstruction algorithms and lesion locations.

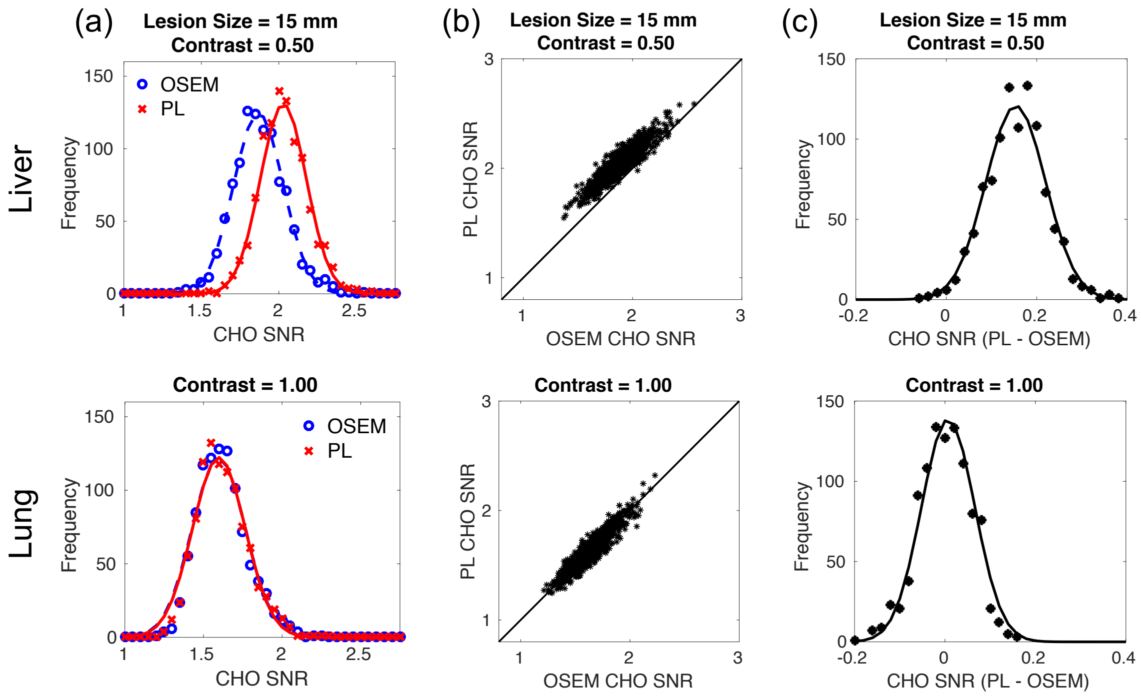


Figure 5.6. (a) Bootstrap distributions for OSEM (o's) and PL (x's) for liver and lung lesion locations. A Gaussian model is fit to each distribution. (b) CHO SNR for PL versus OSEM for each of the 1000 bootstrap samples, showing the detectability correlation between algorithms. Data points above or below the line of identity indicate improved performance of PL or OSEM, respectively. (c) PL and OSEM difference distribution, where statistical significance increases the further away the distribution is from zero.

We performed a statistical analysis to better characterize the impact of the OSEM and PL algorithms on detectability. The 95% confidence intervals and p-values in Figure 5.4 were determined from a bootstrapping statistical analysis. Example bootstrap distributions using 1000 samples are shown in Figure 5.6(a). The same patient sinograms were used for both algorithms, resulting in a significant correlation between the detectabilities, as seen in Figure 5.6(b). We were interested in the performance difference beyond these correlations, so we evaluated the difference in the bootstrap distributions. The distribution of the performance difference between algorithms is shown in Figure 5.6(c). PL detectability performance improved for increasing lesion size and contrast. PL SNR was statistically significantly higher than OSEM for 20 mm lesions in the liver when contrast ≥ 0.5 ($p < 0.05$ two-sided test), and TOF PL SNR was statistically significantly higher than TOF OSEM for 15 mm lesions in the liver with contrast ≥ 0.5 and 20 mm lesions with contrast ≥ 0.25. For all other cases, there was no statistically significant difference between PL and OSEM ($p > 0.05$).

We explored detectability as a function of image noise as measured by image roughness, shown in Figure 5.7, where image roughness is impacted by PL beta value or OSEM iteration number and magnitude of post-filtering. The impact of OSEM axial and transaxial filters is shown separately. Smoother images were favorable for detection in the liver for both PL and OSEM, with increasing detectability with decreasing image noise. Lung detectabilities showed less dependence on image roughness. Transaxial post-filtering with a 6 mm full-width at half-maximum (FWHM) Gaussian produced smoother images with generally slightly higher detectability compared to no post-filtering. Axial post-filtering also produced smoother OSEM images while improving detectability.

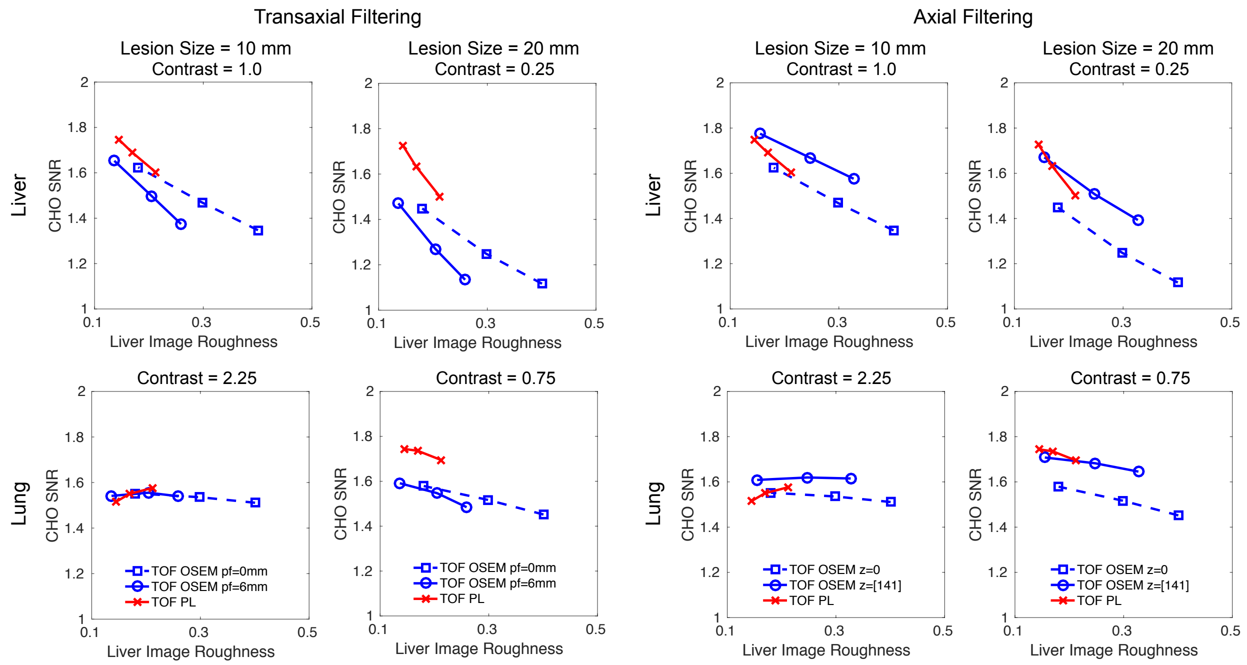


Figure 5.7. CHO SNR as a function of image roughness in the liver, and impact of transaxial and axial filtering applied to OSEM images on detectability. Transaxial and axial filtering results are shown without filtering in the other direction. Representative results are shown for TOF reconstruction and one contrast. Each OSEM curve has 3 markers representing 3, 2 and 1 iterations from left to right, and each PL curve has 3 markers representing beta=450, 350 and 250 from left to right.

5.5 DISCUSSION

We demonstrated in previous studies [94,122] that our convergent PL algorithm improved quantitation accuracy of lesion standardized uptake value (SUV) compared to OSEM at matched image roughness, particularly in cold background regions such as the lung due to the edge-preserving property of RDP and the slow convergence rates of OSEM in cold regions. While the properties of the RDP were favorable for quantitation and higher contrast lesions, we questioned how properties of the RDP would impact lesion detectability for low contrasts. In particular, we were concerned that the RDP might smooth out low contrast lesions, subsequently worsening detectability. Therefore, we performed a study to evaluate lesion detectability over a range of lesion and reconstruction algorithm parameters, including lesion location, size, and contrast, non-TOF/TOF OSEM and PL image reconstruction, and image post-filtering. The results indicate that, for clinically relevant count density and range of reconstruction parameters and lesion properties considered, PL does not decrease lesion detectability compared to OSEM. PL maintained equivalent detectability in the lung and improved detectability in the liver for some cases.

First considering lesion location, lesion detectability was higher in the liver than the lung at equal contrast, due to higher contrast recovery in the liver compared to the lung for an equivalent number of iterative updates in OSEM and for an equivalent beta parameter in PL. It is known that the lung region converges more slowly [94], which was likely the reason for the lower detectability in the lung for OSEM; to control image noise, we only ran the algorithm for 3 iterations with 24 subsets and post-filtered the image. For PL, lower detectability in the lung was likely due to the RDP applying more smoothing in this low-activity region.

Comparing lesion detection using PL versus OSEM for the two lesion locations, we found PL performed better in the liver than the lung. In contrast, in our previous quantitation study [94], results suggested that PL instead outperformed OSEM in the lung. To investigate this difference, we calculated the contrast recovery coefficient (CRC) over a wide range of lesion contrasts as shown in Figure 5.8. The CRC showed greater improvement for PL with increasing contrast, particularly for the smallest lesion size, illustrating the contrast dependent impact of the RDP on quantitation. These results highlight the important difference between detection and quantitation tasks, which should be well defined and considered in study design.

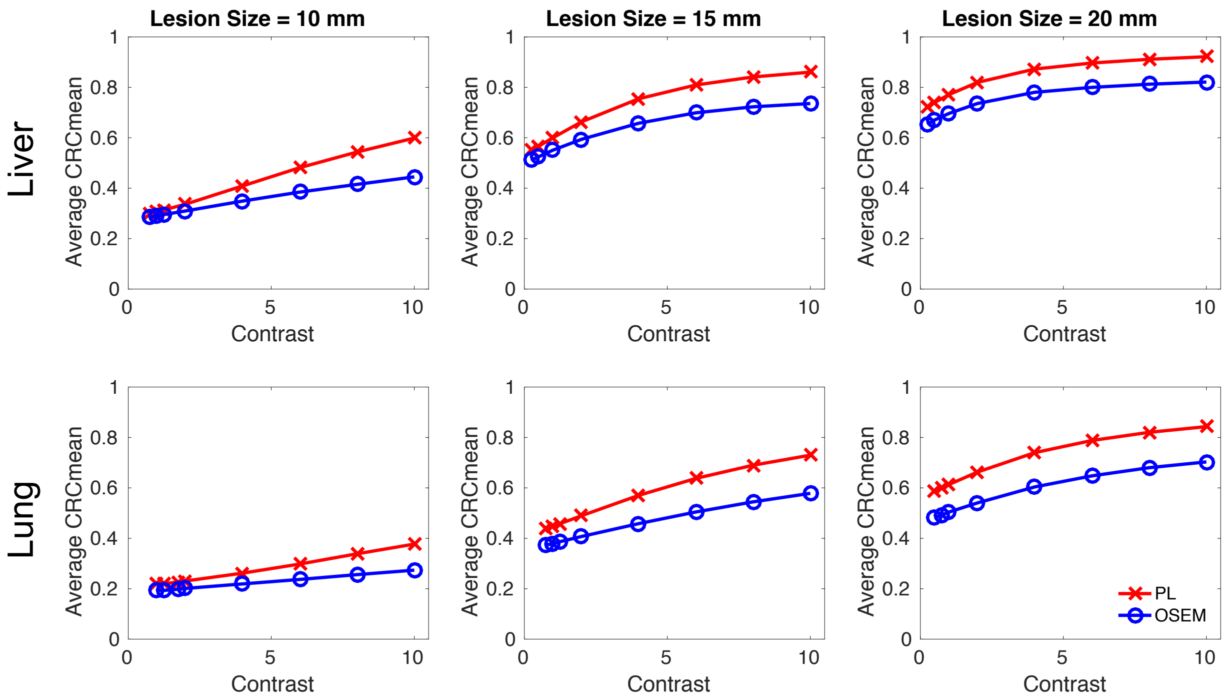


Figure 5.8. Contrast recovery coefficient for a wide range of inserted lesion contrasts in the liver and lung for the three lesion sizes, showing the improved performance of the RDP with increasing contrast, particularly for the 10 mm-sized lesion.

Considering lesion size, PL showed greater relative performance improvement with increasing lesion size, a trend also seen in Qi [123]. PL performance compared to OSEM was lowest for the 10 mm lesion, though the difference between algorithms was not statistically significant. For the 15 and 20 mm lesions and certain contrasts, PL outperformed OSEM. For the small lesions, it is likely that the edge-preserving nature of the prior is less sensitive. The detectability of lesions smaller than 10 mm should be investigated in a future study, especially for higher resolution scanner that will improve contrast recovery.

Finally, we considered detectability as a function image roughness, which is a function of OSEM iteration and post-filtering or PL beta value. The detectability in the liver increased with decreasing image roughness (lower OSEM iteration or higher PL beta), a trend observed by many previous studies [124,125,126,127,128]. Detectability in the lung was less dependent on image roughness, and for the 10 mm lesion, PL detectability decreased with decreasing image roughness, a trend also seen in previous studies [129,118]. These opposite trends illustrate the dependence of lesion detectability on different noise correlations that arise with different reconstruction algorithms and parameters. Further, these trends indicate that OSEM post-filtering or PL beta should be applied cautiously for smaller lesion sizes and lower contrasts, due to the potential to blur the lesion into the background.

In a previous relevant study, Qi [130] compared lesion detectability for PL with quadratic penalties and edge-preserving non-quadratic penalties, such as Huber and Geman-McClure, using a channelized Hotelling observer (CHO), and showed the edge-preserving property of the non-quadratic penalties did not improve lesion detectability. In another similar study, Nuyts et al. [128] found that detectability for PL with RDP using a small or large edge-preserving parameter ($\gamma = 0$ or $\gamma = 200$) was superior to post-smoothed maximum likelihood expectation maximization (MLEM), and when $\gamma = 10$ or $\gamma = 20$, RDP-based PL yielded reduced detectability.

Prior to our previous preliminary study and this current work, there had not yet been a comparison of the detectability of the RDP-based PL to the standard clinical algorithm, OSEM. Additionally, our choice of the RDP parameter, $\gamma = 2$, which resulted in superior lesion quantitation accuracy compared to OSEM [94,98,99,100,101,102,103], was different from $\gamma = 0$ and $\gamma = 10$, which showed very different performance from each other in the study of Nuyts et al. [128]. Due to these differences, it was not straightforward to extrapolate previous study results to the performance of our PL algorithm.

There were some limitations in this study. The model observers were applied to a single axial slice of cropped PET images (30 x 30 pixel ROI). We did not explore larger volumes or location uncertainty. We averaged the results over all patients and did not perform analyses or separate the results based on patient characteristics. Finally, we did not specifically optimize algorithm parameters for detection but instead applied typically clinical parameters for OSEM and the penalty parameters optimized in the quantitation study for PL. Further model observer studies could be conducted to evaluate a broader range of algorithm parameters on detectability.

5.6 CONCLUSION

We evaluated lesion detectability in images reconstructed using a PL algorithm with RDP implemented with BSREM in comparison to images reconstructed using the clinical standard OSEM algorithm. Our CHO model observer showed that at matched image roughness, our RDP-based PL algorithm achieved matched or improved lesion detectability compared to OSEM for the clinically relevant parameter space investigated. The smoothing characteristics of the RDP for low activity backgrounds and low contrast lesions smaller than 10 mm and the resulting impact on detection tasks should be investigated.

Chapter 6. EFFECT OF ^{18}F-FDG UPTAKE TIME ON LESION DETECTABILITY IN PET IMAGING OF EARLY STAGE BREAST CANCER

6.1 ABSTRACT

Prior reports have suggested that delayed FDG-PET oncology imaging can improve the contrast-to-noise ratio (CNR) for known lesions. Our goal was to estimate realistic bounds for lesion detectability for static measurements with one to four hours between FDG injection and image acquisition. We combined three components: a kinetic model of ^{18}F-FDG radiotracer uptake, a forward model of PET data acquisition, and computer model observers to estimate lesion detectability. Model parameters were estimated from dynamic PET studies of patients with early stage breast cancer. These were used to generate time-activity curves (TACs) out to four hours, for which we assumed both nonreversible and reversible models with different rates of FDG dephosphorylation (k_4). For each TAC, 600 PET sinogram realizations were generated, and images were reconstructed using OSEM. Test statistics output from the computer model observers were evaluated using an ROC analysis with the calculated AUC providing a measure of lesion detectability. For the nonreversible model ($k_4 = 0$), the AUC increased in 11/23 (48%) of patients for one to two hours after the current standard post-radiotracer injection imaging window of one hour. This improvement was driven by increased tumor/normal tissue contrast, before the impact of increased noise due to radiotracer decay began to dominate the imaging signal. As k_4 was increased from 0 to 0.01 min-1, the time of maximum detectability shifted earlier, as the decreasing FDG concentration in the tumor lowered the CNR. These results imply that delayed PET imaging may reveal low-conspicuity lesions that would have otherwise gone undetected.

6.2 INTRODUCTION

Positron emission tomography (PET) is clinically used to determine the spread of breast cancer as well as response to therapy [131]. Clinical ^{18}F-fluorodeoxyglucose (FDG) PET imaging protocols typically state that images are to be acquired 45 to 60 minutes post-radiotracer injection [132,133]. However, it has been shown that FDG uptake in malignant tumors increases for many hours after injection, while benign and normal tissue uptake begins to decrease at an earlier time point, potentially allowing for better identification of malignant lesions [134,135,136,137,138,139]. Thie et al. [140] discussed the potential for improved detection with PET imaging, stressing that contrast is a time-dependent quantity and concluding that choosing an optimum imaging time is a complex issue. Some studies have investigated dual time-point imaging to exploit these time-dependent trends, comparing lesion SUV at early and later times [141,142].

The goal of this work was to investigate if and when there is an improved imaging time point for a single static PET scan. Instead of looking at discrimination between benign and malignant lesions, we focused on what imaging time might improve lesion detection, particularly for lesions with low contrast on the threshold of detectability. We hypothesized that imaging at a later time point than current clinical practice, such as two or three hours post-injection versus one hour, can improve the contrast-to-noise ratio (CNR). We tested our hypothesis in a virtual clinical trial, combining measured dynamic PET data with kinetic modeling and PET data generation simulations. We reconstructed images using an ordered subsets expectation maximization (OSEM) reconstruction algorithm. We then applied model observers with a specific detection task to characterize the reconstructed images and determine the time point of maximum lesion conspicuity.

6.3 METHODOLOGY

We tested our hypothesis by combining measured dynamic PET data with kinetic modeling and PET data generation simulations. We reconstructed images using an ordered subsets expectation maximization (OSEM) reconstruction algorithm. We then applied detection model observers to characterize the reconstructed images. We also investigated the effects of post-reconstruction image processing that affects background properties such as noise texture on lesion detectability. A summary of the chain of simulations is shown in Figure 6.1.

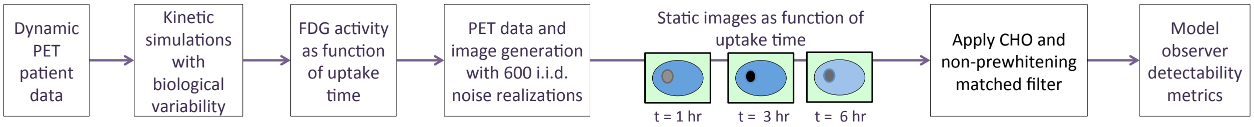


Figure 6.1. Summary flow diagram of simulation methods.

Table 6.1. Patient characteristics (n = 24).

	Mean Value	Range
Age (yr)	61.7	51.6 – 80.0
Weight (kg)	80.9	43.6 – 141.8
Injected dose (MBq)	315	230 – 366
Blood glucose (mg/dL)	101.5	82 – 125
Tumor diameter (cm)	1.6	0.6 – 3.7
Tumor grade	1.6	1 – 2
Biopsy ER (Allred score)	8	7 – 8
Biopsy PgR (Allred score)	5.9	0 – 8
Ki-67 (% staining)	19.5	5 – 70

ER = estrogen receptor
PgR = progesterone receptor

6.3.1 *Patient data*

We obtained 60-minute dynamic PET datasets from 23 patients with ER+/HER2- grade 1 and 2 early stage breast cancer tumors before receiving therapy. Two patients had a second tumor in the contralateral breast. The institutional review board (IRB) approved the prospective study that acquired the patient data. Patient characteristics are described in Table 6.1. Patient data were acquired between April 2010 and November 2013 using a GE Discovery STE Scanner (GE Healthcare, Waukesha, WI). The injected doses ranged from 222 to 370 MBq. Tumor size was measured by ultrasound or mammography.

Images were reconstructed using varying time bin durations post-injection, ranging from five-second time intervals initially to five-minute time intervals from 25 to 60 minutes. The OSEM algorithm used 28 subsets over six iterations. Image voxel size was 3.2 x 4.3 x 4.3 mm. An example reconstructed image for Patient 11 with a Grade 2, 1.8 cm sized tumor is shown in Figure 6.2A. The tumor and normal breast tissue TACs are shown for all patients in Figure 6.2B. Both the tumor and normal breast tissue regions-of-interest (ROIs) were 3 x 3 x 3 voxels, and the hottest tumor pixel was within the ROI. The tumors were not partial volume corrected. The normal breast tissue ROI was placed in the most homogeneous portion in successive tissue in the contralateral breast. One patient with breast implants was excluded from this virtual study due to abnormally high normal tissue data. Another tumor with high activity (>40 kBq/cc) was also excluded, as it was outside the range for this detectability study.

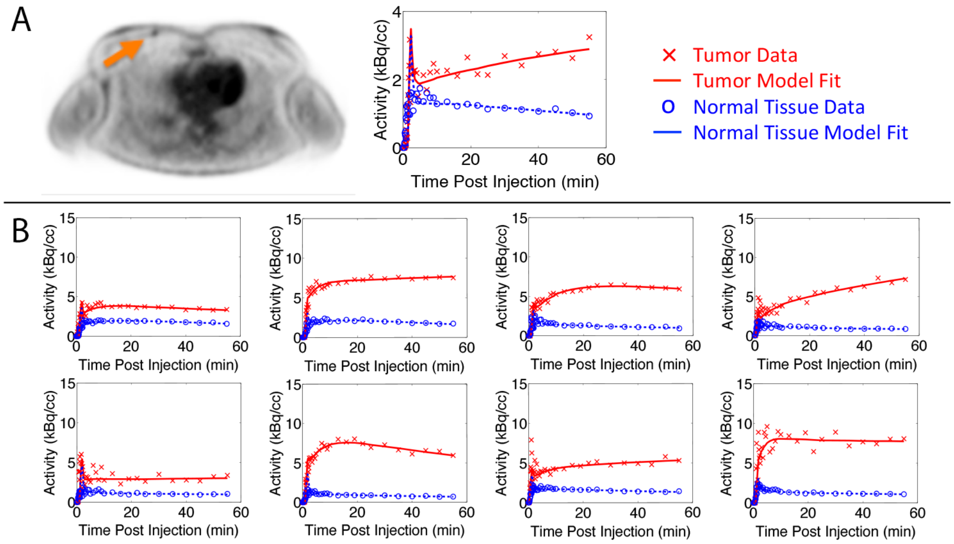


Figure 6.2. (A) Reconstructed patient image along with measured and modeled decay-corrected TACs for Patient 11 with a grade 2, 1.8 cm diameter tumor. The acquired dynamic data are shown as data points, and the TACs generated using the parameters estimated with the kinetic model are shown as lines. (B) TACs and model curves for all patients. The data are plotted based on the middle time point of the time bin.

6.3.2 *Kinetic modeling*

We used a two-tissue compartment model [143] in PMOD kinetic modeling software (PMOD Technologies Ltd, Zurich, Switzerland) [144] to estimate the kinetic parameters. The model was fit to the dynamic tissue data through non-linear optimization of the kinetic parameters. Patient-specific arterial input functions (AIF) were extracted from the dynamic time-course [145,146] and then scaled using the activity concentration measured from a left ventricular ROI drawn over the heart. The vascular blood fraction of tissue was fixed at 0.04 mL/mg based on reported experimental measurements of human breast and normal tissues [147,148]. Measured TACs and the fit model curves for Patient 11 are shown in Figure 6.2B. The estimated parameters are summarized in Table 6.2.

Table 6.2. Estimated kinetic parameters for all patients for nonreversible model ($k_4 = 0$).

	Mean ± SD	Range
K_1 (mL/g/min)	0.043 ± 0.024	0.008 – 0.083
k_2 (min^{-1})	0.144 ± 0.086	0.039 – 0.448
k_3 (min^{-1})	0.020 ± 0.015	0.0026 – 0.062
Ki	0.005 ± 0.003	0.0006 – 0.013

SD = standard deviation

The estimated kinetic parameters were used to generate simulated TACs out to many hours post-radiotracer injection. For forward modeling of the simulated curves, we used an averaged AIF obtained from arterial blood sampling of 13 patients over 5.8 hours [149]. The AIF was scaled so that the values at 60 minutes matched those of each individual patient's AIF at 60 minutes. To simulate biological variability, we generated 600 TAC realizations by adding 10% coefficient of variation to K_1, which controls the rate of uptake of the FDG into the tissue compartment.

We first assumed that k_4, the dephosphorylation rate constant of FDG-6-P, is zero. This is supported by evidence that FDG becomes trapped in the tumor and is only slowly dephosphorylated, due to the low concentration of glucose-6-phosphatase in most cancers with some notable exceptions, such as liver cancer [150]. In a number of prior studies ranging to one or two hours post-radiotracer injection, FDG metabolism was assumed to be irreversible, and k_4 was assumed to be zero [137,151,152,153,154]. However, realizing that k_4 has a greater impact on uptake curves at later time points, we investigated the case of non-zero k_4. The impact of non-zero values of k_4 is not evident during the 60-minute time period due to the small pool size and low parametric rate. This results in very noisy dephosphorylation estimates from the kinetic model. Consulting published values in the literature [151, 155], we assumed three dephosphorylation rates of 0.001, 0.005, and 0.01 min^{-1} in order to evaluate the impact of k_4 on detectability at later time points.

6.3.3 *Image data generation*

The simulated TACs were sampled at five time points (0.5, 1, 2, 3, and 4 hrs) and the tumor and normal tissue FDG activity values were input into a modified NEMA Image Quality phantom with six spheres in different axial slices with a given size and contrast. It was important that the simulated data be based on measured patient data. However, the tumors in the patient data were all readily detectable. In order to generate lesions at the limit of detectability, we simulated the measured radiotracer uptake in a 5 mm diameter tumor. ASIM, an analytical PET simulator [156], was used to generate sinogram data based on a GE Discovery STE PET/CT scanner.

Poisson noise was added to each sinogram based on the expected number of prompt counts as a function of time. The ratio of trues:scatter:randoms coincident events was 1:1:0.35. The initial count rate was determined using the raw patient data following a 370 MBq injection; for a five-minute bed position, 30 and 60 minutes post-injection, 500 and 350 million counts would be acquired, respectively. The number of counts as a function of time beyond the acquired patient data was then dependent on two factors; the first factor was radioactive decay, and the second was the impact of changing activity concentration (due to biological decay or washout) in the field-of-view. The number of acquired counts was assumed to scale proportionately to the normal tissue activity, as the tumor tissue activity has little impact on the overall number of counts in the sinogram.

6.3.4 *Image reconstruction*

Images were reconstructed using OSEM with up to six iterations and 32 subsets, and post-filtered with a 6 mm FWHM Gaussian filter. The results shown here are after two iterations. The image voxel size was 3.2 x 4.3 x 4.3 mm.

6.3.5 *Model observers*

We used both a nonprewhitening matched filter (NPWMF) and channelized Hotelling (CHO) observer model estimate lesion detectability [157,158]. The CHO observer used four non-overlapping rectangular filters and has been validated against human observers [159]. The CHO template was a 20 x 20 pixel region centered over the tumor based on six training images from each set of simulation parameters (2760 images total). The template was then compared separately to the tumor and normal tissue patches to generate test statistics for each over all noise realizations. The histograms of the 600 tumor and 600 normal tissue test statistics for each simulated case were used to generate receiver operating characteristic (ROC) curves. From the ROC curves, we calculated the area under the curve (AUC) for a particular lesion size and contrast and acquisition time post-injection. A summary of all simulation steps and parameters is shown in Table 6.3.

Table 6.3. Simulation summary

Parameter	Value
Number of patients	23
Kinetic parameters	K_1, k_2, k_3 estimated k_4 = 0, 0.001, 0.005, 0.01 min^{-1}
Biological noise	10% CV added to K_1
PET data noise	Poisson noise added to sinogram
Noise realizations	600 per patient
Time points	0.5, 1, 2, 3, 4 hours post-injection
Regions of interest	Tumor and normal tissue

CV = coefficient of variation

6.4 RESULTS

The extrapolated noise-free TACs are shown in Figure 6.3 for Patient 11. In the nonreversible model, the tumor uptake increased for many hours before it began to plateau, which agrees with other reported results that showed FDG uptake can increase for up to six hours before plateauing [134,136,137]. The normal tissue activity concentration decreased with time. In the reversible model, the impact of non-zero k_4 was loss of activity from the bound tissue compartment (C_m) and was observed as a lower tissue TAC.

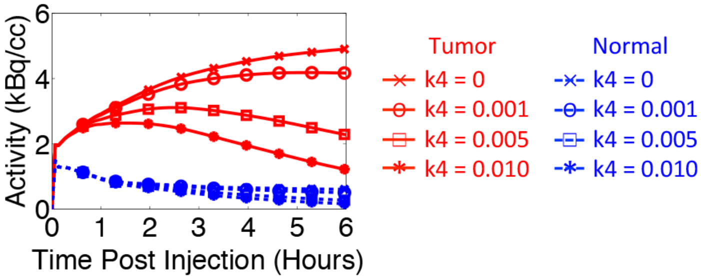


Figure 6.3. Synthetic TACs for tumor and normal tissues assuming k_4 = 0, 0.001, 0.005, and 0.01 min^{-1}.

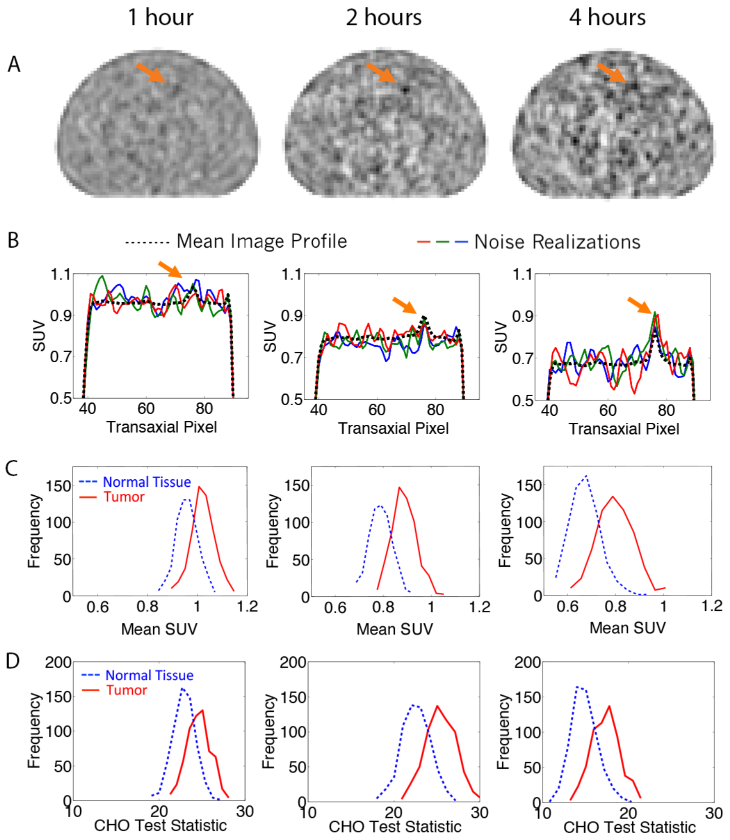


Figure 6.4. For the nonreversible model at 1, 2, and 4 hours post-injection: (A) OSEM reconstructed images for Patient 11 at one, two, and four hours post-injection with $k_4 = 0$. The arrow points to the location of the lesion. (B) Horizontal profiles for three noise realizations as well as the mean profile over all realizations. (C) Histograms of normal and tumor tissue SUVs, with increasing separation between peaks with time as well as a broadening of the distributions indicating increasing noise. (D) Histograms of normal and tumor tissue CHO detectability metrics for Patient 11 with increasing imaging time post-injection and the result ROC curves. The greater the separation between peaks, the higher the detectability.

For the nonreversible model, reconstructed decay-corrected OSEM images are shown in Figure 6.4A for Patient 11. The images are for one, two, and four hours post-injection and are normalized by the mean. Using visual assessment of this one case, the lesion at one hour could be difficult to identify without *a priori* knowledge of the lesion location. With increasing time, the lesion contrast increased but so did the background noise.

The horizontal profiles are shown in Figure 6.4B for three noise realizations as well as the mean image over all 600 realizations. The horizontal profiles across the image and lesion reflect both the increasing contrast and noise. Histograms of the normal and tumor tissue SUVs are shown in Figure 6.4C, where the ROIs matched the size and shape of the tumor.

The ability to differentiate between the tumor and normal tissue is visualized in the histograms of the CHO statistic in Figure 6.4D. The increase in peak separation, and consequently lesion detectability, was due to increased tumor to normal tissue contrast. The increasing noise with time post-radiotracer injection is represented in the increasing width as a result of radiotracer decay.

ROC curves plotting true positive fraction (TPF) versus false positive fraction (FPF) were generated from the CHO histogram distributions. The calculated CHO AUC is shown for Patient 11 in Figure 6.5A. NPWMF trends (data not shown) matched that of the CHO. The AUC results for all patients showed that they divided into three groups. Some tumors showed improvement in detectability with later imaging times, whereas others showed decreasing detectability with increasing time. The third group included the tumors that, despite the small 5 mm size, were always detectable. The results for the first two groups are shown in Figure 6.5B. The curves were divided based on the time of peak AUC; 11/23 (48%) of patients' AUCs peaked after one hour, and 7/23 (30%) of patients' AUCs peaked before or at one hour. The remaining five patients (22%) had an AUC was saturated at one for all time points, indicating that the tumor was always detectable.

The kinetic modeling simulations and analyses were repeated for the reversible model. The AUC results for Patient 11 are shown in Figure 6.5A. Figure 6.5C shows the patients whose AUC curve peaked after one hour in Figure 6.5B. The AUC decreased with increasing k_4 and with increasing time post-injection, shifting maximum detectability to an earlier time point; of the 11 patients who benefited from delayed imaging when $k_4 = 0$, seven benefited with $k_4 = 0.005$ min^{-1} and only two benefited when $k_4 = 0.01$ min^{-1}.

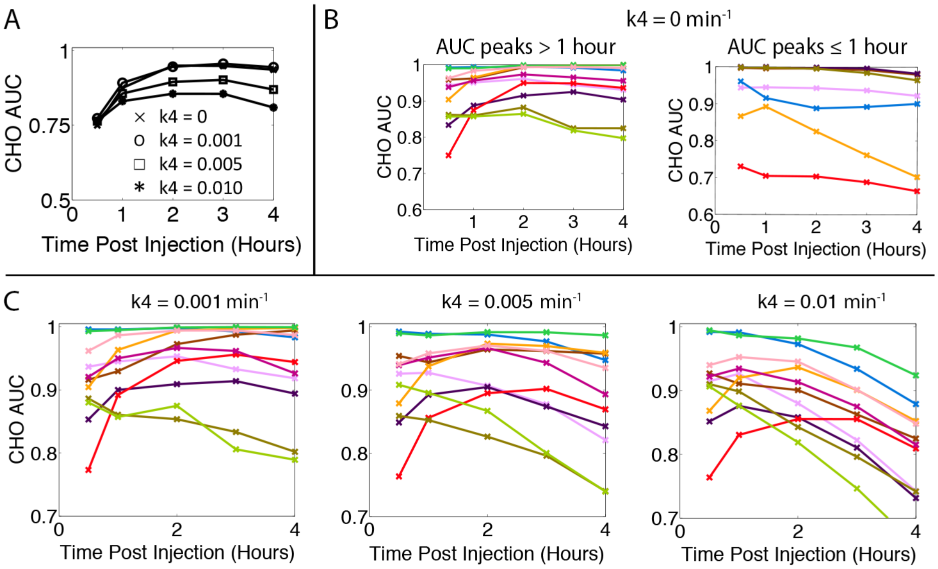


Figure 6.5. (A) AUC as a function of time calculated from CHO ROC curves for Patient 11. (B) AUC results for all patients using the nonreversible model, divided by the AUC peaking after 1 hour (left) or at or before 1 hour (right). Each patient is represented as a different line and color. (C) AUC for the reversible model and patients who benefited from delayed imaging when $k_4 = 0$. As k_4 increases, the peak AUC, or lesion detectability, shifts earlier.

6.5 DISCUSSION

We have assessed the effect of uptake time on tumor detectability, evaluating both a non-reversible and reversible FDG phosphorylation model. As was hypothesized, the SUV and CHO detectability results showed a trade-off between increasing tumor-to-background ratio and increasing image noise due to radiotracer decay. Higher tumor contrast is reflected with increased separation between the peaks of the SUV and CHO detectability metric histograms, whereas increased noise is reflected in the broadened width of the histograms. The AUCs from the CHO metrics quantify this trade-off.

This virtual clinical trial was based on measured patient data in early state breast cancer to generate TACs out to four hours post-radiotracer injection. Assuming the available dynamic data have been acquired to equilibrium, the kinetic parameters (K_1-k_3) do not change significantly, and synthetic curves generated beyond those time points should be realistic [154]. In the non-reversible model, it is assumed that FDG becomes trapped in the tumor and cannot be hydrolysized by the enzyme glucose-6-phosphatase. Whereas hexokinase phosphorylates FDG into FDG-6-phosphate (FDG-6-P), glucose-6-phosphatase dephosphorylates FDG-6-P back to FDG that can leave the cell. Although this enzyme is found in many organs, it is most active in the liver and kidney [160], and phosphatase activity levels are low in most cancer cells [161]. We first investigated the effects of imaging time on tumor detectability following the assumption of $k_4 = 0$ and then applied the same analyses to the reversible kinetic model, as it is likely that the effects of k_4 are non-negligible after the first one or two hours.

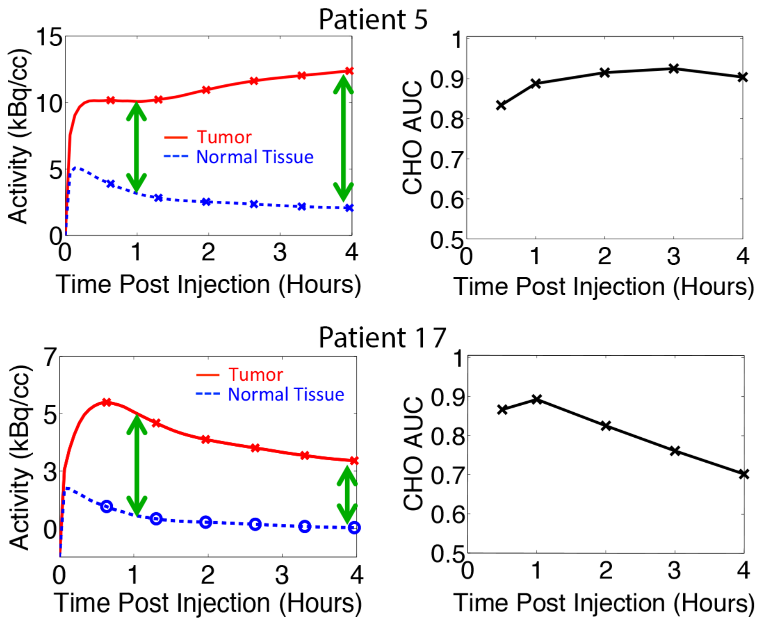


Figure 6.6. Comparison of Patient 5 and Patient 17 TACs to the resulting CHO AUC curves when $k_4 = 0$. Patient 17 showed the most significant decrease in tumor concentration as well as decrease in AUC as a function of time.

We observed that the patients divided into three groups based on the detectability trends reflected by the CHO AUC. To better understand why some tumors appeared to benefit from delayed imaging, we related the AUC trend back to the original TACs and the contrast between the tumor and normal tissues. The TACs for two example patients are shown in Figure 6.6. In the first group who would have benefited from delayed imaging, the AUC curves kept increasing two to three hours post-injection. The TACs for this time period show the tumor concentration increased while the normal tissue concentration decreased. The increasing contrast thus outweighed the increasing noise until later time points when the noise began to dominate and the AUC decreased. In the second group who would not have benefited from delayed imaging, the AUCs decreased with time due to the impact of noise outweighing the contrast. The tumor concentration was instead flat or decreased with time. Therefore, tumors with low initial contrast and/or a significant increase in contrast with time showed increased detectability with delayed imaging. The third group of patients comprised a ‘neutral’ group, with no detectability trend as a function of imaging time. These five patients had tumors with a very large tumor to normal ratio, such that the tumors were always detectable (AUC = 1) despite the small tumor size. As k_4 was increased, the concentration of FDG in the tumor decreased more rapidly with time than in the normal tissue as expected. This difference resulted in decreased tumor contrast and thus decreased detectability, which is reflected in the decreasing AUC trend. With the larger $k_4 = 0.01$ min^{-1}, most AUC curves peaked within one hour.

The detectability trends are dependent upon the rates of FDG accumulating in the tumor, leaving the tumor, and clearing from the blood. The separation of these groups can be predicted by looking at the kinetic rate constants. The overall metabolic flux of FDG into the tumor is calculated as $K_i = K_1 k_3/(k_2 + k_3)$, with a higher flux distinguishing group one versus group two with a two-sided significance of $p = 0.0001$. A higher rate of phosphorylation, k_3, predicted group one versus group two with $p = 0.012$. The rate constants also help explain why the FDG concentration peaked early (between 10 and 30 minutes) in seven of the tumors. These tumors comprised 6/7 (86%) of the patients who did not benefit from delayed imaging, with the remaining patient part of the neutral group. This early peak could be due to dephosphorylation ($k_4 > 0$), and Lodge et al. [134] have noted that there are other causes of decreasing activity concentrations, such as unmetabolized FDG clearing more rapidly from the tissue precursor pool back into the blood (k_2), possibly due to a low fixation rate (k_3). Vriens et al. [162] have reported that low metabolic tumor regions have a higher blood volume and therefore higher uptake of FDG in the early time points post-injection. With a low phosphorylation rate in these regions, the FDG was able to clear back into the blood pool.

It is clear that the detectability trends as a function of time are unique. Ideally, the optimum uptake time for detection could be predicted by tumor type before imaging. We explored possible associations between the detectability trends with tumor characteristics, and only one significant correlation was found when we performed a multivariate analysis of tumor size, grade, hormone receptor status (HER2neu, estrogen, progesterone), Ki-67, and the metabolic rate; a higher metabolic rate predicted tumors to be in group one versus group two with a significance of $p = 0.046$. It was also expected that the groups would separate based on Ki-67, and the observed lack of correlation could be due to the low sample size. The correlation between the biology and kinetic parameters may also have been obscured to the small tumor size and lack of partial volume correction.

There is still the open question of why some patients would benefit from delayed imaging and how to potentially identify them beforehand. This could become more clear by testing delayed imaging on patients, simply asking them to return for a delayed scan after their scheduled clinical scan. Recommending a patient-specific imaging time may ultimately be impractical, and so the goal would be to better understand the trade-offs in imaging time and instead recommend a single time point based on the average response for a group of patients. Even for group two who would not have benefited from delayed imaging, the AUC at later imaging times was relatively flat or only gradually decreasing, and delayed imaging may not have compromised detectability in those cases. Further, the data acquisition time could be lengthened at later time points to compensate for the increased noise due to radiotracer decay and take advantage of the increased contrast. Modern 3D PET systems and time-of-flight technology could also reduce the impact of noise. Finally, although we focused on single time point imaging that requires minimal change to the current workflow and would be more feasible to implement clinically, it may be that dynamic or multiple time point imaging would be helpful by providing the extra uptake information.

Our study had some limitations. Patient data is limited to 60-minute dynamic scans for 23 patients with a lower to intermediate grade tumor and favorable (ER+/HER2-) subtype of breast cancer, which may or may not be similar to other tumor types and subtypes. All data were from actually detectable tumors and so assumptions were made to adjust the tumors toward the limit of detectability. It is not intended to imply that the same tumor would have had the same uptake characteristics at a smaller size but, rather, that a different tumor could have these uptake characteristics at this size. The two-tissue compartment kinetic model simplifies the biological processes and assumes homogeneity in the compartments [132]. Detectability performance was estimating using a model observer in the SKE-BKE case, which may not predict human observers. In this case, we previously validated our model observer against human observers and found good agreement [159].

There is the outstanding question of FDG dephosphorylase activity and when k_4 can be assumed negligible. If FDG has only minimal reversibility, it can adequately be modeled as irreversible for short-duration studies, and the tumor uptake will increase for many hours. In this study, if the acquired 60 minutes of data showed increasing activity, the synthetic curves showed increasing detectability with time before plateauing at much later times. This result agrees with that of other studies of breast and other cancers [134,135,136,138,139] that have shown the tumor FDG concentration continues to increase for many hours, indicating that k_4 is indeed small. For this reason, it is not uncommon for studies to assume $k_4 = 0$. For example, Hamberg et al. [137] acquired non-small cell lung cancer data to 90 minutes and then assumed $k_4 = 0$ when extrapolating to infinity. However, both Lodge et al. [134] and Spence et al. [136] acquired data up to six and eight hours and reported non-zero values of k_4 in brain (gliomas and gray matter) and soft tissue masses (benign and malignant), respectively. Lodge et al. showed that dephosphorylation becomes evident in soft tissue sarcoma TACs after four hours post-injection, and the results of Spence et al. showed a decrease in brain tumor between 3.5 and 5 hours. In the case of brain tissue, it has been shown the assumption of a non-reversible model is valid for up to only two hours post-injection [133,154]. Therefore, it was important to consider the possible effects of $k_4 > 0$ in this study.

Most dynamic studies are conducted to 60 minutes, as were our breast cancer patient data. However, Lucignani et al. [154] showed that even estimating parameters in the reversible model is not accurate when using experimental data less than two hours. In particular, the estimation of k_4 is difficult from only 60 minutes of dynamic data because the effects of a small k_4 and initial small FDG-6-P pool size do not significantly impact the uptake curves over that time frame. Not surprisingly we found our k_4 estimates to be highly uncertain. We instead decided to fix the dephosphorylation rate constant for three values larger than zero based on values published in the literature. Unfortunately, there is little data available on k_4 estimates for cancer and even less so for breast cancer specifically. Krak et al. [151] reported an average value of $k_4 = 0.003 \pm 0.022$ min^{-1} in locally advanced breast cancer (LABC). Tseng et al. [155] estimated a much higher value also in LABC of 0.014 ± 0.012 min^{-1}, but still with a large uncertainty, confirming our experience that k_4 is indeed difficult to estimate. The k_4 values we simulated spanned a range of plausible values, accounting for the impact of FDG dephosphorylation at later times, although the level of dephosphorylation remains undetermined. The above effects are potential explanations for the decrease in the tumor uptake shown for Patient 17 (Figure 6.6) even when k_4=0.

6.6 CONCLUSIONS

Using lesion properties from breast cancer FDG studies, it was shown that delayed imaging at two to three hours post-radiotracer injection may improve tumor detectability in some patients. We used TACs that captured the biology of the tumor to simulate realistic PET images in a virtual clinical trial. Tumor detectability depended on the tumor and normal TACs and trade-off of increasing contrast and increasing noise with time post-radiotracer injection. Although FDG dephosphorylation is often considered negligible, we have shown that the imaging time for maximizing lesion detectability is dependent on the rate of FDG leaving the tumor, either by dephosphorylation or other pathways, which has not yet been accurately characterized in breast cancer. This method of a virtual clinical trial can be used to guide a future clinical trial in patients to further evaluate the impact of imaging time and other parameters on tumor detection, response to therapy, and outcome measures.

Chapter 7. A VIRTUAL CLINICAL TRIAL COMPARING THE ABILITY OF STATIC VERSUS DYNAMIC PET IMAGING METRICS TO MEASURE TUMOR RESPONSE THERAPY

7.1 ABSTRACT

We developed a method to evaluate variations in the PET imaging process in order to characterize the ability of static and dynamic metrics to measure breast cancer response to therapy in a clinical trial setting. We performed a virtual clinical trial by generating 540 i.i.d. PET imaging study realizations for each of 22 original fluorodeoxyglucose (^{18}F-FDG) breast cancer patient studies pre- and post-therapy. Each noise realization accounted for known sources of variability in the imaging process. We then performed a ROC analysis on the resulting SUV and kinetic parameter uncertainty distributions to assess the impact of the variability on the measurement capabilities of each metric. The kinetic macro parameter, Ki, showed more variability than SUVmax (CV of 16.6% compared to 13.5%). However, for the patients who did not show perfect separation between the pre- and post-therapy parameter uncertainty distributions (AUROC<1), dynamic imaging outperformed SUVmax in distinguishing metabolic change in response to therapy (14/16 patients, $p<0.05$). For the patient cohort in this study, which comprised of non-high-grade ER+ tumors, Ki outperformed SUVmax in an ROC analysis of the parameter uncertainty distributions pre- and post-therapy. This methodology can be applied to many different scenarios with the ability to inform the design of clinical trials.

7.2 INTRODUCTION

Positron emission tomography (PET) imaging is a diagnostic tool used to both quantify and verify the extent of disease (e.g. cancer staging) and to monitor treatment response or measure disease progression. Research and development efforts aimed at improving the performance of PET imaging in these tasks span a complex space of hardware, software, and clinical protocols. Assessing the impact of modifications to any of these areas on image-based tasks can be time-consuming and difficult.

In this study, we focused on the application of PET imaging to assess response to therapy and its ability to measure change in radiotracer uptake. This ability depends upon a number of factors, including both the bias and uncertainty in the measurements pre- and post-therapy. Numerous sources of variability in the imaging process lead to uncertainties in image-derived metrics that are often large relative to the underlying signal. Some of the physiological and technical sources of uncertainty have been summarized in Thie et al. [163] and Boellaard et al. [164]. These include dose and scanner calibrations, blood glucose level, inflammation, uptake time, and image reconstruction parameters.

Both kinetic parameters and standardized uptake value (SUV) have been used to monitor response, though SUV is more common owing to its easier estimation [165,166]. While simplified methods such as SUV are more clinically practical to assess treatment response, studies have found results and conclusions could differ from those derived from a full kinetic analysis using dynamic image data [166,167]. Static SUV measurements are subject to variability in uptake time, body habitus, administered dose measurements, and other factors [168]. Kinetic parameter estimates may be more robust against such variations, but estimating the parameters is also more complex, requiring an arterial input function and dynamic data acquisition, and the range of error can be higher for kinetic parameters than for SUV [169]. Under certain conditions it has been found that estimates of kinetic parameters are better than SUVs in distinguishing response [170,171], particularly for lower uptake tumors [172]. Despite such studies comparing SUV and kinetic parameters, the inherent bias and variance of these parameter estimation processes and the impact on response assessment have not been fully characterized.

Virtual clinical trials (VCTs) can provide a detailed characterization of such output metrics by evaluating the impact of any part of the imaging process, including patient protocol, scanner design, image reconstruction, and image analysis methods. Harrison et al. [173] utilized a VCT to evaluate the impact of tumor size, scan duration, and image reconstruction filter on SUV variability and the ability of SUV to measure change. Other studies have used VCT simulation techniques to aid in the design of clinical studies by determining study power [174], sample size [175], and drug dosing [176,177]. Through VCTs, we can improve our understanding of any parameter of interest and inform the design of prospective clinical trials, with the goal to improve the success rate, reduce the cost of clinical development, and more quickly translate imaging advancements into the clinic. In this work, we conducted a VCT to characterize the variation of the SUV measurements and kinetic parameters estimates and then evaluate the ability of these metrics to measure a change in radiotracer uptake in response to therapy.

7.3 METHODOLOGY

An overview of the simulation process and which uncertainties impact each step is shown in Figure 7.1. Each step is described in more detail in the following sections.

7.3.1 *Patient data*

The patient datasets were 60-minute dynamic acquisitions of 22 patients with non-high-grade (grades 1 and 2), estrogen receptor positive (ER+) primary breast cancer pre- and post-therapy. Patient characteristics are shown in Table 7.1. The scans were performed on a GE Discovery STE PET/CT Scanner (GE Healthcare, Waukesha, WI). The data were reconstructed using 3D OSEM with 28 subsets, six iterations, and a 7 mm transaxial Gaussian post-filter. The image voxel size was 3.3 x 4.3 x 4.3 mm. The tumor and normal breast tissue regions-of-interest (ROIs) were 3 x 3 x 3 voxels (1.6 cc). The normal breast tissue ROI was placed in the most homogeneous portion in successive tissue in the contralateral breast. The patient dataset is the same as in reference [178], where it is described in more detail. We note here that we did not consider partial volume effects in the estimate of SUV, as the tumor was not expected to significantly change in size between the pre- and post-therapy scan (average time = 20 days) when treated with hormonal therapy.

Table 7.1. Patient characteristics pre-therapy (n = 22).

	Mean Value	Range
Age (yr)	61.7	51.6 – 80.0
Weight (kg)	80.9	43.6 – 141.8
Injected dose (MBq)	315	230 – 366
Blood glucose (mg/dL)	101.5	82 – 125
Tumor diameter (cm)	1.6	0.6 – 3.7
Biopsy PgR (Allred score)	5.9	0 – 8
Ki-67 (% staining)	27.3	3.9 – 70.8

PgR = progesterone receptor

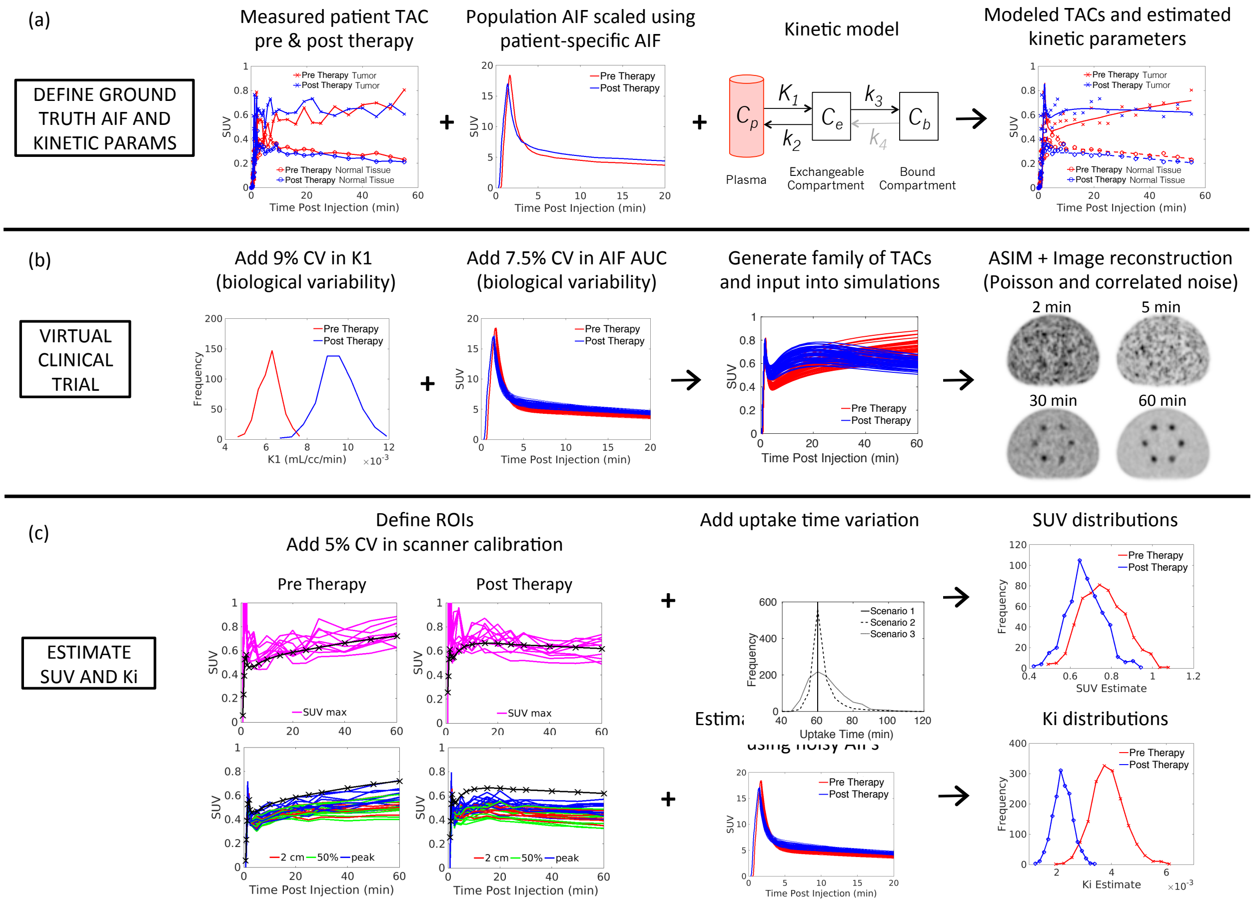


Figure 7.1. Virtual clinical trial simulation of the imaging process and results for patient 11. (a) Measured tumor and normal TACs and the arterial input function were input into a two-compartment model to estimate the kinetic parameters and generate model TACs. (b) A virtual clinical trial was performed by adding uncertainty to K_1 and the AIF, which were used to generate a family of TACs. Each TAC was used to simulate FDG uptake as a function of time for a tumor. While this process was performed for both the tumor and normal tissues, only the tumor data are shown. (c) SUVmax at 60 minutes post-injection and dynamic ROIs were measured from the tumors in the reconstructed images. Uptake time uncertainty was added to SUVmax to generate the final SUVmax uncertainty distributions pre- and post-therapy. The dynamic ROIs were re-input into the kinetic model to generate the Ki uncertainty distributions.

A two-tissue compartment model with irreversible trapping [179] in PMOD kinetic modeling software (PMOD Technologies Ltd, Zurich, Switzerland) [180] was used to estimate the kinetic parameters. Patient-specific arterial input functions (AIF) were extracted from the dynamic time-course [181,182] and then scaled using the activity concentration measured from a left ventricular ROI drawn over the heart. The vascular blood fraction of tissue was fixed at 0.04 based on reported experimental measurements of human breast and normal tissues [183,184]. We assumed irreversible trapping, in which k_4, the dephosphorylation rate constant of FDG-6-P, is zero. This is supported by evidence that FDG becomes trapped in breast tumors with slow retrograde metabolism due to the low concentration of glucose-6-phosphatase in most cancers [185,186]. In other studies ranging to one or two hours post-radiotracer injection, FDG metabolism was assumed to be irreversible, and k_4 was assumed to be zero [187,188,189,190]. Additionally, we compared the estimate Ki from our model with that from a Patlak graphical analysis [191,192], where k_4 is assumed to be 0, and found the correlation to be 0.99.

Example patient images are shown in Figure 7.2. Measured tumor and normal tissue time-activity curves (TACs) with model curve fits pre- and post-therapy are shown in Figure 7.1(a). The model curve fits were generated by parameter estimation using non-linear least squares optimization fitting the model to the TAC data. Statistics on the kinetic modeling parameter estimates for all patients are summarized in Table 7.2.

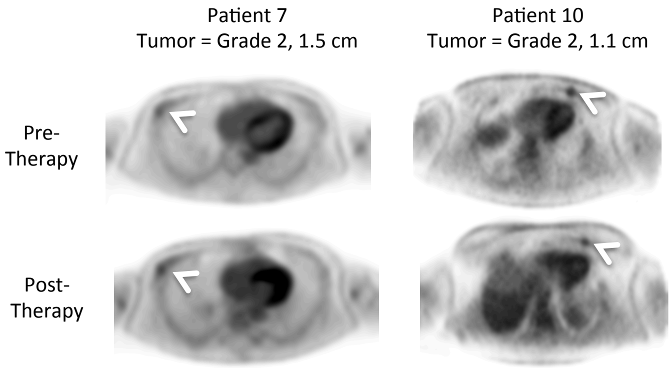


Figure 7.2. Reconstructed patient images (summed 30-60 minutes of dynamic data) pre- and post-therapy.

Table 7.2. Descriptive statistics for estimated tumor kinetic parameters (n = 22).

	Pre-therapy		Post-therapy	
	Mean ± SD[†]	Range	Mean ± SD	Range
K_1 (mL/g/min)	0.043 ± 0.024	0.008 – 0.083	0.035 ± 0.020	0.005 – 0.084
k_2 (min^{-1})	0.144 ± 0.086	0.039 – 0.448	0.119 ± 0.059	0.033 – 0.268
k_3 (min^{-1})	0.020 ± 0.015	0.004 – 0.062	0.013 ± 0.007	0.001 – 0.030
Ki	0.005 ± 0.003	0.0006 – 0.013	0.003 ± 0.002	0.0005 – 0.008

[†] SD = standard deviation

7.3.2 *Synthetic TAC generation*

The patient-specific estimated kinetic parameters were used to generate simulated tumor and normal tissue TACs to two hours post-radiotracer injection for each patient pre- and post-therapy. For forward modeling of the simulated curves, we used an averaged AIF obtained from arterial blood sampling of 57 cancer patients [193]. The population-based AIF was scaled based on the late measurements (30-60 min) of the image-derived AIF for each individual patient.

We added biological variability and generated 540 TAC noise realizations for each pre- and post-therapy patient study. We added the variability based on the findings of Weber et al. [194], who investigated the variability of SUV and kinetic measures for repeat PET scans and found the coefficient of variation (CV) of the kinetic parameter, K_1, and area under the AIF (AUAIF) to be

9% and 7.5%, respectively. K_1 describes the rate FDG delivery to the tissue compartment from the blood, and the AIF reflects the time-course of the FDG through the arterial plasma. We assumed that the kinetic parameters k_2 and k_3 do not change significantly in a test-retest scenario. To add noise to K_1, we randomly sampled from a normal distribution with a CV of 9%. We added noise to the AIF by modifying its peak and shape. A three-exponential decay model [195] was first fit to the AIF. Noise was added to each parameter of the AIF model by sampling from a normal distribution with a mean and standard deviation matched to the population values. Random variation of six parameters resulted in AIFs that were unrealistic, so three constraints were enforced; any point in the new AIF could not deviate more than 15% from the original AIF, the curve must be monotonically decreasing, and the AUAIF must be within the lower and upper limits defined by the CV of 7.5%. The tumor K_1 uncertainty distributions, noisy AIF curves, and new simulated tumor TACs are shown in Figure 7.1(b) for patient 11.

7.3.3 *Image data generation*

The simulated tumor and normal tissue TACs were sampled at 22 time points, matching the timing of the dynamic imaging used in the original image acquisition. The tumor and normal tissue FDG activities were used to define the signal to background ratio of each tumor as a function of time. Considering 540 noise realizations, 22 time points, and 22 patients pre- and post-therapy, over 500,000 sinograms needed to be generated and then reconstructed. To make this task more tractable, we designed a test phantom with 36 spheres, where each sphere represented one noise realization. Therefore, only 15 phantoms were needed to generate the noise realizations for one time point for one patient, reducing the number of simulated sinograms to 14,520. The impact of varying tumor size was not a parameter in this study, so a uniform 20 mm sphere size was used for all patients. An example of an axial slice with six spheres is shown in Figure 7.1(b) for four dynamic time points.

ASIM, an analytical PET simulator [32] generated noise-free sinogram data from each input phantom based on a GE Discovery STE PET/CT scanner. Poisson noise was added to each noise-free sinogram based on the number of prompt coincidence events in the acquired patient data for each time bin as a function of time post-injection. The number of events in each time bin was dependent on the time bin duration, the amount of activity in the field of view, and radiotracer decay. For example, following a 300 MBq injection, 172 million prompt events would be acquired 60 minutes post-injection in a 5-minute duration time frame. The prompt events were assumed to have a trues:scatter:randoms ratio of 1:1:0.35.

7.3.4 *Image reconstruction*

Images were reconstructed using OSEM with two iterations and 32 subsets, and they were post-filtered with a Gaussian kernel having 6 mm and 4 mm full-width-at-half-maximum in the transaxial and axial directions, respectively. The image voxel size was 3.3 x 4.3 x 4.3 mm^3. Reconstructed images for four dynamic imaging frames are shown in Figure 7.1(b). The noise decreased with time due to increasing number of detected events, and the signal to background ratio increased due to increasing tracer uptake over time.

7.3.5 *Estimation of SUV and kinetic parameters*

The reconstructed images for each patient comprised 540 independently and identically distributed (i.i.d.) noise realizations for each dynamic patient scan pre- and post-therapy. The static SUV was measured from a 5-minute time bin at 60 minutes using the maximum SUV in each ROI, as SUVmax can be more robust than SUVmean and has become the standard in the clinic [196,197]. Another source of uncertainty in the static SUV measurement is uptake time, which is the time between radiotracer injection and imaging. Although the standard time is usually around 60 minutes [198], clinical workflow can result in the scan being performed earlier or later. Three different uptake time scenarios were used, as shown in Figure 7.1(c). Scenario 1 was perfect uptake time for a clinical protocol of imaging 60 minutes post-radiotracer injection. The other two scenarios reflect rigorous and less rigorous adherence to the clinical protocol [199], and they are plotted for 1000 samples. For each of the three scenarios, the SUVmax was re-estimated for all of the 540 image realizations.

From each time-course of simulated dynamic images, TACs were measured for tumor and normal tissue (i.e. feature and background) using the mean SUV in a given ROI. To account for ROI variation in practice, three ROI definitions were used. The ROI was defined using the static reconstructed image from 30 to 60 minutes and then applied to the individual dynamic images. The first ROI was an exact match of the tumor size and location, and was based on the known dimension of the simulated tumor (ROI mean). The second ROI was the mean of pixels in a 1 cc sphere centered on the maximum pixel (ROI peak). The third ROI used a 50% threshold of the maximum pixel value (ROI 50%). Additional Poisson distributed measurement noise was added to the AIF based on the activity in the each time bin [200]. The kinetic parameters were re-estimated, and the overall FDG flux constant, Ki, was calculated.

Finally, to account for additional uncertainties in scanner calibrations, an additional 5% CV was added to all measurements by randomly sampling from a normal distribution, based on measurements of variability for a carefully calibrated scanner over one month [201]. The same CV is added to all time points in one TAC. The final TACs constructed from SUVmax and the three ROIs are shown in Figure 7.1(c).

7.3.6 *Receiver operating characteristic curve analysis*

The SUVmax and Ki distributions for each patient pre- and post-therapy were used to generate receiver operating characteristic (ROC) curves. SUVmax comprised 540 measurements for three uptake times, and Ki comprised 540 estimates for three ROI definitions. The area under the ROC (AUROC) curves was calculated using the Wilcoxon rank-sum text.

AUROC is the metric of choice for assessing the ability of an imaging metric to classify patients as responders versus non-responders. However, as with many patient data, we did not know ground truth of the original tumor SUV or Ki. We also did not have outcomes for this patient cohort, so we could not assess or compare the abilities of SUV or Ki to predict patient response. Therefore, we characterized the variability of the meaurements and not the prognostic value of either metric. We assumed our measured FDG uptake patient data pre- and post-therapy to be ground truth and quantified the ability of the SUV and Ki parameters as quantitative imaging biomarkers to determine response to therapy. For a given change in SUV or Ki due to change in radiotracer uptake between pre- and post-therapy imaging, we used the AUROC to compare how well each metric could measure that change after including known sources of variation in the imaging process.

7.4 RESULTS

The AUROC for SUVmax using perfect uptake time and Ki estimated using ROI mean are compared for each patient. A bootstrapping analysis was performed individually for each patient using the 540 noise realizations to obtain confidence intervals and statistical significance of the difference between SUVmax and Ki. Figure 7.3(a) plots the AUROC SUVmax versus AUROC Ki, where the error bars repesent the confidence intervals and the line of identity separates which method better detected change between pre- and post-therapy tracer uptake. Figure 7.3(b) plots the difference between Ki and SUVmax along with the confidence interval. An AUROC difference with error bars that do not cross below zero indicates statistically significant higher Ki performance. The AUROC of the two metrics are also plotted for reference using the right axis. Six patients (27%) had perfect separation between pre- and post-therapy SUV and Ki distributions (AUROC = 1). Ki outformed SUV in measuring change in tumor FDG uptake between pre- and post-therapy for the majority of the remaining patients (14/16 (88%), p<0.05). Twelve patients showed an AUROC improvement greater than 0.05%. For two patients, patient 5 and 12, there was no significant difference in AUROC (p>0.05).

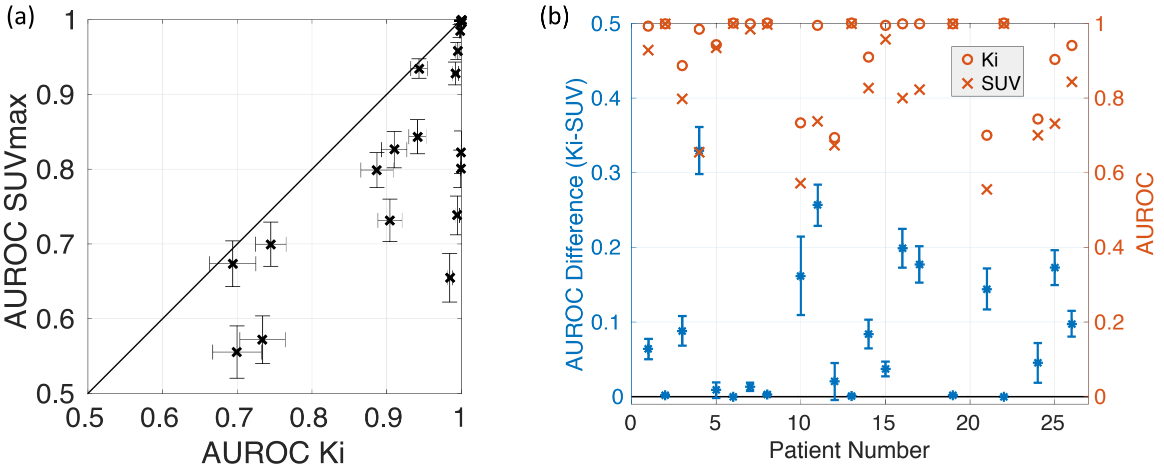


Figure 7.3. (a) Correlation plot of AUROC SUVmax versus AUROC Ki, which shows the benefit of Ki over SUVmax in measuring change in response to therapy. (b) AUROC difference between Ki and SUV with confidence intervals and corresponding AUROC values. Data are plotted for SUVmax using perfect uptake time (scenario 1) and Ki using ROI mean.

Measuring SUVmax with increasing uptake time variability and estimating Ki using different ROI definitions impacted both the variability as well as the calculated AUROC for each metric, as shown in Figure 7.4. The variation in SUVmax increased with increasing uptake time uncertainty, impacting the calculated AUROC. The AUROC increased or decreased depending on the shape of the time activity curves pre- and post-therapy. Similarly, the use of different ROIs impacted both the CV and calculated AUROC of Ki, as the tracer uptake estimates across different ROI estimates depend on the tumor size and spatial uptake distribution. The uncertainty in Ki estimation over all patients was higher than that for SUV with a CV of 16.6% compared to 13.5%.

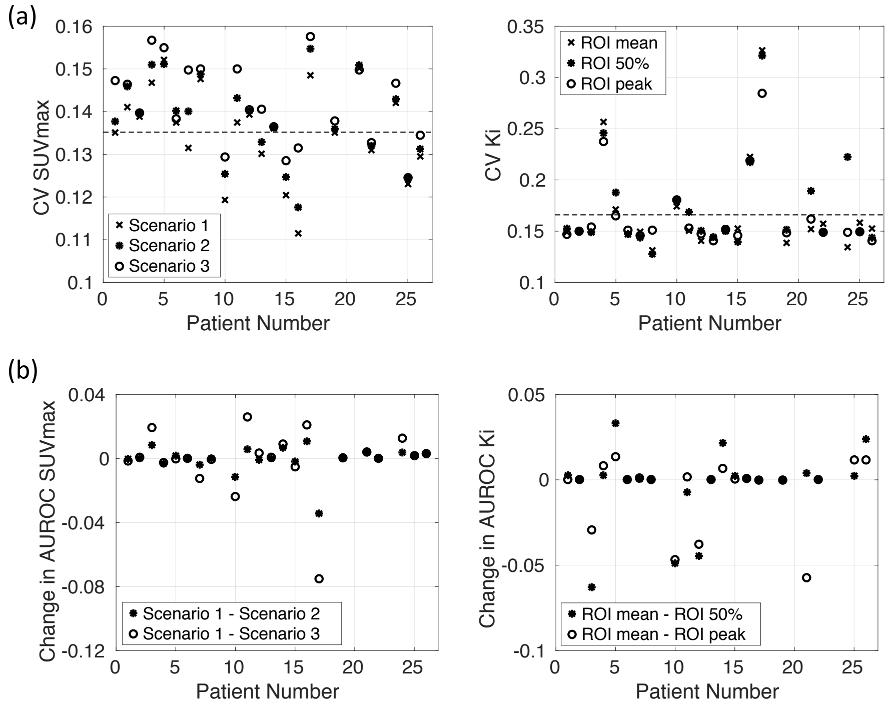


Figure 7.4. Increasing uptake time variability for SUVmax and different ROI definitions for Ki (a) increased the variability in the measurements and (b) impacted the ability of the AUROC to measure change in response to therapy. The mean CV over all patients and scenarios is shown as the dashed horizontal line. The change in AUROC was calculated with respect to perfect uptake time (scenario 1) for SUVmax and ROI mean for Ki.

We assessed the difference of AUROC SUVmax and AUROC Ki as a function of tumor phenotype, including tumor grade, pre-therapy Ki-67, and oncotype DX score, and there was little correlation (data not shown).

7.5 DISCUSSION

We performed a virtual clinical trial to compare the performance of static and dynamic PET measures while evaluating the impact of known sources of uncertainty in the imaging process, including biological, arterial input function, scanner calibration, quantum noise, image noise correlations, uptake time for SUVmax, and ROI selection method for Ki. Variabilities in PET imaging can confound the classification of tumor status as response, stable disease, or progressive disease.

We performed an ROC analysis to determine whether SUV or Ki could better measure change in FDG uptake when accounting for all uncertainties that impact each metric. For the majority of patients in this study, Ki was superior to SUV as measured by the AUROC, which quantifies the separation and variability in the pre- and post-therapy measurement uncertainty distributions. Of note, this patient cohort included non-high-grade ER+ breast cancer, which tends to have modest FDG uptake [202,203]. Doot et al. [172] found Ki to be superior to SUV in lower uptake tumors, and our findings are consistent with the results of this prior study. With dynamic imaging, we could discern smaller changes in lesion uptake, which is related to the ability of the kinetic parameters to estimate the rate of phosphorylated FDG (FDG-6-P) at low uptake. In contrast, a large fraction of the SUV measurement includes free FDG in the vascular and tissue spaces, which is not related to tumor metabolism.

We also found the coefficient of variation of Ki to be higher than that for SUV. Kinetic parameters are estimated from a time-course of images, where the images immediately following radiotracer injection suffer from low count rates in the field of view and are thus very noisy. Therefore, estimation of the kinetic parameters is less robust and subject to increased variability compared to the measurement of SUV from a single image at a later time point. Our results agree with those of Shields et al. [169], who reported that the uncertainty in kinetic measurements was higher than that seen with SUV for patients who underwent test/re-test studies. Taking into account known uncertainties, the CV for SUVmax was 13.5% and is similar to reports in the literature. Kinahan et al. [196] consulted many previous studies and found the estimated variability of SUVmax to be 10% in a test-retest scenario but at least 15% in practice. The slightly lower CV in this VCT could be due a conservative estimate of scanner calibration uncertainty as well as exclusion of body habitus or dose measurement error.

Finally, we explored if there was a correlation in the performance of SUV versus Ki in measuring response to therapy by comparing the differences in AUROC to tumor grade and Ki-67. Previous studies have found a correlation between pre-therapy SUV or Ki and tumor pathology, such as tumor grade and Ki-67 [204,205,206], and results of other studies have suggested that dynamic analysis may provide superior information in assessing patient response [170,171,172]. However, these previous studies included a range of tumor phenotypes, and the tumors in this patient cohort (non-high grade and ER+) were very similar, which was likely the resason that we found Ki to almost unanimously outperform SUV. It is anticipated that there would be a distinction between the performance of SUV and Ki metrics for different tumor characteristics if there had been a wider range of tumor phenotpyes (and pre-therapy FDG uptake). Further, we did not have other outcomes like clinical response, pathologic response, or survival to further evaluate the predictive value of SUV or Ki in measuring tumor uptake change pre- and post-therapy. A future VCT study could explore the linking of PET parameter changes to the biology of the tumor and patient outcomes to evaluate which metric has better prognostic value.

The limited patient cohort size, range of tumor phenotypes, and outcome information were the main limitations of this study. Therefore, in this study we focused on the impact of uncertainites in the imaging process. By identifying the dominant sources of uncertainty in the measurements, steps can be taken to reduce the uncertainties and improve measurement capabilities, whether for measuring tumor response to therapy or comparing two metrics such as SUV and Ki. Finally, this versatile VCT methodology could be readily employed to assess the impact of factors not considered here. With slight modifications we could, for instance, determine the impact of image resolution or injected dose on the classification tasks discussed above.

7.6 CONCLUSIONS

A virtual clinical trial was conducted to compare the ability of static SUV versus dynamic PET imaging to measure response to breast cancer therapy. For this patient cohort, we found that Ki showed a superior ROC performance compared to SUVmax, suggesting that kinetic analysis, which accounts for TAC change in magnitude as well as shape, may better detect or enable earlier assessment of response to therapy. This approach could be extended to other PET tracers or diseases where there is access to prior dynamic imaging studies.

Chapter 8. DISSERTATION DISCUSSION

This chapter summarizes the impact of the VCT methodology developed and the results of the performed research studies. The overall outcomes of the studies are described, the original contributions are distinguished from those that were learned or applied, the limitations are summarized, and motivated future work is presented.

8.1 OUTCOMES

The outcome of this work was the development of a VCT methodology and the application of it to multiple clinically relevant PET imaging questions. In PET imaging, the definition of image quality is very broad but can generally be defined as how well the image serves in accomplishing the desired task. The impact of imaging parameters on such tasks as detection and quantitation can be difficult to assess due to uncertainties in the imaging process. Additionally, the dependence of each step in the imaging process on the outcome is very difficult to extract from patient scans alone, partially because ground truth is unknown. Through VCTs, we can refine our understanding of underlying questions about the PET imaging process that could otherwise challenging to answer.

A wide range of PET imaging simulation and image analysis tools were linked, forming an advanced, complete methodology. Each simulation was founded on retrospective patient data. Then each step in the imaging process was simulated, accounting for variabilities caused by differing individual patient characteristics, incorrect scanner calibrations, estimated data corrections, selection of image reconstruction algorithm parameters, and application of image analysis metrics. Many noise realizations were generated in the addition of these uncertainties, representing many patient scans. As a result, the impact of any given parameter and its uncertainty on the image-based task could be evaluated.

The methods were then applied to four studies that evaluated a new reconstruction algorithm, optimum patient imaging time, and measurement of response to therapy. The first two studies showed that lesion detectability in images using a new penalized likelihood reconstruction (PL) was maintained for low-contrast lesions in the liver and the lung. For larger lesion sizes in the lung, detectability was improved, particularly when using TOF image reconstruction. These results were positive given that the motivation for the study was that the penalty term could overly smooth small, low-contrast lesions into the background. In the third study, it was found that the optimum imaging time was dependent on the metabolic pathway of the radiotracer in both the tumor and normal tissues, and for some cases, imaging at a later time than the current clinical standard imaging time improved lesion detection. Finally, the fourth study showed that for the non-high-grade breast cancer patient cohort, the kinetic macro parameter, Ki, outperformed SUVmax in measuring response to therapy.

The studies in this work demonstrated that the VCT methodology is a versatile tool, and it can be expanded and adapted to investigate a wide range of unanswered questions beyond those considered in this work. The generation of many noise realizations enables full statistical system modeling and analyses of any input and output parameters. For example, proposed system changes or data acquisition protocols can be tested to characterize inherent uncertainties and evaluate clinical feasibility. The results can then be used to inform the design of prospective clinical trials, as improved understanding of key parameters is essential to a successful design. It is not suggested that this methodology is a replacement for clinical trials, but rather that it is a supplement to provide insight into questions clinical trials cannot easily answer.

8.2 CONTRIBUTIONS

The simulation and analysis tools in this PET imaging VCT methodology required a wide range of development. Although none of the tools were individually new, the combination of them to perform a comprehensive imaging simulation was unique. Expertise was required to link the tools together and then design and execute each of the studies. Further, understanding the relationship of how different factors in the imaging process affected others was critical in the interpretation of study results.

The studies performed were based on patient data, but these data required pre-processing before proceeding with the scanner and data simulations and image reconstruction. To characterize lesion detection using the PL algorithm, synthetic lesions were inserted into the sinograms of the original whole-body patient datasets. The lesion insertion code was adapted for use in this study. For example, the software needed to insert the lesions into specific locations of the liver and the lung.

For the breast cancer studies, kinetic parameters needed to be estimated from the original dynamic patient datasets. The time activity curves from these datasets were analyzed using a kinetic model written in both PMOD and MATLAB software; the model in PMOD allowed for more accurate initial parameter estimation from the data, and the model in MATLAB provided easier generation of many synthetic TAC realizations. An algorithm was also written to add uncertainty to the kinetic parameters and AIFs. The TACs were then input into an existing PET imaging data simulator, ASIM, to create sinograms. An input phantom was designed to meet the goals of each study; the locations of multiple spheres in the phantom were chosen to efficiently simulate many noise realizations, and the sizes were chosen to reflect an appropriate size when evaluating detectability or measurement of response to therapy. The scanner simulation parameters were specified to represent a specific PET scanner. Finally, the tumor and background sinograms were generated separately and then combined in data space, enabling the efficient generation of tumor data with any arbitrary contrast and signal-to-noise ratio.

The sinogram data for each study was reconstructed in MATLAB. Many pieces of the data processing and reconstruction algorithms were acquired, modified, and linked together. Poisson noise was added to the sinograms, scatter and random coincident events were estimated and added to the sinograms, and the reconstruction algorithm parameters were specified to match the data simulation parameters. The image reconstruction algorithms were implemented.

After reconstruction, the images were post-processed and analyzed. Algorithms were written to apply filtering, calculate image roughness, specify ROIs, and estimate SUVs. The kinetic model in MATLAB was also critical in efficiently re-estimating the kinetic parameters from the many TACs that included uncertainties in the imaging process. For the observer studies, the human observer study was designed, and the two-alternative forced choice graphical user interface (GUI) was developed. An existing basic model observer program was also expanded and integrated into the studies. Finally, software to perform the ROC analysis was written.

In summary, many tools were implemented and developed, including data generation algorithms, a two-tissue compartment kinetic model, a user interface to acquire human observer data, and image analysis software. The developed algorithms and software were applied to four original research studies and could be applied to the future studies.

8.3 LIMITATIONS

The VCT methodology enabled the inclusion of many realistic details of the imaging process, such as the use of retrospective patient data and Monte Carlo photon tracking simulations.

However, the studies in this work had limitations, as described in the individual study chapters and as discussed here.

The first two studies investigated lesion detectability using human and model observers. These studies were performed using a region of interest within the entire image, with the lesion location known to be in the center of each image patch. In a clinical setting, however, the number of lesions and the locations of them are unknown. The studies performed here could be considered perception studies, as a detection study would involve searching for and detecting the lesion in an unknown location. This study was conducted in a research setting using medical imaging scientists as observers, similar to other studies [207,208,209], but the results are assumed to translate to a clinical setting, where radiologists are the observers reading the images. Finally, though the agreement between the human and CHO model observers was good, the number of human observers was only four.

The second two studies were performed using dynamic data acquisitions from 24 breast cancer patients. The tumors from these patients were all of similar morphology (non-high-grade and ER+), which limited further investigation into the results of optimum imaging time and differences between SUV and Ki in measuring response to therapy. This patient dataset also did not contain any outcome information, and so the ability of either optimum uptake time or change in SUV or Ki between pre- and post-therapy scans to predict patient response could not be investigated. Additionally, the tumors of this patient cohort were already known and detectable. In order to perform the optimum imaging time for detection study, the simulated tumor characteristics needed to be modified to bring the tumor to the limit of detectability. Instead of altering the tumor to normal tissue uptake ratio, the lesion size of the tumors was reduced to 5 mm.

A two-tissue compartment kinetic model was assumed both when estimating kinetic parameters from measured TACs as well as when generating synthetic TACs from estimated kinetic parameters. In the optimum imaging time study, in order to study the impact of imaging at a later time, the synthetic time activity curves needed to be generated out to many hours post radiotracer injection. When estimating kinetic parameters, it is often assumed that k_4 is negligible. It is likely that this assumption holds for shorter time periods, but it may not be negligible for later times beyond two hours. In these studies, only 60 minutes of dynamic data was available to estimate the kinetic parameters for generating the synthetic TACs, making it difficult to obtain accurate estimates of k_4. Instead, a plausible range of k_4 values was assumed and four specific k_4 values tested.

Finally, to reduce the computation time required to generate many noise realizations for both breast cancer patient studies, the simulations were conducted using a digital phantom containing 36 spheres, where each sphere represented a unique tumor imaging scenario. To capture more realistic image background texture, ground-truth lesions could have been inserted into the patient data, as was done for the PL algorithm studies.

8.4 FUTURE WORK

This section describes how the developed tools and study results in this work motivate and provide the framework and preliminary data for future studies using similar virtual methods. The anticipation is that the use of the VCT methodology will support the development of PET imaging in clinical trials and in the clinic to improve detection of tumors, response to therapy, and ultimately improve patient care.

Lesion detection studies to evaluate the PL algorithm could be expanded to include lesion sizes smaller than 10 mm for both whole-body and high-resolution PET imaging systems.

Human observer studies could be repeated using the TOF data and for the smaller lesion sizes to confirm agreement with the CHO model observer over a wider range of simulation parameters. Finally, the model observer studies could be expanded to include more complicated tasks, such as localization tasks or scanning models, which may be even more predictive of human observers [210]. The expansion of the model observer analyses to include localization tasks should also be applied to the study of delayed imaging time to more accurately test lesion conspicuity as a function of time.

In the study measuring response to therapy, the impact of each source of uncertainty in the imaging process could be individually explored, as some sources likely dominate the resulting impact on SUV and Ki. This study could be repeated using the penalized likelihood algorithm, as the consistent quantitation of this algorithm may reduce uncertainties in assessing response to therapy.

In both the delayed imaging time and measurement of response to therapy studies, realistic background noise should be considered, either by inserting tumors into patient data or by adding noise textures, such as a lumpy background, to the reconstructed phantom.

The breast cancer patient cohort did not have patient survival or time to progression information. The two studies based on this data should be repeated using patient datasets with such outcome information to investigate if there is a correlation of optimum imaging time in predicting patient outcome or whether change in SUV or Ki is more predictive of patient prognosis. For example, a tumor with peak detectability at one-hour post injection might have a favorable prognosis over a tumor with peak detectability at two hours, even if they are both graded equivalently based on the pathology. It could also be determined whether change in SUV or Ki can be linked with other clinically relevant variables known to be important in breast cancer outcome.

The tumor characteristics in this patient cohort were very uniform, and the studies should be repeated for a broader range of tumors, cancer types, and radiotracers to determine if the results are repeatable over diverse patient populations and imaging scenarios. Specifically, the results of these studies will contribute to the submission of a grant proposal to develop optimal imaging protocols for the detection and assessment of response to therapy in breast and lung cancers using FDG and ^{18}F-fluorothymidine (FLT), which is a measure of DNA synthesis and cell proliferation.

Finally, it would be clinically advantageous if the optimum delayed imaging time could be determined *a priori*. It is unlikely that the tumor pathology could provide enough information, as the optimum imaging time depends not only on the magnitude of tracer uptake in the tumor, but also the signal-to-background ratio and the shape of the TACs; however there may be another predictive metric.

Chapter 9. CONCLUSIONS

A virtual clinical trial methodology was developed and applied to four PET imaging studies. The first two studies showed that a new penalized likelihood reconstruction algorithm, originally developed to improve lesion quantitation over the standard clinical algorithm, OSEM, maintained or improved lesion detectability for low-contrast lesions in the lung and liver. In the third study, it was found that delayed imaging beyond the standard imaging time of one-hour post injection improved lesion detectability in some non-high-grade tumors of breast cancer patients. In the final study, it was found that the kinetic macro parameter, Ki, outperformed SUV in measuring tumor response to therapy. The studies in this work demonstrated that the VCT methodology is a versatile tool, and it can be expanded and adapted to investigate a wide range of unanswered questions beyond those considered in this work to inform the design of clinical trial studies.

BIBLIOGRAPHY

[1] Wolf G and Abolmaali N. Preclinical molecular imaging using PET and MRI. Recent Results Cancer Res 187, 257-310 (2013).

[2] Schulz RB and Semmler W. Fundamentals of Optical Imaging. In: Semmler W and Schwaiger M, eds. Molecular Imaging I (Handbook of Experimental Pharmacology). Berlin: Springer-Verlag; 2008.

[3] Weber WA. Assessing tumor response to therapy. J Nucl Med. 2009;50(S1):1S-10S.

[4] Wahl, et al. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. J Nucl Med. 50(S1), 122S-150S (2009).

[5] Linden H and Dehdashti F. Novel Methods and Tracers for Breast Cancer Imaging. Sem Nucl Med 43, 324-329 (2013).

[6] Landau SM, Thomas BA, Thurfjell L, Schmidt M, Margolin R, Mintun M, Pontecorvo M, Baker SL, Jagust WJ, Initiative ADN. Amyloid PET imaging in Alzheimerâ□™s disease: a comparison of three radiotracers. European journal of nuclear medicine and molecular imaging. 2014;41(7):1398-1407.

[7] Kaira K, et al. Correlation Between 18F-FDG Uptake on PET and Molecular Biology in Metastatic Pulmonary Tumors. J Nucl Med. 52, 705-711 (2011).

[8] Budd GT, Cristofanilli M, and Ellis MJ. Circulating Tumor Cells versus Imaging: Predicting Overall Survival in Metastatic Breast Cancer. Clin Cancer Res 12, 6403-6409 (2006).

[9] Soret M, Bacharach SL, and Buvat I. Partial-volume effect in PET tumor imaging. JNM 48, 932-945 (2007).

[10] Kalinyak JE, et al. Breast cancer detection using high-resolution breast PET compared to whole-body PET or PET/CT. EJNMMI 41(2), 260-275 (2014).

[11] Kumar V, et al. Variance of SUVs for FDG-PET/CT is Greater in Clinical Practice Than Under Ideal Study Settings. Clin Nucl Med 38, 175-182 (2013).

[12] Kinahan PE and Fletcher JW. PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy. Semin Ultrasound CT MR 31(6), 496-505 (2010).

[13] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: a cancer journal for clinicians. 2016;66(1):7-30.

[14] Subramaniam R. Preface PET/CT: Adding Value to Patient Outcomes and Health Care Delivery. PET Clin 10 xi (2015).

[15] Milano A, et al. Targeted–Therapy and Imaging Response: A New Paradigm For Clinical Evaluation? Reviews on Recent Clinical Trials 6, 1-6 (2011).

[16] Zhu A, Lee D, and Shim H. Metabolic PET Imaging in Cancer Detection and Therapy Response. Semin Oncol. 38(1), 55-69 (2011).

[17] Levin CS and Hoffman EJ. Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution. Phys Med Biol 44, 781-799 (1999).

[18] Saha GB. (2010). *Basics of PET Imaging: Physics, Chemistry, and Regulations*. New York, NY: Springer Science+Business Media.

[19] van der Veldt, A. A. M., E. F. Smit, and A. A. Lammertsma. 2014. Positron emission tomography as a method for measuring drug delivery to tumors in vivo: the example of [11 C] docetaxel. Ways to improve tumor uptake and penetration of drugs into solid tumors 50.

[20] Knoll GF. Radiation detection and measurement. John Wiley & Sons; 2010.

[21] Ollinger JM. Model-based scatter correction for fully 3D PET. Physics in medicine and biology. 1996;41(1):153.

[22] D Wilson. 18F-fluorodeoxyglucose (18F-FDG) Positron Emission Tomography in Oncology. British Columbia Cancer Agency. R05-0076. https://clinicaltrials.gov/show/NCT00207298

[23] Strauss HW, Mariani G, Volterrani D, Larson SM. (Eds). (2013). Nuclear Oncology: Pathophysiology and Clinical Applications. New York, NY: Springer Science+Business Media.

[24] Weber, WA, et al. Quantitative Assessment of Tumor Metabolism Using FDG-PET Imaging. Nucl Med Biol 27, 683-687 (2000).

[25] Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. science. 2009;324(5930):1029-1033.

[26] Kaarstad K, et al. Metabolic Fate of 18F-FDG in Mice Bearing Either SCCVII Squamous Cell Carcinoma or C3H Mammary Carcinoma. JNM 43, 940-947 (2002).

[27] Bender D, et al. Metabolites of 18F-FDG and 3-O-11C-Methylglucose in Pig Liver. JNM 42, 1673-1678 (2001).

[28] National electrical manufacturers association. (2007). NEMA Standards Publication NY 2-2007: Performance measurements of positron emission tomographs. Rosslyn, VA: NEMA.

[29] Segars P and Tsui BMW. MCAT to XCAT: The Evolution of 4-D Computerized Phantoms for Imaging Research. IEEE Proceedings 97(12), 1954-1968 (2009).

[30] Muzi M, et al. Quantitative assessment of dynamic PET imaging data in cancer imaging. Mag Res Imaging 30, 1203-1215 (2012).

[31] Harrison RL and Lewellen TK. (2013). The SimSET Program. In M Ljungberg, S-E Strand, MA King (Eds.), Monte Carlo Calculations in Nuclear Medicine (2nd ed.) (87-108). Boca Raton, FL: CRC Press.

[32] Comtat, C, Kinahan PE, Defrise M, Michel C, Lartizien C, Townsend DW. Simulating whole-body PET scanning with rapid analytical methods. IEEE Nucl Sci Symp Conf Rec 1999;1260-1264.

[33] Henkin RE, et al. (2006). Nuclear Medicine: 2-Volume Set (2nd ed.). St Louis, MO: Mosby.

[34] Barrett HH, Myers KJ. Foundations of image science. John Wiley & Sons; 2013.

[35] Vandenberghe S, et al. Iterative reconstruction algorithms in nuclear medicine. Comp Med Imaging and Graphics 25, 105-111 (2001).

[36] Fessler JA. Penalized weighted least-squares image reconstruction for positron emission tomography. Medical Imaging, IEEE Transactions on. 1994;13(2):290-300.

[37] Asma E, Ahn S, Qian H, Gopalakrishnan G, Thielemans K, Ross SG, Manjeshwar RM. Ganin A. Quantitatively accurate image reconstruction for clinical whole-body PET imaging. Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific; 2012; IEEE; 2012.

[38] Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. Medical Imaging, IEEE Transactions on. 1982;1(2):113-122.

[39] Snyder DL, Miller MI, Thomas Jr LJ, Politte DG. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. IEEE Trans Med Imag. 1987;6(3):228-238.

[40] Cho JH, Fessler JA. Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-D x-ray CT. Medical Imaging, IEEE Trans Med Imag. 2015;34(2):678-689.

[41] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society Series B (methodological). 19771-38.

[42] Hudson HM and Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. Trans Med Imaging 13(4), 601-609 (1994).

[43] Angelis GI, Thielemans K, Tziortzi AC, Turkheimer FE, Tsoumpas C. Convergence optimization of parametric MLEM reconstruction for estimation of Patlak plot parameters. Computerized Medical Imaging and Graphics. 2011;35(5):407-416.

[44] Boellard R, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 37, 181-200 (2010).

[45] Harrison RL, et al. A Virtual Clinical Trial of FDG-PET Imaging of Breast Cancer: Effect of Variability on Response Assessment. Translational Oncology 7(1), 138-146 (2014).

[46] Vanderhoek M, Perlman SB, and Jeraj R. Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response. JNM 53, 4-11 (2012).

[47] Macmillan NA Creelman CD. Detection theory: A user's guide. New York: Cambridge University Press; 1991.

[48] Burgess A. E, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," Med. Phys. 22(5), 643-655 (1995).

[49] He X and Park S. Model observers in medical imaging research. Theranostics 3(10), 774-786 (2013).

[50] Barrett H, et al. Model observers for assessment of image quality. Proc Natl Acad Sci 90, 9758-9765 (1993).

[51] Gallas BD and Barrett HH. Validating the use of channels to estimate the ideal linear observer. J Opt. Soc. Am. A 20(9), 1725-1738 (2003).

[52] Platisa L, et al. Channelized Hotelling observers for the assessment of volumetric imaging data sets. J Opt. Soc. Am. A 28(6), 1145-1163 (2011).

[53] Zhang L, Cavaro-Menard C, and Le Callet, P. An overview of model observers. IRBM 35 214-224, (2014).

[54] Gifford H. C., King M. A., de Vries D. J. and Soares E. J., "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," J. Nucl. Med. 41(3), 514-521 (2000).

[55] Abbey CK and Bochud FO. (2000). Modeling Visual Detection Tasks in Correlated Image Noise. In J Beutel, HL Kundel, and RL Van Metter (Eds.), *Handbook of Medical Imaging: Vol 1. Physics and Psychophysics* (629-654). Bellingham, WA: SPIE.

[56] ICRU Report 54. (1996). Medical Imaging-The Assessment of Image Quality. Bethesda, MD: ICRU.

[57] Wunderlich A, Noo F. Confidence intervals for performance assessment of linear observers. Med Phys. 38(7), S57-68 (2011).

[58] Carson RE. (2003). Tracer Kinetic Modeling in PET. In PE Valk, DL Bailey, DW Townsend, MN Maisey (Eds.), *Positron Emission Tomography: Basic Science and Clinical Practice*. London, UK: Springer-Verlag (147–179).

[59] Morris ED, et al. (2004). Kinetic modeling in positron emission tomography. In MN Wernick and JN Aarsvold (Eds.), *Emission tomography: The fundamentals of PET and SPECT*, London, UK: Elsivier (499-540).

[60] Hudson H. M. and Larkin R. S., "Accelerated image reconstruction using ordered subsets of projection data," Trans. Med. Imag. 13(4), 601-609 (1994).

[61] Qi J., and Leahy R. M., "Iterative reconstruction techniques in emission computed tomography," Phys. Med. Biol. 51, 541-578 (2006).

[62] Nuyts J., Beque D., Dupont P. and Mortelmans L., "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," Trans. Nuc. Sci. 49(1), 56-60 (2002).

[63] Asma E., Ahn S., Ross S., Chen A. and Manjeshwar R. M., "Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images," IEEE NSS/MIC Conference Proceedings, 4062-4066 (2012).

[64] Ma H., Asma E., Ahn S., Ross S., Manjeshwar R. M., Wilson D., Tonseth R. P., Tran A., Celler A. and Bénard F., "Clinical evaluation of penalized likelihood reconstruction in whole-body PET studies," Annual Congress of the European Association of Nuclear Medicine (2013).

[65] Sah B.-R., Veit-Haibach P., Delso G., Wollenweber S., Tarrade X., Licato P., von Schulthess G. and Pietsch C., "Clinical evaluation of a new block sequential regularized expectation maximization (BSREM) reconstruction algorithm in PET/CT studies," J. Nucl. Med. 55(S1), 2097 (2014).

[66] Passalaqua S., Kappadath S., Branch D., Ross S., Stearns C., Schomer D. and Mawlawi O., "Qualitative and quantitative evaluation of regularized PET image reconstruction," J. Nucl. Med. 55(S1), 579 (2014).

[67] Barrett H. H., Yao J., Rolland J. P. and Myers K. J, "Model observers for assessment of image quality," Proc. Natl. Acad. Sci. 90, 9758-9765 (1993).

[68] De Pierro A. R. and Yamagishi M. E. B., "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," Trans. Med. Imag. 20(4), 280–288 (2001).

[69] Ahn S. and Fessler J. A., "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," Trans. Med. Imag. 22(5), 613-626 (2003).

[70] Abbey C. K., Barrett H. H., and Eckstein M. P., “Practical issues and methodology in assessment of image quality using model observers,” SPIE Proceedings 3032, 182-194 (1997).

[71] Tong S., Alessio A. M. and Kinahan P. E., “Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation,” Phys. Med. Biol. 55(5), 1453-1473 (2010).

[72] Surti S. and Karp J. S., “Experimental evaluation of a simple lesion detection task with time-of-flight PET,” Phys. Med. Biol. 54(2), 373-384 (2009).

[73] Kim J. S, Kinahan P. E., Lartizien C., Comtat C. and Lewellen T. K., ”A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging,” Trans. Nucl. Sci. 51(1), 34-40 (2004).

[74] Panin V. Y., Chen M. and Casey M. E., “Numerical observer evaluation of 3D statistical PET reconstruction method with system response modeling,” IEEE NSS/MIC Conference Proceedings, 2143-2147 (2005).

[75] Qi J., “Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability,” Trans. Med. Imag. 23(3), 321-329 (2009).

[76] Nuyts J., Michel C., Brepoels L., De Ceuninck L., Deroose C., Goffin K., Mottaghy F. M., Stroobants S., Van Riet J. and Verscuren R., “Performance of MAP reconstruction for hot lesion detection in whole-body PET/CT: An evaluation with human and numerical observers,” Trans. Med. Imag. 28(1), 67-73 (2009).

[77] Petschke A. and La Rivière P. J., “Comparison of photoacoustic image reconstruction algorithms using the channelized Hotelling observer,” J. Biomed. Optics 18(2), 026009 (2013).

[78] El Fahkri G., Surti S., Trott C. M., Scheuermann J. and Karp J. S., “Improvement in lesion detection with whole-body oncologic time-of-flight PET,” J. Nucl. Med. 52(3), 347–353 (2011).

[79] Kadrmas D. J., Casey M. E., Conti M., Jakoby B. W., Lois C. and Townsend D. W., “Impact of time-of-flight PET on tumor detection,” J. Nucl. Med. 50(8), 1315–1323 (2009).

[80] Li K., Garrett J. and Chen G. H., ”Correlation between human observer performance and model observer performance in differential phase contrast CT,” Med. Phys. 40, 111905 (2013).

[81] H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” IEEE Trans. Med. Imag. 13(4), 601-609 (1994).

---

[82] J. Qi, R. M. Leahy, “Iterative reconstruction techniques in emission computed tomography,” Phys. Med. Biol. 51, 541-578 (2006).

[83] J. A. Fessler, “Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography,” IEEE Trans. Imag. Proc. 5(3), 493-506 (1996).

[84] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs,” IEEE Trans. Imag. Proc. 5(9), 1346-1358 (1996).

[85] J. Qi and R. M. Leahy, “Resolution and noise properties of map reconstruction for fully 3-D PET,” IEEE Trans. Med. Imag. 19(5), 493-506 (2000).

[86] S. Ahn and R. M. Leahy, “Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET,” IEEE Trans. Med. Imag. 27(3), 413-424 (2008).

[87] J. Dutta, S. Ahn, Q. Li, “Quantitative statistical methods for image quality assessment,” Theranostics. 3(10), 741-756 (2013).

[88] J. Qi, “Analysis of lesion detectability in bayesian emission reconstruction with nonstationary object variability,” IEEE Trans. Med. Imaging 23(3), 321-329 (2004).

[89] J. Qi, R. H. Huesman, “Penalized maximum-likelihood image reconstruction for lesion detection,” Phys. Med. Biol. 51(16), 4017 (2006).

[90] J. Qi, R. H. Huesman, “Theoretical study of penalized-likelihood image reconstruction for region of interest quantification,” IEEE Trans. Med. Imaging. 25(5), 640-648 (2006).

[91] S. Geman and D. E. McClure, “Bayesian image analysis: an application to single photon emission tomography,” Proc. Am. Stat. Assoc., 12-18 (1985).

[92] W. Chlewicki, F. Hermansen, and S. B. Hansen, “Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior,” Phys. Med. Biol. 49, 4717-4730 (2004).

[93] E. Asma and R. M. Manjeshwar, “Analysis of organ uniformity in low count density penalized likelihood PET images,” IEEE Med. Imag. Conf. Rec. 6, 4426-4432 (2007).

[94] S. Ahn, et al., “Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET,” Phys. Med. Biol. 60, 5733-5751 (2015).

[95] J. Nuyts, et al., “A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography,” IEEE Trans. Nuc. Sci. 49(1), 56-60 (2002).

---

[96] A. R. De Pierro and M. E. B. Yamagishi, “Fast EM-like methods for maximum “a posteriori” estimates in emission tomography,” IEEE Trans. Med. Imaging 20, 280-288 (2001).

[97] S. Ahn and J. A. Fessler, “Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms,” IEEE Trans. Med. Imaging 22, 613-626 (2003).

[98] E. J. Teoh, et al., “Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules,” Eur. Radiol. (2015).

[99] E. J. Teoh, et al., “Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system,” J. Nucl. Med. 56(9), 1447-1452 (2015).

[100] H. Ma, et al., “Clinical evaluation of penalized likelihood reconstruction in whole-body PET studies,” Annual Congress of the European Association of Nuclear Medicine (2013).

[101] B.-R. Sah, et al., “Clinical evaluation of a new block sequential regularized expectation maximization (BSREM) reconstruction algorithm in PET/CT studies,” J. Nucl. Med. 55(S1), 2097 (2014).

[102] S. Passalaqua, et al., “Qualitative and quantitative evaluation of regularized PET image reconstruction,” J. Nucl. Med. 55(S1), 579 (2014).

[103] J. Lantos, et al., “Standard OSEM vs. regularized PET image reconstruction: qualitative and semi-quantitative comparison,” J. Nucl. Med. 56(S3), 1805 (2015).

[104] N. Parvizia, et al., “Does a novel penalized likelihood reconstruction of 18F-FDG PET-CTimprove signal-to-background in colorectal liver metastases?,” Eur. J. Radiol. 84(10), 1873-1878 (2015).

[105] S. Y. Kang, et al., “Increased detectability of small lesions by BSREM reconstruction of PET in breast cancer,” J. Nucl. Med. 56(S3), 1832 (2015).

[106] K. A. Wangerin, et al., “Improving lesion detectability in PET imaging with a penalized likelihood reconstruction algorithm,” Proc. SPIE Medical Imaging 9416, 94160W (2015).

[107] C. K. Abbey, M. P. Eckstein, F. O. Bochud. Estimation of human-observer templates in two-alternative forced-choice experiments. Proc. SPIE Medical Imaging 3663. 284-295 (1999). doi: 10.1117/12.349653.

[108] A. E. Burgess. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. Med. Phys. 22(5):643-655 (1995).

[109] J. Yao, H. H. Barrett. Predicting human performance by a channelized Hotelling observer model. Proc. SPIE Mathematical Methods in Medical Imaging 1768. 161-168 (1992). doi:10.1117/12.130899.

[110] V. Bettinardi, et al. Physical performance of the new hybrid PET/CT Discovery-690. Medical physics. 2011;38(10):5394-5411.

[111] A. M. Alessio, et al., “Application and evaluation of a measured spatially variant system model for pet image reconstruction,” IEEE Trans. Med. Imaging, 29, 938-949 (2010).

[112] H. C. Gifford, et al., “Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging,” J. Nucl. Med. 41(3), 514-521 (2000).

[113] C. K. Abbey, H. H. Barrett, and M. P. Eckstein, “Practical issues and methodology in assessment of image quality using model observers,” Proc. SPIE Medical Imaging 3032, 182-194 (1997).

[114] S. Tong, A. M. Alessio, and P. E. Kinahan, “Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation,” Phys. Med. Biol. 55(5), 1453-1473 (2010).

[115] H. H. Barrett, et al., “Model observers for assessment of image quality,” Proc. Natl. Acad. Sci. 90, 9758-9765 (1993).

[116] K. Li, J. Garrett, and H. G. Chen, ”Correlation between human observer performance and model observer performance in differential phase contrast CT,” Med. Phys. 40, 111905 (2013).

[117] J. G. Brankov, “Evaluation of the channelized Hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection,” Phys. Med. Biol. 58(20), 7159 (2013).

[118] D. J. Kadrmas, et al., “Impact of time-of-flight on PET tumor detection,” J. Nucl. Med. 50(8), 1315-1323 (2009).

[119] J. Schaefferkoetter, et al., “Clinical impact of time-of-flight and point response modeling in pet reconstructions: A lesion detection study,” Phys. Med. Biol. 58(5), 1465 (2013).

[120] Y. Zhang, B. T. Pham, M. P. Eckstein, “The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds.,” IEEE Trans. Med. Imaging. 25(10), 1348-1362 (2006).

[121] A. Wunderlich, et al., “Exact confidence intervals for channelized Hotelling observer performance in image quality studies.,” IEEE Trans. Med. Imaging. 34(2), 453-464 (2015).

---

[122] E. Asma, et al., “Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images,” IEEE Med. Imag. Conf. Rec., 4062-4066 (2012).

[123] J. Qi, “Investigation of lesion detection in map reconstruction with non-gaussian priors,” IEEE Med. Imag. Conf. Rec. 3, 1704-1708 (2005).

[124] S. Surti and J. S. Karp, “Experimental evaluation of a simple lesion detection task with time-of-flight PET,” Phys. Med. Biol. 54(2), 373-384 (2009).

[125] V. Y. Panin, M. Chen, and M. E. Casey, “Numerical observer evaluation of 3D statistical PET reconstruction method with system response modeling,” IEEE Med. Imag. Conf. Rec., 2143-2147 (2005).

[126] J. Qi, “Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability,” IEEE Trans. Med. Imag. 23(3), 321-329 (2009).

[127] A. Petschke and P. J. La Rivière, “Comparison of photoacoustic image reconstruction algorithms using the channelized Hotelling observer,” J. Biomed. Optics 18(2), 026009 (2013).

[128] Nuyts J., Michel C., Brepoels L., De Ceuninck L., Deroose C., Goffin K., Mottaghy F. M., Stroobants S., Van Riet J. and Verscuren R., “Performance of MAP reconstruction for hot lesion detection in whole-body PET/CT: an evaluation with human and numerical observers,” IEEE Trans. Med. Imaging 28(1), 67-73 (2009).

[129] G. El Fahkri, et al., “Improvement in lesion detection with whole-body oncologic time-of-flight PET,” J. Nucl. Med. 52(3), 347–353 (2011).

[130] J. Qi, “Comparison of lesion detection and quantification in MAP reconstruction with Gaussian and non-Gaussian priors,” Int. J. Biomed. Imaging 87567 (2006).

[131] Weber WA. Positron emission tomography as an imaging biomarker. J Clin Oncol. 2006;24(20):3282-3292.

[132] Boellaard R, O’Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, Oyen WJ, Kotzerke J, Hoekstra OS, Pruim J, Marsden PK, Tatsch K, Hoekstra CJ, Visser EP, Arends B, Verzijlbergen FJ, Zijlstra JM, Comans EF, Lammertsma AA, Paans AM, Willemsen AT, Beyer T, Bockisch A, Schaefer-Prokop C, Delbeke D, Baum RP, Chiti A, Krause BJ. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging. 2010;37(1):181-200.

[133] Schmidt K, Mies G, Sokoloff L. Model of kinetic behavior of deoxyglucose in heterogeneous tissues in brain: a reinterpretation of the significance of parameters fitted to homogeneous tissue models. J Cereb Blood Flow Metab. 1991;11(1):10-24.

---

[134] Boerner AR, Weckesser M, Herzog H, Schmitz T, Audretsch W, Nitz U, Bender HG, Mueller-Gaertner HW. Optimal scan time for fluorine-18 fluorodeoxyglucose positron emission tomography in breast cancer. Eur J Nucl Med. 1999;26(3):226-230.

[135] Lodge MA, Lucas JD, Marsden PK, Cronin BF, O'Doherty MJ, Smith MA. A PET study of 18FDG uptake in soft tissue masses. Eur J Nucl Med. 1999;26(1):22-30.

[136] Spence AM, Muzi M, Mankoff DA, O'Sullivan SF, Link JM, Lewellen TK, Lewellen B, Pham P, Minoshima S, Swanson K, Krohn KA. 18F-FDG PET of gliomas at delayed intervals: improved distinction between tumor and normal gray matter. J Nucl Med. 2004;45(10):1653-1659.

[137] Hamberg LM, Hunter GJ, Alpert NM, Choi NC, Babich JW, Fischman AJ. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification. J Nucl Med. 1994;35(8):1308-1312.

[138] Basu S, Kung J, Houseni M, Zhuang H, Tidmarsh GF, Alavi A. Temporal profile of fluorodeoxyglucose uptake in malignant lesions and normal organs over extended time periods in patients with lung carcinoma: implications for its utilization in assessing malignant lesions. Q J Nucl Med Mol Imaging. 2009;53(1):9-19.

[139] Matthiessen LW, Johannesen HH, Skougaard K, Gehl J, Hendel HW. Dual time point imaging fluorine-18 flourodeoxyglucose positron emission tomography for evaluation of large loco-regional recurrences of breast cancer treated with electrochemotherapy. Radiol Oncol. 2013;47(4):358-365.

[140] Thie JA, Hubner KF, Smith GT. Optimizing imaging time for improved performance in oncology PET studies. Mol Imaging Biol. 2002;4(3):238-244.

[141] Kumar R, Loving VA, Chauhan A, Zhuang H, Mitchell S, Alavi A. Potential of dual-time-point imaging to improve breast cancer diagnosis with (18)F-FDG PET. J Nucl Med. 2005;46(11):1819-1824.

[142] Zytoon AA, Murakami K, El-Kholy MR, El-Shorbagy E. Dual time point FDG-PET/CT imaging. Potential tool for diagnosis of breast cancer. Clin Radiol. 2008;63(11):1213-1227.

[143] Sokoloff L, Reivich M, Kennedy C, Des Rosiers MH, Patlak CS, Pettigrew KD, Sakurada O, Shinohara M. The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. J Neurochem. 1977;28(5):897-916.

[144] Burger C, Buck A. Requirements and implementation of a flexible kinetic modeling tool. J Nucl Med. 1997;38(11):1818-1823.

[145] Huang J, O’Sullivan F. An analysis of whole body tracer kinetics in dynamic PET studies with application to image-based blood input function extraction. IEEE Trans Med Imaging. 2014;33(5):1093-1108.

[146] O’Sullivan F, Kirrane J, Muzi M, O’Sullivan JN, Spence AM, Mankoff DA, Krohn KA. Kinetic quantitation of cerebral PET-FDG studies without concurrent blood sampling: statistical recovery of the arterial input function. IEEE Trans Med Imaging. 2010;29(3):610-624.

[147] Beaney RP, Lammertsma AA, Jones T, McKenzie CG, Halnan KE. Positron emission tomography for in-vivo measurement of regional blood flow, oxygen utilisation, and blood volume in patients with breast carcinoma. Lancet. 1984;1(8369):131-134.

[148] Tromberg BJ, Coquoz O, Fishkin JB, Pham T, Anderson ER, Butler J, Cahn M, Gross JD, Venugopalan V, Pham D. Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration. Philos Trans R Soc Lond B Biol Sci. 1997;352(1354):661-668.

[149] Olshen A, O’Sullivan F. Camouflaged deconvolution with application to blood curve modeling in FDG PET studies. J Am Stat As. 1997;92(440):1293-1303.

[150] Weber G. Enzymology of cancer cells (second of two parts). N Engl J Med. 1977;296(10):541-551.

[151] Krak NC, van der Hoeven JJ, Hoekstra OS, Twisk JW, van der Wall E, Lammertsma AA. Measuring [(18)F]FDG uptake in breast cancer during chemotherapy: comparison of analytical methods. Eur J Nucl Med Mol Imaging. 2003;30(5):674-681.

[152] Zasadny KR, Tatsumi M, Wahl RL. FDG metabolism and uptake versus blood flow in women with untreated primary breast cancers. Eur J Nucl Med Mol Imaging. 2003;30(2):274-280.

[153] Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. J Clin Oncol. 1993;11(11):2101-2111.

[154] Lucignani G, Schmidt KC, Moresco RM, Striano G, Colombo F, Sokoloff L, Fazio F. Measurement of regional cerebral glucose utilization with fluorine-18-FDG and PET in heterogeneous tissues: theoretical considerations and practical procedure. J Nucl Med. 1993;34(3):360-369.

[155] Tseng J, Dunnwald LK, Schubert EK, Link JM, Minoshima S, Muzi M, Mankoff DA. 18F-FDG kinetics in locally advanced breast cancer: correlation with tumor blood flow and changes in response to neoadjuvant chemotherapy. J Nucl Med. 2004;45(11):1829-1837.

---

[156] Comtat C, Kinahan PE, Defrise M, Michel C, Lartizien C, Townsend DW. Simulating whole-body PET scanning with rapid analytical methods. NSS Conference Record 1999; 1260-1264.

[157] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci U S A. 1993;90(21):9758-9765.

[158] Gifford HC, King MA, de Vries DJ, Soares EJ. Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. J Nucl Med. 2000;41(3):514-521.

[159] Wangerin K, Ahn S, Ross SG, Kinahan PE, Manjeshwar RM. Improving lesion detectability in PET imaging with a penalized likelihood reconstruction algorithm. Proc SPIE. 2015;94160W:doi:10.1117/12.2082301.

[160] Weber G, Cantero A. Glucose-6-phosphatase activity in normal, precancerous, and neoplastic tissues. Cancer Research. 1955;15(2):105-108.

[161] Caracó C, Aloj L, Chen LY, Chou JY, Eckelman WC. Cellular release of [18F]2-fluoro-2-deoxyglucose as a function of the glucose-6-phosphatase enzyme system. J Biol Chem. 2000;275(24):18489-18494.

[162] Vriens D, Disselhorst JA, Oyen WJ, de Geus-Oei LF, Visser EP. Quantitative assessment of heterogeneity in tumor metabolism using FDG-PET. Int J Radiat Oncol Biol Phys. 2012;82(5):e725-31.

[163] Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. J Nucl Med. 2004;45(9):1431-1434.

[164] Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50(S1):11S-20S.

[165] Freedman NM, Sundaram SK, Kurdziel K, Carrasquillo JA, Whatley M, Carson JM, Sellers D, Libutti SK, Yang JC, Bacharach SL. Comparison of SUV and Patlak slope for monitoring of cancer therapy using serial PET scans. Eur J Nucl Med Mol Imaging. 2003;30(1):46-53.

[166] Muzi, M., F. O'Sullivan, D. A. Mankoff, R. K. Doot, L. A. Pierce, B. F. Kurland, H. M. Linden, and P. E. Kinahan. Quantitative assessment of dynamic PET imaging data in cancer imaging. Magn Reson Imaging. 2012;30:1203-1215.

[167] Cheebsumon P, Velasquez LM, Hoekstra CJ, Hayes W, Kloet RW, Hoetjes NJ, Smit EF, Hoekstra OS, Lammertsma AA, Boellaard R. Measuring response to therapy using FDG PET: semi-quantitative and full kinetic analysis. Eur J Nucl Med Mol Imaging. 2011;38(5):832-842.

[168] Mankoff D, Muzi M, Zaidi H. Quantitative analysis in nuclear oncologic imaging. In: Zaidi H, editor. Quantitative analysis in nuclear medicine imaging. Singapore: Springer Science+Business Media, Inc.; 2006. p. 446-447.

[169] Shields AF, Lawhorn-Crews JM, Briston DA, Zalzala S, Gadgeel S, Douglas KA, Mangner TJ, Heilbrun LK, Muzik O. Analysis and Reproducibility of 3 -Deoxy-3 -[18F] Fluorothymidine Positron Emission Tomography Imaging in Patients with Non–Small Cell Lung Cancer. Clin Cancer Res. 2008;14:4463-4468.

[170] Dimitrakopoulou-Strauss A, Strauss LG, Burger C, Rühl A, Irngartinger G, Stremmel W, Rudi J. Prognostic aspects of 18F-FDG PET kinetics in patients with metastatic colorectal carcinoma receiving FOLFOX chemotherapy. J Nucl Med. 2004;45(9):1480-1487.

[171] Römer W, Hanauske AR, Ziegler S, Thödtmann R, Weber W, Fuchs C, Enne W, Herz M, Nerl C, Garbrecht M, Schwaiger M. Positron emission tomography in non-Hodgkin’s lymphoma: assessment of chemotherapy with fluorodeoxyglucose. Blood. 1998;91(12):4464-4471.

[172] Doot RK, Dunnwald LK, Schubert EK, Muzi M, Peterson LM, Kinahan PE, Kurland BF, Mankoff DA. Dynamic and static approaches to quantifying 18F-FDG uptake for measuring cancer response to therapy, including the effect of granulocyte CSF. J Nucl Med. 2007;48(6):920-925.

[173] Harrison RL, Elston BF, Doot RK, Lewellen TK, Mankoff DA, Kinahan PE. A Virtual Clinical Trial of FDG-PET Imaging of Breast Cancer: Effect of Variability on Response Assessment. Transl oncol. 2014;7(1):138-146.

[174] Zin CS, Nissen LM, O’Callaghan JP, Duffull SB, Smith MT, Moore BJ. A randomized, controlled trial of oxycodone versus placebo in patients with postherpetic neuralgia and painful diabetic neuropathy treated with pregabalin. J Pain. 2010;11(5):462-471.

[175] Graesslin O, Abdulkarim BS, Coutant C, Huguet F, Gabos Z, Hsu L, Marpeau O, Uzan S, Pusztai L, Strom EA. Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer. J Clin Oncol. 2010;28(12):2032-2037.

[176] Jones B, Dale RG. The potential for mathematical modelling in the assessment of the radiation dose equivalent of cytotoxic chemotherapy given concomitantly with radiotherapy. Br J Radiol. 2005;78:939-944.

[177] Agur Z. From the evolution of toxin resistance to virtual clinical trials: the role of mathematical models in oncology. Future Oncol. 2010;6:917-927.

[178] Wangerin KA, Muzi M, Peterson LM, Linden HM, Novakova A, O’Sullivan F, Kurland BF, Mankoff DA, Kinahan PE. Effect of 18F-FDG uptake time on lesion detectability in PET imaging of early stage breast cancer. Tomography. 2015;1:53.

[179] Sokoloff L, Reivich M, Kennedy C, Des Rosiers MH, Patlak CS, Pettigrew KD, Sakurada O, Shinohara M. The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. J Neurochem. 1977;28(5):897-916.

[180] Burger C, Buck A. Requirements and implementation of a flexible kinetic modeling tool. J Nucl Med. 1997;38(11):1818-1823.

[181] Huang J, O'Sullivan F. An analysis of whole body tracer kinetics in dynamic PET studies with application to image-based blood input function extraction. IEEE Trans Med Imaging. 2014;33(5):1093-1108.

[182] O'Sullivan F, Kirrane J, Muzi M, O'Sullivan JN, Spence AM, Mankoff DA, Krohn KA. Kinetic quantitation of cerebral PET-FDG studies without concurrent blood sampling: statistical recovery of the arterial input function. IEEE Trans Med Imaging. 2010;29(3):610-624.

[183] Beaney RP, Lammertsma AA, Jones T, McKenzie CG, Halnan KE. Positron emission tomography for in-vivo measurement of regional blood flow, oxygen utilisation, and blood volume in patients with breast carcinoma. Lancet. 1984;1(8369):131-134.

[184] Tromberg BJ, Coquoz O, Fishkin JB, Pham T, Anderson ER, Butler J, Cahn M, Gross JD, Venugopalan V, Pham D. Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration. Philos Trans R Soc Lond B Biol Sci. 1997;352(1354):661-668.

[185] Weber G. Enzymology of cancer cells (second of two parts). N Engl J Med. 1977;296(10):541-551.

[186] Okazumi S, Isono K, Enomoto K, Kikuchi T, Ozaki M, Yamamoto H, Hayashi H, Asano T, Ryu M. Evaluation of liver tumors using fluorine-18-fluorodeoxyglucose PET: characterization of tumor and assessment of effect of treatment. J Nucl Med. 1992;33(3):333-339.

[187] Hamberg LM, Hunter GJ, Alpert NM, Choi NC, Babich JW, Fischman AJ. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification. J Nucl Med. 1994;35(8):1308-1312.

[188] Zasadny KR, Tatsumi M, Wahl RL. FDG metabolism and uptake versus blood flow in women with untreated primary breast cancers. Eur J Nucl Med Mol Imaging. 2003;30(2):274-280.

[189] Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. J Clin Oncol. 1993;11(11):2101-2111.

[190] Lucignani G, Schmidt KC, Moresco RM, Striano G, Colombo F, Sokoloff L, Fazio F. Measurement of regional cerebral glucose utilization with fluorine-18-FDG and PET in heterogeneous tissues: theoretical considerations and practical procedure. J Nucl Med. 1993;34(3):360-369.

[191] Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J Cereb Blood Flow Metab. 1983;3(1):1-7.

[192] Patlak CS, Blasberg RG. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations. J Cereb Blood Flow Metab. 1985; 5(4):584-590.

[193] Olshen AB, O'sullivan F. Camouflaged deconvolution with application to blood curve modeling in FDG PET studies. Journal of the American Statistical Association. 1997;92(440):1293-1303.

[194] Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med. 1999;40(11):1771-1777.

[195] Feng D, Huang S-C, Wang X. Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography. Int J Biomed Comput. 1993;32(2):95-110.

[196] Kinahan PE, Fletcher JW. PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. Semin Ultrasound CT MR 2010; 31(6):496-505.

[197] Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 2009; 50(S1):122S-150S.

[198] Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, Oyen WJ, Kotzerke J, Hoekstra OS, Pruim J, Marsden PK, Tatsch K, Hoekstra CJ, Visser EP, Arends B, Verzijlbergen FJ, Zijlstra JM, Comans EF, Lammertsma AA, Paans AM, Willemsen AT, Beyer T, Bockisch A, Schaefer-Prokop C, Delbeke D, Baum RP, Chiti A, Krause BJ. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging. 2010;37(1):181-200.

[199] Kurland BF, Muzi M, Peterson LM, Doot RK, Wangerin KA, Mankoff DA, Linden HM, Kinahan PE. Multicenter clinical trials using 18F-FDG PET to measure early response to oncologic therapy: effects of injection-to-acquisition time Variability on required sample size. J Nucl Med. 2016;57(2):226-230.

[200] Graham MM. Physiologic smoothing of blood time-activity curves for PET data analysis. J Nucl Med. 1997;38(7):1161.

[201] Lockhart CM, MacDonald LR, Alessio AM, McDougald WA, Doot RK, Kinahan PE. Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. J Nucl Med. 2011;52(2):218-224.

[202] Fusselman DDD, Cutler PD, Katzenellenbogen JA, Welch MJ. Positron Tomographic Assessment of Estrogen Receptors in Breast Cancer: Comparison with with FDG-PET and In Vitro Receptor Assays. J Nucl Med 1995; 36:1766-1774.

[203] Osborne JR, Port E, Gonen M, Doane A, Yeung H, Gerald W, Cook JB, Larson S. 18F-FDG PET of locally invasive breast cancer and association of estrogen receptor status with standardized uptake value: microarray and immunohistochemical analysis. J Nucl Med. 2010;51(4):543-550.

[204] Buck A, Schirrmeister H, Kühn T, Shen C, Kalker T, Kotzerke J, Dankerl A, Glatting G, Reske S, Mattfeldt T. FDG uptake in breast cancer: correlation with biological and clinical prognostic parameters. Eur J Nucl Med Mol Imaging. 2002;29(10):1317-1323.

[205] Shimoda W, Hayashi M, Murakami K, Oyama T, Sunagawa M. The relationship between FDG uptake in PET scans and biological behavior in breast cancer. Breast Cancer. 2007;14(3):260-268.

[206] Koo HR, Park JS, Kang KW, Han W, Park IA, Moon WK. Correlation between 18F-FDG uptake on PET/CT and prognostic factors in triple-negative breast cancer. Eur Radiol. 2015;25(11):3314-3321.

[207] Gang, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. Med Phys 38(4), 1754-1768 (2011).

[208] Yu L, et al. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: Impact of radiation dose and reconstruction algorithms. Med Phys 40(4), 041908-1-9 (2013).

[209] Hernandez-Giron I, et al. Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms. Br J Radiol 87, 1-10 (2014).

[210] Gifford HC. Efficient visual-search model observers for PET. The British journal of radiology. 2014;87(1039):20140017.

VITA

Kristen A. Wangerin was born July 28, 1983. She attended Lake Central High School in Indiana, during which she was the trombone section leader of the wind ensemble and marching band. She participated in the annual Indiana State School Music Association solo and ensemble competitions playing trombone, euphonium, and piano. She found a love of running as a 100 m and 300 m hurdler on the track and field team.

She obtained her undergraduate degree in Nuclear Engineering from Purdue University in 2005. While at Purdue, she was an active member in the Society of Women Engineers, where she held positions of newsletter chair, publicity chair, and secretary; the American Nuclear Society; and Mortar Board Honor Society. She also ran two Chicago marathons as part of a charity team for Opportunity Enterprises in Valparaiso, Indiana.

Kristen then joined General Electric's Global Research Center in upstate New York as part of their Edison Engineering Development Program. She completed three rotations in medical imaging technologies while obtaining her Master's degree from Rensselaer Polytechnic Institute in Nuclear Engineering in 2008. She joined GE Runners and competed in the annual corporate track and field meet hosted by the United States Corporate Athletic Association, and she also found a passion for the mountains, triathlons, figure skating, and playing the violin.

Kristen returned to graduate school to expand her biomedical and clinical experience and has found her time as part of the Imaging Research Laboratory invaluable. Beyond her graduate studies, she could be found outside running, hiking, biking, skiing, or climbing.