Quantitative Accuracy of Iterative Reconstruction Algorithms in Positron Emission Tomography

A thesis submitted to the
University of Manchester
for the degree of Doctor of Philosophy in
the Faculty of
Biology, Medicine and Health

2016

Ian Armstrong

School of Health Sciences

Contents

List of figures 4
List of tables 6
List of abbreviations 7
Abstract 9
Declaration 10
Copyright statement 11
Acknowledgements 12
The Author 13
1. Introduction 14
1.1 Cancer 14
1.2 The birth of Positron Emission Tomography 16
1.3 Quantification of uptake of radio-labelled compounds in tumours 16
1.4 The growth of PET and PET/CT imaging and its applications 20
1.5 PET/CT in the UK and NHS 26
1.6 PET in Manchester 28
1.7 Scope of this project 30
1.8 Thesis structure 32
1.9 References 42
2. Principles of Positron Emission Tomography 50
2.1 Overview 50
2.2 Gamma-ray interactions 51
2.3 PET detector design 52
2.4 Development of PET hardware 55
2.5 Factors that degrade PET data 57
2.6 Advances in PET design 62
2.7 Time of flight PET 64
2.8 References 67
3. PET image acquisition and reconstruction 72
3.1 Overview 72
3.2 PET data collection 73
3.3 Data corrections 77
3.4 Image reconstruction 84
3.5 Analytical reconstruction 84

3.6 Iterative reconstruction 87
3.7 References 100
4. First paper 109
5. Second paper 138
6. Third paper 157
7. Fourth paper 168
8. Summary and conclusions 177
8.1 Impact on quantification 177
8.2 Improvements in image quality 180
8.3 Conclusions 182
8.4 References 184
9. Future work after PhD 185
9.1 Correlation with lung nodule and lymph node histology 185
9.2 Continuation of development of extended NEMA phantom 187
9.3 Assessment of image quality with next-generation PET scanners 189
9.4 References 191

Thesis word count: 25,790 (excluding papers and references)

LIST OF FIGURES

Figures in Chapter 1: Introduction

Figure 1.1 Illustration of activity concentration of FDG in blood and tissue over time 19

Figure 1.2 NEMA image quality phantom 34

Figures in Chapter 2: Principles of Positron Emission Tomography

Figure 2.1 Block detector design 56

Figure 2.2 Different coincidence event types 60

Figure 2.3 Illustration of the statistical noise observed in images 61

Figure 2.4 Representation of depth of interaction and inter-crystal scatter 62

Figures in Chapter 3: PET image acquisition and reconstruction

Figure 3.1 An example PET image 72

Figure 3.2 Example of a sinogram 74

Figure 3.3 Example of direct-plane and cross-plane sinograms 75

Figure 3.4 Illustration of 2D and 3D data acquisition 76

Figure 3.5 Assignment of events into TOF sinograms 77

Figure 3.6 Distribution of scattering angle cross-section for different gamma-ray energies 80

Figure 3.7 Example of backprojection for a varying range of projection angles 85

Figure 3.8 Schematic representation of the ramp filter 86

Figure 3.9 Illustration of filtered backprojection 86

Figure 3.10 Representation of depth of interaction and inter-crystal scatter and resultant image 94

Figure 3.11 Schematic illustration of the Gibbs artefact 96

Figures in Chapter 4: First paper

Figure 1 Transaxial PET images of the NEMA IQ phantom 116

Figure 2 Max-AC recovery for each of the six spheres filled in the 8:1 phantom 118

Figure 3 Max-AC values where the reconstruction parameters are set to minimize the variation of recovery across the different sphere sizes 120

Figure 4 AC recovery for the 10 mm and 37 mm sphere for max-AC, mean-AC and peak-AC 122

Figure 5 Coefficient of variation of AC measurements for the 10 mm and 37 mm sphere for max-AC, mean-AC and peak-AC 124

Figure 6 Max-AC for the 10 mm and 37 mm sphere filled with 2:1, 4:1 and 8:1 (bottom row) concentration ratio 126

Figure 7 Activity concentration recovery for each sphere after 3 iterations of each reconstruction algorithm 129

Figures in Chapter 5: Second paper

Figure 1	Phantom sphere profiles	146
Figure 2	Coronal PET images	146
Figure 3	Bland-Altman plots of relative differences for SUV_{max} relative to OSEM	149
Figure 4	Bland-Altman plots of relative differences for SUV_{peak} relative to OSEM	149
Figure 5	Bland-Altman plots of relative differences for TLG-40 relative to OSEM	150
Figure 6	Percentage differences versus lesion volume	150

Figures in Chapter 6: Third paper

Figure 1	Coronal and transaxial PET images for patient with BMI of 26.6 kg/m2	162
Figure 2	Coronal and transaxial PET images for patient with BMI of 42.4 kg/m2	163
Figure 3	Plots of measured SNR in the liver	164
Figure 4	Categorical visual assessments of image quality	165
Figure 5	Noise-equivalent count rate against ^{18}F-FDG activity in the NEMA phantom	166

Figures in Chapter 7: Fourth paper

Figure 1	Graphical representation of the estimation of ensemble variance	170
Figure 2	Sections through various 3-D regions that were evaluated	170
Figure 3	Coefficient of variation of EV values of spherical regions	171
Figure 4	Estimated covariance kernels from 200 replicate images	172
Figure 5	Comparison of estimated and measured EV for spherical regions	172
Figure 6	Repeatability of EV measures using data from separate phantoms	173
Figure 7	Spatial variation of covariance kernels	174
Figure 8	EV values calculated from different locations in the phantom	173
Figure 9	Comparison of estimated EV from NEMA background	173
Figure 10	Ratios of measured EV in background and NEMA spheres with various reconstruction parameters	175
Figure 11	Covariance kernels derived from four patient livers	175
Figure 12	Comparison of estimated EV values using covariance kernels from four patient livers	174

Figures in Chapter 8: Summary and conclusions

Figure 8.1	Gains in SNR using TOF for 88 routine patients	181

LIST OF TABLES

Tables in Chapter 1: Introduction

Table 1.1	Top ten causes of registered male deaths in England and Wales for 2014	14
Table 1.2	Top ten causes of registered female deaths in England and Wales for 2014	15
Table 1.3	T, N and M classifications based on disease characterisation for lung cancer	23
Table 1.4	Disease staging of lung cancer according to T, N and M classifications	24

Tables in Chapter 2: Principles of Positron Emission Tomography

Table 2.1	Characteristics of positron-emitting radionuclides that are commonly used in PET imaging	51
Table 2.2	Physical characteristics of scintillation crystals for PET detectors	55

Tables in Chapter 4: First paper

Table 1	Reconstruction parameters to give minimum dependency of max-AC on sphere size	119

Tables in Chapter 5: Second paper

Table 1	Phantom recovery data in unfiltered images	145
Table 2	Phantom recovery data	145
Table 3	Patient liver noise	147
Table 4	Relative uptake differences for matched voxel COV	148
Table 5	Relative uptake differences for matched SUV_{max}	148
Table 6	SUV_{max} changes for lesions with borderline values for malignancy	151

Tables in Chapter 6: Third paper

Table 1	Patient demographics in study group	160
Table 2	Liver SNR, mediastinum SUV_{mean} and lesion SUV_{max} for image data	163
Table 3	Relative changes of liver SNR, mediastinum SUV_{mean} and lesion SUV_{max} for image data	164
Table 4	Summary of reductions in prescribed activity and scanning time	166

LIST OF ABBREVIATIONS

AC	Activity concentration
ATP	Adenosine triphosphate
BGO	Bismuth germinate
CMUH	Central Manchester University Hospitals
COV	Coefficient of variation
CRG	Clinical Reference Group
CT	Computed tomography
DG	Deoxyglucose
DG-6-P	Deoxyglucose-6-phosphate
EV	Ensemble variance
FBP	Filtered back-projection
FDEV	Fourier-derived ensemble variance
FDG	Fluorodeoxyglucose
FOV	Field of view
FWHM	Full-width half-maximum
GE	General Electric
HD	HD·PET (*Siemens reconstruction algorithm*)
ICSCNM	Intercollegiate Standing Committee on Nuclear Medicine
IQ	Image quality (*PET phantom*)
LAC	Linear attenuation coefficient
LOR	Line of response
LSO	Lutetium oxyorthosilicate
MLEM	Maximum likelihood expectation maximisation
MR	Magnetic resonance
MRC	Medical Research Council
NEC	Noise equivalent counts
NEMA	National Electrical Manufacturers Association
NHS	National Health Service
NICE	National Institute for Clinical Excellence
OP	Ordinary Poisson
OSEM	Ordered subset expectation maximisation
PERCIST	PET Response Criteria in Solid Tumours
PET	Positron emission tomography
PET/CT	PET with integrated CT
PET/MR	PET with integrated MR
PMT	Photomultiplier tube
PSF	Point spread function
RM	Resolution modelling
ROI	Region of interest
SNR	Signal-to-noise ratio

SUV	Standardised uptake value
TLG	Total lesion glycolysis
TOF	Time of flight
TNM	Tumour, Node, Metastases (*cancer scoring system*)
UHD	ultraHD·PET (*Siemens reconstruction algorithm*)
VOI	Volume of interest
WMIC	Wolfson Molecular Imaging Centre

ABSTRACT

The University of Manchester

Ian Armstrong. Doctor of Philosophy.

Quantitative Accuracy of Iterative Reconstruction Algorithms in Positron Emission Tomography

2016

Positron Emission Tomography (PET) plays an essential role in the management of patients with cancer. It is used to detect and characterise malignancy as well as monitor response to therapy. PET is a quantitative imaging tool, producing images that quantify the uptake of a radiotracer that has been administered to the patient. The most common measure of uptake derived from the image is known as a Standardised Uptake Value (SUV). Data acquired on the scanner is processed to produce images that are reported by clinicians. This task is known as image reconstruction and uses computational algorithms to process the scan data. The last decade has seen substantial development of these algorithms, which have become commercially available: modelling of the scanner spatial resolution (resolution modelling) and time of flight (TOF). The Biograph mCT was the first scanner from Siemens Healthcare to feature these two algorithms and the scanner at Central Manchester University Hospitals was the first Biograph mCT to go live in the UK. This PhD project, sponsored by Siemens Healthcare, aims to evaluate the effect of these algorithms on SUV in routine oncology imaging through a combination of phantom and patient studies.

Resolution modelling improved visualisation of small objects and resulted in significant increases of uptake measurements. This may pose a challenge to clinicians when interpreting established uptake metrics that are used as an indication of disease status. Resolution modelling reduced the variability of SUV. This improved precision is particularly beneficial when assessing SUV changes during therapy monitoring.

TOF was shown to reduce image noise with a conservation of FDG uptake measurements, relative to non-TOF algorithms. As a result of this work, TOF has been used routinely since mid-2014 at the CMUH department. This has facilitated a reduction of patient and staff radiation dose and an increase of 100 scans performed each year in the department.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

ACKNOWLEDGEMENTS

There are many people that have played an invaluable role in supported me over the course of this PhD project. I wish to express my sincere gratitude to all of them.

My three supervisors: Dr Julian Matthews, Dr Heather Williams and Dr Jerome DeClerck of the University of Manchester, Central Manchester University Hospitals and Siemens Healthcare respectively. Their guidance and support has been fantastic over the course of this project.

Siemens Healthcare for sponsoring this project along with the travel budget, which has allowed me to present my work in some amazing parts of the world.

My current and past colleagues in the nuclear medicine department at Central Manchester University Hospitals who supported me enormously in getting the project established and for their continued support throughout. In particular, I would like to thank Dr Jackie James who provided essential clinical input in assessing hundreds of PET/CT images.

Matt Kelly, formally of Siemens Healthcare, for being extremely helpful with many aspects of data analysis. In particular, for his patience with my complete lack of Matlab experience!

Finally my wife Rosie, and children Aaron and Megan. They have been very patient when I have been away at conferences and especially over the last few months when I have been writing this thesis.

THE AUTHOR

Ian Armstrong graduated from the University of Manchester with a first class MPhys Physics degree in 2004. It was a final year project, based at the University's Wolfson Molecular Imaging Centre, which provided his first exposure to Positron Emission Tomography. During this project, Ian designed and manufactured cyclotron targetry for producing fluorine-18. This provided the motivation to choose a career in medical physics. Ian completed a two-year training post at The Christie Hospital before moving to Nuclear Medicine department at Central Manchester University Hospitals in 2006 as a medical physicist. Since October 2013, Ian has been in the position of Principal Physicist in the department.

In his medical physics role, Ian developed an interest in image reconstruction and optimisation. This was initially in the area of cardiology, where he had substantial input in the implementation of a dosing protocol based upon BMI alongside routine use of advanced iterative reconstruction, which resulted in significant reductions in administered activity and radiation to patients undergoing myocardial perfusion scans. Alongside the work in this thesis that focuses on oncological applications in PET/CT, Ian has also had substantial involvement in the characterisation and optimisation of reconstruction in cardiac PET using rubidium-82 in the department.

Since starting this 6-year part-time PhD, Ian has become father to two children. He is also a keen photographer and a member of South Manchester Camera Club. Some of the photographs taken during overseas travel when presenting work from this project have received high acclaim in regional photographic organisations.

1. INTRODUCTION

1.1 Cancer

Cancer (malignant neoplasm) of any form is the biggest cause of mortality in England and Wales accounting for 29% of deaths in 2014, ahead of circulatory diseases (27%) and respiratory diseases (13%) [1]. The top ten underlying causes of registered deaths in England and Wales in 2014 for males and females are summarised in Table 1.1 and Table 1.2 respectively.

Rank	Underlying cause of death	Number of deaths registered	Percentage of all male deaths
1	Ischaemic heart diseases	36,319	14.8
2	Dementia and Alzheimer disease	17,177	7.0
3	Malignant neoplasm of trachea, bronchus and lung	16,959	6.9
4	Chronic lower respiratory diseases	14,565	5.9
5	Cerebrovascular diseases	14,194	5.8
6	Influenza and pneumonia	11,242	4.6
7	Malignant neoplasm of prostate	10,153	4.1
8	Malignant neoplasm of colon, sigmoid, rectum and anus	7,718	3.1
9	Malignant neoplasms, stated or presumed to primary of lymphoid, haematopoietic and related tissue	6,454	2.6
10	Diseases of liver	4,737	1.9

Table 1.1 Top ten causes of registered male deaths in England and Wales for 2014 [1].

Rank	Underlying cause of death	Number of deaths registered	Percentage of all female deaths
1	Dementia and Alzheimer disease	34,321	13.4
2	Ischaemic heart diseases	24,190	9.4
3	Cerebrovascular diseases	19,963	7.8
4	Chronic lower respiratory diseases	14,467	5.6
5	Influenza and pneumonia	14,212	5.5
6	Malignant neoplasm of trachea, bronchus and lung	13,909	5.4
7	Malignant neoplasms of female breast	10,097	3.9
8	Malignant neoplasm of colon, sigmoid, rectum and anus	6,569	2.6
9	Diseases of the urinary system	5,032	2.0
10	Malignant neoplasms, stated or presumed to primary of lymphoid, haematopoietic and related tissue	5,025	2.0

Table 1.2 Top ten causes of registered female deaths in England and Wales for 2014 [1].

Cancer is the uncontrolled proliferation and potential spread of abnormal cells, with most types leading to the development of *malignant neoplasms* – a mass of cancer cells – which are commonly referred to as tumours or lesions. Cancerous cells exhibit a number of biological traits that give rise to this sinister behaviour [2]. One particular trait is the disease's ability to infiltrate normal tissues and spread through the body to form new deposits, referred to as *metastases*, which also grow. As the disease progresses, it can lead to the function of individual organs affected being critically impaired or ceasing, and ultimately this can lead to death [3].

Otto Warburg is recognised as one of the pioneers in the study of the characteristics of cancer, with his early work on metabolic rates of normal and cancer cells. Both normal and cancer cells require energy to survive, with the major source of cellular energy originating from the compound adenosine triphosphate (ATP). ATP is created by two cellular processes. The first is glycolysis - a process in cells that converts glucose to pyruvate and lactic acid, which produces 2 ATP molecules from every glucose molecule. The second process is oxidative phosphorylation in the Krebs cycle, which uses the pyruvate produced by glycolysis to produce 36 molecules of ATP per glucose molecule [4]. By measuring blood glucose and lactic acid concentration in an artery and vein either side of normal tissue and cancer cells [5],

Warburg demonstrated glucose utilisation and lactic acid production was greater in cancer cells. Warburg concluded that during the transition from normal to cancer cells irreversible damage occurs to the cell's ability to produce ATP during the Krebs cycle [6]. In order for the damaged cells to survive, the balance between the two processes that produce ATP changes dramatically in favour of glycolysis and hence glucose utilisation is greater in cancer cells. This is known as the Warburg Effect [4].

1.2 The birth of Positron Emission Tomography

In the 1950s, work by William Sweet [7] and Frank Wrenn [8] began on using positron-emitting radionuclides to assess the localisation of brain tumours. This technique took advantage of the emission of two 511 keV gamma-rays in opposing directions that are produced by positron annihilation with an electron in matter. Using two individual opposing gamma-ray detectors, it is possible to define the line, referred to as the *Line of Response* (LOR), created by the gamma-rays and hence infer the possible location of the positron-electron annihilation from which they resulted. In the early 1970s, Burnham and colleagues had begun to construct multi-detector systems to map out a 2-dimensional distribution of positron-emitting radionuclides [9]. This was achieved by mapping multiple LORs over a range of angles and positions such that the distribution of the radionuclide could be derived. In 1975, the first *Positron Emission Tomography* (PET) system for use in human subjects was built by Michael Phelps and Ed Hoffman [10] and the first images from human studies published the following year in the Journal of Nuclear Medicine [11]. These early studies used short-lived radionuclides such as carbon-11, nitrogen-13 and oxygen-15 and focussed mainly on neurological disorders. Following the collection of many LORs on the PET scanner, a process known as image reconstruction is required, which takes the spatial information of the LORs that is acquired on the scanner and creates an image. The image consists of discrete elements, known as voxels (short for volume element), which typically have units of Bq/ml to reflect the activity concentration of the radiotracer within the patient.

1.3 Quantification of uptake of radio-labelled compounds in tumours

The measurement of enhanced glucose utilisation, as demonstrated by Warburg, was studied using radioactivity initially with radio-labelled glucose. This used either carbon-11 and gamma-ray detectors [12] or carbon-14 for detection by

autoradiography [13]. However, the mathematical analysis of the process is complicated as glucose is not trapped within tissues and the short half-life of carbon-11 increases the complexity. A new compound was formulated by Lou Sokoloff [14] known as 2-Deoxyglucose (DG), an analogue of glucose, differing only by the replacement of a hydroxyl group by a hydrogen atom. Like glucose, it is transported by glucose transporter proteins into cells. Within the cell, it is phosphorylated by the hexokinase enzyme during glycolysis to form DG 6-phosphate (DG-6-P). The removal of the hydroxyl group prevents further metabolism down the glycolytic pathway so the compound remains essentially trapped in the cells. A multi-compartmental kinetic analysis of the DG molecule was used to quantify the uptake in rat brains [14]. The method derived three transfer rate constants of K_1 (DG in plasma into DG in tissue), k_2 (DG in tissue into DG in plasma), k_3 (phosphorylation of DG in tissue into DG-6-P in tissue). The DG molecule in this work by Sokoloff was radio-labelled with carbon-14 and detection was performed by autoradiography and not by PET. This changed in the late 1970s when, according to a historical account by Ronald Nutt [15], Lou Sokoloff was at a wine tasting event with Martin Reivich, a neurologist from the University of Pennsylvania, and discussed the idea of radio-labelling the DG molecule with the positron-emitting radionuclide fluorine-18. The 110 minute half-life of fluorine-18 alleviated the need for an onsite cyclotron, meaning the compound could be transported to another centre for use in PET imaging. After discussions with colleagues at Brookhaven, 2-[^{18}F]-*fluorodeoxyglucose* (FDG) was synthesised [16] and is now the most commonly used radiotracer in PET imaging [17].

A key use of PET is to derive activity concentrations of the injected radiotracer that reflects the biological processes determined by the chemical form of the radiotracer. When using FDG for oncology imaging, the main clinical application is to assess the glucose metabolic rate of tissues with the greater glucose utilisation of malignant cells providing a means of detection. Given that FDG is a glucose analogue, there is general systemic uptake of the radiotracer. However, the higher glucose utilisation of malignant cells leads to a greater than normal accumulation of FDG. On a PET image, the visual intensity of the voxels reflects the magnitude of activity concentration and so a lesion can be identified from the higher visual intensity compared with the expected intensity of surrounding tissue.

Quantifying FDG uptake can be performed by several methods. Kinetic analysis, similar to that of Sokoloff, was performed in the brain using FDG, radio-labelled with fluorine-18 and data derived from PET images [18]. This study observed the very slow reversibility of FDG and led to an inclusion of an additional rate constant, k_4, representing the dephosphorylation of DG-6-P in tissue into DG in tissue. Early examples have demonstrated this dynamic image approach for oncology imaging [19–22]. To perform this analysis, a PET image acquisition is acquired using multiple time frames from the point of FDG injection, producing multiple images over the course of the acquisition, which lasts for an hour or more. The image field of view must be positioned such that it covers a single known tumour region within the patient. Arterial blood samples are taken from the patient throughout the image acquisition to measure in order to measure the blood plasma concentration of FDG, making the procedure complex, relatively invasive and lengthy. This technique is used to determine the transport rate constants of the FDG. From this, the rate of FDG uptake from plasma into the cell as trapped FDG-6-P, K_i, is determined as:

$$K_i = \frac{K_1 k_3}{\left(k_2 + k_3\right)}. \tag{1.1}$$

Noting that k_4 has been omitted as it is assumed to be very small in comparison with the other rate constants [18].

An alternative method was proposed by Clifford Patlak, which derived the rate of FDG uptake from activity concentration measurements of the blood plasma and in tissue from an image [23]. While alleviating the need for blood samples, the procedure still required a long PET scan. In both of these methods, the area of the patient that is covered by the imaging is relatively small due to the limited axial field of view of the PET scanner, which at the time was typically 15 cm or less. As a result, lesions within the body that lie outside the field of view may be missed so this approach is unsuitable if the purpose of the PET scan is to detect and quantify the tumour burden as a whole.

For routine clinical PET, a measurement of uptake derived from a wholebody image acquired at a single time-point after the radiotracer injection is desirable. After the

time of injection of FDG, the activity concentration in the plasma and tissue will vary as the FDG is transferred from plasma and trapped in tissue. Figure 1.1 gives a schematic of the time-activity curves of FDG in these two compartments.

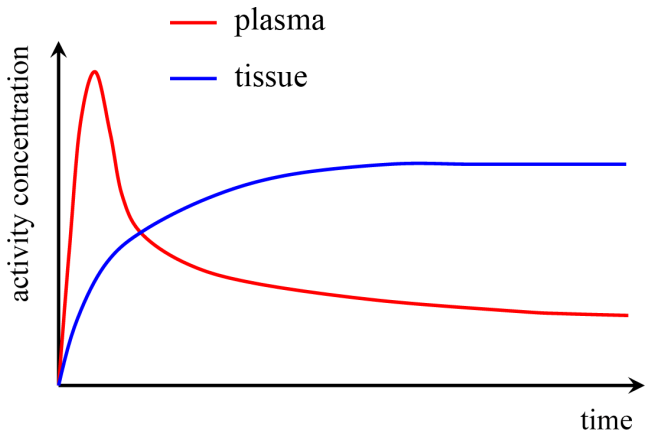


Figure 1.1 Schematic to illustrate the activity concentration of [^{18}F]FDG in blood and tissue over time.

After a time of approximately 45 minutes, the uptake in tissue is relatively stable and the activity in the plasma has fallen to low levels [21, 24, 25]. Therefore after this point in time, a measurement of activity concentration in a tumour can be taken, which is proportional to the rate of uptake. However, this measurement is dependent on the quantity of FDG injected and the volume of dilution for the patient in which it is distributed. By normalising according to these factors, a semi-quantitative measurement of glucose metabolic rate can be performed. Two normalisation methods can be used: either to take the ratio of uptake in malignant cells to normal tissue [26, 27] or, more commonly, to normalise to the injected activity concentration and patient weight. The latter technique dates back to the 1940s where studies assessed the uptake of phosphorus-32 in tumours [28]. Tumours were excised and the radioactivity per weight of the sample was determined. This was normalised to the radioactivity administered divided by the patient weight. This consequently resulted in a unit-less value and was referred to as the Differential Absorption Ratio. This method was used to quantify tracer uptake in hepatic tumours [29] and lung tumours [30]. Through the 1990s, the measure was also referred to as Differential Uptake Ratio and Standardised Uptake Ratio before the term *Standardised Uptake Value* (SUV) was universally accepted.

For normalisation to patient body weight, SUV is expressed as

$$\mathrm{SUV} = \frac{\text{tissue activity concentration}}{\text{injected activity}/\text{body weight}} \tag{1.2}$$

Given the transient nature of tracer distribution and uptake in tissue, the measurement of SUV should be performed at a standard time post-injection to provide standardised and meaningful data [31]. For FDG oncology scans, PET images are typically performed at 60 minutes after injection of the radiotracer [17], as recommended in guidelines [31]. Comparisons of SUV with rate constants from full compartmental analysis have shown that, when the imaging time is controlled, SUV is generally a good surrogate for the metabolic rate of glucose utilisation [21, 24].

A tumour will almost certainly exhibit heterogeneous FDG uptake and extend over several voxels in the PET image. There is a need to derive a single parameter from these voxel values, to produce a metric such as SUV, to reflect the tumour uptake. One begins by defining a region that encompasses the tumour volume using the voxel intensity for guidance and then defining which voxels within the tumour should contribute to the uptake measurement for the tumour. The current most commonly used method is to only include maximum voxel value within the region surrounding this tumour, which produces a SUV variant referred to as $\mathrm{SUV}_{\mathrm{max}}$ [17]. Other SUV variants can be derived from an average of several voxels and are known as $\mathrm{SUV}_{\mathrm{mean}}$. Voxels contribute to $\mathrm{SUV}_{\mathrm{mean}}$ if they fall above a particular threshold, which may be absolute in value or relative to the maximum voxel value within the region.

1.4 The growth of PET and PET/CT imaging and its applications

Through the 1980s, FDG PET was used primarily for the assessment of myocardial viability and as a research tool. The utilisation of FDG for oncology was slow to become established and it was not until 1992 that the first wholebody FDG PET images from oncology patients were published [32]. Through the 1990s, several studies were published evaluating the use of FDG in oncology, primarily in lung cancer [27, 33–36] but also other cancers including lymphoma [37] and hepatic

tumours [38]. In these studies, quantification of the FDG uptake in tumours used either lesion-to-normal ratios or SUV. The quantification of the FDG uptake was used to differentiate malignant lesions from benign. Early work on lung cancer showed that lesions confirmed to be benign by biopsy did not exhibit an SUV of greater than 2.5 and it was suggested that this value provided a useful discriminator between benign and malignant lesions [39]. In this work, SUV_{mean} was extracted from regions that were outlined around lesions on a single image at the point of the most intense lesion uptake. This lead to the notion of FDG PET being used as a "metabolic biopsy" reflecting how an SUV of 2.5 could be used to discriminate malignancy from benign lesions [40, 41]. These two studies employed an SUV_{max} technique and assumed malignancy in lesions that exhibited an SUV_{max} of greater than 2.5. Despite the age of these studies and advances in scanner technology, this threshold is a widely used approach, which has been criticised [42]. Measurement of SUV suffers from partial volume effects that result in an increasing negative bias in SUV with decreasing object size [43]. Clinicians have developed an appreciation for this and can estimate the severity of this using measurements of tumours derived from an anatomical image from *Computed Tomography* (CT).

Based on the evidence from the 1990s, the American Food and Drug Administration provided approval in 1997 for the use of FDG in oncology [44]. In 1998 the United States Health Care Financing Administration, now known as the Centers for Medicare & Medicaid Services, granted reimbursement for PET imaging for the diagnosis of lung cancer. Over the next three years, the range of reimbursed indications expanded to include colorectal cancer, oesophageal cancer, melanoma and lymphoma [15, 45].

At the turn of the millennium, another major milestone was reached with PET imaging. That was the development of a PET scanner combined with CT [46]. This provided co-registered images of function (glucose metabolism if using FDG) from the PET and anatomy from the CT. The wholebody anatomical image, acquired at the high-speed of a CT scan, provided the ability to accurately locate an area of increased FDG uptake and produced a huge step in terms of diagnostic confidence making it probably one of the pivotal technological developments for routine oncology PET/CT imaging.

The range of applications of PET/CT in oncology has continued to increase. All of which are based upon the magnitude and extent of FDG uptake that is determined from the PET images and localisation, which is enhanced by using the CT images. The applications can be broadly classified as follows:

Initial diagnosis. PET/CT can be used to help determine whether a suspicious tissue mass identified from previous anatomical imaging may be malignant. As stated, normal physiological uptake of FDG is seen on PET/CT images and clinicians will look to identify any areas that exhibit greater than normal uptake or changes in shape or symmetry. Further assessment is achieved by measuring the FDG uptake usually by using SUV in the lesion in question.

Staging of disease. If a primary cancer is either confirmed or suspected then PET/CT can be used to assess the extent, and hence stage, of the disease. Metastatic deposits can occur in either the lymphatic or other organs. These may be undetectable using anatomical imaging but are likely to exhibit FDG uptake. If the FDG uptake that is assumed malignant is detected beyond the primary tumour then the spread of the cancer can be classified. Total disease burden is based upon a three classification scoring system commonly referred to as the TNM system, which is currently in its 7th version [47]. The T classification is for tumour assessment and is based on size and invasion of neighbouring tissue. The N classification is to describe the degree of metastatic spread through for lymph nodes. The M is used to describe the presence of metastases elsewhere in the body. Criteria are defined for each type of cancer that state how each of the three classifications should be scored. Table 1.3 shows how T, N and M classifications are assigned for lung cancer. Based upon the TNM score, the disease stage is defined with each type of cancer having its own staging [47]. Table 1.4 shows the disease staging for lung cancer and illustrates how different TNM classification scores can result in the same disease staging. The overall disease staging will influence the clinical decisions as part of the patient management.

Classification	Description
T classifications	
T1a	Tumour < 2 cm in greatest dimension
T1b	Tumour 2 to 3 cm in greatest dimension
T2a	Tumour 3 to 5 cm in greatest dimension
T2b	Tumour 5 to 7 cm in greatest dimension
T3	Tumour > 7 cm in greatest dimension or invasion into parietal pleural chest wall, diaphragm, phernic nerve, mediastinal pleura,
T4	Tumour of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus, vertebral body, carina, separate tumour nodule(s) in different ipsilateral lobe
N classifications	
N0	No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes,
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
M classifications	
M0	No distant metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe tumour with pleural nodules or malignant pleural effusion
M1b	Distant metastasis – in extrathoracic organs

Table 1.3 T, N and M classifications based on disease characterisation for lung cancer [47].

Stage	T classification	N classification	M classification
1A	T1a	N0	M0
	T1a	N0	N0
1B	T2a	N0	M0
2A	T2b	N0	M0
	T1a	N1	M0
	T1b	N1	M0
	T2a	N1	M0
2B	T2B	N1	M0
	T3	N0	M0
3A	T1a	N2	M0
	T1b	N2	M0
	T2a	N2	M0
	T2b	N2	M0
	T3	N1	M0
	T3	N2	M0
	T4	N0	M0
	T4	N1	M0
3B	T1a	N3	M0
	T1b	N3	M0
	T2a	N3	M0
	T2b	N3	M0
	T3	N3	M0
	T4	N2	M0
	T4	N3	M0
4	Any T	Any N	M1a
	Any T	Any N	M1b

Table 1.4 Disease staging of lung cancer according to T, N and M classifications [47].

Grading of disease. Confirmation of malignant cells within a tumour is commonly achieved by either surgical biopsy or resection of suspicious tissue. Analysis of tissue samples is performed by a pathologist and the cell type of the tumour is determined. Two examples of malignant cell types are adenocarcinoma and squamous cell carcinoma. Squamous cell carcinomas tend to be more aggressive and grow faster, resulting in a worse prognosis [48, 49]. The other outcome from the pathology is the level of cell differentiation, which is a measure of how similar the cells in the sample are compared with the origin cell type. The magnitude of FDG uptake within a

tumour has been shown to be suggestive of both cell type and differentiation and therefore linked to prognosis [50–53].

Monitoring response to cancer therapy. Response to cancer therapy is classified using four categories: progressive disease; no change; partial response and complete response [54]. Traditionally, the classification of tumour response is determined by a change in tumour size, derived from anatomical imaging such as CT or *Magnetic Resonance* (MR) imaging and follows a scheme known as *Response Evaluation Criteria for Solid Tumours* (RECIST) [55]. However, data from PET/CT imaging can also be incorporated into this scheme where the degree of change in FDG uptake within the tumour specifies the response [56]. In 2009, a new protocol for assessment of response to therapy for cancers forming solid tumours was established, known as *PET Response Criteria in Solid Tumours* (PERCIST) [57].

Checking for disease recurrence. Patients that are in remission following cancer treatment will be monitored to check for signs of disease recurrence. This monitoring is usually in the form of clinical assessment, periodic blood tests to identify blood-markers that are indicative of malignancy, and anatomical imaging. If markers are seen to rise or imaging reveals suspicious pathology, a PET/CT scan can be performed to detect any potential returning disease.

Establishing primary tumour. Patients may exhibit symptoms and possibly blood markers that are highly suggestive of malignancy but the primary tumour has not been located on anatomical imaging. In these cases, PET/CT can be used to possibly identify a primary tumour.

Radiotherapy planning. External beam radiotherapy uses very high energy X-rays to target and kill the cancerous cells within the body. The X-rays are delivered from various directions and a treatment planning step will determine the optimum configuration to deliver the desired dose to the tumour while sparing healthy tissue as much as possible. Traditional treatment planning is based upon anatomical images from CT. By including information from the PET/CT images as well, the most metabolically active areas can be preferentially targeted, which may result in sparing of healthy tissue. In addition, the concept of "dose-painting" is a growing area of

research. Here the distribution of the radiotherapy dose delivery within the tumour is heterogeneous and guided by the intra-tumour uptake of the radiotracer such that areas of greater uptake, hence with greater metabolic rate and considered more aggressive, receive the greater radiotherapy dose.

1.5 PET/CT in the UK and NHS

The first PET facility in the UK was the Medical Research Council (MRC) Cyclotron Unit, which opened in 1979 at the Hammersmith Hospital in London [58]. This was a research facility with a primary focus on neurological conditions, with research groups investigating oncology and cardiology. The UK's first clinical PET centre opened in 1992 at Guy's and St. Thomas' Hospital in London with an initial vision of imaging neurological and cardiological conditions to reflect activities in the US at the same time [45]. Within two years of opening, the proportion of referrals for oncology indications had risen to 48% of the work and today the proportion is over 90% [45].

Despite the growing evidence of the clinical utility of PET in oncology, especially lung cancer, the development of clinical PET within the UK and its National Health Service (NHS) was slow in the late 1990s and early 2000s compared with European countries. The low number of clinical PET centres were either partly hospital-funded or privately-funded and there was no national strategy on the role of PET in the NHS. In 2005 there were seven NHS PET scanners operational in the UK for clinical use rather than research, corresponding to one scanner per 8.6 million of the UK population [59]. This was some way behind the Belgium, Germany, Austria, Sweden and Denmark, which all had one PET scanner serving less than 2 million population. In addition, five of these systems were located in London so provision was all but absent outside the capital. The limiting factor in the UK was perceived to be cost to the NHS [60].

In 2003, an intercollegiate committee (ICSCNM) produced a report outlining a potential strategy for PET within the NHS [61].

Key points from the document were

- recommendations of the number of PET scans that were likely to be required each year;
- that PET scans should be funded centrally by the NHS;
- that delivery should be based around the proposed 37 cancer networks within the UK;
- a list of suggested evidence-based clinical indications for which PET imaging should be funded.

In 2005, the first version of the NHS National Institute for Clinical Excellence (NICE) evidence-based guidelines for lung cancer was published [62]. The guidelines stated that:

- every cancer network should have rapid access to FDG PET;
- patients with lung cancer should be treated within 62 days from the initial referral by a general practitioner;
- FDG PET should play an essential role in their management.

The role of FDG PET was recommended for:

- the assessment of single pulmonary nodules;
- the assessment of mediastinal lymph nodes for determining whether patients undergo resection or biopsy of lymph nodes depending on the nodal (N score) disease staging derived from the PET images;
- the assessment for patients who are candidates for radical radiotherapy.

Later in the same year, two companion documents were published that set out the strategy for PET within the NHS: the Department of Health *Framework for PET in the UK* [59] and the collaborative *PET in the UK Strategy* published by the Royal College of Radiologists, British Nuclear Medicine Society and the ICSCNM [63]. From here, models were proposed to provide approximately 40,000 PET scans per year over a three to five year period. It was also stated that all new scanners should

be integrated PET/CT systems and not PET-only machines. With agreed funding for scans and investment, clinical PET/CT for oncology finally became established in the UK. Additional capacity was provided by the independent sector by way of mobile systems. Recent analysis has shown that, from 2008, the number of PET/CT scans performed in the UK and Ireland has seen an annual increase of 14% [64]. In 2012, the Royal College of Radiologists and Royal College of Physicians prepared a document summarising the evidence for clinical indications in PET/CT for the UK. The document forms the basis for NHS funding of PET/CT examinations according to the clinical indication and is now in its third edition [65]. It is likely that the evidence for the use of PET/CT will continue to increase and consequently the funded indications and number of scans performed will continue to expand.

In 2015, the delivery of PET/CT in England was split between the independent provider Alliance Medical and NHS hospitals, with both routes receiving NHS funding. Recommendations for use of PET/CT were made according to the PET/CT Clinical Reference Group (CRG) which fell under the larger NHS Commissioning Board and had the role of developing specialist services within the NHS [66]. In 2013, the PET/CT CRG established a national contract that specified standards of service required for the delivery of PET/CT [67]. A key requirement of the contract is that a provider of PET/CT will perform the scan and return the diagnostic report and images to the referring clinician within seven business days. With increasing numbers of referrals, these requirements exert considerable pressure on PET centres to increase capacity.

1.6 PET in Manchester

In Manchester, PET began at the Christie hospital with the installation of a General Electric (GE) Advance PET-only system that was for research purposes in 1999. In 2001, a clinical PET imaging service using coincidence imaging on a traditional GE gamma camera was set up by the Manchester Royal Infirmary nuclear medicine department. After two years, the department was performing almost 200 oncology scans per year with referrals primarily from Manchester Royal but also several other hospitals around the region. In 2004, and in light of the developments for PET within the NHS at the time, the Christie hospital established the Oncology PET FDG service and hence initiated a clinical PET service across the Manchester region. The main

clinical indication for imaging, for which funding was agreed, was for staging lung cancer patients. As the evidence base for more indications grew, the number of funded PET scans increased. In 2007, the PET-only system at the Christie hospital was replaced with an integrated GE Discovery STE PET/CT system. This greatly increased throughput and, consequently, the gamma camera PET service at Manchester Royal ceased in the same year. The next three years saw annual increases of 30% in the number of patients scanned under the service.

In 2006, the University of Manchester Wolfson Molecular Imaging Centre (WMIC), which is located next to the Christie hospital site started PET scanning. This imaging research centre was equipped with a GE PETtrace cyclotron and two PET scanners – a High Resolution Research Tomograph (HRRT) and Siemens/CTI Biograph PET/CT system. The HRRT is a PET-only dedicated brain imaging scanner while the Biograph is a full-size PET/CT for wholebody patient scans. In 2009, Central Manchester University Hospitals (CMUH) began operating in a new hospital building, bringing together adult (Manchester Royal Infirmary), children's, women's and eye hospitals into a single building. A new and expanded nuclear medicine department was formed in the new hospital and a state-of-the-art Siemens Biograph mCT PET/CT scanner installed.

By the time the Biograph was installed at CMUH, the demand for clinical oncology scans had exceeded the capacity of the single scanner at the Christie hospital and so scans were also being performed at the WMIC. Capacity was increased significantly when the clinical oncology imaging service went live in early 2010 at CMUH. A target annual throughput of 1500 scans at CMUH was agreed between the two hospitals.

In May 2010, CMUH became the second centre in the UK to launch a clinical PET/CT myocardial perfusion imaging service using rubidium-82. Within three months, the service was performing 100 scans per month and the working day was split approximately 50:50 between the cardiac work and the oncology work.

Currently, the CMUH PET/CT scanner operates at maximum capacity providing a clinical service for 56 hours, Monday to Friday. Additional monthly Saturday

services were established in 2016 to increase scanning capacity, but these are currently dedicated to the rubidium cardiac service. In 2015, 3321 studies were performed on the PET/CT. These are broken down into rubidium myocardial perfusion (1746), FDG oncology (1428), inflammation and infection (87), brain imaging (26), myocardial viability (24), paediatric congenital hyper-insulinism (9) and brain amyloid plaque imaging (1).

In late 2016, a GE SIGNA PET with a combined MR scanner will be installed at CMUH. Funding for the system has been awarded by the MRC Dementias Platform UK. The PET/MR will be operated four days per week by The University of Manchester and one day by CMUH.

1.7 Scope of this project

To put the scope of this project into context, it is beneficial to briefly summarise the use of PET in oncology, particularly in Manchester where this project has been undertaken. PET/CT plays a major role in the management of patients with cancer, and the most common clinical indication in the region is lung cancer. SUV provides considerable input into the diagnosis, with SUV_{max} being used exclusively in the Manchester region. Local agreement has been established on the clinical interpretation of SUVs derived from images. Therefore, as the Biograph mCT at CMUH works within the oncology service in collaboration with the scanner at the Christie hospital, the quantitative performance of the two scanners needs to be relatively consistent to provide commonality. This was established during acceptance testing of the Biograph at CMUH.

The Biograph mCT from Siemens Healthcare, used in this project, was the first commercial system from Siemens to provide *Time of Flight* (TOF) functionality (with image reconstruction using this information referred to as ultraHD·PET (UHD)) and also iterative reconstruction with resolution modelling (referred to as HD·PET (HD)). The scanner was the first of its type to go live in the UK and, as such, there was an urgent need to evaluate the impact and potential benefits of the advances in technology for routine oncology imaging. After the system was commissioned these advanced algorithms were not used due to lack of familiarity

and so the standard iterative reconstruction (referred to as OSEM) was used for all patient scans.

There were clear mutual benefits to both CMUH and Siemens Healthcare in evaluating the performance of the technological advances within a high throughput oncology imaging service. Consequently, this PhD project was sponsored by Siemens Healthcare. During the planning of the project, an outline of mutual benefits to both CMUH and Siemens Healthcare was defined as the basis for agreement of funding. However, the agreement was that the direction of the project was relatively fluid, and hence the aims could be modified according to the progress and findings throughout the project.

The original aims discussed as part of the project outline were to:

- Investigate variation in quantitative accuracy with lesion size for OSEM, HD and UHD reconstructions under various, clinically-relevant conditions;
- Select a subset of promising reconstruction parameters for OSEM, HD and UHD reconstructions under imaging conditions comparable to those seen clinically and assess reproducibility of the region-based quantitative measures;
- On basis of accumulated evidence, select preferred reconstruction parameters and Region of Interest (ROI) analyses for OSEM, HD and UHD;
- Extend application of preferred reconstruction parameters and ROI analyses to patient studies;
- Establish relationship between selected ROI-based quantitative measures from OSEM, HD and UHD, using preferred reconstruction parameters and ROI-analysis methods.

There are two areas of mutual interest that are evaluated in this PhD thesis, in the context of addressing the original project aims but also the influence of the demand to perform ever-increasing numbers of scans as part of a clinical oncology service in an NHS hospital. The work performed is divided into two main themes.

The first of these is the impact of the advanced reconstruction algorithms on measurements of FDG uptake derived from the images. This is divided into an assessment of both the accuracy and variability of uptake measurements. To assess quantitative accuracy, a ground truth of activity concentration is required and hence the project first focussed on phantom studies before moving on to clinical work. The second theme is to explore the potential of exploiting anticipated gains in image quality from these advanced reconstruction algorithms to reduce the imaging time, administered activity, or both, for routine oncology scanning to meet the NHS service requirements for oncology PET/CT [67].

1.8 Thesis structure

Following on from this introduction are two review chapters. The first will give an overview of the physical principles of positron emission and detection and a technical description of the development of the PET scanner. The second review will discuss the various image reconstruction techniques that are employed on PET/CT systems.

The thesis then consists of four papers of work that aim to describe the impact of advanced iterative reconstruction and TOF. At the time of thesis submission, two of the studies have been published in peer-reviewed journals. The rationale and aims of the four papers are summarised as follows. The findings from the four pieces of work and follow-on work will be discussed in the final chapter of the thesis.

1.8.1 First paper: Impact of advances in PET reconstruction algorithms and uptake measurement technique on the accuracy and precision of SUV

This first piece of work was performed in 2010 soon after the Biograph was installed at CMUH. There were three iterative reconstruction algorithms available on the Biograph. These were:

- the standard iterative algorithm (OSEM) that has been used to standardise the Biograph image quality with that of the scanner at the Christie hospital since the start of the routine oncology service;
- an iterative algorithm with spatial resolution modelling (HD);
- an iterative algorithm with combined spatial resolution modelling and time of flight (UHD).

The scope of this work was to characterise these two new advanced algorithms using the traditional algorithm as a reference with a focus on quantitative accuracy. To allow accuracy and variability of activity concentration to be assessed, a phantom was used to provide the ground truth.

The aims of the work are summarised as follows:

- to compare the impact of the new reconstruction algorithms on quantitative measurements with the standard algorithm;
- to assess the performance of these reconstruction algorithms on objects with a range of contrast;
- to assess the choice of region method used to derive the activity concentration measurement;
- to assess whether the new reconstruction algorithms provided beneficial outcomes regarding accuracy or variability of measurements;

PET images were acquired using a standard phantom, known as the NEMA image quality phantom [68], which is comparable in cross section to a slim torso and

consists of six spheres, with diameters ranging from 10 mm to 37 mm as shown in Figure 1.2.

Figure 1.2 NEMA image quality phantom showing fillable spheres.

The compartment surrounding the spheres can be filled and images were acquired on the Biograph with the spheres filled with three different contrast levels of 2:1, 4:1 and 8:1. A specific aim here was to demonstrate whether the advanced algorithms provided any advantage in the low contrast objects. Three region-based methods were evaluated when deriving the activity concentration measurements of the spheres. Two were chosen to reflect measurements in clinical practice: a measurement derived from the maximum voxel as used for SUV_{max} and a "peak" measure as defined according to the SUV_{peak} measure proposed in the PERCIST protocol [57]. The third region was purely geometric, including all voxels within the region with equal diameter to the sphere. This final region technique is akin to the method that is prescribed by the NEMA test protocol [68], which is a technique that allows for a standardised comparison of PET systems from various manufacturers.

As well as assessing the accuracy, multiple images were created of each phantom allowing the variability of each measurement to be assessed. This is an important factor as it allows a degree of confidence to be assigned to single measurements for clinical applications of staging and grading and for serial measurements for assessing response to therapy where a change is indicative of response.

1.8.2 Second paper: Impact of point spread function modelling and time-of-flight on FDG uptake measurements in lung lesions using alternative filtering strategies.

This second piece of work was performed in 2013, following a software update to the Biograph that enabled an additional reconstruction of TOF without the resolution modelling, resulting in three alternative reconstruction options compared with the traditional algorithm: TOF, HD and UHD. From findings of the first piece of work, it was clear that the use of resolution modelling (HD and UHD) resulted in a significant increase in SUV_{max} in small objects. It was anticipated that this would translate into clinical imaging, resulting in significant changes in interpretation of lesions if using the existing local agreement for SUV_{max}. The TOF reconstruction had not been evaluated during the phantom study and so no knowledge of its impact relative to the standard iterative algorithm was available.

The work evaluated the impact of these three advanced reconstructions relative to the standard algorithm with clinical data. Two alternative objective criteria were defined based upon clinically relevant requirements. These two criteria were visual image noise, with the assumption that image appearance should be comparable, and SUV_{max} quantification. For both alternative criteria, the intention was to establish reconstruction parameters for the TOF, HD and UHD algorithms that provided a "matching" of each criterion using the standard algorithm as a reference.

The aims of the work were:

- to establish two sets of reconstruction parameters for each algorithm to match each of the two criteria described using a phantom image;
- to apply the parameters derived from the phantom data to FDG oncology clinical images to assess the validity of this technique;
- to evaluate several FDG uptake measurements in lesions for each of the two discrete sets of reconstruction parameters;
- to evaluate any benefits of either of the matched images by assessing the signal-to-noise ratio of lesions.

The matching was performed using the application of low-pass Gaussian filter kernels to NEMA phantom images, with the full-width-half-maximum (FWHM) of the kernel being adjusted. Given that, unlike patients, the phantom is a fixed size, it was not apparent whether the filter kernel choices established from phantom data could be translated to clinical imaging to provide the same results.

It was anticipated that substantially different Gaussian filter kernel widths would be required for each of the advanced reconstruction algorithms to match a given criterion. From the first study, the use of resolution modelling was shown to reduce noise in the background compartment of the phantom. It was expected that to match image noise for the HD and UHD reconstructions, a relatively narrow Gaussian kernel would be required, resulting in images with a relatively consistent appearance but a likely increase in SUV_{max} in smaller lesions. It was expected that to match image SUV_{max} for the HD and UHD reconstructions, a relatively broad Gaussian kernel would be required, resulting in images with a potentially "over-smoothed" appearance that may be unpopular with reporting clinicians.

In the clinical data, FDG uptake in lesions was quantified by SUV_{max}, SUV_{peak} and Total Lesion Glycolysis (TLG). TLG is the product of the mean activity concentration measured from a group of voxels and the total grouped volume of these voxels, and therefore has units of activity. Image noise was assessed in an area of homogenous FDG uptake within the liver by calculating the coefficient of variation of voxel values within a spherical region.

1.8.3 Third paper: The assessment of time-of-flight on image quality and quantification with reduced administered activity and scan times in FDG PET.

Over the course of this PhD, the range of clinical indications for patients undergoing PET scans with FDG at CMUH has expanded, which has generated an increasing challenge of scheduling patients on the camera in the available time. The two indications accounting for the increased use on the scanner are imaging of suspected implanted cardiac device infection and large vessel vasculitis. Both of these scans employ a dual-time point imaging protocol, with images acquired at 90 minutes and 180 minutes post-injection of FDG, with vasculitis studies also requiring a vertex to

feet scan at the 90 minute scan. Combining this with the need to comply with the requirements of the NHS PET/CT oncology service contract [67], there became a strong need to increase scanner throughput to meet demand. For all of these indications, the existing imaging protocol at CMUH defined a prescribed radioactivity of 350 MBq and 400 MBq of FDG for patients below and above 100 kg body weight respectively.

The expected increase in the number of patients undergoing FDG PET/CT scans also put emphasis on the need to address increases in radiation doses to the nuclear medicine staff. The dose rates to staff from FDG patients are approximately 10-fold that of patient administered a comparable level of a technetium-99m radiotracer commonly used for gamma camera imaging [69].

To address these needs, the third piece of work, performed in 2014, aimed to evaluate expected reductions of image noise with TOF reconstruction to allow for a reduction in imaging time and also administered activity. A second requirement was that SUV_{max} was conserved with any change in protocol to maintain the continuity of image interpretation. From the second piece of work, the levels of post-filtering necessary for the reconstructions with resolution modelling to maintain SUV_{max}, produced images with an appearance that was not acceptable to our clinicians. As such, the use of resolution modelling was not considered in this work. A software upgrade to the system allowed the PET image from each bed position to be reconstructed using a fraction of the acquired data or counts.

There were four main aims to the work:

- to evaluate the validity of using reduced-count reconstructions as a surrogate for a reduction in the administered activity;
- to evaluate the reductions in image noise observed with TOF reconstruction using clinical reconstruction parameters compared with the current standard iterative reconstruction;
- to conserve SUV_{max} for lesions in the reduced-count TOF images compared with the full-count data;
- to produce a more consistent level of image quality across the patient population scanned at CMUH.

This would allow the reduced-count data to be reconstructed with the TOF algorithm and compared with the standard non-TOF iterative algorithm that had been used routinely. As a metric of image quality, the signal-to-noise ratio in the liver (ratio of mean to standard deviation of voxel values within a region) was used. This has been shown to be a good surrogate for qualitative visual image quality [70].

1.8.4 Fourth paper: Evaluation of the utility of estimated covariance kernels for predicting regional ensemble variance

Changes in SUV over two or more PET scans acquired either after or through the course of cancer therapy are used as an assessment of response to the therapy. It appears that SUV_{max} is used predominantly for this purpose [57] but an alternative metric was suggested as part of the PERCIST protocol, SUV_{peak}. Whereas SUV_{max} is derived from the single maximum voxel value within a tumour, SUV_{peak} is derived from the mean value of voxels contained within a 1.0 ml spherical region. To assess the reliability of any treatment-related changes, an appreciation of the inherent variability of these two metrics is required. The choice of reconstruction algorithm is likely to influence this variability considerably. SUV_{max} will be heavily influenced by the voxel variance, while the voxel averaging in SUV_{peak} will reduce this sensitivity but introduce a greater dependence on ensemble noise properties. The ensemble noise, or Ensemble Variance (EV), is the variability of a group of voxels

across multiple images and is related to the voxel covariance. It has been shown that resolution modelling reduces the voxel variance, giving the visual impression of reduced noise [71], which was shown to reduce the variation of SUV_{max} considerably in the first piece of work in this project. This voxel variance reduction is achieved primarily through increased voxel covariance (or correlation). Increased correlation leads to a greater EV [72], particularly in small regions [73] such as those used for SUV_{peak}. It has been suggested that resolution modelling is detrimental and should not be used for such application in PET [72, 74]. There is therefore a need to investigate the impact of voxel correlation on ensemble variance with clinically relevant reconstruction parameters.

This piece of work focuses on the assessment of voxel correlation and the impact of EV and expands from work presented at the IEEE Molecular Imaging Conference in October 2015 [75]. Two reconstruction algorithms were investigated: iterative reconstruction with and without resolution modelling. It was hypothesized from theory that a voxel covariance kernel could be used to predict the EV of the region shapes that were evaluated. Covariance kernels are acquired from uniform areas within an image, which are not typically representative of the heterogeneous radiotracer distribution within a tumour. To be useful in clinical imaging, it would also be necessary to generate a covariance kernel from an area of homogeneous FDG uptake, such as the liver, on a single wholebody image.

There were four main aims to the work:

- to assess the accuracy of EV estimations, in a range of region shapes and sizes, produced from voxel correlation data, using measured EV across multiple image replicates;
- to compare the impact of reconstructions with and without resolution modelling on EV;
- to compare relative changes in EV, with and without resolution modelling, within homogenous background areas of phantoms with regional high contrast areas;
- to compare the correlation kernels generated across multiple replicate images with those generated from a limited number of sub-samples extracted from a single image.

The work used simulation data and also 200 replicate images from a Ge-68 uniform phantom to measure the EV of several different regions shapes with varying sizes. Covariance kernels were generated from cubic sub-samples from the data. The formation of each kernel initially used sub-samples from all 200 samples. Given the aim of deriving a covariance kernel from a limited number of sub-samples extracted from a single clinical image, the robustness of the covariance kernels generated from fewer images was assessed. The measurement of EV in a background region is of limited clinical importance and so data from a NEMA phantom were included, with EV measured in the background and also in each of the six hot spheres across multiple image replicates. Finally, covariance kernels were generated from four clinical FDG oncology scans, using sub-samples extracted from the liver.

Development of a new test object for modern PET scanners

In addition to the four papers stated above, Ian Armstrong and Heather Williams have been collaborating with Leeds Test Objects on the development of a new test phantom for PET systems with advanced reconstruction algorithms. The evaluation has used analysis methods that have been developed as part of this PhD project. The phantom design is based around the existing NEMA phantom with the option to replace the standard six spheres (diameters ranging from 10 mm to 37 mm) with

smaller spheres (diameters ranging from 7 mm to 28 mm). The phantom aims to fully establish performance gains of the ever-improving modern PET systems. Initial findings from the work have been presented at the British Nuclear Medicine Society Annual Meeting in 2012 [76], winning first prize for posters, and European Associate of Nuclear Medicine Annual Congress in 2014 [77].

1.9 References

1, Office for National Statistics. *Deaths registered in England and Wales (Series DR): 2014*. ONS, 2015.

2, Hanahan D, Weinberg RA. *The hallmarks of cancer*. Cell 2000; **100**:57-70

3, Cancer Research UK. *How can cancer kill you*? http://www.cancerresearchuk.org/about-cancer/cancers-in-general/cancer-questions/how-can-cancer-kill-you. Accessed 1st September 2016

4, Kim J-w, Dang CV. *Cancer's Molecular Sweet Tooth and the Warburg Effect*. Cancer Res 2006; **66**:8927-30

5, Warburg O, Wind F, E N. *The Metabolism of Tumors in the Body*. J Gen Physiol 1927; **8**:519-530

6, Warburg O. *On the Origin of Cancer Cells*. Science 1956; **123**:309-314

7, Sweet WH. *The uses of nuclear disintegration in the diagnosis and treatment of brain tumor*. N Engl J Med 1951; **245**:875-878

8, Wrenn FR, Good ML, Handler P. *The Use of Positron-emitting Radioisotopes for the Localization of Brain Tumors*. Science 1951; **113**:525-527

9, Burnham CA, Brownell GL. *A Multi-Crystal Positron Camera*. IEEE Trans Nucl Sci 1972; **19**:201-205

10, Phelps ME, Hoffman EJ, Mullani NA, Ter-Pogossian MM. *Application of Annihilation Coincidence Detection to Transaxial Reconstruction Tomography*. J Nucl Med 1975; **16**:210-224

11, Hoffman EJ, Phelps ME, Mullani NA, Higgins CS, Ter-Pogossian MM. *Design and Performance Characteristics of a Whole-Body Positron Transaxial Tomograph*. J Nucl Med 1976; **17**:493-502

12, Raichle ME, Larson KB, Phelps ME, Grubb RL, Jr., welch MJ, Ter-Pogossian MM. *In vivo measurement of brain glucose transport and metabolism employing glucose- -11C*. Am J Physiol 1975; **228**:1936-1948

13, Hawkins R, Hass WK, Ransohoff J. *Measurement of regional brain glucose utilization in vivo using [2(-14)C] glucose*. Stroke 1979; **10**:690-703

14, Sokoloff L, Reivich M, Kennedy C, Des Rosiers MH, Patlak CS, Pettigrew KD, et al. *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**:897-916

15, Nutt R. *The History of Positron Emission Tomography*. Mol Imag Biol 2002; **4**:11-26

16, Gallagher BM, Ansari A, Atkins H, Casella V, Christman DR, Fowler JS, et al. *Radiopharmaceuticals XXVII. 18F-Labeled 2-Deoxy-2-Fluoro-D-Glucose as a Radiopharmaceutical for Measuring Regional Myocardial Glucose Metabolism In Vivo: Tissue Distribution and Imaging Studies in Animals*. J Nucl Med 1977; **18**:990-996

17, Beyer T, Czernin J, Freudenberg LS. *Variations in Clinical PET/CT Operations: Results of an International Survey of Active PET/CT Users*. J Nucl Med 2011; **52**:303-310

18, Huang SC, Phelps ME, Hoffman EJ, Sideris K, Selin CJ, Kuhl DE. *Noninvasive determination of local cerebral metabolic rate of glucose in man*. Am J Physiol 1980; **238**:E69-E82

19, Wu H-M, Hoh CK, Huang S-C, Yao W-J, Phelps ME, Hawkins RA. *Quantification of Serial Tumor Glucose Metabolism*. J Nucl Med 1996; **37**:506-513

20, Kole AC, Nieweg OE, Pruim J, Paans AMJ, Plukker JTM, Hoekstra HJ, et al. *Standardized Uptake Value and Quantification of Metabolism for Breast Cancer Imaging with FDG and L-[1-11C]Tyrosine PET*. J Nucl Med 1997; **38**:692-696

21, Eary JF, Mankoff DA. *Tumor Metabolic Rates in Sarcoma Using FDG PET*. J Nucl Med 1998; **39**:250-254

22, Kole AC, Nieweg OE, Hoekstra HJ, van Horn JR, Koops HS, Vaalburg W. *Fluorine-18-Fluorodeoxyglucose Assessment of Glucose Metabolism in Bone Tumors*. J Nucl Med 1998; **39**:810-815

23, 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**:1-7

24, Sadato N, Tsuchida T, Nakaumra S, Waki A, Uematsu H, Takahashi N, et al. *Non-invasive estimation of the net influx constant using the standardized uptake value for quantification of FDG uptake of tumours*. Eur J Nucl Med 1998; **25**:559-564

25, Huang SC. *Anatomy of SUV*. Nucl Med Biol 2000; **27**:643-646

26, Som P, Atkins HL, Bandoypadhyay D, Fowler JS, MacGregor RR, Matsui K, et al. *A Fluorinated Glucose Analog, 2-fluoro-2-deoxy-D-glucose (F-18): Nontoxic Tracer for Rapid Tumor Detection*. J Nucl Med 1980; **21**:670-675

27, Kubota K, Matsuzawa T, Fujiwara T, Ito M, Hatazawa J, Ishiwata K, et al. *Differential Diagnosis of Lung Tumor with Positron Emission Tomography: A Prospective Study*. J Nucl Med 1990; **31**:1927-1932

28, Kenney JM, Marinelli LD, Woodard HQ. *Tracer Studies with Radioactive Phosphorus in Malignant Neoplastic Disease*. Radiology 1941; **37**:683-690

29, Fukuda H, Ito M, Matsuzawa T, Abe Y, Yoshioka, Kubota K *et al*. *Clinical Evaluation of Cancer Diagnosis with ^{18}F-2-fluoro-D-glucose. Its usefulness in Liver and Pancreas Cancers*. CYRIC Annual Report, Tohoku University 1983 244-250

30, Kubota K, Matsuzawa T, Ito M, Ito K, Fujiwara T, Abe Y, et al. *Lung Tumor Imaging by Positron Emission Tomography Using C-11 L-Methionine*. J Nucl Med 1985; **26**:37-42

31, Boellaard R, Delgado-Bolton R, Oyen WG, Giammarile F, Tatsch K, Eschner W, et al. *FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0*. European Journal of Nuclear Medicine and Molecular Imaging 2015; **42**:328-354

32, Dahlbom M, Hoffman EJ, Hoh CK, Schiepers C, Rosenqvist G, Hawkins RA, et al. *Whole-Body Positron Emission Tomography: Part I. Methods and Performance Characteristics*. J Nucl Med 1992; **33**:1191-1199

33, Gupta NC, Frank AR, Dewan NA, Redepenning LS, Rothberg ML, Mailliard JA, et al. *Solitary pulmonary nodules: detection of malignancy with PET with 2-[F-18]-fluoro-2-deoxy-D-glucose*. Radiology 1992; **184**:441-444

34, Dewan NA, Gupta NC, Redepenning LS, Phalen JJ, Frick MP. *Diagnostic Efficacy of PET-FDG Imaging in Solitary Pulmonary Nodules: Potential Role in Evaluation and Management*. Chest 1993; **104**:997-1002

35, Lowe VJ, Hoffman JM, DeLong DM, Patz EF, Coleman RE. *Semiquantitative and Visual Analysis of FDG-PET Images in Pulmonary Abnormalities*. J Nucl Med 1994; **35**:1771-1776

36, Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB, et al. *Staging non—small cell lung cancer by whole-body positron emission tomographic imaging*. Ann Thorac Surg 1995; **60**:1573-1582

37, Carr R, Barrington SF, Madan B, O'Doherty MJ, Saunders CAB, van der Walt J, et al. *Detection of Lymphoma in Bone Marrow by Whole-Body Positron Emission Tomography*. Blood 1998; **91**:3340-3346

38, Delbeke D, Martin WH, Sandler MP, Chapman WC, Wright, Jr J, et al. *Evaluation of benign vs malignant hepatic lesions with positron emission tomography*. Arch Surg 1998; **133**:510-516

39, Patz EF, Lowe VJ, Hoffman JM, Paine SS, Burrowes P, Coleman RE, et al. *Focal pulmonary abnormalities: evaluation with F-18 fluorodeoxyglucose PET scanning*. Radiology 1993; **188**:487-490

40, Hain SF, Curran KM, Beggs AD, Fogelman I, O'Doherty MJ, Maisey MN. *FDG-PET as a "metabolic biopsy" tool in thoracic lesions with indeterminate biopsy*. Eur J Nucl Med 2001; **28**:1336-1340

41, Beggs A, Hain S, Curran K, O'Doherty M. *FDG-PET as a "metabolic biopsy" tool in non-lung lesions with indeterminate biopsy*. Eur J Nucl Med Mol Imaging 2002; **29**:542-546

42, Kwee T, Cheng G, Lam MEH, Basu S, Alavi A. *SUVmax of 2.5 should not be embraced as a magic threshold for separating benign from malignant lesions*. Eur J Nucl Med Mol Imaging 2013; **40**:1475-1477

43, Hoffman EJ, Huang SC, Phelps ME. *Quantitation in positron emission computed tomography: 1. Effect of object size*. J Comput Assist Tomogr 1979; **3**:299-308

44, Smith WJM. *Medicare to reimburse for lung cancer PET imaging*. J Nucl Med 1998; **39**:22N,25N

45, Maisey M. *The Introduction and Development of Clinical PET in the United Kingdom*. In: McCready R, Gnanasegaran G, Bomanji J. *A History of Radionuclide Studies in the UK: 50th Anniversary of the British Nuclear Medicine Society*. Springer, 2015

46, Beyer T, Townsend DW, Brun T, Kinahan PE, Charron M, Roddy R, et al. *A Combined PET/CT Scanner for Clinical Oncology*. J Nucl Med 2000; **41**:1369-1379

47, Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. *AJCC cancer staging manual (7th ed)*. Springer 2010

48, Siewert JR, Ott K. *Are Squamous and Adenocarcinomas of the Esophagus the Same Disease?* Semin Radiat Oncol 2007; **17**:38-44

49, Yano M, Yoshida J, Koike T, Kameyama K, Shimamoto A, Nishio W, et al. *The Outcomes of a Limited Resection for Non-Small Cell Lung Cancer Based on Differences in Pathology*. World J Surg 2016:1-10

50, Higashi K, Ueda Y, Sakurai A, MingWang X, Xu L, Murakami M, et al. *Correlation of Glut-1 glucose transporter expression with [^{18}F]FDG uptake in non-small cell lung cancer*. Eur J Nucl Med Mol Imaging 2000; **27**:1778-1785

51, Cerfolio RJ, Bryant AS, Ohja B, Bartolucci AA. *The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival*. J Thorac Cardiovasc Surg 2005; **130**:151-159

52, de Geus-Oei L-F, Krieken JHJMv, Aliredjo RP, Krabbe PFM, Frielink C, Verhagen AFT, et al. *Biological correlates of FDG uptake in non-small cell lung cancer*. Lung Cancer 2007; **55**:79-87

53, Shimizu K, Hirami Y, Saisho S, Yukawa T, Maeda A, Yasuda K, et al. *Maximal Standardized Uptake Value on FDG-PET Is Correlated With Cyclooxygenase-2 Expression in Patients With Lung Adenocarcinoma*. Ann Thorac Surg 2012; **93**:398-403

54, World Health Organisation. *WHO Handbook for Reporting Results of Cancer Treatment*. World Health Organisation Publication No. 48, 1979

55, Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. *New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)*. Eur J Cancer 2009; **45**:228-247

56, Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. *Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group*. Eur J Cancer 1999; **35**:1773-1782

57, 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**:122S-150S

58, Ghosh PR. *The International State of PET*. J Nucl Med 1991; **32**:28N,32N-33N,51N

59, Department of Health. *A Framework for the Development of Positron Emission Tomography Services in England*. Department of Health, 2005

60, Bedford M,. Maisey M. *Requirements for clinical PET: comparisons within Europe*. Eur J Nucl Med Mol Imaging 2004; **31**:208-221

61, Intercollegiate Standing Committee on Nuclear Medicine. *Positron emission tomography. A strategy for provision in the UK*. Intercollegiate Standing Committee on Nuclear Medicine 2003

62, National Institute for Clinical Excellence. *Lung cancer. The diagnosis and treatment of lung cancer*. NICE Clinical Guideline 24; 2005

63, The Royal College of Radiologists. *PET-CT in the UK. A strategy for development and integration of a leading edge technology within routine clinical practice*. The Royal College of Radiologists, 2005

64, Credo Consulting. *A White Paper investigation into the proposed commissioning of new PET-CT services in England*. Credo Consulting, April 2014

65, The Royal College of Radiologists. *Evidence-based indications for the use of PET-CT in the United Kingdom 2016*. The Royal College of Radiologists 2016

66, NHS Commissioning Board. *Clinical Reference Groups for Specialised Services: A Guide for Clinicians*. NHS England 2013

67, NHS England. *2013/14 NHS Standard Contract for Positron Emission Tomograph – Computed Tomograph (PET_CT) Scanning (All Ages)*. B02/s/a. NHS England, 2013

68, NEMA. *NEMA Standards Publication NU 2-2007: Performance Measurements of Positron Emission Tomographs*. NEMA, 2007

69, Harding LK, Mostafa AB, Roden L, Williams N. *Dose rates from patients having nuclear medicine investigations*. Nucl Med Commun 1985; **6**:191-194

70, McDermott G, Chowdhury F, Scarsbrook A. *Evaluation of noise equivalent count parameters as indicators of adult whole-body FDG-PET image quality*. Ann Nucl Med 2013; **27**:855-861

71, Alessio AM, Stearns CW, Shan T, Ross SG, Kohlmyer S, Ganin A, et al. *Application and Evaluation of a Measured Spatially Variant System Model for PET Image Reconstruction*. IEEE Trans Med Imaging 2010; **29**:938-949

72, Rahmim A, Qi J, Sossi V. *Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls*. Med Phys 2013; **40**:064301-064315

73, Blinder SAL, Dinelle K, Sossi V. *Scanning rats on the high resolution research tomograph (HRRT): A comparison study with a dedicated micro-PET*. Med Phys 2012; **39**:5073-5083

74, Alessio AM, Rahmim A, Orton CG. *Resolution modeling enhances PET imaging*. Med Phys 2013; **40**:120601

75, Armstrong IS, Williams HA, Matthews JC. *Evaluation of the utility of estimated covariance kernels for predicting regional ensemble variance*. 2015 IEEE Nuclear Science Symp Conf Record, 2015

76, Armstrong IS, Williams HA. *A new "extended NEMA" image quality phantom to illustrate the impact of Time-of-Flight (TOF) PET*. Nuc Med Commun 2012; **33**:P11

77, Armstrong IS, Williams HA. *A new PET test phantom to evaluate performance of modern PET systems*. Eur J Nucl Med Mol Imaging 2014; **41**(Suppl 2):S377

2. PRINCIPLES OF POSITRON EMISSION TOMOGRAPHY

2.1 Overview

The discovery of the positron (β^+) in 1933 by Carl Anderson [1] gave birth to the potential of PET imaging. The positron is a particle with equal mass to an electron but with a positive charge. Like electrons, a positron will travel along a tortuous path through matter, losing energy as it interacts with atoms within the matter. These interactions are with orbital electrons of the atoms of matter and can lead to excitation and ionisation of the atoms and molecules. The positron travels a finite distance in matter before it annihilates with an electron from one of the atoms in the matter. This *annihilation* produces two gamma-rays, each with 511 keV of energy, that are produced simultaneously and travel in almost opposite directions. They can be detected using a pair of opposing gamma-ray detectors using the time of detection to associate the pair with a common annihilation event and hence form the LOR. By collecting many LORs over a range of angles and positions, one can determine the distribution and biokinetics of a radiotracer in the patient [2].

A *tracer* is a substance that is designed to mimic endogenous compounds within the body. This allows functional processes to be measured whilst not interfering with the underlying system function. A *radiotracer* is a tracer that has been chemically labelled with a radionuclide. PET is a non-invasive radionuclide imaging technique that determines the distribution of a radiotracer within a patient. In PET, radiotracers are labelled with positron-emitting radionuclides. Some radiotracers have identical chemical structure to endogenous compounds. An example of such a radiotracer is [^{15}O]water, where the stable oxygen atom is replaced with oxygen-15. This emits positrons with a mean energy of 735 keV and has a half-life of two minutes. Other radiotracers are compounds that differ chemically from endogenous molecules but mimic their behaviour. Table 2.1 shows characteristics of a selection of positron-emitting radionuclides and radiotracers that are used in PET imaging. PET is a functional imaging modality reflecting biological processes rather than anatomical images, such as those produced by X-ray imaging.

RN	β^+ energy (keV) Max	Mean	$T_{1/2}$ (min)	Production	BF	Example radiotracers
C-11	960	386	20	Cyclotron	100%	[^{11}C]Methionine
N-13	1190	492	10	Cyclotron	100%	[^{13}N]Ammonia
O-15	1720	735	2.0	Cyclotron	100%	[^{15}O]Water
F-18	635	250	109	Cyclotron	97%	[^{18}F]FDG
Ga-68	1899	836	68	Generator	88%	[^{68}Ga]Peptide agents
Rb-82	3350	1535	1.25	Generator	82%	[^{82}Rb]Rb-Cl

Table 2.1 Characteristics of positron-emitting radionuclides that are commonly used in PET imaging. RN: radionuclide; $T_{1/2}$: physical half-life; BF: positron decay branching fraction.

2.2 Gamma-ray interactions

After a positron-emitting radiotracer has been administered to the patient, many pairs of 511 keV gamma-rays will be produced from positron-electron annihilation. These gamma-rays can interact with the tissue in the patient, primarily with orbital electrons of the tissue atoms and fall into two categories: photoelectric absorption and Compton scattering. During Photoelectric absorption, a gamma-ray will transfer all of its energy to an orbital electron, which is ejected from the atom and the gamma-ray ceases to exist. For photons with energy above the highest atomic electron binding energy, the probability of photoelectric absorption is approximately proportional to Z^3/E^3, where Z is the atomic number of the matter and E is the energy of the gamma-ray [3]. During Compton scatter the gamma-ray transfers only a portion of its energy to the orbital electron, which is ejected. The gamma-ray energy is reduced and, due to conservation of momentum, the direction of travel of the gamma-ray is changed. The energy of the scattered gamma-ray, E_{sc}, for a gamma-ray of initial energy E_γ that is scattered by an angle of θ is expressed as [4]

$$E_{sc} = \frac{E_\gamma}{1 + \left(E_\gamma / 0.511\right)\left(1 - \cos\theta\right)} \tag{2.1}$$

This scattered gamma-ray may undergo no further interaction with the tissue, additional Compton scatter or photoelectric absorption. By far the most dominant type of interaction for the 511 keV gamma-rays in tissue is Compton scattering [4].

2.3 PET detector design

The PET detection process for gamma-rays being emitted from the patient uses an array of detectors that are arranged in a ring in modern systems. The PET detector is bombarded by millions of individual gamma-rays every second. The array of detectors is connected to a coincidence processing unit which performs the coincidence event association. If two 511 keV gamma-rays are detected at the same time they are associated with each other and the event is known as a coincidence event. Individual gamma-ray events that are detected but not associated with a paired event are referred to as *single events*. Due to the low geometric detector efficiency in comparison to the patient size, they constitute the majority of events.

The coincidence processing unit uses a defined time duration, which is referred to as the *coincidence timing window*. When one singles event is detected, the coincidence timing window is opened and one or more singles events may be detected before the window is closed. If two gamma-rays are detected within the coincidence window, the event is known as a *prompt* or coincidence event [5], and defines the LOR. It is possible that multiple, i.e. > 2, gamma-ray detections occur within the coincidence window. In these situations the true LOR cannot be determined and the event is rejected.

Gamma-ray detection is typically performed using *scintillation detectors* [6], which comprise of two principal components, a *scintillation crystal* and a *photomultiplier tube* (PMT). The scintillation crystal absorbs the gamma-ray energy by either photoelectric absorption or Compton scatter. As gamma-ray energy is lost, optical *scintillation photons* are produced within the crystal. The number of scintillation photons produced is proportional to the energy lost by the gamma-ray. A common scintillation material is sodium iodide, doped with thallium (NaI(Tl)), which will produce approximately 19,000 optical photons when a 511 keV gamma-ray is completely absorbed [6] and was used in the first PET systems [7].

The scintillation photons are detected by a PMT that is coupled to the crystal, sometimes via a light-guide. The front surface of a PMT is called a photocathode and converts the incident scintillation photons from the crystal to photoelectrons via the photoelectric effect. The number of photoelectrons produced is proportional to the

number of incident scintillation photons. These photoelectrons are accelerated towards an electrode within a vacuum, called a dynode, which has a positive voltage applied. When colliding with the dynode, the increase in kinetic energy results in the ionisation and displacement of a greater number of electrons into the vacuum. A series of dynodes in a chain with increasing positive voltage applied results in a cascade of ever increasing number of electrons with the electrons emitted by the final dynode transferring a charge to an anode connected to a pre-amplifier, which is read as a voltage from the PMT. The magnitude of the voltage is proportional to the energy of the gamma-ray that was absorbed in the crystal and gave rise to the scintillation photons. The variability of the signal from a PMT follows Poisson statistics [8]. Therefore if a greater number of scintillation photons are incident on the photocathode, the output from the PMT will be more precise.

An energy calibration of the PMT output voltage is performed to allow the incident gamma-ray energy to be determined. An acceptance window, referred to as an *energy window* is defined so that only gamma-rays that are detected with an energy falling within the window are collected. The lower level discriminator (LLD) of the energy window is used to exclude low energy gamma rays that are likely to have lost energy through Compton scatter in the patient. While one would not expect to detect gamma-rays with energies greater than 511 keV, the upper level discriminator (ULD) of the energy window must be set to account for the finite energy resolution of the system due to factors such as the statistical nature of the PMT output. The ULD also rejects events that arise due to *pulse pile-up*. Pulse pile-up occurs when the scintillation photons in the crystal have not fully decayed before a new gamma-ray interaction occurs in the crystal. Consequently, the output voltage from the PMT will not have returned to zero and the output voltage from the new gamma-ray will be added to any residual output voltage from the previous gamma-ray interaction. The resultant summed output voltage will infer that the apparent energy of the new gamma-ray is artificially high, falling above the ULD and will be rejected. Pulse pile-up is particularly problematic in systems that use scintillation crystals with long decay times. The Siemens Biograph mCT has an energy window of 435 keV to 650 keV [9].

The choice of scintillation crystal has a substantial impact on the performance of the PET detector. There are several factors to consider when evaluating the choice of scintillation crystal:

- *Light output*. The Poisson statistical nature of the PMT output means that a greater output of incident scintillation photons from the crystal improves the precision of the PMT output. The PMT output is proportional to gamma-ray energy and hence energy resolution is improved with crystals that have a greater light output. A crystal that produces a greater number of scintillation photons per unit of volume allows smaller crystals to be fabricated while maintaining acceptable light output [10]. A smaller crystal allows for a greater number of LORs and hence improved spatial resolution.
- *Density*. As discussed previously, the probability of photoelectric absorption, and hence the stopping power for the gamma-rays, is proportional to Z^3/E^3. A crystal with a greater density will increase the stopping power and hence detector efficiency. Despite this, Compton scattering is still the dominant interaction for 511 keV gamma-rays [4].
- *Decay time*. The time that the scintillation photons take to decay within the crystal heavily dictates the temporal length of the coincidence timing window. It also influences the count-rate capability, from pulse pile up, and temporal resolution of detectors.
- *Emission wavelength*. The efficiency of a PMT photocathode is dependent on the scintillation photon wavelength. It is therefore advantageous that this wavelength is close to this optimal efficiency.
- *Mechanical robustness*. Crystals should ideally be robust and easy to manufacture into small pieces. Some materials are hygroscopic, requiring encapsulation, which is not ideal.

Data for several materials used for scintillation crystals in PET systems are shown in Table 2.2.

	Relative light output (%)	Decay time (ns)	Density (g cm^{-3})	Peak wavelength (nm)	Rugged	Hygroscopic
NaI(Tl)	100	230	3.67	415	No	Yes
BGO	15	300	7.13	480	Yes	No
LSO	75	40	7.40	420	Yes	No
GSO	40	50	6.71	430	Yes	No
$LaBr_3$	175	15	5.29	360	No	Yes
CsF	6	2.5	4.61	390	No	Very
BrF_2	5	0.6	4.89	210	Yes	No

Table 2.2 Physical characteristics of scintillation crystals that have been or are used in PET detectors [11–13].

The basic design of the detector in all modern PET scanners is fairly similar, consisting of several rings of scintillation detectors stacked together along the length of the patient to increase the size of area that can be imaged (known as the axial *field of view* (FOV)). The length of the axial FOV along the patient ranges between 15 and 25 cm on various modern PET scanners. Each detector element can detect in coincidence with a number of detectors on the opposing side of the ring. The spatial separation of LOR is determined by the size of the scintillation detectors.

2.4 Development of PET hardware

The first PET scanner was built by Michael Phelps and Ed Hoffmann, with results published in 1975 [7]. This early system used large NaI(Tl) for the scintillation crystal material which, at the time, was the only available suitable material [14]. The system consisted of twenty-four 5.1 cm NaI(Tl) detectors arranged in a hexagonal array. Each set of four detectors was in coincidence with the opposite four with a coincidence timing window of 30 ns.

Sodium iodide however does not make the best scintillator for PET due primarily to its low density and hence detection efficiency of 511 keV gamma-rays (Table 2.2). In 1977 an alternative scintillation crystal, bismuth germinate (BGO), was evaluated [15]. BGO has a notably greater density than NaI(Tl) and hence allowed for smaller

crystals yet provided the same levels of stopping power. This improved spatial resolution, whilst maintaining the same overall detection efficiency.

Taking advantage of the increased density of BGO, a new detector design was proposed in 1986, known as the *Block Detector* [16]. The design employed a single block of BGO divided into subsections by slots that are cut partially down the depth of the crystal within which light reflective material is packed – see Figure 2.1. This allowed the location of the gamma-ray absorption to be identified within one of these subsections of the single BGO crystal using a limited number of PMTs. Logic is performed on the PMT outputs mounted to the crystal to determine which crystal subsection the gamma-ray was absorbed in. The first example of this design used a 30 mm thick BGO crystal divided into 8×4 (5.6 mm × 13.5 mm respectively) subsections and was backed by four PMTs. While the number of crystal subsections has increased and manufacturing processes have progressed, nearly all modern PET scanners still use detectors with this same underlying design.

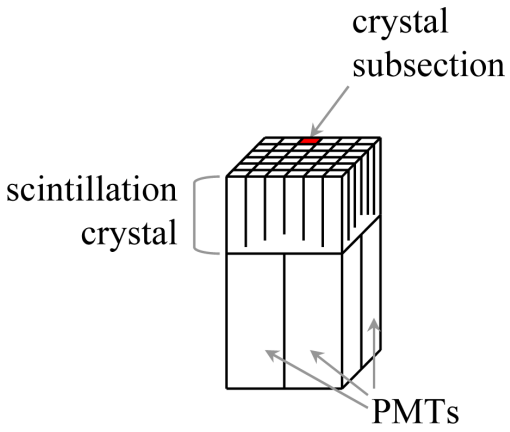


Figure 2.1 Block detector design with four photomultiplier tubes (PMTs) and a crystal subsection highlighted in red [16].

Despite its advantages over NaI(Tl), BGO had two significant negative aspects limiting its performance in PET. Firstly, the low light output degrades the energy resolution performance and limits the size of subsections, and hence the spatial resolution [10]. Secondly, the slow scintillation decay time limits the count-rate performance of a system due to count losses through pulse pile-up as described. It also inhibits the use of narrow coincidence timing windows giving rise to a greater number of random coincidences and also does not allow for time of flight data acquisition, which shall be discussed later in this review.

In 1992, a new scintillation material was discovered by Melcher and Schweitzer called lutetium oxyorthosilicate (LSO) [11]. The faster decay time of LSO allows for higher count rate systems and a reduction in the temporal width of the coincidence window, which is discussed later in this chapter. However, LSO does result in a modest reduction in stopping power compared with BGO and so the detection efficiency is slightly reduced [17].

The first commercial scanner with LSO crystals was introduced in 2001 by Siemens/CTI, with a coincidence timing window of 6 ns compared with 12 ns for the comparable system using BGO [18]. In 2005, performance data were published on a system with an updated LSO block design and improved electronics [19]. This "HI-REZ" detector developed by Siemens consisted of a LSO block divided into 13×13 subsections with dimensions 4×4×20 mm [20]. The block design employed on the PET/CT system at CMUH is very similar to that used in the original HI-REZ design.

2.5 Factors that degrade PET data

PET aims to locate the radiotracer and so the ideal LOR should consist of two gamma-rays that originate from the radiotracer location, escape from the patient without undergoing any interaction and are detected with 100% efficiency in the PET scanner. In reality, there are many factors that act to degrade the collection of PET data. These are due to four main categories: the physics of the positron decay and annihilation, gamma-ray interactions, erroneous LORs and detector limitations.

2.5.1 Physics of positron decay and annihilation

Positron range

A positron that has been emitted from a radionuclide will travel a finite distance through matter due to its energy. The positron range is dependent on the initial energy and the electron density of the matter. The initial positron energy can be plotted as a probability function up to the maximum possible energy [21] and is specific to a given radionuclide. Combining the initial energy distribution with the random path, the path length from positron emission to annihilation forms a probability distribution that decreases approximately exponentially [22] for radionuclides with relatively low positron energy such as fluorine-18 but has a more

complex distribution for radionuclides with higher energy [23]. In soft tissue, the mean positron range is stated as 0.6 mm for fluorine-18, which has a mean positron energy of 248 keV and 5.9 mm for rubidium-82, which has a mean positron energy of 1551 keV [23, 24]. This means that the LOR indicates the position of the annihilation and not the radiotracer from which the positron originated as would be desired.

Gamma-ray acolinearity

When the positron annihilates with an electron it has a small amount of momentum remaining. This momentum is conserved following the annihilation and the gamma-rays do not travel exactly at 180° to one another and instead have a variation of ±0.25° [25]. From simple trigonometry, the FWHM of the positional uncertainty about the annihilation can be expressed as

$$FHWM = 0.0022 \times D, \tag{2.2}$$

where D is the diameter of the PET detector ring. This shows that acolinearity effects can be reduced by decreasing the diameter of the detector. However, for clinical scanners, there is clearly a minimum acceptable diameter for accommodating the patient in the scanner.

2.5.2 Gamma-ray interactions

Gamma-ray attenuation

The interactions that gamma-rays undergo, as discussed previously, can lead to a loss of gamma-rays detected from the patient, which is known as *attenuation*. Attenuation is an exponential process, with the probability, P, of both gamma-rays emerging from matter that constitutes the patients body being

$$P = e^{-\int_{LOR} \mu(L)dL} \tag{2.3}$$

where L is the distance along the LOR and $\mu(L)$ is the *linear attenuation coefficient* (LAC) [4]. The LAC for 511 keV gamma-rays in water, which is approximately equivalent to soft tissue, is 0.096 cm^{-1} [26]. If the LOR passes through 40 cm of soft

tissue, which is not unlikely in the human body, the probability of both gamma-rays emerging from an object is 2.15%. Fewer gamma-rays from annihilations occurring within the centre of the patient will escape in comparison to annihilations towards the edge. This can be compensated for by measuring the attenuating properties of the patient, which will be covered in the next chapter.

2.5.3 Erroneous LORs

Ideally, every prompt coincidence event should represent the LOR that passes through the annihilation point of the positron within the patient that gave rise to the two detected 511 keV gamma-rays. In reality, there are three types of events that can constitute a prompt coincidence event:

- *True coincidence*. Two gamma-rays originating from the same annihilation pass through the patient and do not change direction through scattering so that the LOR intercepts the true origin of annihilation;
- *Scattered coincidence*. One or both of the gamma-rays originating from the same annihilation event undergo Compton scatter on one or more occasions within the patient and change direction such that the LOR is erroneous as it does not intercept the true origin of annihilation;
- *Random coincidence*. Two singles events from two unrelated annihilations are detected within the coincidence timing window leading to an erroneous LOR.

These three events are illustrated in Figure 2.2.

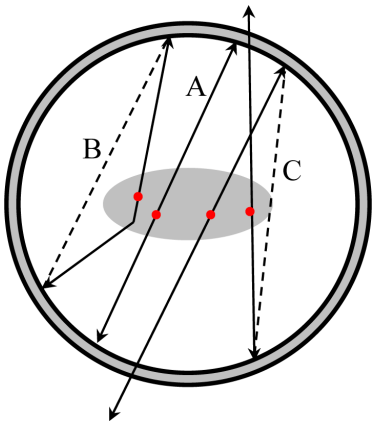


Figure 2.2 Different coincidence event types that are detected by a PET detector. A: a true coincidence event, B: a scattered event where the lower gamma-ray is scatter before leaving the patient and C: a random coincidence event arising from two independent annihilations. Positron annihilation locations are represented by red circles, gamma-rays paths by solid lines and erroneous LORs by dashed lines.

There is no temporal correlation between two annihilations that give rise to random coincidence events. As such, the temporal length of the coincidence timing window dictates the probability of the number of random events that are detected. It is dictated primarily by hardware limitations and processing time of the electronics (see section 2.3). The mean rate of random coincidences, R_{ij}, creating a LOR between two detectors i and j can be modelled as

$$R_{ij} = 2\tau \cdot r_i \cdot r_j \,, \tag{2.4}$$

where τ is the temporal width of the coincidence timing window and r_i and r_j are the singles detection rates of gamma-rays on two detectors i and j [25]. Hence the rate of random coincidence events will increase quadratically as radioactivity increases.

2.5.4 Detector limitations

Detector efficiency

Detection of radioactive events follows Poisson statistics. Consequently, if the number of detected events is low due to poor efficiency, the relative precision on the number of events is low. In images, this is manifested as statistical noise, as shown in Figure 2.3. To compensate, the time for the scan needs to be extended to collect sufficient events to provide a useful image.

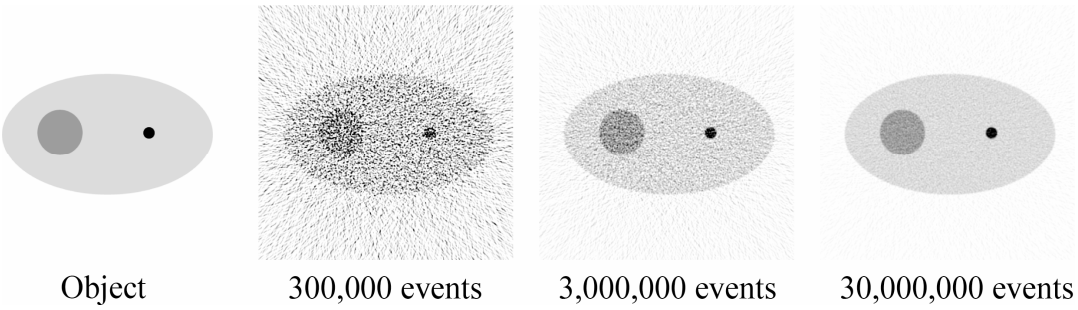


Figure 2.3 Illustration of the statistical noise observed in images produced from different levels of coincidence events.

Detector size

If one ignores the factors of positron range and acolinearity, the limiting factor on the spatial uncertainty of the LOR is the size of the detector elements. For a pair of detectors in coincidence, the annihilation event can be positioned anywhere within a column that is created by the two detector faces. As the detector becomes smaller, so does the column width and the localisation of the annihilation becomes more precise. Smaller detectors will enable a greater number of unique LORs and hence will be able to provide greater spatial sampling of the annihilation locations [15].

Pulse pile-up and count losses

Systems with slow scintillation crystals are susceptible to pulse pile-up. This limits the maximum count rate that a scanner can operate at. There are two degrading factors that result from pulse pile-up. The first is a loss of detected events when two or more gamma-rays are incident on a detector and the energy measurement falls above the ULD. The second is mis-positioning of events within a block detector. This occurs when multiple gamma-rays with reduced energy, typically due to scatter within the patient, are absorbed in different locations in the block and the energy measurement falls within the energy window. Here an average absorption position is incorrectly calculated from the PMT logic. This results in a loss of spatial resolution that has been shown to worsen with increasing activity in the FOV [27].

This factor is particularly problematic for dynamic studies that are used for kinetic analysis where count rates can be very high during the injection of the radiotracer.

Depth of interaction and inter-crystal scatter

For gamma-rays that are incident at a normal angle to the face of the scintillation crystal, the length of crystal required to totally absorb 50% of the gamma-rays is 10 mm and 11 mm for BGO and LSO scintillator crystals respectively [17]. For a gamma-ray that strikes a block detector at an oblique angle, the path inside the crystal subsection is likely to result in the gamma-ray passing into the adjacent subsection of the crystal.

The incident gamma-ray may not undergo photoelectric absorption within the scintillation crystal and instead may undergo Compton scatter. As such, there is a possibility that the scattered gamma-ray will leave the crystal subsection and pass through into an adjacent subsection before depositing sufficient energy to enable detection. Examples of depth of interaction and inter-crystal scatter are shown in figure 2.4.

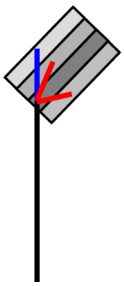


Figure 2.4 Representation of depth of interaction and inter-crystal scatter problem in a PET detector. The blue line illustrates un-scattered crystal penetration due to insufficient path length in the incident crystal subsection while the red lines illustrate inter-crystal scatter. The degree of greyscale on each crystal subsection is an indication of the proportion of events registered in each subsection after collecting multiple instances of the given LOR.

2.6 Advances in PET design

2.6.1 Development of PET/CT

PET images provide an excellent representation of the distribution of administered radiotracer, but lack anatomical features that would be useful to clinicians. This is particularly problematic in the abdominal region. In the early 1990s, some PET cameras did not consist of a complete ring of scintillation detectors around the patient [28]. In 1991 it was suggested by an oncology surgeon that the gaps within

the ring could be used to house an X-ray tube and detectors to enable acquisition of a CT image to provide complimentary anatomical images for the PET. This design concept was abandoned due to the infeasibility of mounting the CT detector on the rotating PET gantry. Instead a far simpler sequential design was developed with a CT detector gantry in front of a PET detector ring. Seven years later, the first prototype system with combined PET and CT detectors was produced [29, 30]. In 2001, the first commercial PET/CT system was produced (GE Discovery LS) and, by 2008, there were 2500 PET/CT systems in use worldwide [31]. By 2006, all commercial PET systems were sold with integrated CT. The CT image also provided a means of correcting for the attenuation of the gamma-rays, and will be discussed in the next chapter. The high speed CT acquisition of a wholebody anatomical image was a key factor in enabling routine clinical oncology scanning to be a reality. In the 2005 Framework, the UK Department of Health recommended that combined PET/CT systems should be the preferred choice of system [32].

2.6.2 Development of PET/MR

The additional radiation dose and lack of intricate fine detail within soft tissue associated with the CT component of a PET/CT has driven the development of PET/MR systems. The MR provides an anatomical image with no additional radiation dose. However, the high magnetic field causes the electrons that are accelerated from dynode to dynode in the PMTs to travel in helical paths and consequently the PMT no longer functions [33]. Early attempts to acquire PET data in the presence of an MR field employed either optical fibres to transfer the scintillation photons from the scintillation crystals to the PMTs that were some distance from the PET detector ring [34] or physical separation of the PET and MR detectors with a common imaging table [35]. Recently, however, there has been a move towards replacing PMTs with solid state detectors such as avalanche photo diodes (APDs) and silicone photomultipliers (SiPMs) [36–38]. Both of these detectors work in the presence of a magnetic field meaning integrated systems are possible. Unlike APDs, SiPMs have the ability to also provide time of flight information.

2.7 Time of flight PET

2.7.1 Overview

Following the positron annihilation, the two 511 keV gamma-rays travel at the speed of light. For annihilation events that are not on the midpoint of the LOR, the arrival time of the two gamma-rays at the PET detector ring will differ due to their differing path lengths. The time difference in arrival times, Δt, is given as:

$$\Delta t = \frac{2x}{c}, \tag{2.5}$$

where x is the distance from the midpoint of the LOR to the point of annihilation and c is the speed of light [39]. If the detector ring can measure the difference in arrival time with a temporal resolution with FWHM σ_t, it follows that the spatial uncertainty of the annihilation event, σ_x, will be:

$$\sigma_x = \frac{c\sigma_t}{2}. \tag{2.6}$$

PET scanners that measure and utilise this time difference with a temporal resolution that is less than the duration of the coincidence timing window are said to the *Time of Flight* (TOF) PET systems. If the measurement of the time difference was perfect, then one would know exactly where along the LOR the annihilation occurred. To locate an annihilation event to within 1 cm, the timing resolution would need to be approximately 70 ps. To put this into context, most current-generation PET/CT systems using PMTs with TOF capability operate with timing resolutions between 500 and 600 ps, translating to a positional uncertainty along the LOR of 7.5 to 9.0 cm. New state-of-the-art systems with solid-state detectors have timing resolutions of approximately 380 to 400 ps translating to a positional uncertainty along the LOR of 5.7 to 6.0 cm. [40].

The theory of TOF PET followed soon after the development of the first PET scanners. The long scintillation decay time of both NaI(Tl) (230 ns) and BGO (300 ns) prohibited both of these materials as choices for TOF PET. Early investigations into TOF PET used pairs of caesium fluoride (CsF) scintillation detectors, which have a 2.5 ns decay time [41]. This was followed on with TOF PET systems using

CsF [42, 43] and using barium fluoride (BaF) crystals which have a 0.6 ns decay time [44]. Despite demonstrating the potential of TOF PET, neither of these scintillation materials are ideal for incorporation into a PET system due to their low density and hence low stopping power for 511 keV gamma-rays and low light output [45].

Arguably the most important technological development for TOF PET was the discovery of LSO [11]. LSO provided a notably faster decay time compared with BGO and combined with high density and high light output. As such, this became the choice of crystal for one of the first wholebody TOF PET systems [46], with the first commercial system using LYSO, a variation on the LSO crystal, and had a timing resolution of approximately 600 ps [47]. The measured timing resolution of the Siemens Biograph mCT, equipped with LSO crystals, at CMUH is 530 ps [9] resulting in a positional uncertainty of the positron annihilation of approximately 8 cm. In late 2016, a new GE SIGNA PET/MR scanner will be installed at CMUH, which employs SiPM detectors and should provide a timing resolution of between 380 to 400 ps, which is a notable improvement on the Biograph mCT [40].

2.7.2 Benefits of TOF PET

The most commonly accepted benefit of TOF is a reduction in the variance of the data, which leads to an increase in *signal-to-noise ratio* (SNR) of the reconstructed images [48, 49]. This has been described by Budinger [39] and results in the following equation:

$$\frac{SNR_{TOF}}{SNR_{non-TOF}} = \sqrt{\frac{D}{\sigma_x}}, \tag{2.7}$$

where D is the diameter of the object in the PET FOV and σ_x is the spatial uncertainty of the annihilation event from equation 2.6.

As discussed previously, the gain in SNR with TOF increases with larger objects. It has been suggested that, for random coincidences, this variance reduction is even greater than expected. This is because the effective diameter of the distribution of random coincidences is considerably larger than the physical object in the FOV

[50, 51]. It has been demonstrated on a first-generation TOF PET system that the improvements in SNR with TOF may exceed those based purely on the object size, using equation 2.7, with high activity levels, and hence greater randoms, in the FOV [51].

2.8 References

1, Anderson CD. *The Positive Electron*. Phys Rev 1933; **43**:491-494

2, Burnham CA, Brownell GL. *A Multi-Crystal Positron Camera*. IEEE Trans Nucl Sci 1972; **19**:201-205

3, Saha GB. *Interaction of Radiation with Matter*. In: Saha GB. *Physics and Radiobiology of Nuclear Medicine 4th Edition*. Springer, 2013

4, Evans RD. *The interaction of electromagnetic radiations with matter. Compton scattering and absorption*. In: Evans RD. *The Atomic Nucleus*. McGraw-Hill, 1955

5, Knoll GL. *Linear and Logic Pulse Functions*. In: Knoll GL. *Radiation Detection and Measurement, 3rd Edition*. Wiley 2000

6, Knoll GL. *Scintillation Detector Principles*. In: Knoll GL. *Radiation Detection and Measurement, 3rd Edition*. Wiley 2000

7, Phelps ME, Hoffman EJ, Mullani NA, Ter-Pogossian MM. *Application of Annihilation Coincidence Detection to Transaxial Reconstruction Tomography*. J Nucl Med 1975; **16**:210-224

8, Knoll GL. *Photomultiplier Tubers and Photodiodes*. In: Knoll GL. *Radiation Detection and Measurement, 3rd Edition*. Wiley 2000

9, Jakoby BW, Bercier Y, Conti M, Bendriem B, Townsend D. *Physical and clinical performance of the mCT time-of-flight PET/CT scanner*. Phys Med Biol 2011; **56**:2375-2389

10, Moses WW. *Time of flight in PET revisited*. IEEE Transactions on Nuclear Science 2003; **50**:1325-1330

11, Melcher CL, Schweitzer JS. *Cerium-doped lutetium oxyorthosilicate: a fast, efficient new scintillator*. IEEE Trans Nucl Sci 1992; **39**:502-505

12, Bailey DL, Karp JS, Surti S. *Physics and Instrumentation in PET*. In: Bailey DL, Townsend DW, Valk PE, Maisey MN. *Positron Emission Tomography*. Springer 2005

13, Nassalski A, Kapusta M, Batsch T, Wolski D, Mockel D, Enghardt W, et al. *Comparative Study of Scintillators for PET/CT Detectors*. IEEE Trans Nucl Sci 2007; **54**:3-10

14, Nutt R. *The History of Positron Emission Tomography*. Mol Imag Biol 2002; **4**:11-26

15, Cho ZH, Farukhi MR. *Bismuth germanate as a potential scintillation detector in positron cameras*. J Nucl Med 1977; **18**:840-844

16, Casey ME, Nutt R. *A Multicrystal Two Dimensional BGO Detector System for Positron Emission Tomography*. IEEE Trans Nucl Sci 1986; **33**:460-463

17, Sanchez-Crespo A, Larsson SA. *The influence of photon depth of interaction and non-collinear spread of annihilation photons on PET image spatial resolution*. Eur J Nucl Med Mol Imaging 2006; **33**:940-947

18, Herzog H, Tellmann L, Hocke C, Pietrzyk U, Casey ME, Kuwert T. *NEMA NU2-2001 guided performance evaluation of four Siemens ECAT PET scanners*. IEEE Trans Nucl Sci 2004; **51**:2662-2669

19, Brambilla M, Secco C, Dominietto M, Matheoud R, Sacchetti G, Inglese E. *Performance characteristics obtained for a new 3-dimensional lutetium oxyorthosilicate-based whole-body PET/CT scanner with the National Electrical Manufacturers Association NU 2-2001 standard*. J Nucl Med 2005; **46**:2083-2091

20, Panin VY, Kehren F, Michel C, Casey M. *Fully 3-D PET reconstruction with system matrix derived from point source measurements*. IEEE Trans Med Imaging 2006; **25**:907-921

21, Levin CS, Hoffman EJ. *Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution*. Phys Med Biol 1999; **44**:781-799

22, Phelps ME, Hoffman EJ, Huang S-C, Ter-Pogossian MM. *Effect of Positron Range on Spatial Resolution*. J Nucl Med 1975; **16**:649-652

23, Jodal L, Le Loirec C, Champion C. *Positron range in PET imaging: an alternative approach for assessing and correcting the blurring*. Phys Med Biol 2012; **57**:3931-3943

24, Jodal L, Le Loirec C, Champion C. *Positron range in PET imaging: non-conventional isotopes*. Phys Med Biol 2014; **59**:7419-7434

25, Budinger TF. *The Coming Age of Pet (Part 1) PET instrumentation: What are the limits?* Semin Nucl Med 1998; **28**:247-267

26, Hubbell JHaS, S.M. . *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients (version 1.4)*. National Institute of Standards and Technology, Gaithersburg, MD 2004

27, Germano G, Hoffman EJ. *A study of data loss and mispositioning due to pileup in 2-D detectors in PET*. IEEE Trans Nucl Sci 1990; **37**:671-675

28, Bailey DL, Young H, Bloomfield PM, Meikle SR, Glass D, Myers MJ, et al. *ECAT ART — a continuously rotating PET camera: Performance characteristics, initial clinical studies, and installation considerations in a nuclear medicine department*. Eur J Nucl Med 1997; **24**:6-15

29, Townsend DW, Beyer T, Kinahan PE, Brun T, Roddy R, Nutt R, et al. *The SMART scanner: a combined PET/CT tomograph for clinical oncology*. 1998 IEEE Nuclear Science Symp Conf Record; 1170-1174

30, Beyer T, Townsend DW, Brun T, Kinahan PE, Charron M, Roddy R, et al. *A Combined PET/CT Scanner for Clinical Oncology*. J Nucl Med 2000; **41**:1369-1379

31, Townsend DW. *Combined positron emission tomography-computed tomography: the historical perspective*. Semin Ultrasound CT MR 2008; **29**:232-235

32, Department of Health. *A Framework for the Development of Positron Emission Tomography Services in England*. Department of Health, 2005

33, Christensen NL, Hammer BE, Heil BG, Fetterly K. *Positron emission tomography within a magnetic field using photomultiplier tubes and lightguides*. Phys Med Biol 1995; **40**:691-697

34, Catana C, Wu Y, Judenhofer MS, Qi J, Pichler BJ, Cherry SR. *Simultaneous acquisition of multislice PET and MR images: initial results with a MR-compatible PET scanner*. J Nucl Med 2006; **47**:1968-1976

35, Zaidi H, Ojha N, Morich M, Griesmer J, Hu Z, Maniawski P, et al. *Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system*. Phys Med Biol 2011; **56**:3091-3106

36, Lecomte R, Cadorette J, Rodrigue S, Lapointe D, Rouleau D, Bentourkia M, et al. *Initial results from the Sherbrooke avalanche photodiode PET scanner*. 1995 IEEE Nuclear Science Symp Conf Record; 1353-1357

37, Herbert DJ, Saveliev V, Belcari N, Ascenzo ND, Guerra AD, Golovin A. *First results of scintillator readout with silicon photomultiplier*. IEEE Trans Nucl Sci 2006; **53**:389-394

38, Kim H, Chen CT, Ronzhin A, Ramberg E, Los S, Murat P, et al. A silicon photomultiplier signal readout using transmission-line and waveform sampling for Positron Emission Tomography. 2012 IEEE Nuclear Science Symp Conf Record; 2466-2468

39, Budinger TF. *Time-of-Flight Positron Emission Tomography: Status Relative to Conventional PET*. J Nucl Med 1983; **24**:73-78

40, Levin CS, Maramraju SH, Khalighi MM, Deller TW, Delso G, Jansen F. *Design Features and Mutual Compatibility Studies of the Time-of-Flight PET Capable GE SIGNA PET/MR System*. IEEE Trans Med Imaging 2016; **PP**:1-1

41, Allemand R, Gresset C, Vacher J. *Potential Advantages of a Cesium Fluoride Scintillator for a Time-of-Flight Positron Camera*. J Nucl Med 1980; **21**:153-155

42, Mullani NA, Gaeta J, Yerian K, Wong WH, Hartz RK, Philippe EA, et al. *Dynamic Imaging with High Resolution Time-of-Flight PET Camera - TOFPET I*. IEEE Trans Nucl Sci 1984; **31**:609-613

43, Ter-Pogossian MM, Ficke DC, Yamamoto M, Hood JT. *Super PETT I: A Positron Emission Tomograph Utilizing Photon Time-of-Flight Information*. IEEE Trans Med Imaging 1982; **1**:179-187

44, Lewellen TK, Bice AN, Harrison RL, Pencke MD, Link JM. *Performance measurements of the SP3000/UW time-of-flight positron emission tomograph*. IEEE Trans Nucl Sci 1988; **35**:665-669

45, Lewellen TK. *Time-of-flight PET*. Semin Nucl Med 1998; **28**:268-275

46, Conti M, Bendriem B, Casey M, Mu C, Kehren F, Michel C, et al. *Implementation of time-of-flight on CPS HiRez PET scanner*. 2004 IEEE Nuclear Science Symp Conf Record; 2796-2800

47, Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS. *Performance of Philips Gemini TF PET/CT Scanner with Special Consideration for Its Time-of-Flight Imaging Capabilities*. J Nucl Med 2007; **48**:471-480

48, Conti M. *Focus on time-of-flight PET: the benefits of improved time resolution*. Eur J Nucl Med Mol Imaging 2011; **38**:1147-1157

49, Surti S. *Update on time-of-flight PET imaging*. J Nucl Med 2015; **56**:98-105

50, Kimdon JA, Qi J, Moses WW. *Effect of random and scatter fractions in variance reduction using time-of-flight information*. 2003 IEEE Nuclear Science Symp Conf Record; 2571-2573

51, Conti M. *Effect of randoms on signal-to-noise ratio in TOF PET*. IEEE Trans on Nucl Sci 2006; **53**:1188-1193

3. PET IMAGE ACQUISITION AND RECONSTRUCTION

3.1 Overview

The purpose of image reconstruction is to generate a volume of image data that reflects the distribution of radiotracer that is within the patient. In the patient, there is a spatially continuous distribution of the radiotracer. Given the digital display nature of medical images, one requirement is a spatial discretization of the radiotracer distribution. Therefore, a reconstructed image typically consists of a 3-dimensional array of contiguous discrete cuboidal elements called voxels. The intensity value for each voxel typically indicates the radioactive concentration of the radiotracer within this voxel volume, with units of Bq/ml. Images are viewed by clinicians by extracting a 2-dimensional array, or slice plane, of voxel values from the reconstructed image volume. These can be transaxial (cross-sectional), coronal (as viewed from the front of the patient) or sagittal (as viewed from the side of the patient) planes. For display purposes, the voxel intensity values are assigned a particular colour using a colour map lookup table. An example of coronal, sagittal and transaxial image planes displayed using a greyscale lookup table are shown in Figure 3.1.

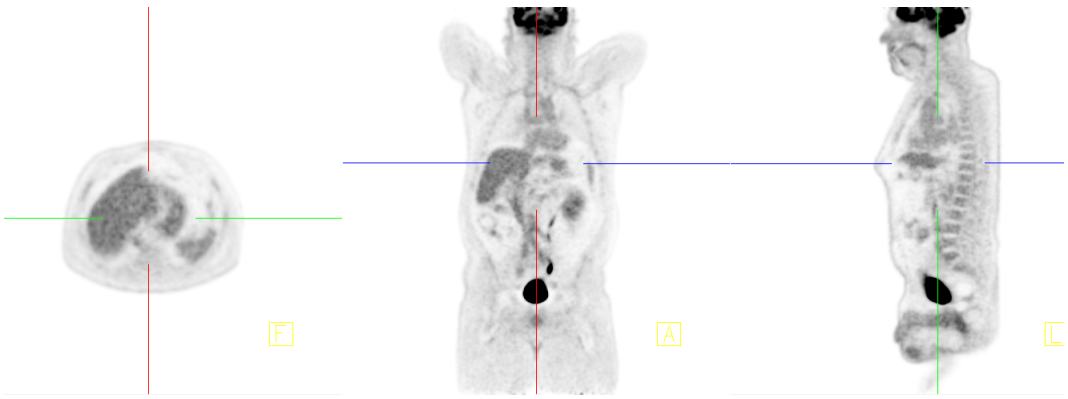


Figure 3.1 An example PET image showing transaxial (*left*), coronal (*centre*) and sagittal (*right*) image planes and displayed in greyscale.

For the purpose of reporting clinical PET images, the voxel values are usually normalised to SUV by the display program that is being used to display the images. As discussed previously, clinicians use SUV to inform their judgement on whether a particular area of FDG uptake is indicative of malignancy. By far the most commonly used variant of SUV is SUV_{max}, which is derived from a single voxel,

which has the greatest value within a given region. Other variants such as SUV_{mean} or total lesion glycolysis are also being used [1].

3.2 PET data collection

As discussed in the first review chapter, a detected PET event is referred to as a *Line of Response* (LOR), which joins the two crystal detector elements that detected the two 511 keV gamma-rays in coincidence. These lines can pass through the patient at different angles.

It is most common that the positions of these LORs are stored in arrays that are referred to as a *sinograms*. A sinogram consists of a number of bins arranged in a 2-dimensional array. One row within a sinogram represents the data collected from parallel LORs at a given angle, and will be referred to as radial offsets. The value contained in a bin represents the line integral, in the form of the ray transform [2], through a particular LOR through the patient. The radiotracer within a 2-dimensional transaxial plane in the patient consists of distribution $A(x,y)$. The radial offset r of a sinogram bin that will be populated from a LOR at an angle θ originating from position x and y is

$$r = x\cos\theta + y\sin\theta. \tag{3.1}$$

Figure 3.2 shows the relation between LOR with different angles and the corresponding sinogram bin locations.

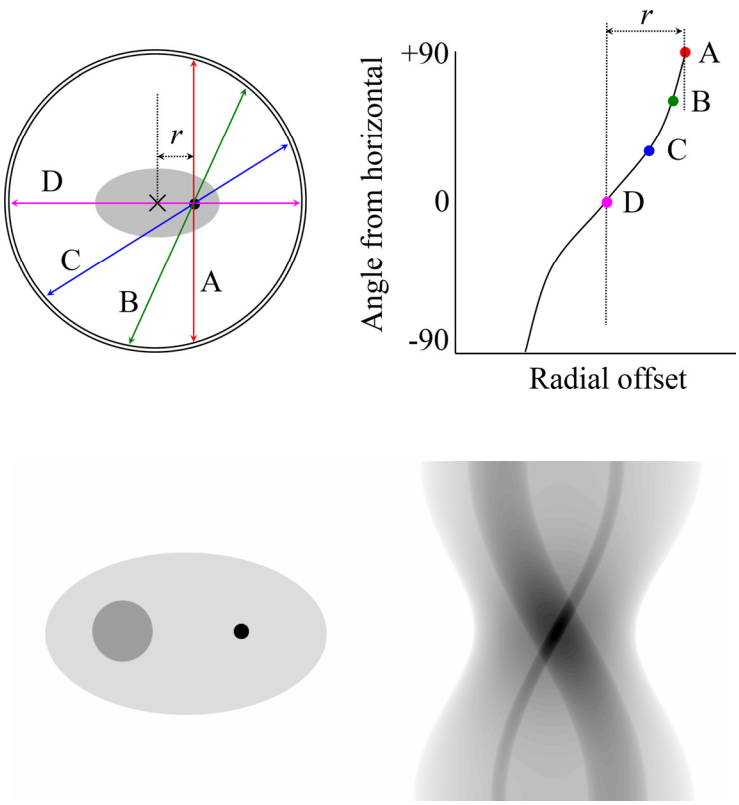


Figure 3.2 *Top*: Schematic to represent the relationship between LOR of different angles originating from a position that is radially offset by a distance *r*, as shown on the left, and the corresponding positions in a sinogram shown on the right. *Bottom*: an example sinogram obtained from the activity distribution shown.

Each sinogram represents the coplanar LORs in a single transaxial plane through the patient, which are detected by a specific ring of detector elements. These are referred to as *direct-plane* sinograms. Additional coincidence events can occur between adjacent rings of detector elements are referred to as *cross-plane* sinograms. Direct-plane and cross-plane sinograms are illustrated in Figure 3.3. Each sinogram therefore represents an imaging plane with a physical thickness of one half of the physical axial dimension of the detector elements within the ring. For example, on the Siemens Biograph mCT, each crystal element in the block assembly is 4 mm in the axial direction and so each transaxial imaging plane is 2 mm thick. There are $(2N - 1)$ sinograms defined on a detector consisting of N detectors.

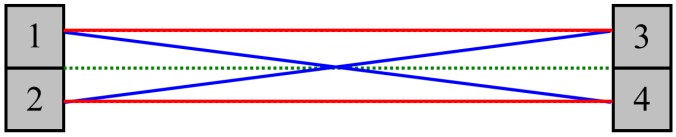


Figure 3.3 Examples of direct-plane sinograms (*solid red lines*), consisting of coincidence events between detectors 1 and 3 or 2 and 4, and cross-plane sinograms (*green dashed line*), consisting of coincidence events between detectors 1 and 4 or 2 and 3 as shown by the two solid blue lines. Note that two LORs can contribute to the same cross-plane sinogram bin.

3.2.1 3-dimensional data collection

From the example shown in Figure 3.3, it follows that the number of events acquired in the cross-plane sinograms will be twice that acquired in the direct-plane sinograms. This is due to two possible LORs contributing to the same cross-plane sinogram bin. In early PET systems, coincidence events between detector rings spaced further apart in the axial direction were not allowed due to physical septa, usually tungsten, placed between detector rings [3]. Removing these septa allows for many more combinations of LORs with detectors over a greater axial separation as can be seen from Figure 3.4. This results in a significant increase in the overall sensitivity of the detector [3, 4]. The axial sensitivity profile changes from the oscillation between direct and cross plane sensitivity in 2D acquisition to a pyramid-shaped profile peaking at the centre of the axial FOV (Figure 3.4).

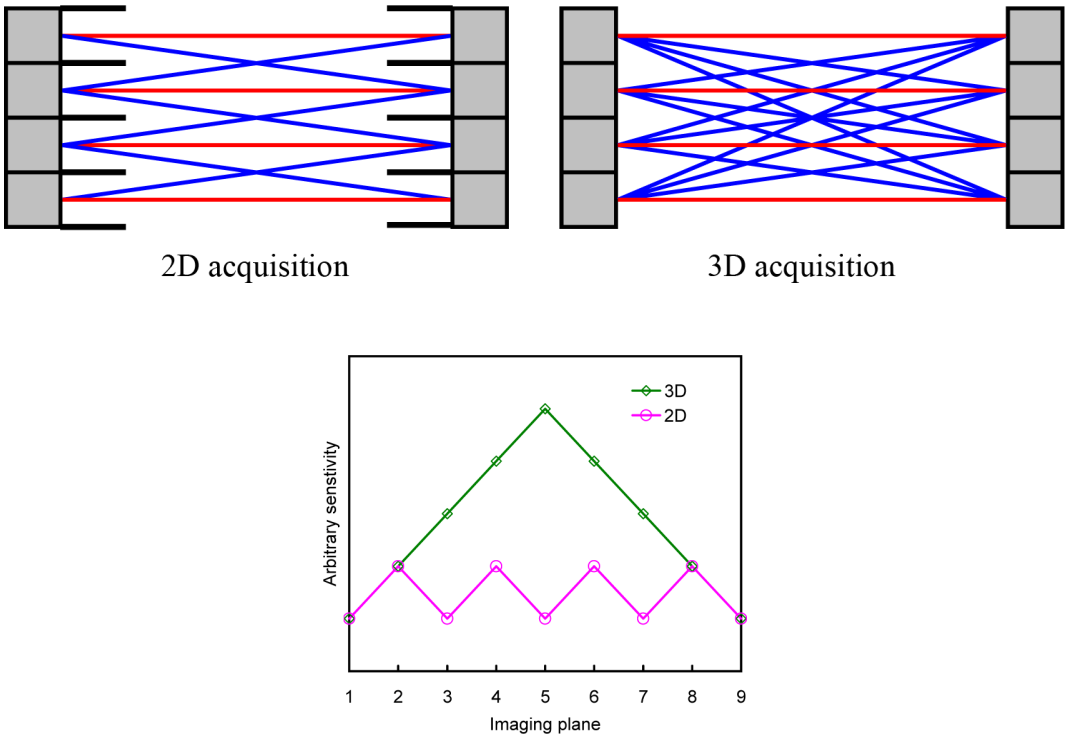


Figure 3.4 Illustration of 2D data acquisition with inter-detector septa in place and 3D data acquisition with septa removed with the relative differences in sensitivity shown the plot.

3.2.2 Wholebody imaging

The coverage of a patient from which coincidence events are acquired is defined by the axial FOV. On modern systems this is typically between 15 cm and 25 cm, with the axial FOV of the Biograph mCT being 21.6 cm [5]. This is insufficient to acquire data from the entire patient and so a wholebody PET acquisition is performed by moving the imaging bed, and hence patient, axially through the PET FOV. This is usually performed as a "step-and-shoot" technique where the bed stops at defined positions, which are referred to as *bed positions*. To achieve uniform image quality, it is necessary that the effective sensitivity is uniform across the patient. This is straightforward with 2D scanning as the axial sensitivity profile can be considered almost uniform whereas the pyramid-shaped 3D profile requires a substantial overlap of adjacent bed positions. The overlap between adjacent bed positions on the Biograph is a 43% of the axial FOV.

3.2.3 Time of flight data collection

For the purpose of this review, it is most sensible to discuss the method employed by the Biograph mCT for TOF data acquisition. The TOF difference of arrival time of the two gamma-rays is binned into one of 13 contiguous timing bins that are 312 ps in temporal width [5], which translates to a spatial rebinning of 4.7 cm along the LOR. These timing bins take the form of sinograms and are referred to as *TOF sinograms* that are indexed relative to the central sinogram, which translates the point TOF differences for events at the midpoint of the LOR as illustrated in Figure 3.5. If perfect timing resolution is assumed, it follows that a point source at the centre of gantry will populate data into only the "0" TOF sinogram over the complete angular range of LORs. The timing resolution of the Biograph mCT is 530 ps [5] and so there is a probability that events will be binned into neighbouring TOF sinograms. An offset point source will contribute to various TOF sinogram bins according to the angle of LOR. The assignment of data into particular TOF sinograms is shown in Figure 3.5.

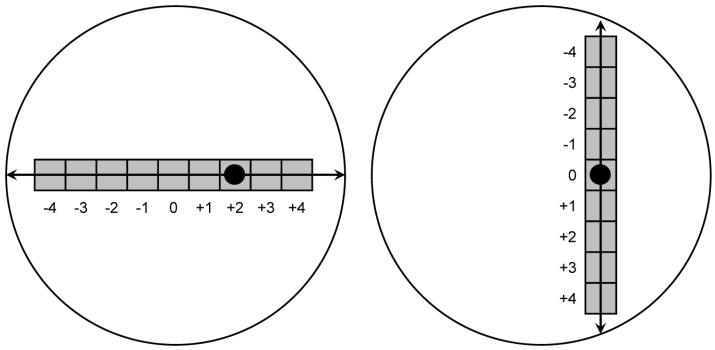


Figure 3.5 Schematic showing the assignment of events arising from a radially offset location into TOF sinograms. The grey boxes represent which TOF sinogram that the LOR should be rebinned into. Assuming perfect timing resolution, the horizontal LOR, shown on the left, would be placed in the "+2" TOF sinogram whereas the vertical LOR would be placed in the "0" TOF sinogram.

3.3 Data corrections

For PET images to be a quantitatively accurate representation of the activity distribution within the patient, the physical and technical factors that degrade the image must be corrected for. The acquired sinograms consist of prompt coincidence events and are comprised of true coincidence events and contaminants in the form of random coincidences and scattered coincidences. These contaminants should be

removed or accounted for to produce a sinogram of true coincidences [6]. These true coincidences can then be corrected for the attenuation within the patient.

3.3.1 Random coincidences

Random coincidences occur when singles events from two unrelated annihilations are detected leading to an erroneous LOR. There is no spatial correlation to the random coincidences and so they contribute to a uniform background throughout the sinogram data [7]. From equation 2.4 in chapter 2, it can be seen that the random coincident rate for a given LOR can be estimated from the singles count rate of the individual detectors forming the LOR [8, 9].

An alternative approach of estimating random coincidences is using a delayed coincidence window [10–12]. Here the detection times of one of the gamma-rays is delayed by an amount sufficient to prevent the detection of true and scattered events but short enough to prevent any redistribution of activity in the body. Provided that the temporal separation of the coincidence window and the delayed window is greater that the width of the coincidence timing window, coincidences between two singles, one from each window, must be random. By collecting coincidences between the two windows, a *randoms sinogram* is formed, which gives an estimation of the distribution of random coincidences within the prompt sinogram.

The estimation of random coincidences from both methods follows Poisson statistics. The delayed technique results in an estimation of random coincidences that is considerably noisier than the estimate from singles rate due to substantially lower number of events [13, 14]. The process of subtracting the random sinograms from the prompt sinograms will inevitably increase the statistical noise in the resultant sinogram regardless of the method of estimating the random coincidences, leading to quantitative bias. The delayed estimation technique results in noisier corrected sinograms due to the noisier random sinograms [6, 15].

The technique of estimating random coincidences from the singles events rate requires knowledge of the singles event rate for each detector pair forming the LOR. This may not always be the case in systems that use block detector design, where the event rate may only be measured at the block level. An alternative method was later

proposed that employed a summation technique across a bank of detectors of the delayed coincidence measurements to produce an estimate of random coincidences with statistical noise comparable to that from the singles rate estimation [6]. This technique is known as a *smoothed randoms* technique and is the method used on the Siemens Biograph mCT.

3.3.2 Scattered events

Gamma-rays that undergo Compton scatter within the patient will change direction and lose an amount of energy according to the angle of scatter as shown by equation 2.1. If both gamma-rays are detected by the detector ring, the LOR position will be incorrect. The removal of inter-plane septa in 2D data acquisition and move to 3D data acquisition means the impact of scattered gamma-rays becomes far more significant [16]. Like random coincidence events, an estimate of scattered events present in the prompt sinograms can be derived for correction. Following a correction for random coincidences, the prompt sinogram should consist of only true and scattered coincidence. There are three main techniques to correct for scattered gamma-rays: convolution, measurement by multiple energy windows and modelled.

The convolution technique can be performed on the prompt sinograms and assumes that there is spatial correlation between the unscattered and scattered events within the prompt sinogram [16, 17]. A pre-determined scaling factor, referred to as the *scatter fraction*, is used to estimate the proportion of the scatter distribution to subtract from the prompt sinogram.

The measurement technique of estimating the scatter component in the prompt sinograms can be performed using one or more additional *scatter energy windows* placed below the standard energy window used to collect the prompt coincidence events. If either one or both detected gamma-rays have an energy that falls within a scatter window, the measured LOR is classified as a scattered event. [18, 19]. There are two problems with this approach. Firstly, like the delayed random estimate technique, the scatter estimate is relatively noisy due to the Poisson nature of the data. Secondly, the spatial distribution of scattered events in the scatter window may not be form a reliable estimate of the spatial distribution of scattered events in the prompt window [20].

The modelled technique has the scatter probability of gamma-rays taking into account the radionuclide distribution and the material in which gamma-rays are scattered [21–23]. The models by Ollinger at al and Watson et al are based on the assumption that only one of the gamma-ray pairs undergoes Compton scatter within the patient.

With the knowledge of the LACs, the probability of scatter along a given LOR through a particular angle can be deduced from the Klein-Nishina formula, which describes the differential cross-section of scatter through a solid angle, Ω is [24]:

$$\frac{d\sigma}{d\Omega} \propto P(E_\gamma,\theta)^2\left[P(E_\gamma,\theta)+\frac{1}{P(E_\gamma,\theta)}-2\sin^2\theta\cos^2\varphi\right] \tag{3.2}$$

$$P(E_\gamma,\theta)=\frac{1}{1+(E_\gamma/0.511)(1-\cos\theta)}$$

With the distribution for various gamma-ray energies shown in Figure 3.6.

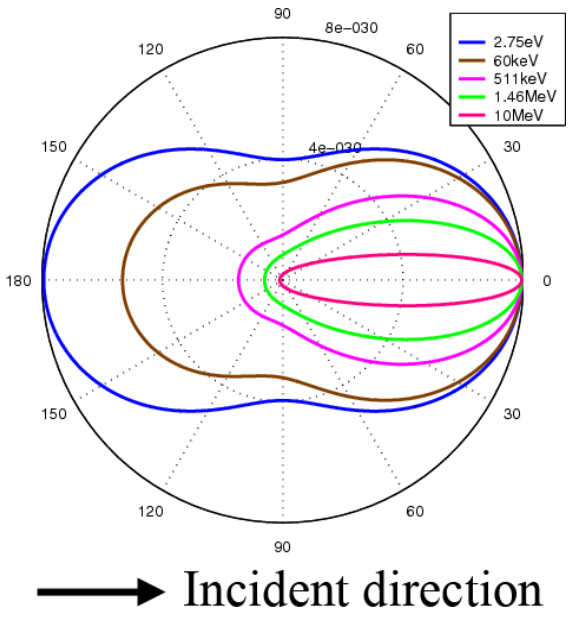


Figure 3.6 Distribution of scattering angle cross-section for different gamma-ray energies [25].

These methods provide a noise-free estimation of the distribution of scattered within the prompt sinogram. The absolute magnitude of the scatter estimate needs to be scaled to match the prompt data. The method employed on the Biograph mCT uses a scaling factor deduced from events that have been measured outside the object boundary, referred to as the *tails*, determined by a CT image on modern PET/CT systems. These events by definition can only be scattered events if a randoms correction has been applied. Scaling using the tails requires co-registration between the PET and the CT data and artefacts can occur if this is not the case [26]. If the patient size is large, the number of events in the tails can be low and lead to suboptimal correction and alternative Monte Carlo methods of scaling have been suggested [27].

3.3.3 Attenuation

Attenuation is the most significant degrading factor of PET data [28] and the level of attenuation for a LOR passing through a given thickness of matter can be calculated using equation 2.3. The attenuation of the LOR is independent of the position of the annihilation along the LOR. Therefore, if the attenuation along the LOR can be deduced it can be corrected for. This requires either knowledge of the LACs along the length of the LOR or a measurement of the attenuation of the LOR [29].

In objects of relatively uniform shape and density, such as the brain, the attenuation can be calculated by determining the object boundary from the sinogram data and assuming a single LAC. This allows the path lengths at various LOR angles to be calculated and hence attenuation can be deduced [30]. This approach breaks down in other body areas due to the non-uniform shape and range of LACs.

Measurement of the LOR attenuation can be deduced from a *transmission scan*. This scan uses a gamma-ray emitting radioactive *transmission source*, which rotates around the patient with the quantity of gamma-rays being detected on the other side of the patient, usually by the PET detector itself. By comparing the quantity of transmission gamma-rays emerging from the patient with a *blank scan* – a transmission scan performed in air – the LOR attenuation can be deduced [31–33]. Typical transmission sources were long-lived positron-emitting radionuclides such as Ge-68 [31, 32, 34], using the 511 keV gamma-rays or Cs-137, which is a single

gamma-ray emitter with energy of 662 keV [35]. If the bins in the emission and transmission sinograms are aligned the PET emission sinogram can simply be corrected for attenuation by dividing by the transmission sinogram [36].

The introduction of PET/CT, with a CT image that was co-registered with the PET image, provided a means of determining the LACs along the LOR. This alleviated the requirement to measure the gamma-ray attenuation using rotating radioactive source [37]. The voxel values in the CT image, in Hounsfield Units, are converted to a map of LACs for 511 keV gamma-rays using a bi-linear look-up table [38].

3.3.4 Normalisation

Ideally, the coincidence detection efficiency in the detector ring would be identical for all LORs so that exposing the detector to a uniform source of radioactivity would produce the same event rate on each and every detector pair. However, factors such as manufacturing imperfections, gamma-ray energy, electronic drift of the PMT output, singles event count rate and the angle of incidence of the gamma-ray will introduce a variation in the intrinsic efficiency of individual crystals. Consequently, there are variations in efficiency for an angular span of LORs between one common crystal on one side of the detector and opposing crystal that are in coincidence [39]. The coincidence detection efficiency of a given LOR is the product of the intrinsic efficiencies of the two crystals that detect each gamma-ray and a geometric factor, which is dependent on the solid angle between the detector pair and the radial offset of the LOR [39–41].

A common method to assess the variation in LOR efficiency is to acquire data from a uniform radioactive source such as a plane source that is rotated [40] or cylinder placed at the centre of the FOV [42]. The count rate, R_{ij}, for a given LOR formed from two detectors is [40]

$$R_{ij} = A\varepsilon_i\varepsilon_j G_{ij} \tag{3.3}$$

where A is the activity that the detectors i and j are exposed to, ε_i and ε_j are the intrinsic efficiencies detectors i and j respectively and G_{ij} is the geometric efficiency between the two detectors. Events are collected for each LOR and the variation in the

number of events collected is used to correct for the differences in efficiency. This is known as direct normalisation.

The introduction of 3D data acquisition introduced difficulties with direct normalisation. The main issues being the need to acquire data for many more LORs, which required a long image acquisition, and the substantial dead-time effects from the increased sensitivity compared with 2D acquisition [43]. In addition, systems may apply on-line compression of the acquired data such that several LORs populate a single sinogram bin. This compression may be variable depending on the particular acquisition parameters and complicates the normalisation process [39]. To address this, a component-based approach was proposed for normalisation. The model comprises of two primary components: the geometric efficiency and intrinsic detector efficiency [39, 43, 44]. The geometric efficiency is dependent on the scanner construction and temporally invariant factors for each LOR with the intrinsic detector efficiencies being subject to electronic drift. By modelling the normalisation components, the normalisation parameters for a given set of acquisition conditions and parameters can be extracted during the reconstruction process. However, as the intrinsic components need to be measured periodically, a measurement of a uniform source is required.

3.3.5 Dead-time

The concept of dead-time of a detection system is the difference in the observed count rate compared to what is expected for the given activity. Once an event has been detected, another cannot be detected in a given time interval – referred to as the *dead-time interval* [45]. For systems that do not suffer from dead-time effects, the observed count rate would follow a linear relationship with the activity that the detector is irradiated with. Due to the mechanisms involved in the detection process, which have a finite dead-time interval, this is not the case. As the rate of incident gamma-rays increases on the detectors, events can be lost due to dead-time and the observed count rate decreases from what is expected. For a crystal or block detector, the increasing count rate causes the loss of events through pulse pile-up. For coincidence events additional dead-time losses occur due to triple (or more) coincidence events where three (or more) gamma-rays are detected within the coincidence timing window [46]. There are two models that can describe the

observed dead-time effects: *non-paralysable* and *paralysable* [47]. A non-paralysable has a fixed dead-time interval after the detection of one event and hence has a maximum permissible counting rate based upon this interval. The observed count-rate plateaus to this maximum permissible rate as the irradiation activity increases. A paralysable requires successive events to arrive at time intervals of at least the dead-time interval and the interval time begins again with each detected event. The observed count-rate is shown to reach a maximum count-rate and then reduce as the irradiation activity continues to increase. Studies on dead-time in PET systems have demonstrated a non-paralysable behaviour [46, 48]. Dead-time can be assessed by recording the observed count rate of a phantom as it decays and this information is stored. For a given count-rate, the dead-time losses can be determined and the correction can be included as part of the normalisation correction [39].

3.4 Image reconstruction

For simplicity, it is useful to start by considering a 2-dimensional reconstruction process of a single transaxial image representing the continuous radiotracer distribution $A(x,y)$ within the patient. A reconstructed transaxial image will consist of an array of discrete voxels in x and y directions. A single row in a sinogram contains line integrals with a discrete spacing of r for a given projection angle θ. It follows that the reconstruction process will populate the voxels with the values from the sinogram bins. There are two classifications of reconstruction techniques: analytical, which involve executing a formula to estimate the image data from the acquired data and iterative, which repeatedly re-estimates the image data from the projection data and a current image estimate, often in order to solve an optimisation problem.

3.5 Analytical reconstruction

3.5.1 Backprojection

Mathematically, the process of *backprojection* is the adjoint of the process by which sinogram bins are populated by line integrals of LORs. The process translates the values from sinograms (referred to as *projection space*) into the transaxial image plane (referred to as *image space*). Since there is no positional information to determine the location of the positron-electron annihilation along the LOR, one approach is to spread the values in each sinogram bin evenly along a line of voxels

that correspond to the LOR location. This is repeated using the data from all projection angles, as shown in Figure 3.7.

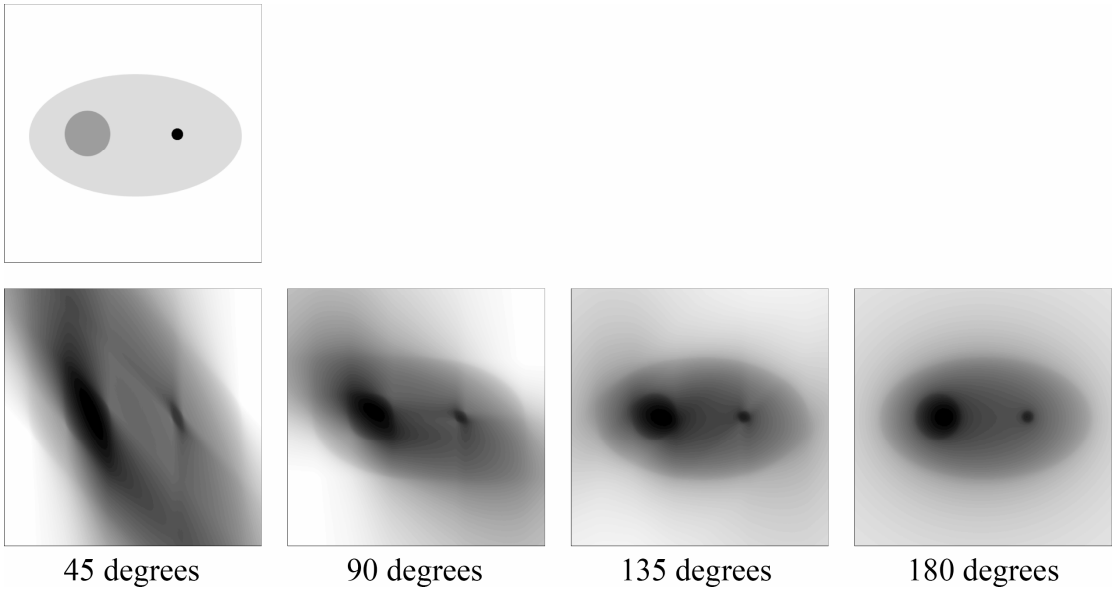


Figure 3.7 A schematic showing how the distribution of activity (*top image*) only becomes evident as backprojection of data is performed over the full 180 degree angular range.

3.5.2 Filtered backprojection (FBP)

As can be seen from Figure 3.7, the process of backprojection gives a fairly poor approximation of the radiotracer distribution as the image is degraded by low frequency radial blurring artefacts. This is resolved by applying a ramp filter [49] to each sinogram row in the radial offset direction before they are back-projected. This filter has a modulation transfer function which is a ramp filter and increases in amplitude proportionally with the absolute value of the spatial frequency as shown in Figure 3.8. The process of applying this filter usually takes advantage of the convolution theorem whereby the discrete Fourier transform of the sinogram data in the radial offset direction is multiplied by the ramp filter in frequency space before the inverse discrete Fourier transform in the radial offset direction is used to create filtered sinogram data. This filtered sinogram is then back-projected as described above.

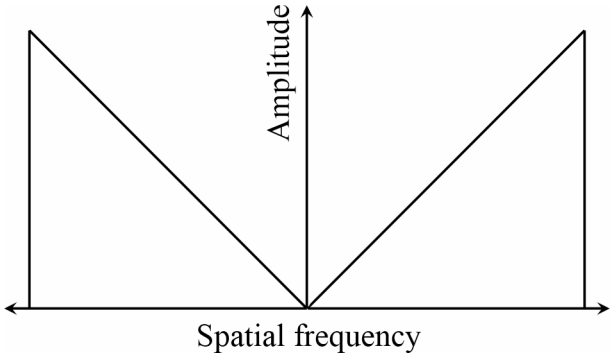


Figure 3.8 A schematic representation of the ramp filter shown in the frequency domain.

This process is known as *filtered-backprojection* (FBP). The ramp filter reduces the low frequencies in the reconstructed image but amplifies the high spatial frequency components; this is manifested in the image as restoring areas of sharp contrast as can be seen in Figure 3.9. However the magnification of high frequencies also magnifies noise in the projection data and with typical numbers of events that are acquired in clinical imaging, the high frequency noise is so dominant that the images are not useful without modification. To reduce the high-frequency noise, the modulation transfer function is modified by combining with a low-pass filter to reduce the high frequency amplitudes. This reduces the high frequency noise components but inevitably degrades the spatial resolution of the projections and hence resultant image (see Figure 3.9).

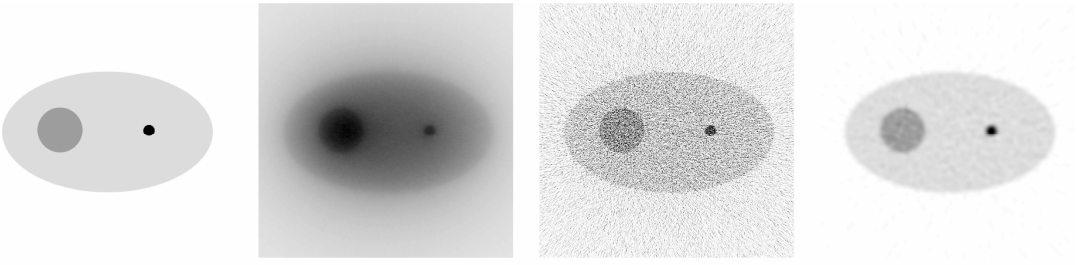


Figure 3.9 Illustration of the application of the ramp filter to a backprojection (BP) image to produce a filtered backprojection (FBP) image.

3.6 Iterative reconstruction

3.6.1 Maximum Likelihood Expectation Maximisation

Iterative reconstruction iteratively updates an estimate of the image from an initial estimate with each update referred to as an iteration. The most common implementation is a statistical image reconstruction algorithm based upon *Maximum Likelihood Expectation Maximisation* (MLEM) [50, 51]. The aim of MLEM is to maximise the likelihood of an image that represents the radiotracer distribution given the data that have been acquired by the scanner.

To appreciate the MLEM algorithm, it is necessary to appreciate the statistical nature of the data being estimated. The emission of radiation is a random event and can be accurately modelled using Poisson statistics. Given the Poisson nature of the radiotracer distribution, we assume that the number of positron-electron annihilations occurring within each voxel can be modelled by independent Poisson distributions with mean λ_j for the j^{th} image voxel [50]. For a given voxel, the LORs will be emitted in all directions. Therefore, given the lack of 100% geometric enclosure of the detector around the patient, possible attenuation of one or both of the gamma-rays, and the possible failure of the system to detect both gamma-rays, the probability of detecting every LOR emitted is less than 1. It is useful to define a system matrix, which states the probability that a LOR emitted from the j^{th} voxel in the image is detected by the system giving rise to the increment of the value in the i^{th} bin of a sinogram [50]

$$S_{ij} = P(\text{emission from } j \mid \text{detection in } i), \tag{3.4}$$

Therefore, the counts in the i^{th} sinogram bin arising from all image voxels j=1, j=2, j=3…j=N with intensities $\lambda_1, \lambda_2 \lambda_3 \ldots \lambda_N$ can be modelled as an independent Poisson distribution with mean Y_i as

$$Y_i = \sum_j S_{ij} \lambda_j \tag{3.5}$$

From this a log-likelihood function can be constructed as [50, 51].

$$\ln(Y,\lambda) = \sum_i \left[-Y_i + Y_i \ln\left(\sum_j S_{ij}\lambda_j \right) - \ln Y_i! \right] \tag{3.6}$$

In MLEM equation 3.6 is maximized using the expectation maximisation algorithm. This results in the iterative algorithm:

$$\begin{aligned} \lambda_j^{n+1} &= \frac{\lambda_j^n}{s_j} \sum_j S_{ij} \frac{Y_i^{measured}}{\sum_j S_{ij}\lambda_j^n} \\ s_j &= \sum_j S_{ij} \end{aligned} \tag{3.7}$$

where $Y_i^{measured}$ is the measured counts in the i^{th} sinogram bin.

Implementation

There are six steps to the MLEM algorithm:

1. Start with an initial image estimate;
2. Forward project the image estimate to estimate the measured data;
3. Ratio the measured data with the estimated measured data;
4. Back project the ratio to calculate an update image;
5. Multiply this update image by the previous image estimate;
6. Repeat from step 2.

The first stage in the MLEM is to begin with an initial estimate of the radiotracer distribution. Given that the emission intensity from a given voxel cannot be negative, it is logical to use only positive values in this initial estimate, which is usually taken to be a uniform image such that voxel intensities $\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_N$ in the estimated image are equal. From this estimate image, the estimated sinograms are created by forward projecting the voxel values in image space into projection space. Unlike FBP, where backprojection is a direct process, the process of forward and backward projecting in iterative reconstruction is done using the system matrix, S as given in equation 3.4.

For the n^{th} iteration in the reconstruction loop, each bin in the estimated sinogram $y_1^n\ y_2^n\ y_3^n$ etc. can be expressed from the forward projection operation as

$$y_i^n = \sum_j S_{ij} \lambda_j^n \,, \tag{3.8}$$

where y_i^n is the value of the i^{th} bin in the estimated sinogram of the n^{th} iteration, S_{ij} is the corresponding element in the system matrix that defines the contribution of the j^{th} voxel to the i^{th} sinogram bin and λ_j^n is the value of the j^{th} voxel in the estimated image. For completeness, it can be seen that the adjoint of this process is the backward projection operation, which is written as

$$\lambda_j^n = \sum_i S_{ij} y_i^n \,. \tag{3.9}$$

The system matrix can be composed of multiple factors, depending on the variant of MLEM that has been applied. In its most basic form, the system matrix comprises solely of the geometric probabilities that the angle and position of a given LOR originating from a voxel will contribute to a given sinogram bin. In this case, the sinograms have had corrections (attenuation, random coincidences, scatter, normalisation) performed prior to the MLEM process and the process is referred to as *unweighted* MLEM [28]. The process of data corrections means that the values in these pre-corrected sinograms are no longer described by Poisson statistics which does not satisfy the assumption of the MLEM algorithm and results in errors within the images [52].

For the purposes of this review, the focus will be on the *Ordinary Poisson* (OP) variant. In OP-MLEM, the system matrix is composed of the geometric, attenuation and normalisation component such that

$$S_{ij} = G_{ij} A_i N_i \tag{3.10}$$

where G_{ij} is the geometric probability that the angle and position of the given LOR, originating from the j^{th} voxel, will correspond to the i^{th} sinogram bin; A_i is the probability that both gamma-rays are not lost through attenuation along the path

length of the LOR detected in the i^{th} sinogram bin, (section 3.3.3) and N_i is probability that both gamma-rays are detected and is obtained from the detector normalisation (section 3.3.4) of the i^{th} LOR.

The estimated sinograms arising from equation 3.8 are compared with the measured sinogram data and the ratio K_i^n of corresponding bins is obtained. The values in the estimated sinogram bins $y_1, y_2, y_3 \ldots y_N$ are intended to represent the *true* coincidence events that would give rise to the image. However, the measured sinogram data are the *prompt* events and are true coincidence events contaminated with random coincidence events and scattered events. As described in sections 3.3.2 and 3.3.3, it is possible to generate sinogram data that represents scattered LORs or random coincidence LORs. In FBP reconstructions, these scatter and random sinograms are subtracted from prompt sinograms prior to reconstruction. This not only destroys the Poisson nature of sinograms but can result in negative values, which can destabilise the MLEM algorithm. The scattered and random coincidence sinograms are independent from the estimated true sinograms and hence can be added [52] as such

$$K_i^n = \frac{Y_i^{measured}}{y_i^n + r_i + s_i} = \frac{Y_i^{measured}}{\sum_j S_{ij}\lambda_j^n + r_i + s_i}. \tag{3.11}$$

where r_i and s_i are the values of the i^{th} bins in the measured or modelled random coincidence and scattered event sinograms respectively. If one starts with an image estimate that contains only positive values in all voxels, it can be seen here that these ratios will always be positive. These ratios in projection space are then back-projected into image space using the system matrix to obtain an array of update coefficients for voxels in the estimated image and is sometimes referred to as the update image. The update coefficient, C_j^n, for the j^{th} voxel, is

$$C_j^n = \frac{1}{\sum_i S_{ij}} \cdot \sum_i S_{ij} K_i^n. \tag{3.12}$$

These update image coefficients are used to update the j^{th} voxel in the estimated image, for j=1, j=2, j=3…j=N, as such:

$$\lambda_j^{n+1} = \lambda_j^n \cdot C_j^n . \tag{3.13}$$

Every update coefficient will be positive, and hence the voxel values in the reconstructed image will also all be positive. This non-negativity is consistent with the Poisson nature of the tracer distribution being estimated [53].

As the number of iterations increase, so does the likelihood that the estimated image is a true representation of the radiotracer distribution within the patient. This process of the estimated distribution becoming closer to the actual distribution is known as *convergence*.

Images reconstructed with MLEM have been shown to contain less noise, with a limited number of image updates, than those reconstructed with FBP [50]. Firstly, there is no need to apply the ramp filter and so there is no inherent noise amplification. Expanding this, it is known that the variance in an image reconstructed with FBP is uniformly distributed [54] and hence in areas the areas of low intensity in the image have low signal-to-noise ratio. The noise properties of the MLEM are far more complex but can be approximated to the voxel variance being approximately proportional to the magnitude of voxel intensity [53, 54] and hence areas of lower intensity have lower noise.

The second, and arguably more significant, advantage over analytical approaches is the ability to include information about the physics of gamma-ray measurement. This can include attenuation, scatter and the geometric response of the system. Analytical approaches assume that the line integrals are 1-dimensional with all LORs having equal probability of detection [39] when in reality the LORs arise between two detectors of finite size and so a LOR represents a volume between the detectors.

Iterative methods are not without disadvantages. They are computationally more demanding; the need to perform both forward and backprojection effectively doubles the number of operations performed compared with FBP. The system matrix will

rarely contain sufficient data to perfectly model the system response and so the forward and backward projection operations will place data in erroneous locations, which is manifested as noise [56–58]. Combining the non-complete system matrix with the discretization of data into voxels and forcing the algorithm to produce data that agrees with the noisy nature of the measured data, the problem is ill-conditioned [53]. As a result, image noise is seen to increase with increasing numbers of iterations such that an image that has converged, using MLEM, to the most likely solution will in reality be of very little use due to the prohibitively high degree of noise in the reconstructed image [56, 59]. There are three main mitigating strategies that may be used. The first is to stop the reconstruction process at a number of iterations before convergence is reached but the degree of noise is acceptable. The consequence of this approach is that quantitative accuracy of the image is compromised. The second is to perform a greater number of iterations, and then apply a low-pass filter (initially referred to as sieves) to reduce the high spatial frequency noise in the reconstructed image [56]. The consequence of this approach is that spatial resolution is reduced in the same way as it is for FBP with the low-pass filter. The final approach is to suppress the amplification of noise during the reconstruction process by constraining the update by including a penalty term in update term in equation 3.12. This shall be discussed in detail later in this review.

While the clinical usefulness of MLEM by way of improved noise suppression is clear, the quantitative accuracy is a more complex issue. Comparisons between the quantitative accuracy of images reconstruction with FBP and MLEM have been performed [60–62]. Findings from these studies demonstrated that several hundred to one thousand reconstruction updates to the reconstructed image were required to obtain comparable quantitative performance with FBP, particularly for small objects. In clinical imaging it is not uncommon for the number of image updates to be in the range of approximately 40 to 60 during reconstruction. This is usually also combined with a low-pass smoothing filter and this would suggest that convergence of the reconstruction is rarely achieved in clinical imaging. The limiting factors are reconstruction times but mainly the need to constrain the noise in the image to a degree that allows for acceptable interpretation of the images.

3.6.2 Ordered Subset Expectation Maximisation

The slow convergence rate of MLEM requires many updates to reach an acceptable solution and this limits its appeal in routine imaging where there is a need to produce good quality images in a short time. In 1994, an alternative accelerated form of the MLEM algorithm was proposed by Hudson and Larkin [63]. This method, known as *Ordered Subset Expectation Maximisation* (OSEM), is probably one of the single most important developments in image reconstruction and made routine use of iterative reconstruction a reality. It is by far the most commonly used algorithm employed on PET scanners.

MLEM and OSEM differ by the number of times that the reconstructed image estimate is updated per reconstruction iteration. In MLEM, the forward and backprojection operations are performed for all projection angles to form the update image that is multiplied with the previous image estimate. On the Biograph mCT, there are 168 discrete angular steps in the sinograms over the 180 degrees of data acquisition [5] and so the algorithm must perform 336 projection operations (168 forward and 168 backward) before the image is updated. OSEM divides the number of projection angles into groups called *subsets*. Now the reconstructed image estimate is updated after forward and backprojecting each subset. The completion of one iteration is defined when all subsets, and hence all projection angles, have been used. For OSEM with 24 subsets, as used clinically on the Biograph, the same 168-angle data translates to seven projection angles per subset and so only 14 projection operations are performed before the image is updated. The computation required by both MLEM and OSEM to complete a single iteration is essentially the same but the image is updated more than once per iteration with OSEM and the process is accelerated notably. The total number of updates of the reconstructed image estimate is the product of the number of iterations and the number of subsets. Hence images produced by OSEM reconstructions employing 2 iterations & 24 subsets, 4 iterations & 12 subsets or 6 iterations & 8 subsets will appear very similar visually [63].

Technically, the maximum number of subsets is equal to the number of discrete projection angles. However, it is has been shown that with high numbers of subsets and noisy projection data can result in a less accurate image estimate when the ground truth of the underlying activity distribution is known [63]. A phantom-based

lesion detection study performed on a Biograph mCT showed that lesion detection performance reduced for high numbers of subsets [64]. In routine use, the choice of both iterations and subsets is chosen based on the pragmatic considerations that make up an efficient clinical imaging service: obtaining suitable quality images in a reasonable time frame.

3.6.3 Resolution modelling

Traditionally, the geometric aspect of the system matrix is equivalent to FBP in that it assumes that the projection of a LOR from a given voxel is a line integral to a sinogram bin that corresponds to the crystal block element that the gamma-ray is incident on. As discussed in chapter 2, there are several factors that degrade the positioning of the LOR, such as depth-of-interaction effects, inter-crystal scatter and gamma-ray acolinearity. These factors are neglected by the line integral method. The result is a degradation of radial spatial resolution that worsens with increasing radial offset due primarily to depth-of-interaction effects. This is illustrated for the example of point source in Figure 3.10.

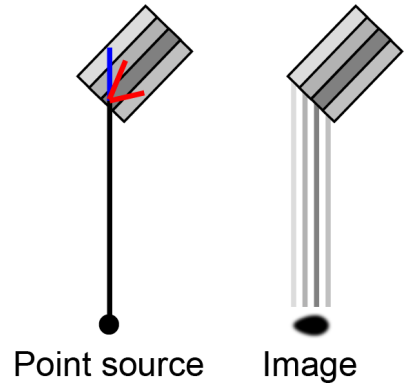


Figure 3.10 Illustration of the degradation of radial spatial resolution due to depth-of-interaction crystal penetration and inter-crystal scatter for a block detector of four crystals. The image on the left shows the incident LOR in black. The blue line illustrates un-scattered crystal penetration due to insufficient path length in the incident crystal while the red lines illustrate inter-crystal scatter. The degree of greyscale on each crystal is an indication of the proportion of events registered in each crystal after collecting multiple instances of the given LOR. The image on the right shows an illustration of the image formed by backprojection of the data from each crystal element.

To resolve this issue, the concept of *resolution modelling* can be employed, which involves the incorporation of the geometric response of the system into the system matrix. This has been proposed by several groups [58, 65, 66]. One method used to establish the resolution model was to acquire data of a point source at a high number of positions throughout the field of view [58, 67] and measure the system response. Due to limitations of data storage, it was not feasible to store the measured point spread response data for all the acquired positions and so a two-component model was established based upon a limited set of the measurements. The first component describes the shape and extent of the radial point spread response and consists of two half-Gaussian distributions with differing widths such that the half-Gaussian on the inner side of the point source has a greater width due to crystal depth of interaction effects. The second component describes how the point spread function varies across the field of view. This model of the point spread response is then incorporated into the system matrix. The benefits of resolution modelling include an improvement in spatial resolution of the reconstructed image, which results in better contrast recovery of both cold and hot focal areas [66, 67]. The improvements in spatial resolution as a result of resolution modelling improve the visualisation of small objects and hence have been shown to improve detection of small lesions [68].

Resolution modelling requires a greater degree of richness within the system matrix. This results in voxels in the reconstructed image contributing to more sinogram bins during forward projection and vice-versa during backprojection. This gives rise to an increased level of correlation of near-neighbouring voxels within the reconstructed image [66]. A consequence is that the spatial variance of voxel values in the reconstructed image, generally perceived visually as high frequency noise, is reduced. This is reinforced by the different rates of convergence of high frequency (i.e. noise) components of the image with and without resolution modelling. With resolution modelling, convergence is slower for high frequencies than without resolution modelling and hence when using typical clinical reconstruction parameters, the high frequency components are more apparent without resolution modelling [69].

There are several potential disadvantages to resolution modelling. The first and most significant is Gibbs artefacts. This artefact causes an overshoot at boundaries of

differing activity concentrations and gives rise to an edge-enhancement [57, 61, 70]. Possible suggested causes of the Gibbs artefact include an overestimation of the width of the point spread function that is incorporated into the system matrix compared with the true point spread function of the system. An alternative explanation is that the high frequency components that are lost during the image acquisition cannot be truly recovered during the reconstruction [57]. This can lead to a positive bias in activity concentration measurements for small areas of high activity concentration. There is also a certain size of object where the two overshoot artefacts constructively interfere and cause a more substantial positive bias [70]. Traditionally SUV, in particular SUV_{max}, has a monotonic response with the size of object [71], that reporting clinicians have been accustomed to for many years. The use of resolution modelling removes this relationship [66], as shown by Figure 3.11, which can cause difficulty with SUV interpretation.

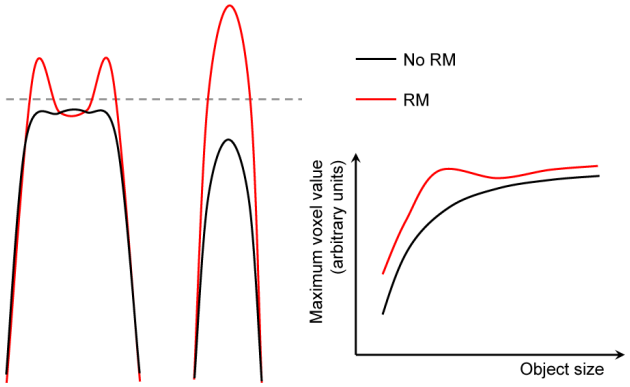


Figure 3.11 Schematic illustration of the Gibbs artefact that is a consequence of iterative reconstruction with resolution modelling (RM). The image on the left represents a profile through two objects of differing sizes. The larger object shows the two Gibbs overshoot artefacts separated at the object boundary. The smaller object shows the two Gibbs overshoot artefacts overlying and hence constructively interfering. The dashed grey line shows the true activity concentration. The schematic plot on the right shows how the maximum voxel value in an object is related the object size and in particular how the application of RM removes the monotonic relation between the two [66, 70].

Alternative SUV variants to SUV_{max} have been suggested, particularly in therapy monitoring, such as SUV_{peak} [72]. This variant of SUV is derived from the mean value of voxels encompassed by a 1 ml spherical region and was proposed

specifically for use in assessment of response to cancer therapy which will be performed over two or more PET scans. With the use of region-mean metrics there is a need to appreciate their variability. This is quantified by ensemble variance (EV), which is the variance of these region-mean measurements over a number of replicate images. A further and more disputed disadvantage to resolution modelling is its impact on EV. For small regions, the increased voxel correlation due to resolution modelling has been suggested to increase the EV [73], making it a less favourable choice of reconstruction for such clinical tasks [74].

3.6.4 Maximum A Posteriori (MAP) algorithms

As stated, the number of image updates for either MLEM or OSEM required to obtain convergence is at least several hundred [62]. The OSEM algorithm accelerates this convergence by reducing the number of required iterations but the number of updates remains unchanged. The visual image quality is degraded by noise and so convergence is not usually reached and instead, the process is terminated early to control noise. This is performed at the expense of quantitative accuracy. Alternatively, more updates can be performed and the final image is smoothed at the expense of resolution, but again it is rare that convergence is achieved. One strategy is to include a penalty function during the update process of the MLEM/OSEM equation. Put simply, the penalty factor acts to reduce the magnitude of the update coefficients such that equation 3.12 is modified to become

$$C_j^n = \frac{1}{\sum_i S_{ij} + \beta\Phi} \cdot \sum_i S_{ij} K_i^n , \tag{3.14}$$

where β is a weighting factor and Φ is a particular function, referred to as a *prior*, which is derived from the voxels within the image. The prior is usually derived from the absolute or relative differences of a voxel with its near-neighbours. A high difference could be indicative of noise and would increase the magnitude of the prior, reducing (or penalizing) the amount of change applied to the voxel. The prior is usually determined from the image estimation of the previous iteration and this is referred to as a *one-step-late* technique [75]. Strictly, there is no right or wrong form of this prior function and several have been proposed [75–77].

The only commercial implementation of a MAP reconstruction algorithm is from GE, referred to as *Q.Clear*, which is a penalized OSEM algorithm that uses a relative difference prior function [77] as the penalty. Initial reports of this algorithm demonstrate substantial improvements in signal-to-noise compared with conventional OSEM algorithms [78, 79]. Like resolution modelling the changes to quantitative FDG uptake measurements, in particular $SUV_{max,}$ in clinical PET pose a challenge to the community. Algorithms that include a prior term are usually referred to as having *regularisation*.

3.6.5 Time of flight reconstruction

TOF PET can be implemented using traditional analytical methods or iterative methods. In analytical methods, the backprojection operator is weighted according to the positional uncertainly along the LOR that is ascertained from the TOF difference that has been measured. This technique has been referred to as confidence weighting [80]. A comparison of performance between FBP-TOF with confidence weighted backprojection and MLEM with TOF reported superior results, by way of reduced voxel variance, with the MLEM algorithm [81]. For the purposes of this review, the focus shall remain on TOF as part of iterative reconstruction (OSEM) and, in particular, the method implemented on the Siemens Biograph mCT.

It has been stated that for TOF acquisition on the Biograph mCT, the LORs are binned into one of 13 TOF sinograms according to the measured TOF difference of the two gamma-rays. The temporal widths of the bins are 312 ps and the scanner has a quoted FWHM timing resolution of 530 ps [5]. The process of forward and backprojection is similar to non-TOF OSEM but an additional probability term is defined, based on TOF difference that is independent to the traditional geometric system matrix that is used for forward and backprojection. This probability function is referred to as the *Time Spread Function* (TSF) [82]. The reconstructed image estimate is forward projected into each TOF sinogram (t=1, t=2…t=13), by a weighting according to the TSF, and compared with the measured data in each TOF sinogram.

Therefore equation 3.8 is modified to contain the term T_{ijt}, which is the probability of a LOR originating from voxel j with a given angle contributing to a given TOF sinogram:

$$y_{it}^{n} = T_{ijt} \cdot \sum_{j} S_{ij} \lambda_{j}^{n} , \tag{3.15}$$

where y_{it}^{n} is now the value of the i^{th} bin in the estimated t^{th} TOF sinogram. Revisiting the diagram in figure 2.4, if we assume that the horizontal LOR corresponds to sinogram bin A and the vertical LOR corresponds to sinogram bin B, it follows that the T_{Ajt} will be greatest for t = +2 TOF sinogram and T_{Bjt} will be greatest for t = 0 TOF sinogram. The process continues along the same principle as standard OSEM such that the sinogram bin ratios of the acquired data in each TOF sinogram compared with each of the corresponding estimated TOF sinograms are determined as.

$$K_{it}^{n} = \frac{Y_{it}^{measured}}{y_{it}^{n} + r_i + s_i} = \frac{Y_{it}^{measured}}{T_{ijt} \cdot \sum_{j} S_{ij} \lambda_{j}^{n} + r_i + s_i}, \tag{3.16}$$

Finally, the update image coefficients are calculated by projecting the data from all of the TOF sinograms using the TSF probability.

$$C_{j}^{n} = \frac{1}{\sum_{i} S_{ij}} \cdot \sum_{t} T_{ijt} \sum_{i} S_{ij} K_{it}^{n} . \tag{3.17}$$

Given that the number of forward and backprojection operations to perform has now increased 13-fold, it is clear that the reconstruction process takes considerably longer and involves considerably greater degrees of computer memory to process. On the Biograph mCT at CMUH, reconstruction using TOF information does take longer but the overhead is not as severe as one may expect. A non-TOF and TOF reconstruction of an image acquired from one single bed position with three iterations takes 45 and 70 seconds respectively.

3.7 References

1, Beyer T, Czernin J, Freudenberg LS. *Variations in Clinical PET/CT Operations: Results of an International Survey of Active PET/CT Users*. J Nucl Med 2011; **52**:303-310

2, Natterer F. *The mathematics of computerized tomography*. Society for Industrial and Applied Mathematics, 2001

3, Cherry SR, Dahlbom M, Hoffman EJ. *3D PET using a Conventional Multislice Tomograph without Septa*. J Comput Assist Tomogr 1991; **15**:655-668

4, Fahey FH. *Data Acquisition in PET Imaging*. J Nucl Med Technol 2002; **30**:39-49

5, Jakoby BW, Bercier Y, Conti M, Bendriem B, Townsend D. *Physical and clinical performance of the mCT time-of-flight PET/CT scanner*. Phys Med Biol 2011; **56**:2375-2389

6, Casey ME, Hoffman EJ. *Quantitation in positron emission computed tomography: 7. A technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration*. J Comput Assist Tomogr 1986; **10**:845-850

7, Bergstrom M, Bohm C, Ericson K, Eriksson L, Litton J. *Corrections for Attenuation, Scattered Radiation, and Random Coincidences in a Ring Detector Positron Emission Transaxial Tomograph*. IEEE Trans Nucl Sci 1980; **27**:549-554

8, Eriksson L, Bohm C, Kesselberg M, Blomqvist G, Litton J, Widen L, et al. *A Four Ring Positron Camera System for Emission Tomography of the Brain*. IEEE Trans Nucl Sci 1982; **29**:539-543

9, Cooke BE, Evans AC, Fanthome EO, Alarie R, Sendyk AM. *Performance Figures and Images from the Therascan 3128 Positron Emission Tomograph*. IEEE Trans Nucl Sci 1984; **31**:640-644

10, Brix G, Zaers J, Adam LE, Bellemann ME, Ostertag H, Trojan H, et al. *Performance evaluation of a whole-body PET scanner using the NEMA protocol. National Electrical Manufacturers Association*. J Nucl Med 1997; **38**:1614-1623

11, Herzog H, Tellmann L, Hocke C, Pietrzyk U, Casey ME, Kuwert T. *NEMA NU2-2001 guided performance evaluation of four Siemens ECAT PET scanners*. IEEE Trans Nucl Sci 2004; **51**:2662-2669

12, Brambilla M, Secco C, Dominietto M, Matheoud R, Sacchetti G, Inglese E. *Performance characteristics obtained for a new 3-dimensional lutetium oxyorthosilicate-based whole-body PET/CT scanner with the National Electrical Manufacturers Association NU 2-2001 standard*. J Nucl Med 2005; **46**:2083-2091

13, Hoffman EJ, Huang SC, Phelps ME, Kuhl DE. *Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences*. J Comput Assist Tomogr 1981; **5**:391-400

14, Badawi RD, Miller MP, Bailey DL, Marsden PK. *Randoms variance reduction in 3D PET*. Phys Med Biol 1999; **44**:941-954

15, Rahmim A, Cheng J-C, Blinder S, Camborde M-L, Sossi V. *Statistical dynamic image reconstruction in state-of-the-art high-resolution PET*. Phys Med Biol 2005; **50**:4887-4912

16, Bailey DL, Meikle SR. *A convolution-subtraction scatter correction method for 3D PET*. Phys Med Biol 1994; **39**:411-424

17, Bentourkia M, Msaki P, Cadorette J, Lecomte R. *Nonstationary Scatter Subtraction-Restoration in High-Resolution PET*. J Nucl Med 1996; **37**:2040-2046

18, Grootoonk S, Spinksy TJ, Sashinx D, Spyrouk NM, Jones T. *Correction for scatter in 3D brain PET using a dual energy window method*. Phys Med Biol 1996; **41**:2757-2774

19, Shao L, Freifelder R, Karp JS. *Triple energy window scatter correction technique in PET*. IEEE Trans Med Imaging 1994; **13**:641-648

20, Thompson CJ. *The problem of scatter correction in positron volume imaging*. IEEE Trans Med Imaging 1993; **12**:124-132

21, Hiltz LG, McKee BTA. *Scatter correction for three-dimensional PET based on an analytic model dependent on source and attenuating object*. Phys Med Biol 1994; **39**:2059-2071

22, Ollinger J. *Model-based scatter correction for fully 3D PET*. Phys Med Biol 1996; **41**:153-176

23, Watson CC, Newport D, Casey ME, deKemp RA, Beanlands RS, Schmand M. *Evaluation of simulation-based scatter correction for 3-D PET cardiac imaging*. IEEE Trans Nucl Sci 1997; **44**:90-97

24, Klein O, Nishina Y. *Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac*. Z Phys 1929; **52**:853-868

25, Dscrages. *The Klein-Nishina distribution of photon scattering angles over a range of energies*. By Dscraggs (Own work) [Public domain], via Wikimedia Commons. https://commons.wikimedia.org/wiki/File%3AKlein-Nishina_distribution.png. Accessed 19th August 2016

26, Miwa K, Umeda T, Murata T, Wagatsuma K, Miyaji N, Terauchi T, et al. *Evaluation of scatter limitation correction: a new method of correcting photopenic artifacts caused by patient motion during whole-body PET/CT imaging*. Nucl Med Commun 2016; **37**:147-154

27, Ye J, Song X, Hu Z. *Scatter correction with combined single-scatter simulation and Monte Carlo simulation for 3D PET*. 2014 IEEE Nuclear Science Symp Conf Record; 1-3

28, Michel C, Liu X, Sanabria S, Lonneux M, Sibomana M, Bol A, et al. *Weighted schemes applied to 3D-OSEM reconstruction in PET*. 1999 IEEE Nuclear Science Symp Conf Record; 1152-1157

29, Huang SC, Hoffman EJ, Phelps ME, Kuhl DE. *Quantitation in positron emission computed tomography: 2. Effects of inaccurate attenuation correction*. J Comput Assist Tomogr 1979; **3**:804-814

30, Siegel S, Dahlbom M. *Implementation and evaluation of a calculated attenuation correction for PET*. IEEE Trans Nucl Sci 1992; **39**:1117-1121

31, Thompson CJ, Dagher A, Lunney DN, Strother SC, Evans AC. *A Technique To Reject Scattered Radiation In PET Transmission Scans*. Proc SPIE 0671, Physics and Engineering of Computerized Multidimensional Imaging and Processing 1986; 244-253

32, Huesman R, Derenzo SE, Cahoon JL, Geyer AB, Moses WW, Uber DC, et al. *Orbiting transmission source for positron tomography*. IEEE Trans on Nucl Sci 1988; **35**:735-739

33, Carson RE, Daube-Witherspoon ME, Green MV. *A Method for Postinjection PET Transmission Measurements with a Rotating Source*. J Nucl Med 1988; **29**:1558-1567

34, Dahlbom M, Hoffman EJ. *Problems in Signal-to-Noise Ratio for Attenuation Correction in High Resolution PET*. IEEE Trans Nucl Sci 1987; **34**:288-293

35, Karp JS, Muehllehner G, He Q, Xiao-Hong Y. *Singles transmission in volume-imaging PET with a 137 Cs source*. Phys Med Biol 1995; **40**:929 - 944

36, Politte DG, Snyder DL. *Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography*. IEEE Trans Med Imaging 1991; **10**:82-89

37, Kinahan PE, Townsend DW, Beyer T, Sashin D. *Attenuation correction for a combined 3D PET/CT scanner*. Med Phys 1998; **25**:2046-2053

38, Burger C, Goerres G, Schoenes S, Buck A, Lonn A, von Schulthess G. *PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients*. Eur J Nucl Med Mol Imaging 2002; **29**:922-927

39, Casey ME, Gadagkar H, Newport D. *A component based method for normalization in volume PET*. Proc 3rd Int Meeting Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine 1995:67-71

40, Hoffman EJ, Guerrero TM, Germano G, Digby WM, Dahlbom M. *PET system calibrations and corrections for quantitative and spatially accurate images*. IEEE Trans Nucl Sci 1989; **36**:1108-1112

41, Ollinger JM. *Detector efficiency and Compton scatter in fully 3D PET*. IEEE Trans Nucl Sci 1995; **42**:1168-1173

42, Kinahan PE, Townsend DW, Bailey DL, Sashin D, Jadali F, Mintun MA. *Efficiency normalization techniques for 3D PET data*. 1995 IEEE Nuclear Science Symp Conf Record; 1021-1025

43, Theodorakis L, Loudos G, Prassopoulos V, Kappas C, Tsougos I, Georgoulias P. *A review of PET normalization: striving for count rate uniformity*. Nucl Med Commun 2013; **34**:1033-1045

44, Badawi RD, Marsden PK. *Developments in component-based normalization for 3D PET*. Physics in Medicine and Biology 1999; **44**:571-594

45, deKemp R, Nahmias C. *Deadtime correction for non-paralyzable block detectors in PET*. 1992 IEEE Engineering Medicine and Biology Society Proceedings; 1825-1826

46, Daube-Witherspoon ME, Carson RE. *Unified deadtime correction model for PET*. IEEE Trans Med Imaging 1991; **10**:267-275

47, Evans RD. *Applications of Poisson Statistics to Some Instruments Used in Nuclear Physics*. In: Evans RD. *The Atomic Nucleus*. McGraw-Hill, 1955

48, Germano G, Hoffman EJ. *A study of data loss and mispositioning due to pileup in 2-D detectors in PET*. IEEE Trans Nucl Sci 1990; **37**:671-675

49, Bruyant PP. *Analytic and iterative reconstruction algorithms in SPECT*. J Nucl Med 2002; **43**:1343-1358

50, Shepp LA, Vardi Y. *Maximum Likelihood Reconstruction for Emission Tomography*. IEEE Trans Med Imaging 1982; **1**:113-122

51, Lange K, Carson R. *EM reconstruction algorithms for emission and transmission tomography*. J Comput Assist Tomogr 1984; **8**:306-316

52, Daube-Witherspoon ME, Carson RE, Yan YC, Yap TK. *Scatter correction in maximum-likelihood reconstruction of PET data*. 1992 IEEE Nuclear Science Symp Conf Record; 945-947

53, Barrett HH, Wilson DW, Tsui BM. *Noise properties of the EM algorithm: I. Theory*. Phys Med Biol 1994; **39**:833-846

54, Nuyts J. *On estimating the variance of smoothed MLEM images*. IEEE IEEE Trans Nucl Sci 2002; **49**:714-721

55, Alpert NM, Chesler DA, Correia JA, Ackerman RH, Chang JY, Finklestein S, et al. *Estimation of the Local Statistical Noise in Emission Computed Tomography*. IEEE Trans Med Imaging 1982; **1**:142-146

56, Snyder DL, Miller MI. *The Use of Sieves to Stabilize Images Produced with the EM Algorithm for Emission Tomography*. IEEE Trans Nucl Sci 1985; **32**:3864-3872

57, Snyder DL, Miller MI, Thomas LJ, Politte DG. *Noise and Edge Artifacts in Maximum-Likelihood Reconstructions for Emission Tomography*. IEEE Trans Med Imaging 1987; **6**:228-238

58, Panin VY, Kehren F, Rothfuss H, Hu D, Michel C, Casey ME. *PET reconstruction with system matrix derived from point source measurements*. IEEE Trans Nucl Sci 2006; **53**:152-159

59, Vardi Y, Shepp LA, Kaufman L. *A Statistical Model for Positron Emission Tomography*. J Amer Statist Assoc 1985; **80**:8-20

60, Liow JS, Strother SC. *Performance of maximum likelihood based image reconstruction for quantitative imaging*. 1991 IEEE Nuclear Science Symp Conf Record; 2078-2083

61, Liow JS, Strother SC. *Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions*. IEEE Trans Med Imaging 1991; **10**:563-571

62, Carson RE, Yuchen Y, Chodkowski B, Yap TK, Daube-Witherspoon ME. *Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm*. IEEE Trans Med Imaging 1994; **13**:526-537

63, Hudson HM, Larkin RS. *Accelerated image reconstruction using ordered subsets of projection data*. IEEE Trans Med Imaging 1994; **13**:601-609

64, Morey AM, Kadrmas DJ. *Effect of varying number of OSEM subsets on PET lesion detectability*. J Nucl Med Technol 2013; **41**:268-273

65, De Bernardi E, Mazzoli M, Zito F, Baselli G. *Resolution Recovery in PET During AWOSEM Reconstruction: A Performance Evaluation Study*. IEEE Trans Nucl Sci 2007; **54**:1626-1638

66, Alessio AM, Stearns CW, Shan T, Ross SG, Kohlmyer S, Ganin A, et al. *Application and Evaluation of a Measured Spatially Variant System Model for PET Image Reconstruction*. IEEE Trans Med Imaging 2010; **29**:938-949

67, Panin VY, Kehren F, Michel C, Casey M. *Fully 3-D PET reconstruction with system matrix derived from point source measurements*. IEEE Trans Med Imaging 2006; **25**:907-921

68, Kadrmas DJ, Casey ME, Black NF, Hamill JJ, Panin VY, Conti M. *Experimental Comparison of Lesion Detectability for Four Fully-3D PET Reconstruction Schemes*. IEEE Trans Med Imaging 2009; **28**:523-534

69, Rahmim A, Tang J. *Noise propagation in resolution modeled PET imaging and its impact on detectability*. Phys Med Biol 2013; **58**:6945-6968

70, Bai B, Esser PD. *The effect of edge artifacts on quantification of Positron Emission Tomography*. IEEE Nuclear Science Symp Conf Record; 2263-2266

71, Hoffman EJ, Huang SC, Phelps ME. *Quantitation in positron emission computed tomography: 1. Effect of object size*. J Comput Assist Tomogr 1979; **3**:299-308

72, 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**:122S-150

73, Rahmim A, Qi J, Sossi V. *Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls*. Med Phys 2013; **40**:064301

74, Alessio AM, Rahmim A, Orton CG. *Resolution modeling enhances PET imaging*. Med Phys 2013; **40**:120601

75, Green PJ. *Bayesian reconstructions from emission tomography data using a modified EM algorithm*. IEEE Trans Med Imaging 1990; **9**:84-93

76, Alenius S, Ruotsalainen U. *Bayesian image reconstruction for emission tomography based on median root prior*. Eur J Nucl Med 1997; **24**:258-265

77, Nuyts J, Beque D, Dupont P, Mortelmans L. *A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography*. IEEE Trans Nucl Sci 2002; **49**:56-60

78, Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV. *Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System*. J Nucl Med 2015; **56**:1447-1452

79, Teoh EJ, McGowan DR, Bradley KM, Belcher E, Black E, Gleeson FV. *Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules*. Eur Radiol 2016; **26**:576-584

80, Snyder DL, Thomas LJ, Ter-Pogossian MM. *A Mathematical Model for Positron-Emission Tomography Systems Having Time-of-Flight Measurements*. IEEE Trans Nucl Sci 1981; **28**:3575-3583

81, Politte DG. *Image improvements in positron-emission tomography due to measuring differential time-of-flight and using maximum-likelihood estimation*. IEEE Trans Nucl Sci 1990; **37**:737-742

82, Conti M, Bendriem B, Casey M, Mu C, Kehren F, Michel C, et al. *Implementation of time-of-flight on CPS HiRez PET scanner*. 2004 IEEE Nuclear Science Symp Conf Record; 2796-2800

4. FIRST PAPER

Impact of Advances in PET Reconstruction Algorithms and Uptake Measurement Methods on the Accuracy and Variability of SUV

Article status:

Submitted to Physics in Medicine and Biology in July 2012

Rejection decision received October 2012 following review

Revised and submitted to Medical Physics in May 2013

Rejection decision received May 2013 following review

The paper included in this chapter is the manuscript that was submitted to Medical Physics in May 2013

Impact of Advances in PET Reconstruction Algorithms and Uptake Measurement Methods on the Accuracy and Variability of SUV

I S Armstrong[1,2], M D Kelly[3], H A Williams[1] and J C Matthews[2]
[1] Nuclear Medicine, Central Manchester University Hospitals, Manchester, UK.
[2] University of Manchester, MAHSC, Institute of Population Health, Manchester UK
[3] Siemens PLC, Healthcare Sector, Molecular Imaging, Oxford, UK.

E-mail: ian.armstrong@cmft.nhs.uk

Abstract. Standardized Uptake Value measurements from FDG PET images are routinely used for staging and monitoring response to therapy. New iterative reconstruction algorithms that include point spread function modeling and time of flight have been shown to improve image voxel signal to noise. This study aims to characterize the impact of these algorithms on the activity concentration accuracy and variability of three commonly used uptake measurement techniques. Replicate PET/CT images were acquired of the NEMA image quality phantom, with three different sphere-to-background activity concentration ratios. Images were reconstructed with conventional OSEM iterative reconstruction, OSEM with point spread function (PSF) modeling and OSEM with combined PSF modeling and time of flight (TOF) and two post-filters were used. Maximum, mean and peak activity concentrations in each sphere were measured mimicking SUV methods used clinically for tumors. Substantial benefits were seen for SUV_{max} measures with OSEM+PSF or PSF+TOF compared with OSEM: firstly considerable reduction in the dependency of recovery on sphere size and secondly a marked reduction of measurement variability. These improvements were also noted for SUV_{peak} but were far less pronounced than those seen in SUV_{max}, which highlighted the more robust nature of SUV_{peak} measurement technique.

I. Introduction

Semi-quantitative uptake measurements of [^{18}F]fluoro-deoxyglucose (FDG) are commonplace in oncology PET/CT imaging. The most frequent measure of FDG uptake in tumors being the Standardized Uptake Value (SUV)[1]. Several volumetric methods can be used to produce SUVs characterizing tumors with the most

frequently used being SUV_{max}[2], which is derived from the maximum voxel usually within a 3D volume of interest (VOI) delineating the tumor extent. Other methods compute the mean of all voxels within a VOI, typically defined with an isocontour corresponding to a particular percentage of SUV_{max}. Recently the SUV_{peak} metric was proposed for use in monitoring response to therapy[2]. SUV_{peak} is computed as the maximum of the mean intensities within a set of 1.0 ml spherical VOI centered at voxels within the tumor volume.

In combination with the pattern of FDG uptake, SUV is used to support discrimination between benign and malignant lesions[3–5], and changes in SUV are used to assess response to therapy[2,6]. For both applications it is desirable that SUVs have a low dependency on uncontrollable factors such as lesion size and contrast but perhaps the most important factor is high precision, ensuring good consistency of SUVs.

Despite being the most common uptake measure, SUV_{max} has been shown to be more susceptible to positive bias in phantom measurements than other SUV methods with voxel averaging – particular as the size of object increases[7]. In patient studies, it has also been shown to perform worse in reproducibility studies than alternative SUV methods[8–9] and recently, it has been shown that SUV_{peak} offers improved reproducibility than SUV_{max}[10]. Phantom studies using multiple replicate images have demonstrated that the variability of maximum voxel (activity concentration) AC measurements (analogous with SUV_{max}) is greater than AC measurements with voxel averaging[11–13]. All of these studies have been derived from images reconstructed with either filtered backprojection or "conventional" iterative reconstruction algorithms – that is, algorithms that do not include point spread function (PSF) modeling or time of flight (TOF), which are now commercially available.

The accuracy and variability of AC measurements is dependent on the levels of statistical noise within the images[7,11–13]. The inclusion of PSF modeling in iterative reconstruction compared with non-PSF reconstruction has been shown to produce lower voxel variance measured in a uniform phantom background when matched and low numbers of iterations typical of clinical scans are used[14–16]. This may reduce the variability and likelihood of positive bias in AC measurements but it has been

demonstrated that PSF modeling can produce overshoot artifacts at object boundaries[16–18]. The impact of these artifacts on clinical uptake measures has yet to be quantified by other studies. It has long been understood that inclusion of TOF improves image voxel signal to noise ratio[19], and this has been demonstrated in phantom and patient studies[20–23].

Several studies have presented contrast recovery results using the NEMA image quality (IQ) phantom[24] for images with PSF modeling and/or TOF reconstruction[15,20–22,25–27]. These studies were performed according to the NEMA Standard[24] and, as such, did not explore a wide range of object contrasts or clinically relevant AC measurement techniques.

PSF and TOF-based reconstruction algorithms have been shown to significantly increase SUVs in patient studies[28–31] with average increases of SUV_{max} being up to 50%. Despite this, no studies to date have investigated the effect of these advanced iterative reconstruction algorithms on accuracy and variability in phantoms where the ground truth AC is known. This work aims to bring together several key areas of investigation that, to our knowledge, have not yet been performed as a single body of work. These are 1) comparing PSF modeling and TOF reconstruction separately on a common dataset; 2) comparing a range of clinically relevant contrasts not previously reported; 3) expanding measures of AC from those based on the NEMA protocol to include clinically relevant measures analogous to SUV_{max} and SUV_{peak} and 4) measuring both the accuracy and precision of these measurements for all algorithms through several independent acquisitions. The intention is to provide guidance to allow controllable factors – reconstruction parameters and measurement technique – to be chosen to minimize the dependence of SUV on the uncontrollable factors, enabling appropriate interpretation of SUVs for clinical data.

II. Materials and methods

II.A PET Scanner

The scanner used in this study was a Biograph mCT with 64 slice CT (Siemens Healthcare). The scanner acquires in 3D mode only and includes an extended axial field of view of 21.6 cm (TrueV) which is covered by four rings of detector blocks, each containing 13 × 13 LSO crystal arrays of size 4 mm × 4 mm × 20 mm. The

iterative reconstruction algorithms available on the scanner are 3D Ordinary Poisson Ordered Subset Expectation Maximization (3D-OP-OSEM), 3D-OP-OSEM that incorporates PSF modeling[14] (HD·PET) and 3D-OP-OSEM that incorporates PSF modeling combined with TOF (UltraHD·PET). The timing resolution of the scanner for TOF is specified as 550 ps. A performance assessment of this scanner has been recently published[27].

II.B Image acquisition and reconstruction

A NEMA IQ phantom (PTW, Freiberg, Germany), containing six spheres of 10, 13, 17, 22, 28 and 37 mm internal diameter, was filled with ^{18}F with a background AC of 5.2 kBq/ml. All six spheres were filled with ^{18}F at contrast ratios of 2:1, 4:1 and 8:1 relative to the background. The phantom was positioned such that the plane through the center of the six spheres was at the center of the PET axial field of view and the center of the lung insert was at the center of the transaxial field of view. A one-hour listmode acquisition was performed for each different phantom concentration ratio. The listmode data were divided into ten replicate sinograms each consisting of 6.0×10^7 net true coincidences, by performing a gated reconstruction using a regular ECG gating trigger from an ECG simulator. The number of counts was selected based on NEMA Standard.

Sinograms were reconstructed using the standard 3D-OP-OSEM (OSEM), 3D-OP-OSEM with PSF modeling (OSEM+PSF) and 3D-OP-OSEM with combined PSF modeling and TOF (OSEM+PSF+TOF). Standard reconstruction parameters were 1 to 12 iterations, 24 subsets for OSEM and OSEM+PSF, 21 subsets for OSEM+PSF+TOF and a 256×256 image matrix giving voxel dimensions of 3.2 mm × 3.2 mm × 2.0 mm. Each image consisted of 109 transaxial slices. Two different Gaussian post-filters, either 2 mm or 5 mm FWHM, were applied to the images. To summarize, the combination of parameters is as follows: 3 contrast ratios × 3 reconstruction algorithms × 2 post-filters × 12 iterations × 10 replicates per iteration = 2160 images produced and analyzed.

II.C Activity concentration measurements

All reconstructed images were processed offline in Matlab (The MathWorks, Massachusetts, USA) using processing routines developed in-house. For each sphere, maximum, peak and mean uptake values were calculated. For the maximum uptake measure, a 3D spherical VOI, equal in diameter to each sphere, was centered on the sphere and the maximum voxel value within the VOI was reported as "max-AC". For the peak uptake measure, a 1.0 ml spherical VOI (1.24 cm diameter) was centered over each voxel within each sphere. The mean of those voxels included within this 1.0 ml VOI was found. The max over the set of 1 ml VOIs was reported as "peak-AC". To illustrate the differences in these clinical values from that defined in the NEMA Standard, a third measure was made with a 2D circular region of interest, equal in diameter to each sphere, centered over the sphere and placed on the transaxial image slice that bisected all six spheres. The mean of all voxels within the ROI was reported as "mean-AC". As the true AC in the spheres is known, the uptake measurements were normalized to this true AC. This recovery measure is an idealized surrogate of the clinical SUV and has the advantage of providing a direct measure of the accuracy of the uptake measurement.

II.D Determination of accuracy and variability

To determine the accuracy of the AC values for each of the six individual spheres and three AC metrics described above, the mean AC measurement over the 10 replicate images was calculated and expressed as a percentage of the true AC.

To determine the variability of the AC values for each of the six individual spheres and three AC metrics, the coefficient of variation (COV) calculated from the mean and standard deviation of the AC values over the 10 replicate images.

To demonstrate the dependency of the AC values on the size of the sphere, the COV over the six individual mean sphere AC measurements was determined. A larger COV demonstrated a greater dependency of AC measurement on the sphere size.

III. Results

III.A Visual assessment of images

Figure 1(a) shows the transaxial PET images through the center of spheres in the three different contrast phantoms for 3 iterations of each reconstruction algorithm and a 2 mm post-filter. In a simple review of the images at 8:1 contrast, all six spheres are easily visualized by the authors with any algorithm. At 4:1 contrast, the smallest (10 mm) sphere can still be seen with OSEM, but its extent is not as well defined as with OSEM+PSF or OSEM+PSF+TOF reconstruction. At the lowest 2:1 contrast, the smallest sphere can just been seen in only OSEM+PSF+TOF images. With OSEM+PSF reconstruction, the second smallest (13 mm) sphere can just be seen whereas with OSEM reconstruction, it is much harder to visualize the spheres and only the larger four spheres can be seen.

The image noise characteristics are very different with OSEM+PSF and OSEM+PSF+TOF compared with OSEM reconstruction. The noise in the OSEM+PSF and OSEM+PSF+TOF images appears more correlated with less high spatial frequency variations in the background region compared with the OSEM images. This is attributed to the inclusion of PSF modeling and has been reported previously[16]. Figure 1(b) illustrates the presence of edge artifacts on the spheres when PSF modeling is used. On the larger three spheres, these can be seen but on the 17 and 13 mm spheres, the artefact overlaps and, in the case of the 13 mm sphere, appears to combine to increase the intensity of the sphere.

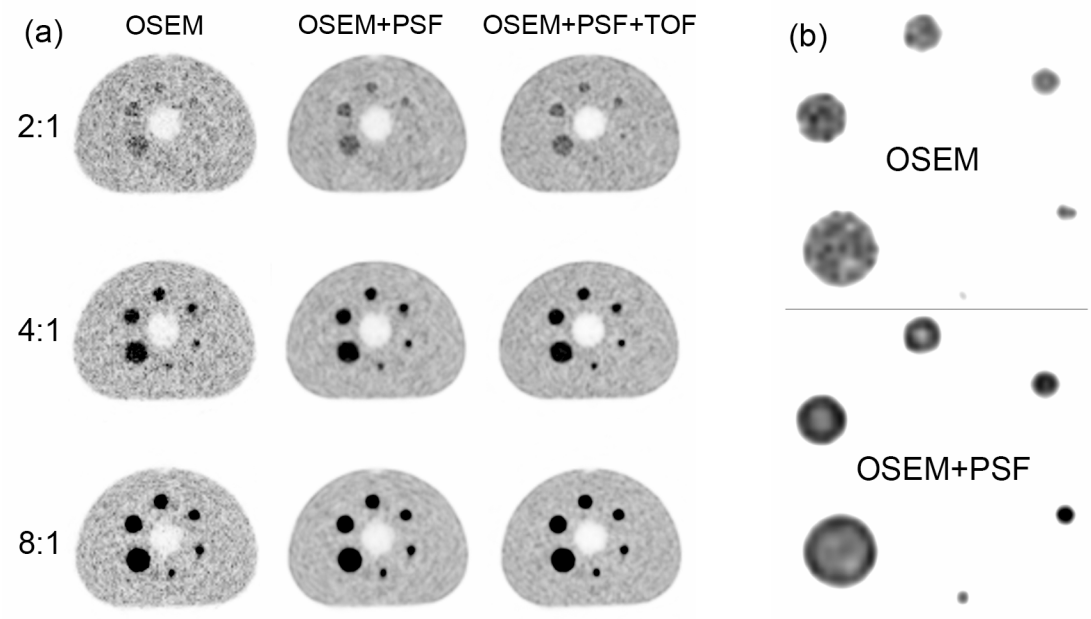


Figure 1. (a) Transaxial PET images through the center of the six spheres in the IQ phantom after three iterations of each reconstruction algorithm and 2 mm post-filter for the three contrast levels and three reconstruction algorithms examined. The white circle in the center of the phantom is the lung insert. The upper and lower window/level values have been set equally on all images. (b) Image of spheres from 8:1 phantom using OSEM and OSEM+PSF reconstruction. The window levels have been set to illustrate the presence of edge artifacts on the larger spheres.

III.B Impact of reconstruction algorithm

Due to the large quantities of data produced in this study, and the clinical prevalence of $SUV_{max,}$ the max-AC data will be used as the reference when showing the impact of reconstruction and sphere contrast as SUV_{max} is the most common clinical uptake measure.

III.B.1 Measurement accuracy

Max-AC recovery plots for the 8:1 phantom are shown in Figure 2. Max-AC increases monotonically with iterations for all algorithms with the 2 mm post-filter but not for the 5 mm post-filter. With the 2 mm post-filter, there is positive bias in all spheres for >2 iterations and >1 iteration of OSEM+PSF and OSEM+PSF+TOF respectively. Recovery with OSEM is lower than OSEM+PSF or OSEM+PSF+TOF in the smaller two spheres for >1 iteration with the 2 mm post-filter. As the sphere size increases, max-AC with OSEM increases, producing the greatest positive bias

across all reconstructions for the 22 mm to 37 mm spheres. Negative bias is seen in the smaller spheres particularly with OSEM, a 5 mm filter and when only a single iteration is used.

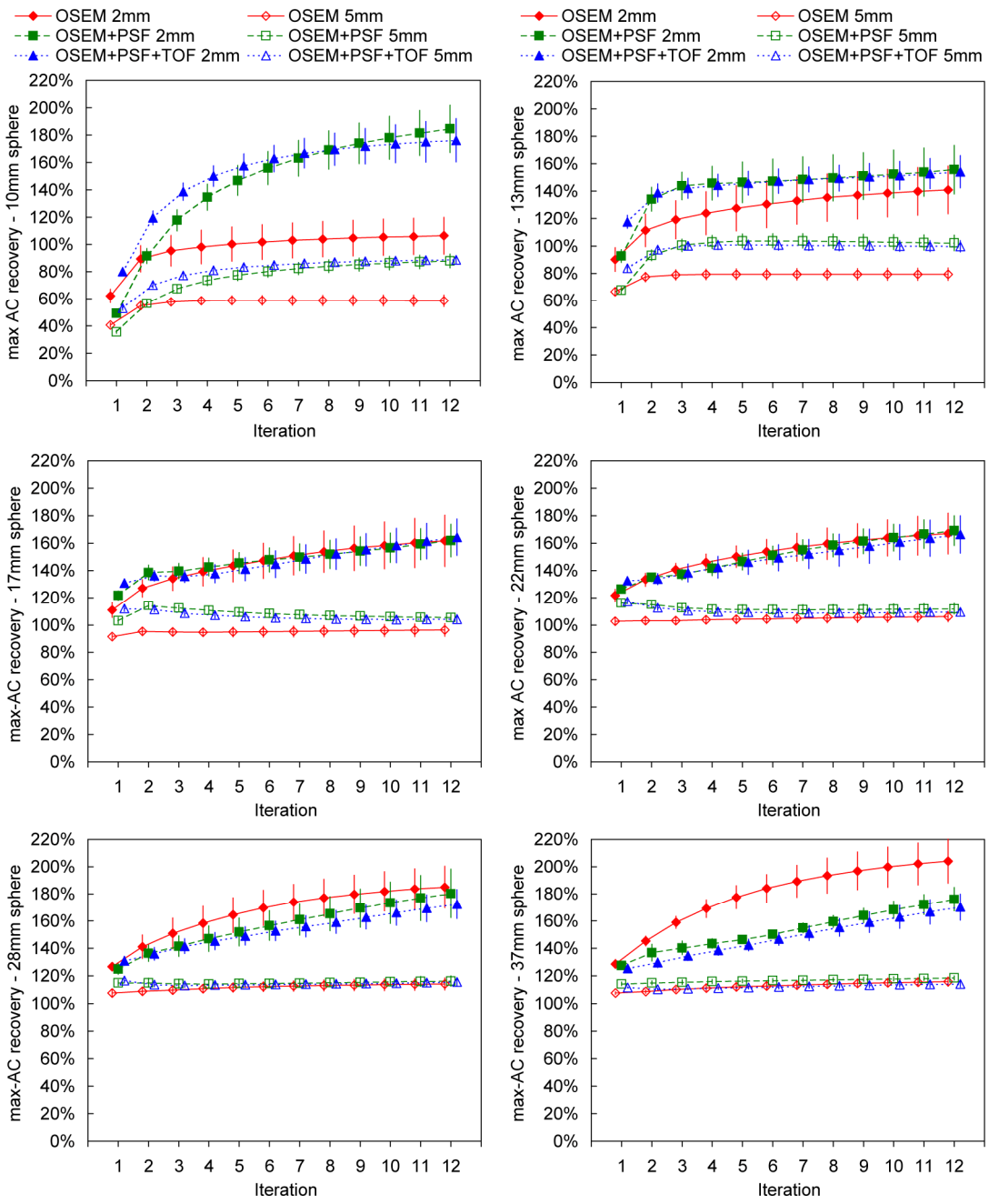


Figure 2. Max-AC recovery for each of the six spheres filled in the 8:1 phantom. Data points are derived from the mean of the measurements from the 10 image replicates and error bars represent the standard deviation from the 10 measurements. Each plot shows the results from the three reconstruction algorithms (color) and with a 2 mm (solid points) and 5 mm (hollow points) FWHM post-filter. To allow easier visualization of the error bars, the OSEM data have been offset to the left of each iteration point and the OSEM+PSF+TOF data have been offset to the right of each iteration point.

III.B.2 Measurement variability

As is shown from the error bars in Figure 2, variability is lower with the 5 mm post-filter than the 2 mm post-filter for all algorithms. Variability in the three smallest spheres is greatest with OSEM for any number of iterations and either post-filter. For < 4 iterations, the differences of COV for OSEM with a 5 mm post-filter are relatively modest compared with OSEM+PSF+TOF with a 2 mm filter. This illustrates that the inherent lower image noise of OSEM+PSF+TOF and results in the ability to use less post-filtering with minimal impact on SUV variability.

III.B.3 Dependency on sphere size

The reconstruction parameters that minimize the dependency of max-AC on sphere size (by minimizing the COV over the six spheres) are shown in Table I. As contrast increases, fewer iterations are required for all algorithms to minimize the dependency on size.

Table I. Reconstruction parameters (number of iterations, from 1–12, and post-filter FWHM of either 2 mm or 5 mm) to give minimum dependency of max-AC on sphere size as quantified by the coefficient of variation values across the six spheres (optimal value in brackets).

	Sphere-to-background concentration		
Algorithm	2:1	4:1	8:1
OSEM	12i + 5 mm	4i + 2 mm (22%)	2i + 2 mm (17%)
OSEM+PSF	12i + 5 mm	8i + 2 mm (12%)	5i + 2 mm (1.6%)
OSEM+PSF+TOF	12i + 5 mm	7i + 2 mm (5.3%)	3i + 2 mm (2.2%)

Max-AC for each sphere and contrast using the parameters in Table I are shown in Figure 3. The plots show that OSEM+PSF and OSEM+PSF+TOF can be used to substantially reduce the dependency of max-AC on sphere size compared with OSEM, particularly for the higher contrast cases. Despite OSEM+PSF offering the least dependency on sphere size (COV of 1.6% for 8:1 contrast), five iterations are required, compared with three using OSEM+PSF+TOF, and the variability on the AC values is greater (error bars in Figure 3).

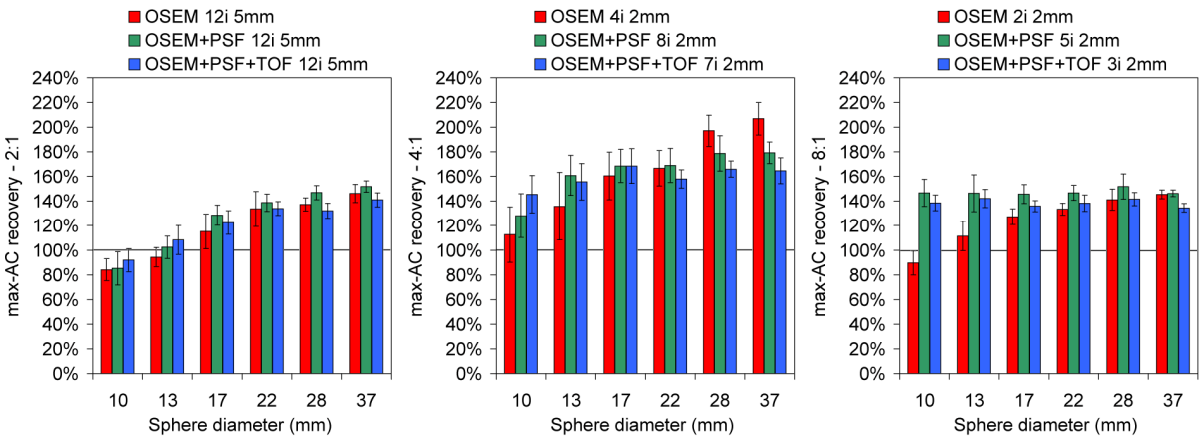


Figure 3. Max-AC values where the reconstruction parameters are set to minimize the variation of recovery across the different sphere sizes. Data are shown for 2:1 (left), 4:1 (center) and 8:1 (right) sphere-to-background concentration ratio. The parameters are different for each contrast ratio and are shown in the legend of each plot. The error bars indicate the standard deviation, obtained from the 10 replicate images, of the percentage recovery.

Despite this very low dependency on sphere size, the 8:1 measurements suffer from considerable positive bias (approximately 40% for OSEM+PSF+TOF 3 iterations, 2 mm post-filter and 45% for OSEM+PSF 5 iterations, 2 mm post-filter). This bias is reduced with the 5 mm post-filter but, as shown in Figure 2, recovery of the smaller spheres is reduced, which has the effect of increasing the dependency of recovery on sphere size.

III.C Comparison of AC measurement techniques

III.C.1 Measurement accuracy

Figure 4 shows recovery of the three AC measurements for the 10 mm and 37 mm sphere in the 8:1 phantom. A negative bias is seen for all mean-AC measurements for all sphere sizes and reconstruction algorithms and, as with max-AC, the mean-AC values are greater in the smallest sphere with OSEM+PSF and OSEM+PSF+TOF compared with OSEM. Mean-AC is seen to decrease as the sphere size reduces, which is most notable with OSEM which has the lowest values of the three algorithms for all sphere sizes at 8:1 contrast. Peak-AC is the most accurate of the measurement techniques in the four largest spheres. However, for the two smaller

spheres, peak-AC has the largest negative bias of the three techniques. The volume of the 10 mm and 13 mm spheres are 0.52 ml and 1.15 ml respectively so this bias is likely due to a combination of the inclusion of background voxels in the peak sphere VOI, and activity spill out from the edge voxels (i.e., partial volume effect).

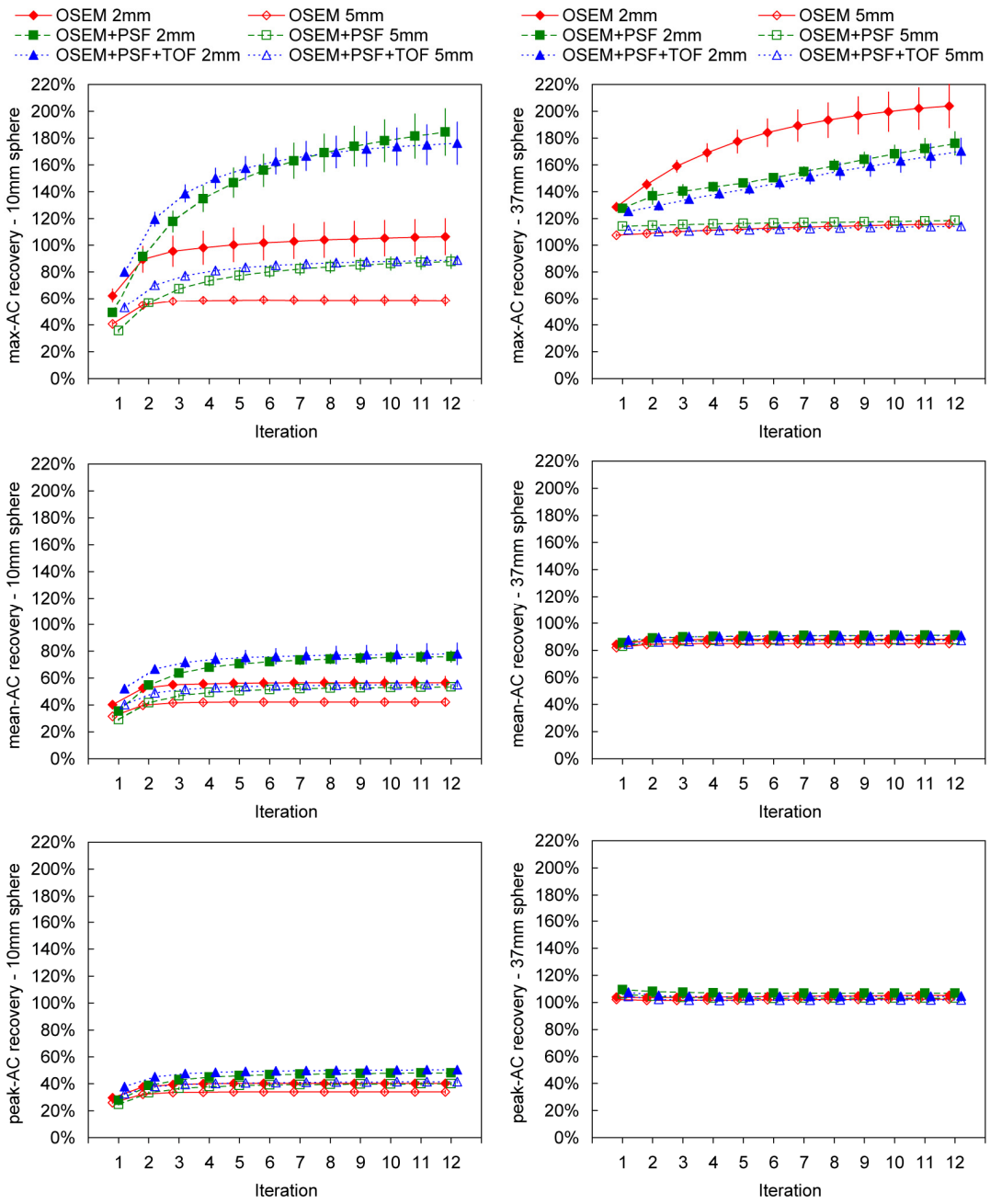


Figure 4. AC recovery for the 10 mm sphere (left column) and 37 mm sphere (right column) filled with 8:1 sphere-to-background concentration ratio using max-AC (top row), mean-AC (middle row) and peak-AC (bottom row). Each plot shows the results from the three reconstruction algorithms and with a 2 mm and 5 mm FWHM post-filter. The error bars indicate the standard deviation, obtained from the 10 replicate images, of the percentage recovery. The line types and x-axis offsets are as in figure 2.

III.C.2 Measurement variability

Figure 5 shows the COV of the three AC measurements for the 10 mm and 37 mm spheres with 8:1 contrast. The plots show that variability of max-AC is the most sensitive to reconstruction algorithm and post-filter. Variability of the mean-AC and peak-AC is comparable for larger spheres with all reconstructions, iteration number and post-filters. In the smaller spheres, variability of mean-AC increased, particularly with minimal post-filtering. Variability of peak-AC was low at 2-3% and showed only modest dependency on sphere size, reconstruction algorithm or post-filter.

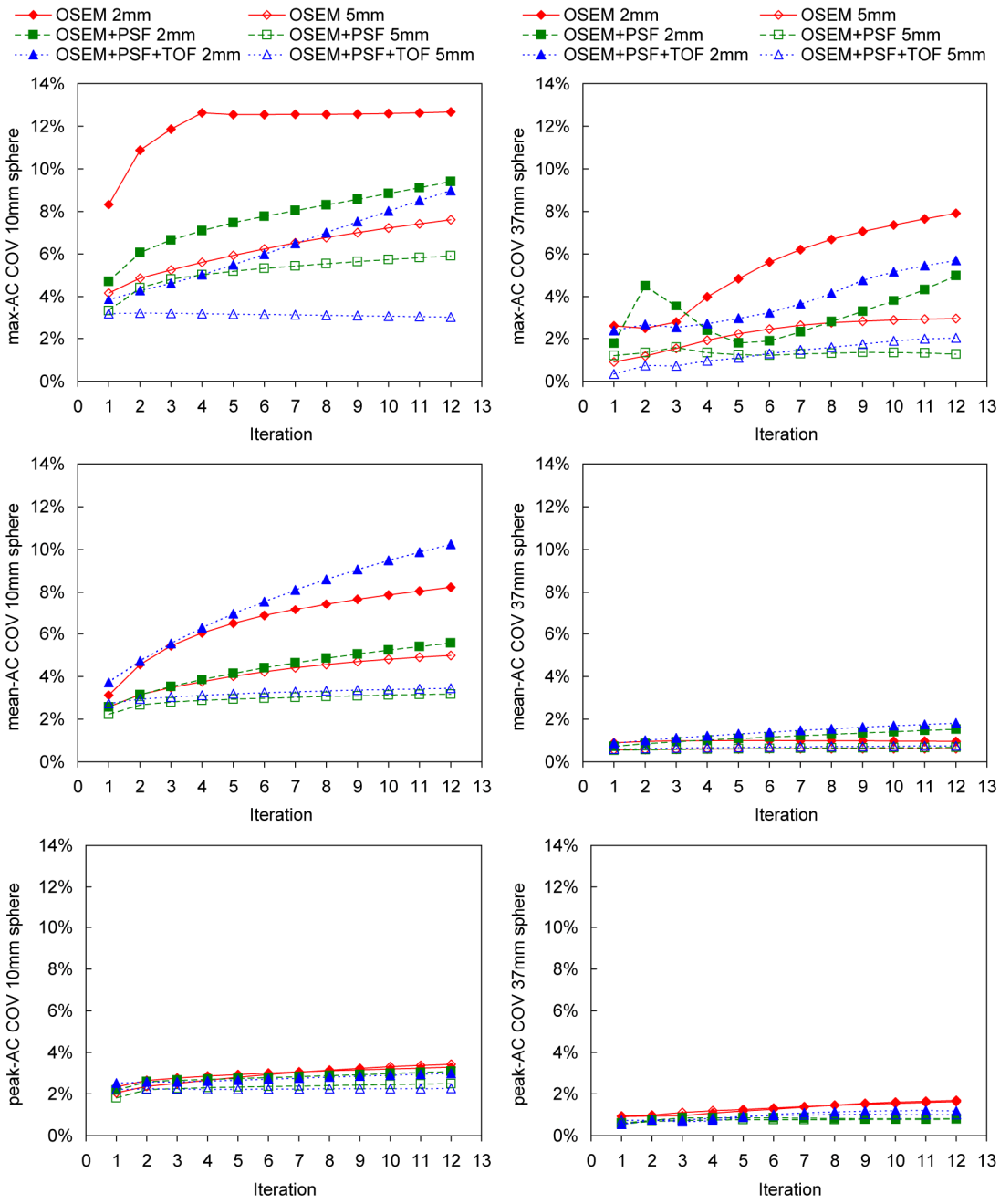


Figure 5. Coefficient of variation (COV) across image replicates of AC measurements for the 10 mm sphere (left column) and 37 mm sphere (right column) filled with 8:1 sphere-to-background concentration ratio using max-AC (top row), mean-AC (middle row) and peak-AC (bottom row). Each plot shows the results from the three reconstruction algorithms and with a 2 mm and 5 mm FWHM post-filter. The line types are as in Figure 2.

III.D Effect of sphere contrast

III.D.1 Measurement accuracy

Figure 6 shows the max-AC plots for the 10 mm and 37 mm spheres filled with the three different contrasts to the background. The plots show that at the 2:1 contrast with a 2 mm post-filter, max-AC is greatest with OSEM reconstruction compared with OSEM+PSF and OSEM+PSF+TOF, with a positive bias in nearly all cases. This is particularly evident as the sphere size increases, which has been shown previously[7]. There is a high dependence of max-AC on iteration number. For example, the max-AC for the 37 mm sphere with 2:1 contrast after 2 and 3 iterations of OSEM with a 2 mm post-filter was 180% and 220% respectively. By comparison, the equivalent recovery at 8:1 contrast was 140% and 160%. This number of iterations is within the range that is likely to be implemented clinically (in this institute, 3 iterations are standard).

As the contrast increases, the positive bias in max-AC reduces with OSEM for all sphere sizes with a 2 mm post-filter. Max-AC for the 10 mm sphere using OSEM+PSF and OSEM+PSF+TOF and 2 mm post-filter increases as the contrast increases. With the 5 mm post-filter, max-AC for the 10 mm sphere is comparable at 2:1 contrast and recovery with both OSEM+PSF and OSEM+PSF+TOF increases compared with OSEM as the contrast increases. It can be seen from the plot how, for all algorithms, the convergence rate of the 10 mm sphere increases as the sphere contrast increases.

III.D.2 Measurement variability

From the error bars on the plots in Figure 6, it can be seen that the lowest variability occurs at the highest contrast. For OSEM reconstruction, variability for 2:1 and 4:1 contrast is comparable in the smaller spheres, while for the OSEM+PSF and OSEM+PSF+TOF, the variability is less at 4:1 contrast.. As has been shown in previous figures, the variability decreases for the larger FWHM post-filters in all three contrast levels. For a particular contrast and post-filter, variability was greatest with OSEM.

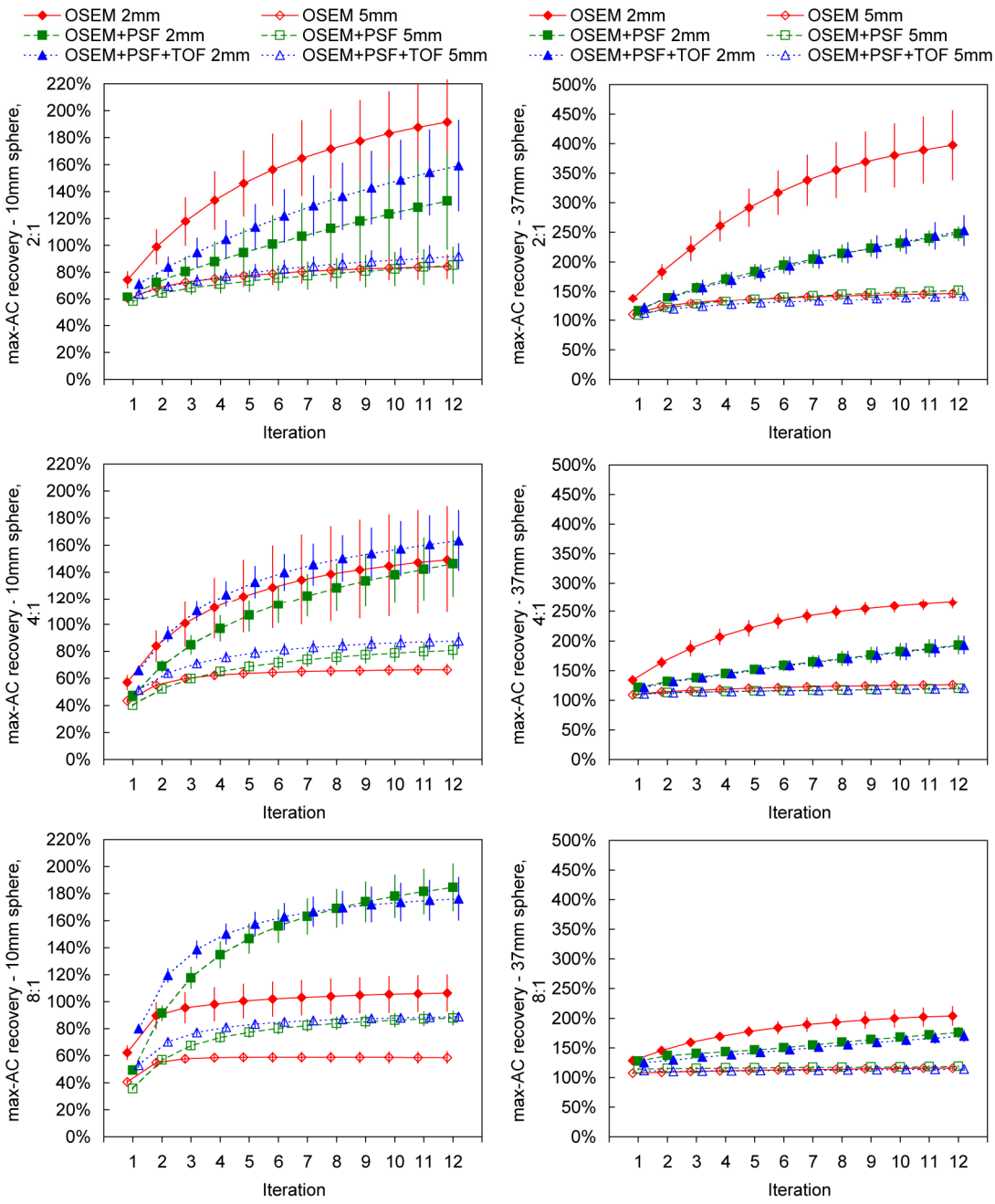


Figure 6. Max-AC for the 10 mm sphere (left column) and 37 mm sphere (right column) filled with 2:1 (top row), 4:1 (middle row) and 8:1 (bottom row) sphere-to-background concentration ratio. Each plot shows the results from the three reconstruction algorithms and with a 2 mm and 5 mm FWHM post-filter. The error bars indicate the standard deviation, obtained from the 10 replicate images, of the percentage recovery. Note that the vertical scales for the 37 mm sphere plots are not the same as the 10 mm plots. The line types and x-axis offsets are as in Figure 2.

III.E Image quality with clinical settings

To compare AC values using clinical parameters, Figure 7 shows recovery of the three AC measurements for this institution's standard parameters for routine FDG scanning: OSEM with 3 iterations, 24 subsets and 5 mm post-filter against OSEM+PSF and OSEM+PSF+TOF with 3 iterations and 2 mm post-filter. We feel the 2 mm post-filter is more appropriate for the latter two reconstructions due to inherently better noise control by these algorithms. The plots show how max-AC is strongly dependent on sphere size with OSEM at all contrast ratios and positive bias increases substantially with OSEM as contrast is reduced. AC values are seen to decrease in the smaller spheres as the contrast is reduced when using OSEM+PSF and OSEM+PSF+TOF, but OSEM+PSF+TOF is least affected.

Figure 7 shows that the mean-AC is much less dependent on reconstruction algorithm, sphere size and contrast than max-AC. The range of mean-AC for all contrasts and sphere sizes is between 41% (8:1 10 mm sphere, OSEM) and 90% (2:1 37 mm sphere, OSEM+PSF+TOF) for all three algorithms compared with the max-AC recovery range, which is from 57% (8:1 10 mm sphere, OSEM) to 156% (2:1 37 mm sphere, OSEM+PSF+TOF). Mean-AC with OSEM still exhibits the greatest dependence on sphere size compared with OSEM+PSF or OSEM+PSF+TOF although the differences are less than with max-AC. The peak-AC dependency on sphere size is interesting in that there is the largest negative bias in the 10 mm and 13 mm spheres of any measurement technique, but the bias for the four larger spheres (17 mm to 37 mm) is the smallest of any measurement technique. All measurements of these four larger spheres have a positive bias of typically 10% or less regardless of the post-filter FWHM. The negative bias in the smaller spheres is likely to be due to the inclusion of background voxels within the peak VOI and partial volume effect and consequently, the 2:1 contrast has the smallest negative bias for all three algorithms as the background voxels do not reduce the VOI mean by the same proportion as they do for the higher contrast ratios. Under close examination of the peak-AC measurements from the 17 mm to 37 mm spheres, it can be seen that, as with other techniques, OSEM is the most dependent on sphere size for all three contrast ratios.

The variability of all measurement techniques is seen to decrease as the contrast increases using any algorithm. Max-AC variability is greatest at three iterations for any particular algorithm. In the smaller spheres, variability is slightly greater in mean-AC than peak-AC, while for the larger spheres, there is very little difference in variability of mean-AC and peak-AC. Variability in the peak-AC measurements is less than 5% for all algorithms and contrasts with only a slight dependence on contrast ratio (variability decreasing with increasing contrast) and is very similar for all three reconstruction algorithms at any given contrast.

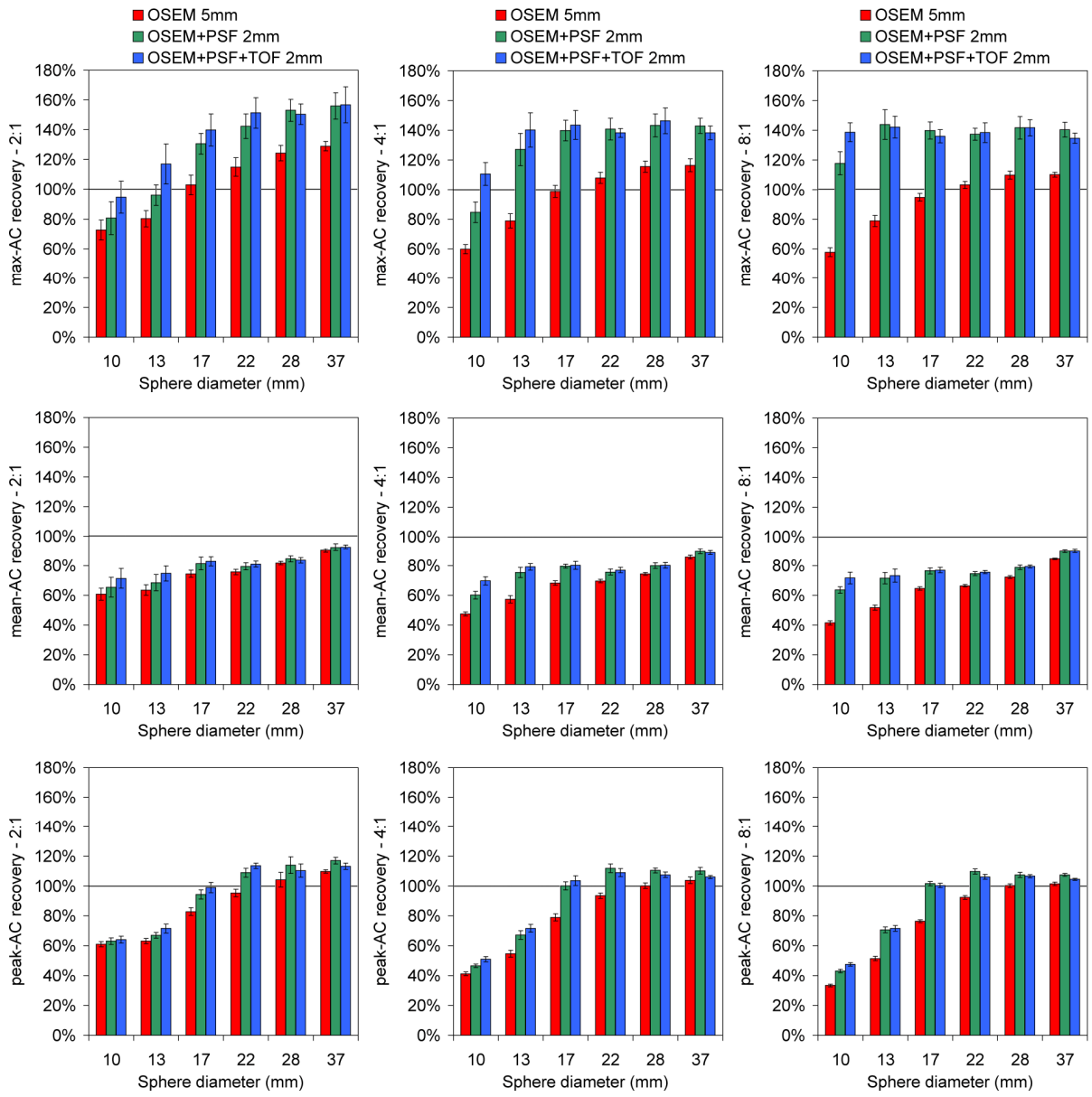


Figure 7. Activity concentration recovery for each sphere diameter after 3 iterations of each reconstruction algorithm with 2:1 (left column), 4:1 (center column) and 8:1 (right column) sphere-to-background concentration ratios. Activity concentration is measured using max-AC (top row), mean-AC (middle row) and peak-AC (bottom row) methods. The error bars indicate the standard deviation, obtained from the 10 replicate images, of the percentage recovery.

IV. Discussion

This work has demonstrated that, even though it is the most common measurement type, SUV_{max} will very rarely give an AC measurement that is accurate using typical reconstruction parameters used in the clinical environment unless the lesion is large

and high levels of post-filtering are applied. By its very nature of being obtained from the maximum value from a distribution of voxel values, it is susceptible to positive bias with noisy data. We have shown that very large positive biases in SUV_{max} are obtained with OSEM reconstruction, particularly as contrast decreases, unless substantial post-filtering is applied to the images. If this bias is suppressed with more post-filtering, this has the detrimental impact of increasing partial volume effect and therefore introducing or increasing a negative bias for small sphere sizes. This study has shown two major advantages of using OSEM+PSF or OSEM+PSF+TOF reconstruction over traditional OSEM for SUV_{max}: firstly the substantial reduction of partial volume effects for high contrast objects (Figure 3) and secondly, a considerable reduction of the variability of measurements (Figure 2).

It was not possible to find a combination of reconstruction parameters that produced the ultimate goal for any of the measurement techniques – that is: recovery that was accurate with little or no dependence on the sphere size or contrast and with low variability. In light of this outcome, the next best solution is to tailor the reconstruction parameters and use an appropriate measurement technique to obtain the desired outcome. These possible outcomes and how to achieve them are described below.

IV.A To minimize dependence on object size

It has been shown that, for high contrast lesions, when using clinical reconstruction parameters, max-AC has the least dependency on the object size for the entire range of sphere sizes used in this study (Figure 7). OSEM+PSF+TOF is the most effective algorithm for this task as it achieves max-AC recovery that is almost independent of sphere size with realistic clinical parameters of 3 iterations and a 2 mm post-filter (Figure 3). If TOF is not available, then the next best option is to use OSEM+PSF, but more iterations are required to increase AC recovery in smaller spheres (Table I, Figure 3). While this may be a beneficial quality of the reconstruction, it should be remembered that the mean bias across all spheres was +40%. One clinical application that may benefit from this outcome is staging of small lung lesions, which are typically high contrast due to the low uptake in surrounding healthy lung tissue, with SUVs that traditionally suffer from negative bias due to partial volume effects due to their small size. As the sphere contrast reduced, it became increasingly difficult to

maintain this independence of recovery on sphere size. This would suggest that, in the application of therapy monitoring, where both lesion contrast and size are likely to change, it will not be possible to find a single set of reconstruction parameters that can be used with SUV_{max} to remove the influence of lesion size. However, in situations where lesion size can be seen to be approximately 20 mm or greater, SUV_{peak} is likely to provide recovery with dependency on size comparable with max uptake without experiencing the large positive bias produced by max uptake measurements (Figure 7).

IV.B To achieve most accurate recovery

Peak-AC has been shown to be the most accurate for the largest 4 spheres at any contrast (Figure 7) and is almost independent of post-filter or reconstruction algorithm (Figure 4). This accurate recovery cannot be achieved with the two smallest spheres as peak-AC had the greatest negative bias compared with either max-AC or mean-AC. This would suggest that in clinical cases where the diameter of the lesion is confidently known to be greater than approximately 20 mm, SUV_{peak} is a far more appropriate measure to use than SUV_{max} as its accuracy is excellent over a wide range of lesion contrasts. If, during therapy monitoring, the lesion is seen to shrink to less than 20 mm, then it is very likely that SUV_{peak} will begin to suffer from negative bias and therefore other factors, such as the change in tumor dimensions assessed by anatomical imaging, need to be considered by clinicians. While the quantitative advantages gained from using OSEM+PSF and OSEM+PSF+TOF for SUV_{peak} are small, the reduced dependency of this measurement on the reconstruction is a positive finding as it allows OSEM+PSF and OSEM+PSF+TOF to be used to improve visualization of small objects (Figure 1(a)) with a smaller impact on the SUV_{peak} measurements compared with SUV_{max}. It should also be noted that the presence of the Gibbs artifacts (Figure 1(b)) is likely to introduce positive bias, particularly in SUV_{max}.

IV.C To achieve lowest variability

Peak-AC has the lowest variability for a particular algorithm and post-filter in nearly all cases. As with the accuracy of peak measurements, there are only small gains to be obtained from using OSEM+PSF or OSEM+PSF+TOF over OSEM reconstruction. When used with OSEM+PSF+TOF reconstruction, peak-AC

measurements have variability that is almost independent of number of iterations for any sphere size even with small FWHM post-filters (Figure 5). This finding is certainly beneficial to therapy monitoring where, if the lesion size can be seen to not change – or at least not fall below 20 mm diameter, then using SUV_{peak} is preferable to SUV_{max} as it will give the greatest confidence of detecting a genuine change in uptake.

If SUV_{peak} functionality is not available in a clinical setting, then it is most likely that SUV_{max} will be used instead. Variability of max-AC can be reduced by using OSEM+PSF+TOF and larger post-filters to the extent that for 8:1 sphere contrast, the variability of max-AC with OSEM+PSF+TOF and a 5 mm filter is almost comparable with peak-AC variability from images with OSEM+PSF+TOF and a 2 mm filter (Figure 5).

IV.D Summary of SUV measurements

Improvements to SUV_{peak} measurements can still be achieved with OSEM+PSF or OSEM+PSF+TOF reconstruction compared with OSEM. It should also be noted that, while these improvements are far less appreciable than those seen with SUV_{max} measurements, the use of OSEM+PSF or OSEM+PSF+TOF reconstruction will have other advantages such as reduced image noise and better lesion visualization.

This study has illustrated the considerable differences between the accuracy and variability of clinical uptake measurement techniques (maximum AC and peak AC) and the measurements made according to the NEMA Standard. The NEMA Standard states that quantitative performance is reported as contrast recovery of the spheres to background rather than explicitly stating the activity concentration within a sphere. However, the method of measuring the sphere activity concentration defined in the Standard would have a similar impact on contrast recovery measurements to the activity concentration measurements in this work. As a consequence, it is highly unlikely that quantitative performance testing made according to the NEMA method alone will fully illustrate the impact of various reconstruction algorithms on clinical SUV measurements in the clinical setting.

IV.E Limitations of this study

As has been demonstrated previously, accuracy and variability of SUVs are dependent on the image noise[7, 26], which increases with decreasing numbers of acquired counts in the images. To keep the number of experimental variables in this study to a manageable limit, it was decided to not vary the acquired counts. It has been shown recently that SUV_{peak} is less sensitive to varying image counts than SUV_{max}[10] but this study did not assess the impact of reconstructions with PSF modeling or TOF. Two studies have measured the robustness of SUV_{max} at varying image noise for OSEM+PSF+TOF reconstruction[32] and OSEM+TOF[26] but there has been no comparative assessment of multiple reconstruction algorithms in the presence of varying levels of image noise. It is also difficult to assess how findings from this work will translate to clinical imaging – particularly when patient movement is considered as this is expected to affect SUVs for small lesions.

V. Conclusion

OSEM+PSF and OSEM+PSF+TOF reconstruction have been shown to offer several advantages over OSEM when reporting either SUV_{max} or SUV_{peak} in a clinical environment – particularly for SUV_{max}. Small improvements with OSEM+PSF and OSEM+PSF+TOF were seen with SUV_{peak} measurements compared with OSEM, but the reduced dependence of SUV_{peak} on both controllable factors such as reconstruction algorithm, post-filter and number of iterations and uncontrollable factors such as object contrast highlight the promising future for this uptake measurement technique. Perhaps the most significant observation is that when using SUV_{max} there is no single combination of reconstruction parameters that will produce optimum images to meet the range of clinical aims – lesion detection and accurate quantification that has a low dependence of the uncontrollable factors of tumor size and contrast.

References

1 S. C. Huang, "Anatomy of SUV," Nucl Med Biol **27**, 643-646 (2000)

2 R. L. Wahl, H. Jacene, Y. Kasamon and M. A. Lodge, "From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors," J Nucl Med **50**, 122S-150 (2009)

3 A. S. Bryant and R. J. Cerfolio, "The maximum standardized uptake values on integrated FDG-PET/CT is useful in differentiating benign from malignant pulmonary nodules," Ann Thorac Surg **82**, 1016-1020 (2006)

4 R. J. Cerfolio and A. S. Bryant, "Maximum standardized uptake values on positron emission tomography of esophageal cancer predicts stage, tumor biology, and survival," Ann Thorac Surg **82**, 391-394; discussion 394-395 (2006)

5 D.-S. Shin, O.-J. Shon, D.-S. Han, J.-H. Choi, K.-A. Chun and I.-H. Cho, "The clinical efficacy of 18F-FDG-PET/CT in benign and malignant musculoskeletal tumors," Ann Nucl Med **22**, 603-609 (2008)

6 W. A. Weber, "Assessing Tumor Response to Therapy," Journal of Nuclear Medicine **50**, 1S-10S (2009)

7 R. Boellaard, N. C. Krak, O. S. Hoekstra and A. A. Lammertsma, "Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study," J Nucl Med **45**, 1519-1527 (2004)

8 N. C. Krak, R. Boellaard, O. S. Hoekstra, J. W. R. Twisk, C. J. Hoekstra and A. A. Lammertsma, "Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial," Eur J Nucl Med Mol Imaging **32**, 294-301 (2005)

9 C. Nahmias and L. M. Wahl, "Reproducibility of Standardized Uptake Value Measurements Determined by 18F-FDG PET in Malignant Tumors," J Nucl Med **49**, 1804-1808 (2008)

10 M. A. Lodge, M. A. Chaudhry and R. L. Wahl, "Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value," J Nucl Med **53**, 1041-1047 (2012)

11 R. K. Doot, P. E. Christian and D. A. Mankoff, presented at the Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE, 2007 (unpublished)

12 R. K. Doot, J. S. Scheuermann, P. E. Christian, J. S. Karp and P. E. Kinahan, "Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT," Med Phys **37**, 6035-6046 (2010)

13 J. Schwartz, J. L. Humm, M. Gonen, H. Kalaigian, H. Schoder, S. M. Larson and S. A. Nehmeh, "Repeatability of SUV measurements in serial PET," Med Phys **38**, 2629-2638 (2011)

14 V. Y. Panin, F. Kehren, C. Michel and M. Casey, "Fully 3-D PET reconstruction with system matrix derived from point source measurements," IEEE Trans Med Imaging on **25**, 907-921 (2006)

15 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**, 1453-1473 (2010)

16 A. M. Alessio, C. W. Stearns, T. Shan, S. G. Ross, S. Kohlmyer, A. Ganin and P. E. Kinahan, "Application and Evaluation of a Measured Spatially Variant System Model for PET Image Reconstruction," IEEE Trans Med Imaging **29**, 938-949 (2010)

17 A. J. Reader, P. J. Julyan, H. Williams, D. L. Hastings and J. Zweit, "EM algorithm system modeling by image-space techniques for PET reconstruction," IEEE Trans Nucl Sci on **50**, 1392-1397 (2003)

18 S. Tong, A. M. Alessio, K. Thielemans, C. Stearns, S. Ross and P. E. Kinahan, "Properties and Mitigation of Edge Artifacts in PSF-Based PET Reconstruction," IEEE Trans Nucl Sci **58**, 2264-2275 (2011)

19 T. F. Budinger, "Time-of-Flight Positron Emission Tomography: Status Relative to Conventional PET," J Nucl Med **24**, 73-78 (1983)

20 S. Surti, A. Kuhn, M. E. Werner, A. E. Perkins, J. Kolthammer and J. S. Karp, "Performance of Philips Gemini TF PET/CT Scanner with Special Consideration for Its Time-of-Flight Imaging Capabilities," J Nucl Med **48**, 471-480 (2007)

21 J. S. Karp, S. Surti, M. E. Daube-Witherspoon and G. Muehllehner, "Benefit of Time-of-Flight in PET: Experimental and Clinical Results," J Nucl Med **49**, 462-470 (2008)

22 C. Lois, B. W. Jakoby, M. J. Long, K. F. Hubner, D. W. Barker, M. E. Casey, M. Conti, V. Y. Panin, D. J. Kadrmas and D. W. Townsend, "An Assessment of the Impact of Incorporating Time-of-Flight Information into Clinical PET/CT Imaging," J Nucl Med **51**, 237-245 (2010)

23 E. Clementel, S. Vandenberghe, J. S. Karp and S. Surti, presented at the Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE, 2011 (unpublished)

24 NEMA, "NEMA Standards Publication NU 2-2007: Performance Measurements of Positron Emission Tomographs," (2007)

25 B. W. Jakoby, Y. Bercier, C. C. Watson, B. Bendriem and D. W. Townsend, "Performance Characteristics of a New LSO PET/CT Scanner With Extended Axial Field-of-View and PSF Reconstruction," IEEE Trans Nucl Sci **56**, 633-639 (2009)

26 I. Murray, A. Kalemis, J. Glennon, S. Hasan, S. Quraishi, T. Beyer and N. Avril, "Time-of-flight PET/CT using low-activity protocols: potential implications for cancer therapy monitoring," Eur J Nucl Med Mol Imaging **37**, 1643-1653 (2010).

27 B. W. Jakoby, Y. Bercier, M. Conti, B. Bendriem and D. Townsend, "Physical and clinical performance of the mCT time-of-flight PET/CT scanner," Phys Med Biol **56**, 2375 (2011).

28 C. Lasnon, R. J. Hicks, J.-M. Beauregard, A. Milner, M. Paciencia, A.-V. Guizard, S. Bardet, R. Gervais, G. Lemoel, G. Zalcman and N. Aide, "Impact of Point Spread Function Reconstruction on Thoracic Lymph Node Staging With 18F-FDG PET/CT in Non–Small Cell Lung Cancer," Clin Nucl Med **37**, 971-976 (2012).

29 I. S. Armstrong, D. Tout, J. M. James, H. A. Williams, M. C. Prescott. "Impact of PSF modelling and time-of-flight on SUV values in lung and head and neck lesions". Nuc Med Commun 2012; **33**:A28

30 G. Akamatsu, K. Mitsumoto, T. Taniguchi, I. Komiya, N. Ohya, K. Abe, S. Baba and M. Sasaki, "The influences of point-spread-function and time-of-flight on SUV of lymph node metastasis in FDG-PET," J Nucl Med **53**, 2509 (2012).

31 M. Gregianin, L. Evangelista, A. R. Cervino, M. Burei and G. Saladini, "Effects of the High Definition technology on Standardized Uptake Values (SUVs) obtained by Fluorodeoxyglucose (FDG) Positron Emission Tomography (PET)/Computed Tomography (CT)," Eur J Nucl Med Mol Imaging **39**, S228 (Suppl 2) (2012).

32 J. M. Marti-Climent, E. Prieto, M. J. García-Velloso, I. Dominguez-Prado, M. Rodriguez-Fraile, J. Arbizu, C. Vigil, C. Caicedo, J. A. Richter, “Time-of-flight and point-spread-function allow shorter time acquisitions or lower activity in PET/CT protocols: a phantom study”. Eur J Nucl Med Mol Imaging **39**, S228 (Suppl 2) (2012)

5. SECOND PAPER

Impact of point spread function modelling and time-of-flight on FDG uptake measurements in lung lesions using alternative filtering strategies

Article status:
Submitted to EJNMMI Physics in February 2014
Accepted decision received September 2014

The paper included in this chapter is the published version:

Armstrong IS, Kelly MD, Williams HA, Matthews JC. *Impact of point spread function modelling and time-of-flight on FDG uptake measurements in lung lesions using alternative filtering strategies*. EJNMMI Physics 2014; **1**:99

Armstrong et al. EJNMMI Physics 2014, 1:99
http://www.ejnmmiphys.com/content/1/1/99

EJNMMI Physics
a SpringerOpen Journal

ORIGINAL RESEARCH Open Access

Impact of point spread function modelling and time of flight on FDG uptake measurements in lung lesions using alternative filtering strategies

Ian S Armstrong[1,2*], Matthew D Kelly[3], Heather A Williams[1] and Julian C Matthews[2]

* Correspondence:
Ian.Armstrong@cmft.nhs.uk
[1]Nuclear Medicine, Central Manchester University Hospitals, Oxford Road, Manchester, UK
[2]Institute of Population Health, MAHSC, University of Manchester, Manchester, UK
Full list of author information is available at the end of the article

Abstract

Background: The use of maximum standardised uptake value (SUV_{max}) is commonplace in oncology positron emission tomography (PET). Point spread function (PSF) modelling and time-of-flight (TOF) reconstructions have a significant impact on SUV_{max}, presenting a challenge for centres with defined protocols for lesion classification based on SUV_{max} thresholds. This has perhaps led to the slow adoption of these reconstructions. This work evaluated the impact of PSF and/or TOF reconstructions on SUV_{max}, SUV_{peak} and total lesion glycolysis (TLG) under two different schemes of post-filtering.

Methods: Post-filters to match voxel variance or SUV_{max} were determined using a NEMA NU-2 phantom. Images from 68 consecutive lung cancer patients were reconstructed with the standard iterative algorithm along with TOF; PSF modelling - Siemens HD·PET (HD); and combined PSF modelling and TOF - Siemens ultraHD·PET (UHD) with the two post-filter sets. SUV_{max}, SUV_{peak}, TLG and signal-to-noise ratio of tumour relative to liver ($SNR_{(T-L)}$) were measured in 74 lesions for each reconstruction. Relative differences in uptake measures were calculated, and the clinical impact of any changes was assessed using published guidelines and local practice.

Results: When matching voxel variance, SUV_{max} increased substantially (mean increase +32% and +49% for HD and UHD, respectively), potentially impacting outcome in the majority of patients. Increases in SUV_{peak} were less notable (mean increase +17% and +23% for HD and UHD, respectively). Increases with TOF alone were far less for both measures. Mean changes to TLG were <10% for all algorithms for either set of post-filters. $SNR_{(T-L)}$ were greater than ordered subset expectation maximisation (OSEM) in all reconstructions using both post-filtering sets.

Conclusions: Matching image voxel variance with PSF and/or TOF reconstructions, particularly with PSF modelling and in small lesions, resulted in considerable increases in SUV_{max}, inhibiting the use of defined protocols for lesion classification based on SUV_{max}. However, reduced partial volume effects may increase lesion detectability. Matching SUV_{max} in phantoms translated well to patient studies for PSF reconstruction but less well with TOF, where a small positive bias was observed in patient images. Matching SUV_{max} significantly reduced voxel variance and potential variability of uptake measures. Finally, TLG may be less sensitive to reconstruction methods compared with either SUV_{max} or SUV_{peak}.

Keywords: PET quantification; PSF modelling; Time-of-flight; SUV; Total lesion glycolysis

Background

[^{18}F]2-Fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) has been shown to play a key role in the management of patients with non-small cell lung cancer in terms of staging and prognosis [1-5] and monitoring response to therapy [6]. In these applications, the uptake of FDG expressed as standardised uptake value (SUV) is of key importance, with SUV_{max} being the most commonly reported measure [7]. The use of SUV_{max} for discrimination between benign and malignancy for soft tissue masses and lymph nodes has been demonstrated for lung cancer patients [8,9] and changes in SUV_{max} used as an indicator of response to therapy [10].

While the use of SUV_{max} is commonplace, it is known to be sensitive to both reconstruction parameters [11] and the amount of statistical image noise, leading to poorer test-retest consistency relative to other SUV-based metrics [12,13]. Consequently, alternative metrics such as SUV_{peak} [14] and total lesion glycolysis (TLG), the product of SUV_{mean} and metabolic tumour volume derived from the PET images, have been suggested for use, particularly in monitoring response to therapy [6,15]. Recently, TLG has also been shown to offer superior prognostic information than SUV_{max} [16-20].

In recent years, there have been significant advances in iterative image reconstruction algorithms and scanner hardware. Consequently, reconstruction algorithms that include point spread function (PSF) modelling [20,21] and time of flight (TOF) [22] have become commercially available on PET/CT scanners, with TOF also available on PET/MR [23].

The use of PSF modelling, with and without TOF, has been shown to improve signal-to-noise ratio (SNR) [24-27] and lesion detectability [28-30] partly through decreasing voxel variance. However, the implementation of PSF modelling, both within projection space and image space, from different manufacturers and also academic institutions has been shown to produce Gibbs artefacts [21,31-35] (Nick Vennart, personal communication). In patient imaging, the Gibbs artefact, combined with reduced partial volume effects, has a significant impact on SUV_{max} [36-38]. This is particularly evident with minimal or no post-reconstruction filtering, which has been shown in phantom studies with numerical observers to provide greater lesion detectability [28-30]. Changes to SUV_{max} as a consequence of PSF modelling present a challenge as changes to defined local practice for reporting may be required such as changing the thresholds used for the discrimination of malignancy. The scanner used in this study has been part of a multi-site network of scanners for routine FDG oncology imaging since 2009. SUV_{max} is the reported uptake metric, and the consensus amongst local reporting clinicians within the network is that lesions with $SUV_{max} > 5.0$ are considered highly suspicious of malignant disease.

It is necessary, in practice, to smooth clinical images to provide image quality that is deemed acceptable for clinical reporting. This degrades the spatial resolution but increases signal to noise. The degree of smoothing applied at any given centre is heavily influenced by the experience and personal preferences of the reporting clinicians, informed by the advice of physicists providing scientific support. Where several PET scanners serve the same patient population, it is also advantageous to match imaging performance across the network in terms of visual image quality and quantitative characteristics.

A trade-off curve of signal enhancement versus noise reduction when using PSF and/or TOF algorithms can be established by applying a range of reconstruction post-filters.

It has been demonstrated that it is possible to match SUV_{max} from PSF-based reconstruction with traditional non-PSF algorithms by applying a particular post-filter. Lasnon et al. [39] showed that a 7.0-mm full-width-half-maximum (FWHM) post-filter with PSF reconstruction gave comparable recovery coefficients in phantom data to non-PSF reconstructions and brought the recovery coefficients in line with European recommendations [40]. Another study proposed the application of a post-filter for the purpose of quantification [41]. This study also demonstrated that despite a spatially dependent PSF, this approach of using a single post-filter choice was adequate for all lesions irrespective of their location in the field of view. The application of a relatively broad post-filter to PSF modelling images may seem counterintuitive as it will undo the improvements in partial volume effect, but there are likely to be other benefits that have not been reported such as a reduction in voxel variance in the images.

Another potential solution may be to use alternative uptake metrics to SUV_{max}. One study [37] suggested that TLG may be more stable when comparing PSF to non-PSF reconstruction, but this study only assessed ten lung lesions. Another study [38] has suggested the move to SUV_{mean} based upon a 50% isocontour of SUV_{max}. To our knowledge, there are currently no studies that investigate the impact of these reconstructions with PSF modelling and TOF on TLG and SUV_{peak}.

The primary aim of this study was to evaluate the impact of PSF modelling and TOF on SUV_{max}-based lesion classification as implemented at the local institution. This was performed using Siemens reconstruction software including implementations for TOF and PSF modelling (HD, UHD). Implementations of reconstruction algorithms can differ, and therefore, the results might be specific to HD and UHD; however, we feel it is likely that findings may be generalisable to other reconstruction implementations with similar philosophies. Any change in FDG uptake measurements across different reconstruction protocols can hopefully allow other centres to assess how such changes may impact their approaches to lesion classification. Two set criteria for post-filtering the images were assessed based upon characteristic locations on a signal enhancement versus noise reduction trade-off curve. These two points are 1) matching image noise (voxel variance) which was expected to enhance signal and 2) matching signal (SUV_{max}) which, based on previous studies [39,41], was anticipated to require greater levels of post-filtering and hence reduce image noise. This latter approach is aimed to be particularly relevant to centres that wish to maintain uptake quantification for practical purposes, which is particularly important in multi-site imaging networks. In addition, this work aimed to expand on the results of previous studies [36-38] with the addition of TOF, evaluation of other uptake metrics such as SUV_{peak} and TLG, and determining gains in SNR for the two strategies.

Methods

PET/CT scanner

The PET scanner used in this study was a Siemens Biograph mCT with 64 slice CT (Siemens Medical Solutions, Erlangen, Germany). The scanner has a four-ring extended axial field of view of 21.6 cm (TrueV) and includes options for PSF modelling (Siemens HD·PET) and combined PSF modelling with TOF (Siemens ultraHD·PET) in the image reconstruction. Performance data for the scanner has been published previously [42].

Phantom acquisitions

A NEMA NU-2 image quality (IQ) phantom (PTW, Freiburg, Germany) was filled with [^{18}F]FDG so that the background compartment and all six hot spheres had activity concentrations of 5.19 and 41.7 kBq/ml, respectively. This 8:1 contrast was chosen to mimic lung lesion contrast, which is generally high. In order to divide the data into ten replicate datasets, a gated 60-min list-mode acquisition was performed using an ECG simulator as the gating input. Each replicate image contained 30 million (±0.2%) net true coincidences as this was typical of the number of counts measured over the thorax in our standard patient acquisitions. Images were reconstructed using four methods: standard 3-D ordinary Poisson ordered subset expectation maximisation (OSEM) reconstruction; OSEM with TOF (TOF); OSEM with PSF modelling - Siemens HD·PET (HD); and OSEM with both PSF and TOF - Siemens ultraHD·PET (UHD). For non-TOF reconstructions, 3 iterations and 24 subsets (3i24s) were used, while for TOF reconstructions, 2 iterations and 21 subsets (2i21s) were used.

Two iterations were chosen for TOF reconstructions as TOF has been shown to provide faster convergence with comparable signal to noise achieved in fewer iterations than non-TOF [27,43], and it has been shown in published performance data for the scanner that one fewer iteration with TOF is optimal [42], providing similar background variability and marginally superior contrast recovery in smaller objects. However, it is not possible to exactly match the number of subsets for TOF and non-TOF reconstructions. All images were reconstructed into a 256×256 matrix with voxel sizes of 3.2 mm × 3.2 mm × 2.0 mm. As is routinely performed with patient data, a 5.0-mm FWHM Gaussian post-filter was applied to the OSEM images. The baseline parameters of 3 iterations and 24 subsets and 5.0-mm post-filter for OSEM reconstruction have been in routine use since the scanner was commissioned in 2009. These parameters were selected to align SUV_{max} quantification and voxel variance with other scanners in the local oncology imaging network.

A variety of post-filters with different kernel widths was applied to the TOF, HD and UHD images with kernel widths ranging from 0 to 10 mm FWHM in step sizes for 0.1 mm.

Noise matching

Twelve circular regions of interest (ROIs) of 37-mm diameter were placed in the phantom background over five separate slices (60 ROIs in total) of the IQ phantom image in accordance with the NEMA NU-2-2007 standard [44]. For each image replicate, the average coefficient of variation (COV) over the 60 ROIs was calculated as

$$\mathrm{COV}_R = \sum_{k=1}^{60} \frac{\sigma_{k,R}}{\mu_{k,R}}, \tag{1}$$

where $\sigma_{k,R}$ and $\mu_{k,R}$ are the voxel standard deviation and mean, respectively, within ROI k and replicate R. The mean and standard deviation of COV_R was determined across all ten replicate images. The OSEM 3i24s 5.0-mm post-filter image was used to compute the reference COV value. For the three other reconstruction methods, the post-filter that gave the smallest difference in COV, relative to the OSEM image, was determined.

SUV_{max} matching

SUV_{max} is the uptake measure used in our routine patient reports and so was the measure chosen to match across the reconstruction algorithms. To achieve this, SUV_{max} was measured in each hot sphere in the phantom for the OSEM images using a 3-D volume of interest, equal in diameter to each true sphere size and centred on the sphere. As with the COV matching, a post-filter was incremented in 0.1-mm steps on the other three reconstructions until the summed squared difference of SUV_{max} for the six hot spheres relative to those in the OSEM image was minimised.

FDG patient acquisitions

Patient preparation

Retrospective data from 68 (33 males; mean [range] weight: 72.5 kg [40 to 136]; mean [range] body mass index: 26.3 kg/m^2 [14.1 to 51.8]) consecutive routine oncology patients referred for assessment of single pulmonary nodule or staging of non-small cell lung cancer were included in this study. All data were fully anonymised before inclusion. Patients fasted for 6 h prior to the injection of FDG and were asked to drink at least 500 ml of water before the scan. Blood glucose was measured with permissible limits of 3.0 to 12.5 mmol/l. Patients with a body weight <100 kg were prescribed 350 MBq of [^{18}F]FDG, while those with body weight >100 kg (two in this study) were prescribed 400 MBq. The mean [range] administered activity of [^{18}F]FDG was 365.5 MBq [242.0 to 423.1]. It can be noted that the minimum dose administered is considerably below the prescribed activity - this was due to a patient arriving late and insufficient remaining activity in the stock vial. The mean [range] time was 64.3 [59 to 87] min from the time of injection to commencing the scan. Advice from the local ethics committee deemed that the use of retrospective anonymised patient data did not require formal ethical approval.

PET/CT acquisitions

The PET acquisition was performed from eyes to mid-thigh for all patients, requiring six or seven bed positions. The acquisition time for each bed position was 2.5 min. Attenuation correction was performed using a non-contrast CT acquisition performed prior to the PET acquisition. Scatter and random corrections were applied to all images. All images were reconstructed with OSEM 3i24s and 5.0-mm post-filter as the reference, along with the phantom-determined TOF, HD and UHD protocols, which match either voxel COV or SUV_{max}.

Uptake measurements

All images were viewed and the uptake quantified using Siemens TrueD image display software (Siemens Medical Solutions, Erlangen, Germany). In each patient, a 3-cm-diameter spherical volume of interest (VOI) was placed within an area of uniform FDG distribution in the liver, and the COV of the voxels within the VOI was calculated. Three FDG uptake measurements were derived for each identified lesion within the lung: SUV_{max}, SUV_{peak} (as defined in the PET response criteria in solid tumours (PERCIST) protocol [14]) and TLG. SUV was normalised to patient body weight only. Volume delineation for TLG was performed using a 40% threshold of SUV_{max} (TLG-40). Recent meta-analyses [16,17] have highlighted several methods for volume delineation - either using percentage or absolute SUV thresholds. The choice of a percentage threshold in this study

was based on a hypothesis that as the magnitude of the partial volume effect varied with different reconstructions, the impact on the tumour volume and SUV_{mean} would be inversely related. This may result in a more stable value for the TLG. It should be noted that other methods of delineation are likely to produce alternative results. Lesion volume was measured on the OSEM image using a 40% threshold of SUV_{max}.

Signal to noise

It is difficult to estimate SNR directly in a lesion due to inhomogeneous uptake; therefore, we have adopted the use of the liver as a source for the background and noise measurement. This technique has been performed previously [25] and is considered a reasonable relative surrogate for SNR in the lesion. For lesions with SUV_{max} above the PERCIST threshold of 1.5 times the mean SUV in the liver VOI + 2 standard deviations of the voxels within the liver VOI [14], the signal-to-noise ratio of the tumour, relative to the liver, ($SNR_{(T\text{-}L)}$) was calculated as

$$SNR_{(T\text{-}L)} = \frac{\text{Tumour} - \text{Liver}}{\sigma_L}, \tag{2}$$

where the *Tumour* refers to SUV_{max} in the lung lesion, *Liver* is the mean SUV measured in the liver VOI and σ_L is the standard deviation of voxel values measured in the liver VOI. This method allows comparison to other studies, which have used the same metric [25,42]. $SNR_{(T\text{-}L)}$ of all qualifying lesions was determined for each reconstruction using the two filtering schemes of matched voxel COV and matched SUV_{max}. The gain in $SNR_{(T\text{-}L)}$ was expressed for the TOF, HD and UHD reconstructions as the ratio to the $SNR_{(T\text{-}L)}$ measurements from the standard OSEM images of the same patient.

Statistical analysis

Relative percentage differences of the uptake metrics relative to OSEM were expressed as mean with 95% confidence intervals. Bland-Altman analysis was also performed on the data. Relative changes of >25% for SUV_{max} and >30% for SUV_{peak} were considered clinically significant based upon EORTC [10] and PERCIST [14] guidelines respectively. In addition, hypothetical changes to patient management as a consequence of SUV_{max} based on local practice were recorded. Differences in voxel COV in the liver VOI and gains in $SNR_{(T\text{-}L)}$ were assessed using a paired *t* test with a *p* value <0.05 considered to be significant.

Results

Phantom images

The FWHM of the post-filters obtained for matching voxel COV to OSEM 3i24s and a 5.0-mm post-filter were 4.4, 3.8 and 2.9 mm for TOF, HD and UHD, respectively. The FWHM of the post-filters obtained for matching SUV_{max} were 4.8, 6.6 and 6.5 mm for TOF, HD and UHD, respectively. To provide an illustration of the underlying impact of each algorithm, SUV_{max}, expressed as a percentage of the true activity concentration, and noise data are first shown with no post-filter in Table 1. Data are then presented with the two post-filter sets as described in Table 2. From the data, it is seen that there is considerable increase in SUV_{max} in the two smallest spheres with HD and UHD with matched voxel COV. The variability of SUV_{max} was greater in the two smallest spheres at matched voxel COV, particularly with HD and UHD; the positive bias in the larger spheres with

Table 1 Phantom recovery data for unfiltered images

	Sphere size						Background voxel COV
	10 mm	13 mm	17 mm	22 mm	28 mm	37 mm	
OSEM (%)	106 (19.5)	136 (20.2)	155 (19.8)	168 (18.0)	184 (9.5)	186 (8.2)	46.3 (0.51)
TOF (%)	90 (12.0)	123 (11.2)	132 (7.2)	144 (11.0)	157 (11.4)	167 (13.7)	37.4 (0.48)
HD (%)	99 (13.7)	155 (7.2)	144 (6.2)	145 (5.3)	144 (7.1)	147 (6.0)	18.2 (0.32)
UHD (%)	103 (9.0)	151 (8.8)	141 (5.9)	136 (4.5)	135 (5.9)	138 (5.6)	14.9 (0.46)

SUV_{max} in each of the image quality spheres expressed as a percentage of the true activity concentration, and voxel COV in the phantom background. Data are shown for all four reconstruction algorithms with no post-filtering applied. Values are mean and standard deviation (SD) obtained from the replicates, with the latter shown in parentheses. For clarity, the SD shown is the SD across the replicates expressed as a percentage of the true activity concentration in the sphere.

OSEM and TOF at matched voxel COV is likely to be due to image voxel variance, while with HD and UHD at matched voxel COV, Gibbs artefacts are also expected to contribute. This can be seen in Figure 1, which shows profiles through the centre of the 37-, 22- and 13-mm spheres.

With post-filters to match SUV_{max} recovery, variability is comparable or less with HD and UHD compared with OSEM. To verify the cross-calibration between the dose calibrator and scanner, the activity concentration, averaged across the 60 background ROIs, was measured as 5.14 ± 0.1 kBq/ml.

Patient images

Figure 2 shows images from a single representative female patient with a BMI of 37 kg/m^2. The image has been cropped to show only the lung lesion and liver. Voxel COV within the liver VOI was 16.3%, 15.0%, 16.5% and 15.4% for OSEM, TOF, HD and UHD, respectively, with matched voxel COV post-filters and 13.5%, 10.8% and 7.95% for TOF, HD and UHD, respectively, with matched SUV_{max} post-filters. SUV_{max} for the lesion in the right lung was 5.4, 6.0, 8.2 and 10.1 for OSEM, TOF, HD and UHD, respectively, with matched noise post-filters and 5.2, 5.7 and 5.7 for TOF, HD and UHD, respectively, with matched SUV_{max} post-filters. The visual reduction in voxel variance within the liver is evident in the HD and UHD images with the matched SUV_{max} protocol.

Table 2 Phantom recovery data

	Sphere size						Background voxel COV
	10 mm	13 mm	17 mm	22 mm	28 mm	37 mm	
OSEM	53	80	96	102	110	110	12.9
Matched voxel COV							
TOF (%)	58 (4.4)	84 (3.4)	97 (3.2)	101 (2.4)	110 (2.4)	113 (3.8)	12.8 (0.23)
HD (%)	73 (7.1)	123 (4.5)	122 (2.9)	121 (5.3)	121 (3.8)	123 (2.8)	12.8 (0.38)
UHD (%)	89 (6.6)	138 (7.5)	129 (3.1)	123 (2.7)	125 (4.0)	127 (4.2)	12.7 (0.43)
Matched SUV_{max}							
TOF (%)	55 (3.8)	80 (2.8)	94 (2.9)	98 (1.8)	107 (2.3)	110 (2.9)	11.2 (0.37)
HD (%)	47 (3.0)	80 (2.3)	103 (2.7)	102 (3.5)	105 (1.3)	106 (0.9)	7.57 (0.17)
UHD (%)	49 (1.9)	81 (1.8)	102 (2.0)	100 (1.2)	104 (2.0)	105 (1.2)	6.27 (0.33)

SUV_{max} in each of the image quality spheres expressed as a percentage of the true activity concentration, and voxel COV in the phantom background. Data are shown for OSEM (reference reconstruction) and the PSF and TOF-based reconstructions with the two post-filter sets. Values are mean and standard deviation (SD) obtained from the replicates, with the latter shown in parentheses. For clarity, the SD shown is the SD across the replicates expressed as a percentage of the true activity concentration in the sphere.

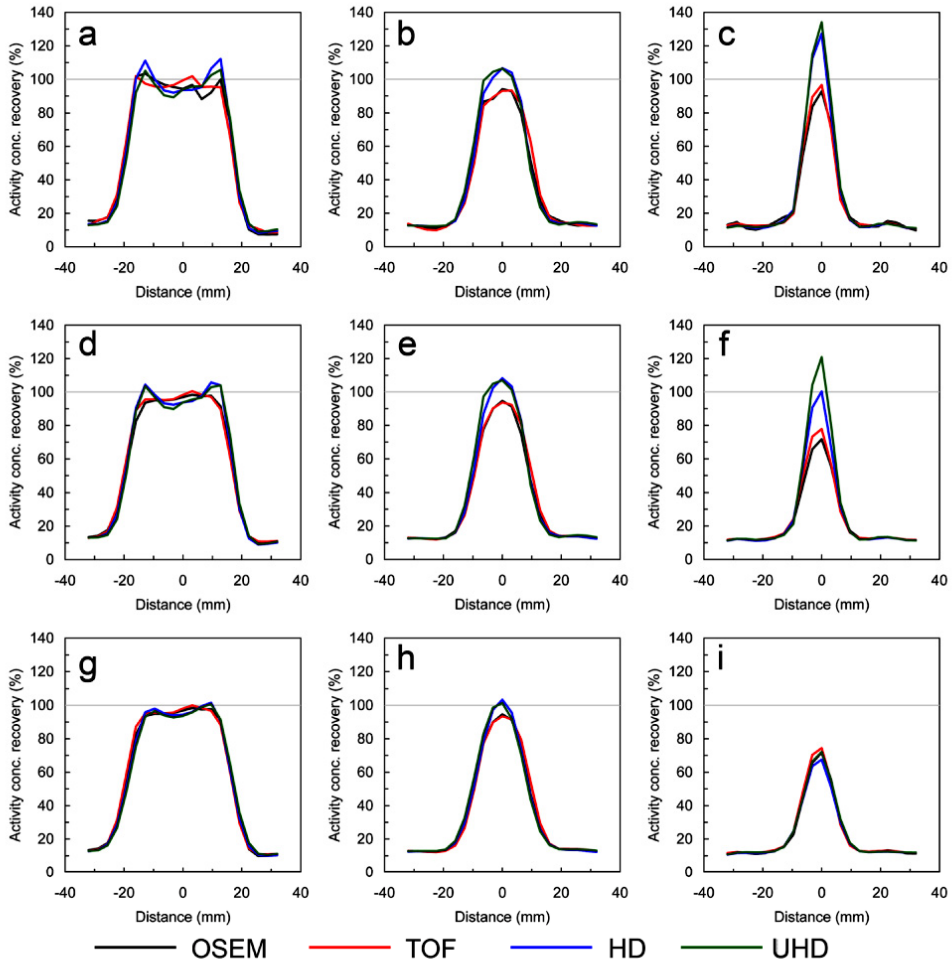


Figure 1 Phantom sphere profiles. Transaxial line profiles through the centre of the 37-mm sphere **(a, d, g)**, 22-mm sphere **(b, e, h)** and 13-mm sphere **(c, f, i)**. Plots in the top row are for unfiltered images, those in the centre row are for post-filters to match voxel COV, and those in the bottom row are for plots to match SUV_{max}.

Liver noise

Table 3 shows the voxel COV data measured in the VOI within the patient livers. There were no significant differences for the PSF and TOF-based reconstructions versus OSEM when using the matched voxel COV post-filters. As with the phantom data, significant reductions of voxel COV were measured for PSF and TOF-based reconstructions compared with OSEM using the post-filters to match SUV_{max} recovery. The mean measurements of

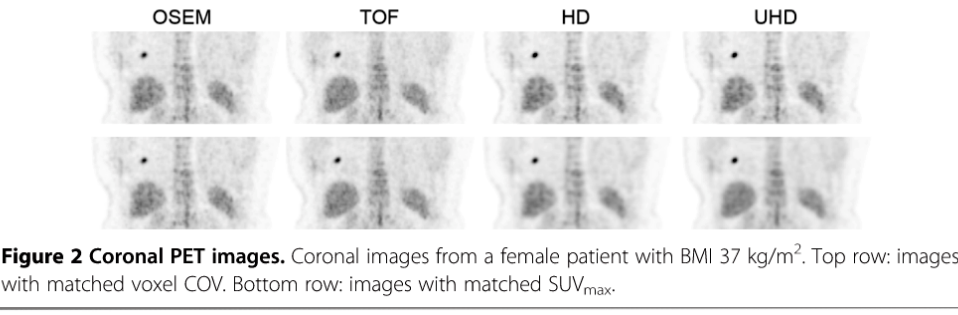


Figure 2 Coronal PET images. Coronal images from a female patient with BMI 37 kg/m^2. Top row: images with matched voxel COV. Bottom row: images with matched SUV_{max}.

voxel COV in the liver VOI for TOF, HD and UHD were 90%, 65% and 56%, respectively, of the value measured using OSEM.

FDG uptake measurements

Tables 4 and 5 summarise the changes of the three uptake measures observed using the PSF and TOF-based reconstructions relative to OSEM. The data in Table 5 for the number of lesions with a change in SUV_{max} greater than 25% occurred in lesions with very low grade uptake (SUV_{max} <2.5). Bland-Altman plots for the relative differences are shown in Figures 3, 4 and 5, which, in addition to data in Tables 4 and 5, show that the smaller values of SUV_{max} and SUV_{peak} experience the greatest increase with matched voxel COV (Figure 3a,b,c and Figure 4a,b,c). For matched SUV_{max} filters, this is still present with TOF algorithms (Figure 3d,f and Figure 4d,f) but not with HD reconstruction.

For matched voxel COV, the increase in both SUV_{max} and SUV_{peak} ratio for PSF and TOF-based reconstructions versus OSEM was inversely related to lesion volume as shown in Figure 6. This reflects what was seen in the image quality phantom measurements. The gains in SUV_{max} were most pronounced with UHD, which is likely to be a consequence of reduced post-filtering compared with HD when voxel COV was matched (2.9 mm for UHD and 3.8 mm for HD). Differences in TLG-40 were not dependent on lesion volume. No relationship between SUV difference and lesion volume was observed for matched SUV_{max} post-filters.

Out of the 74 lesions, 59 had a SUV_{max} of >5.0 using OSEM reconstruction. No change to patient management would occur in these instances as a result of an increase of SUV_{max} when using the PSF and TOF-based reconstructions. A key group of ten patients was identified with low or borderline SUV_{max} (<5.0) for suspicion of malignancy using this institute's practice. The SUV_{max} for these 15 lesions in each of the reconstruction algorithms are shown in Table 6. The table shows that, with matched voxel COV, several of these lesions would change classification with HD and UHD, as would be expected from data in previous tables and figures. With matched SUV_{max} filters, there is only one lesion that would have changed classification according to local practice and only with the TOF reconstruction.

Signal-to-noise gains

Fifty-nine lesions were found to have SUV_{max} above the threshold based on the liver uptake as measured on the OSEM images. Significant $SNR_{(T-L)}$ gains were found for PSF and TOF-based reconstructions with both matched voxel COV and matched SUV_{max}. With the addition of PSF modelling, either to OSEM or OSEM + TOF images, there is a more marked gain in $SNR_{(T-L)}$. For matched voxel COV, $SNR_{(T-L)}$ ratios relative to OSEM were 1.10 ± 0.11, 1.43 ± 0.23 and 1.67 ± 0.41 for TOF, HD and UHD, respectively, and for matched SUV_{max}, they were 1.19 ± 0.12, 1.58 ± 0.16, and 1.94 ± 0.29, respectively. For each reconstruction algorithm, the improvement in $SNR_{(T-L)}$ with matched SUV_{max} versus matched noise was also significant.

Table 3 Patient liver noise

Liver COV (%)	OSEM	TOF	HD	UHD
Matched COV	10.6 (2.3)	10.5 (2.4)	10.6 (2.2)	10.4 (2.1)
Matched SUV_{max}		9.5 (2.1)	6.8 (1.5)	5.9 (1.3)

Image noise, expressed as coefficient of variation (COV), measured in the liver for each reconstruction for matched voxel COV and matched SUV_{max} post-filters. Values are mean and standard deviation, with the latter shown in parentheses.

Table 4 Relative uptake differences for matched voxel COV

Uptake		TOF	HD	UHD
SUV_{max}	Mean% change	+8.8%	+32%	+49%
	95% CI	−7.2% to +25%	+7.8% to +55%	+1.0% to +97%
	n > 25% change	4/74	50/74	63/74
SUV_{peak}	Mean% change	+6.9%	+17%	+27%
	95% CI	−5.7% to +20%	+5.3% to +29%	+1.6% to +51%
	n > 30% change	1/74	1/74	25/74
TLG-40	Mean% change	+1.9%	−8.4%	−7.5%
	95% CI	−16% to +20%	−29% to +12%	−37% to +22%

Mean percentage changes and 95% confidence intervals of the three uptake measures relative to OSEM reconstruction. Also shown are the number of lesions with a greater than 25% and 30% increase in SUV_{max} and SUV_{peak}, respectively. Data in the table are from images using post-filters to match image voxel COV.

Discussion

The deployment of PSF and TOF-based reconstruction methods into routine clinical practice for FDG imaging presents a challenge, particularly in centres or collaborative imaging networks with a defined protocol for classification of malignancy based upon SUV data. To our knowledge, this is the first study that has evaluated the performance of PSF and TOF-based reconstruction algorithms with two post-filtering strategies based on the objective criteria of matched image noise (voxel COV) or matched SUV_{max}, quantifying the impact on SUV_{max}, SUV_{peak}, TLG and $SNR_{(T\text{-}L)}$. Specific findings are applicable to Siemens HD and ultraHD reconstruction algorithms using the parameters applied in the study.

It is clear from the data in Tables 1 and 2 and Figure 3 that quantification differences occur in the phantom data for all algorithms applied in this study. There are several factors that will contribute to the differences: the effect of statistical noise, partial volume effect, the size (and hence number of voxels) of the region of interest and, for the HD and UHD algorithms, Gibbs artefacts. The contributions from these factors to the measurements of SUV_{max} will differ as reconstruction parameters are varied. We believe that the interactions between the various factors are complex and not completely separable. As such, we do not feel that it is possible to identify one single phenomenon as the source of quantification differences for any of the algorithms used.

Table 5 Relative uptake differences for matched SUV_{max} recovery

Uptake		TOF	HD	UHD
SUV_{max}	Mean% change	+5.4%	+0.6%	+5.3%
	95% CI	−8.5% to +19%	−8.1% to +9.4%	−9.5% to +20%
	n > 25% change	1/74	0/74	2/74
SUV_{peak}	Mean% change	+5.3%	+1.7%	+6.3%
	95% CI	−6.4% to +17%	−3.0% to +6.5%	−4.9% to +17%
	n > 30% change	0/74	0/74	1/74
TLG-40	Mean% change	+3.8%	+2.4%	+4.0%
	95% CI	−13% to +20%	−13% to +18%	−12% to +20%

Mean percentage changes and 95% confidence intervals of the three uptake measures relative to OSEM reconstruction. Also shown are the number of lesions with a greater than 25% and 30% increase in SUV_{max} and SUV_{peak}, respectively. Data in the table are from images using post-filters to match SUV_{max}.

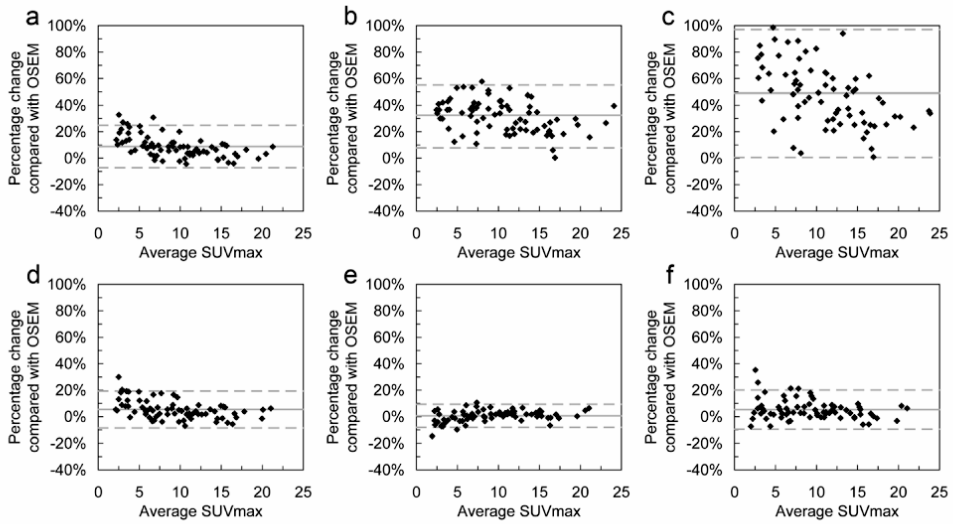


Figure 3 Bland-Altman plots showing relative percentage differences for SUV_{max} relative to OSEM. Plots on the top row are for images with matched image voxel COV post-filters, and plots on the bottom row are for images with matched SUV_{max} recovery post-filters. Reconstructions are TOF **(a, d)**, HD **(b, e)** and UHD **(c, f)**. In each plot, the solid grey line shows the mean percentage difference and the two dashed lines show the 95% confidence intervals.

It can be seen that overestimation occurs for all four reconstruction algorithms (Table 1) and requires the application of a post-filter to reduce this (Table 2). The smaller filter kernel applied to HD and UHD to match noise combined with voxel correlation leads to a lesser reduction of this overestimation. It can be seen that there appears to be a particular size of object where an overestimation with HD and UHD is particularly prominent with no or minimal levels of post-filtering, which, in part, may be due to overlapping Gibbs edge artefacts. Despite this, it can be seen from HD recovery data in Table 2 that, with matched voxel variance, there is very little dependence of

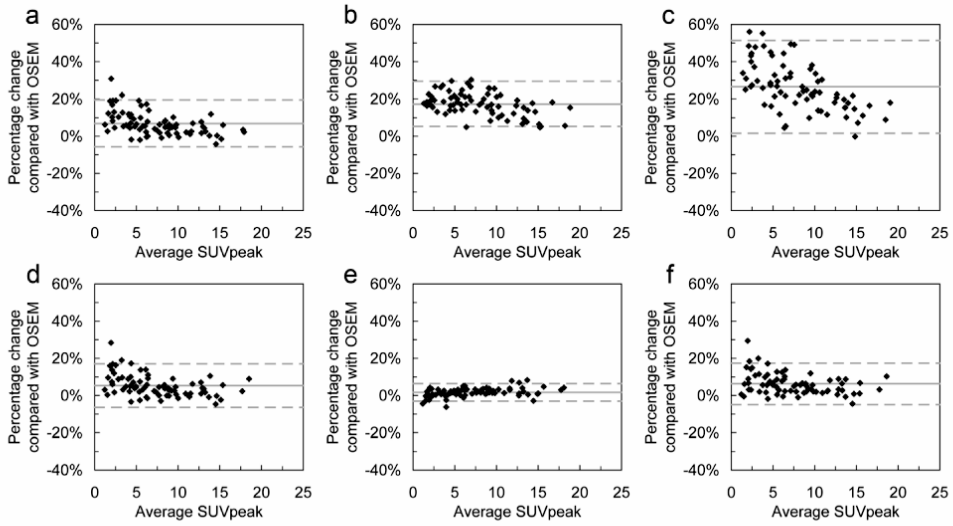


Figure 4 Bland-Altman plots showing relative percentage differences for SUV_{peak} relative to OSEM. Plots on the top row are for images with matched image voxel COV post-filters, and plots on the bottom row are for images with matched SUV_{max} recovery post-filters. Plot layout **(a-f)** and line styles are as per Figure 3.

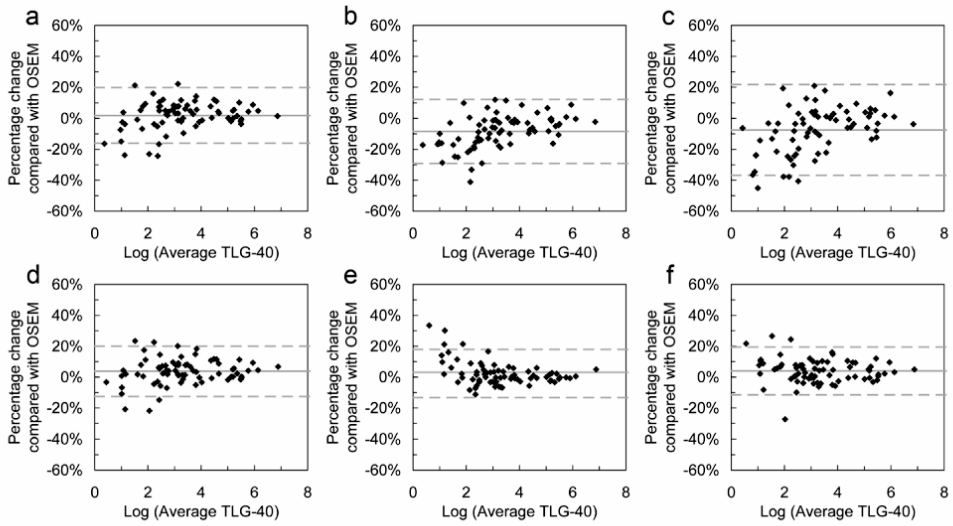


Figure 5 Bland-Altman plots showing relative percentage differences for TLG-40 relative to OSEM. Plots on the top row are for images with matched image voxel COV post-filters, and plots on the bottom row are for images with matched SUV_{max} recovery post-filters. Plot layout **(a-f)** and line styles are as per Figure 3.

recovery on sphere size for the 13- to 37-mm spheres, which is a desirable property. This highlights the importance of establishing a full understanding of the impact of these algorithms, and it is the duty of medical physics experts to educate clinicians on changes expected to quantification.

Ideally, the implementation of PSF modelling would not lead to Gibbs artefacts, but given the necessary compromises for PET imaging with limited statistics, an improvement in one area such as in image resolution is almost certainly going to lead to a deterioration in other aspects. Overall, whether the changes are desirable is application dependent, with our data showing smaller absolute errors for smaller spheres (but not for large spheres) and reduced dependency on quantification with lesion size.

Matching image noise produces marked increases in SUV_{max}, particularly with PSF reconstructions, that are potentially clinically significant, depending on local practice. This highlights the pitfalls of using uptake metrics such as $SUV_{max,}$ that are so sensitive to partial volume effects and reconstruction parameters, with fixed thresholds for malignancy. The largest increases in SUV_{max} occur for small lesions, which typically have

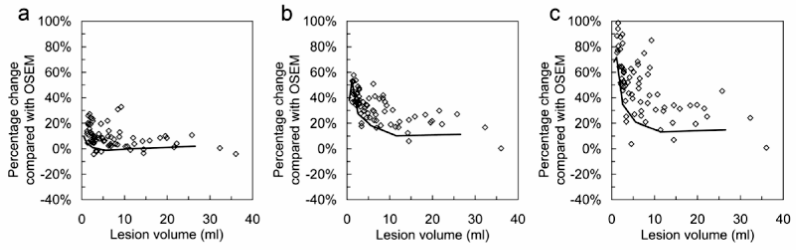


Figure 6 Percentage differences versus lesion volume. Relative percentage differences of SUV_{max} for TOF **(a)**, HD **(b)** and UHD **(c)** compared with OSEM for the 74 lung lesions against lesion volume as measured in the OSEM images using a 40% threshold. All plots are for matched voxel COV. The black line illustrates the percentage differences measured for the six spheres of the image quality phantom.

Table 6 Changes in SUV_{max} with lesions with borderline values for malignancy

Volume	SUV_{max}	Matched noise			Matched SUV_{max}		
(ml)	OSEM	TOF	HD	UHD	TOF	HD	UHD
1.26	2.1	2.4	*2.8*	*3.6*	2.2	1.8	1.9
9.26	2.2	*2.9*	*2.9*	*4.0*	*2.8*	2.3	*2.9*
2.80	2.2	2.5	*3.2*	*3.6*	2.3	2.2	2.2
1.99	2.4	2.8	*3.2*	*4.2*	2.7	2.2	2.4
6.04	2.5	3.1	*3.3*	*4.3*	3.0	2.7	*3.2*
1.83	2.7	3.4	*3.8*	***5.5***	3.2	2.5	3.0
3.99	2.8	3.1	*3.6*	*4.0*	3.0	2.8	3.0
1.52	3.2	4.0	*4.5*	***6.3***	3.8	2.9	3.4
8.90	3.2	3.6	*4.3*	***5.2***	3.4	3.0	3.3
1.56	3.4	4.0	*4.9*	***6.8***	3.8	3.1	3.6
2.51	3.4	4.2	*4.6*	***6.4***	4.0	3.3	4.0
6.33	3.5	4.0	4.3	**5.3**	3.8	3.4	3.7
1.40	3.9	4.1	***6.0***	***6.9***	3.9	3.8	3.8
14.22	4.4	4.6	4.9	**5.2**	4.5	4.3	4.5
2.06	4.5	**5.6**	***7.0***	***8.5***	**5.4**	4.7	4.2

SUV_{max} values for patients with low uptake (<5.0) for suspicion of malignancy and the SUV_{max} data obtained from the PSF and TOF-based algorithms using the two post-filter strategies. The left column shows the lesion volume as measured from the OSEM PET image using a 40% threshold. Values in bold represent lesions that would have changed classification using a strict SUV_{max} cut-off of 5.0. Values in italics represent increases of greater than 25%.

low SUV_{max} (less than 5), which is consistent with other studies [36,37]. One potential solution may be to modify thresholds based on estimated tumour volume. It would be useful to extend the matching of SUV_{max} to smaller objects, but this is not possible due to the limitation of the current NEMA phantom, with 10 mm being the diameter of the smallest sphere insert. It is these small lesions, with SUV_{max} close to the typical cut-offs for discrimination of benign and malignant disease, that are arguably the most critical lesions for lung cancer staging as they are likely to be possible additional pulmonary nodules or lymph nodes. Determining whether a lymph node is malignant, particularly those in the mediastinum, has a considerable influence on the overall staging and will play a major role in patient management. This change in SUV_{max} is expected to require an adaptation of locally used thresholds for discrimination of disease. It was also noted from the phantom studies that variability of SUV_{max} was worse for PSF-based algorithms in the small spheres, which suggests worse test-retest performance in clinical data. This is suspected to be due to increased inter-voxel correlation that is introduced when using PSF-based algorithms [21]. This increased correlation results in a reduction of voxel variance (and hence the voxel COV as used in this study as a noise metric), but it has been shown to potentially result in larger variability of uptake metrics within small ROIs [45]. We feel that the impact of PSF modelling on variability for clinical data has yet to be explored fully, and while this is beyond the scope of this study, it is recommended that caution is observed when applying PSF modelling for assessing response to treatment with follow-up scans. Despite this, the reduced levels of post-filtering required with PSF and PSF + TOF have been shown to improve lesion visualisation [28-30].

With matched voxel COV, SUV_{peak} experiences similar differences to those seen for SUV_{max}, albeit to a lesser extent. Quantification of peak uptake implicitly includes an

additional filtering operation with a spherical kernel. The small mean relative differences for TLG suggests that it is a relatively robust uptake metric when comparing against OSEM images for either filtering strategies. The large degree of variability seen in the relative changes, as highlighted by the confidence intervals in Tables 4 and 5, may be concerning. However, it should also be noted that the total range of TLG observed in this study is approximately a full order of magnitude greater than SUV_{max} and SUV_{peak}. The use of TLG has been reported in assessment of therapy response and, recently, for prognosis in a small number of studies. The increased stability of TLG with a volume delineation based on a percentage of SUV_{max} suggests the metric may be more appropriate than SUV_{max} for staging and prognosis as the evidence base for this metric is established. We believe that this is the first time that the dependence of TLG on reconstruction algorithm has been explored in the literature.

Alternatively, post-filters for PSF and TOF-based algorithms can be determined to give SUV_{max} that, according to this institute's practice, would not alter the outcome of the study. For all lesions with borderline SUV_{max} for suspicion of malignancy, relative changes with PSF and TOF-based reconstructions were less than 20%.

Matching SUV_{max} between PSF-based algorithms and OSEM has been demonstrated previously [39]. However, our study has also shown that matching SUV_{max} will significantly reduce the voxel variance in the image compared with OSEM, which we believe has yet to be demonstrated quantitatively. Combined with increased voxel correlation, this reduction of voxel variance alters the image appearance quite considerably and may be perceived as over-smoothing of images. Findings from this study are based upon an image matrix of 256×256 voxels, whereas other centres may use different parameters such as 200×200 or 400×400 voxels, which are common choices on the mCT due to the system's intrinsic 400×400 matrix. We believe that, when Gaussian post-filtering is applied, the dependence of both image noise and SUV_{max} on matrix choice is diminished. It has also been shown that the thickness of the walls of the fillable spheres of the NEMA phantom has an impact on SUV_{max} quantification [46,47]. This is only seen to cause appreciable error with low sphere-to-background contrast and small spheres, and hence, we expect that the impact on the test objects used in this study is likely to be minimal.

It is noted that the degree of post-filtering for the HD and UHD algorithms (6.6 and 6.5 mm, respectively) will reduce spatial resolution for these PSF-based algorithms that are intended to provide superior spatial resolution. However, we feel that this approach may be beneficial when deploying a new PET/CT scanner to an existing clinical setting, comparing patient scans for follow-up with other systems or supporting the transition to a 'new' imaging facility with a catalogue or library of images with higher resolution.

In this study, the addition of TOF increased the variation in ratio values of image voxel variance for both phantom and patient data with either matched noise or matched SUV_{max}. In the patient data only, TOF appeared to introduce a slight positive bias and greater distribution of differences in the SUV_{max} data. This was not seen in the phantom studies and the cause of this is unclear. It could be due to a dependence on patient size, as TOF is associated with SNR gains proportional to the diameter of object [48]. However, in this study and others [20], this did not appear

to apply in lung images where the majority of tissue in the image has low density with very low uptake of FDG.

We believe this is the first study to demonstrate SNR gains with PSF and/or TOF using lesion uptake as a measure of signal with two different criteria for choosing post-filtering. A recent study has shown reductions in voxel variance and gains in SNR but measured only in uniform areas of uptake with patient livers [27]. One study has evaluated SNR gains using lesion uptake as the signal [25] but only comparing images reconstructed with PSF and PSF + TOF, with the intention to demonstrate the SNR gains brought on by TOF. It was expected that SNR gains would be seen for PSF and TOF-based algorithms compared with conventional OSEM. However, it was not anticipated that the gains in SNR would be greater when parameters are chosen to match SUV_{max}. This may be of particular relevance for low-contrast lesions elsewhere in the body, such as the abdomen, which do not have the inherent high lesion to background contrast of lung lesions. The notion that increased levels of post-filtering may be superior in terms of SNR gains seems slightly at odds with published work on lesion detection that suggest less post-filtering results in optimal lesion detection [28,29]. This may be due to fact that the definition of SNR in this study is not a direct indicator of lesion detectability.

There are two limitations with this study where future work is planned. Firstly, no histological correlation with FDG uptake measured in the lesions was performed as in other studies [36]. Therefore, it is not possible to determine cut-off values and diagnostic accuracy of the uptake metrics in the two strategies of implementation. This is arguably outside the scope of this study as the purpose was not to determine such data. Secondly, we have only assessed lung lesions, and from other studies [25], it is likely that reconstruction will perform differently in other areas of the body.

The effect of PSF and TOF-based reconstruction on quantification, particularly SUV_{max}, has limited their introduction into routine clinical use despite demonstrated improvements in lesion detectability. This study extends existing studies [39] which have shown that the impact on SUV_{max} can be addressed with appropriate post-filters, by demonstrating that the same approach can be used for reconstructions with TOF reconstructions and also with alternative uptake metrics such as SUV_{peak} or TLG. Furthermore, we have demonstrated that this additional filtering to match SUV_{max} actually provides added gains in SNR over parameters to match image voxel COV. However, if the additional smoothing is visually undesirable, an alternative methodology can be used which performs the additional filtering required to match SUV_{max} only for quantification and is not visualised [41].

Conclusions

This work evaluated the impact of reconstructions that include PSF modelling and/or TOF on lesion classification according to a local protocol by assessing changes in FDG uptake measurements. Two objective strategies for post-filtering were investigated: matching image voxel COV versus matching SUV_{max}. For matched voxel COV, considerable increases in SUV_{max} and SUV_{peak} were observed compared with OSEM. Using post-filters to match SUV_{max} reduced the discrepancies of either SUV_{max} or SUV_{peak} across reconstructions, particularly with PSF modelling. This

also resulted in a considerable reduction in voxel variance. Some small discrepancies in patient data still remained when TOF was incorporated, which was not seen in phantom data, warranting further investigation. The TLG metric appears to be more robust in either scheme of post-filtering despite a slightly larger variation in the amount of change, which may be less of a problem considering the large range of TLG data observed. This suggests TLG may be a more suitable metric to adopt instead of SUV_{max} as the evidence base develops. Gains in SNR were seen in both implementations with the greatest gains seen for matched SUV_{max} post-filters.

Abbreviations

COV: coefficient of variation; EORTC: European Organization for Research and Treatment of cancer; FDG: 2-fluoro-2-deoxy-D-glucose; FWHM: full width at half the maximum; HD: Siemens HD-PET reconstruction; OSEM: ordered subset expectation maximisation; PERCIST: PET response criteria in solid tumours; PET: positron emission tomography; PSF: point spread function; ROI: region of interest; SNR: signal-to-noise ratio; SUV: standardised uptake value; TLG: total lesion glycolysis; TOF: time of flight; UHD: Siemens ultraHD-PET reconstruction; VOI: volume of interest.

Competing interests

This study was performed as part of the first author's (IA) PhD project, which receives financial support (course fees) from Siemens Healthcare that is paid to the nuclear medicine department and then directly to the University of Manchester.

Authors' contributions

IA managed and processed all image data and wrote the manuscript. MK assisted with data analysis (MATLAB code) and critically appraised and modified the draft manuscript. HW critically appraised and modified the draft manuscript. JM is a PhD supervisor and critically appraised and modified the draft manuscript. All authors read and approved the final manuscript.

Author details

[1]Nuclear Medicine, Central Manchester University Hospitals, Oxford Road, Manchester, UK. [2]Institute of Population Health, MAHSC, University of Manchester, Manchester, UK. [3]Molecular Imaging, Healthcare Sector, Siemens PLC, Oxford, UK.

Received: 28 February 2014 Accepted: 2 September 2014
Published online: 30 November 2014

References

1. Cerfolio RJ, Bryant AS, Ohja B, Bartolucci AA: **The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival.** *J Thorac Cardiovasc Surg* 2005, **130:**151–159.
2. Cerfolio RJ, Bryant AS, Ojha B, Eloubeidi M: **Improving the inaccuracies of clinical staging of patients with NSCLC: a prospective trial.** *Ann Thorac Surg* 2005, **80:**1207–1214.
3. Subedi N, Scarsbrook A, Darby M, Korde K, Mc Shane P, Muers MF: **The clinical impact of integrated FDG PET–CT on management decisions in patients with lung cancer.** *Lung Cancer* 2009, **64**(3):301–307.
4. Dijkman B, Schuurbiers O, Vriens D, Looijen-Salamon M, Bussink J, Timmer-Bonte J, Snoeren M, Oyen W, van der Heijden H, de Geus-Oei L-F: **The role of 18F-FDG PET in the differentiation between lung metastases and synchronous second primary lung tumours.** *Eur J Nucl Med Mol Imaging* 2010, **37**(11):2037–2047.
5. Gregory DL, Hicks RJ, Hogg A, Binns DS, Shum PL, Milner A, Link E, Ball DL, Mac Manus MP: **Effect of PET/CT on management of patients with non-small cell lung cancer: results of a prospective study with 5-year survival data.** *J Nucl Med* 2012, **53**(7):1007–1015.
6. Erdi YE, Macapinlac H, Rosenzweig KE, Humm JL, Larson SM, Erdi AK, Yorke ED: **Use of PET to monitor the response of lung cancer to radiation treatment.** *Eur J Nucl Med Mol Imaging* 2000, **27**(7):861–866.
7. Beyer T, Czernin J, Freudenberg LS: **Variations in clinical PET/CT operations: results of an international survey of active PET/CT users.** *J Nucl Med* 2011, **52**(2):303–310.
8. Bryant AS, Cerfolio RJ: **The maximum standardized uptake values on integrated FDG-PET/CT is useful in differentiating benign from malignant pulmonary nodules.** *Ann Thorac Surg* 2006, **82**(3):1016–1020.
9. Nambu A, Kato S, Sato Y, Okuwaki H, Nishikawa K, Saito A, Matsumoto K, Ichikawa T, Araki T: **Relationship between maximum standardized uptake value (SUVmax) of lung cancer and lymph node metastasis on FDG-PET.** *Ann Nucl Med* 2009, **23**(3):269–275.
10. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, Pruim J, Price P: **Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group.** *Eur J Cancer* 1999, **35:**1773–1782.
11. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA: **Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study.** *J Nucl Med* 2004, **45:**1519–1527.
12. Nahmias C, Wahl LM: **Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors.** *J Nucl Med* 2008, **49:**1804–1808.

13. Lodge MA, Chaudhry MA, Wahl RL: **Noise considerations for PET quantification using maximum and peak standardized uptake value.** *J Nucl Med* 2012, **53:**1041–1047.
14. 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:**122S–150S.
15. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, Casilla C, Fazzari M, Srivastava N, Yeung HW, Humm JL, Guillem J, Downey R, Karpeh M, Cohen AE, Ginsberg R: **Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis.** *Clin Positron Imaging* 1999, **2:**159–171.
16. Wiele C, Kruse V, Smeets P, Sathekge M, Maes A: **Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours.** *Eur J Nucl Med Mol Imaging* 2013, **40:**290–301.
17. Pak K, Cheon GI, Nam H-Y, Kim S-J, Kang KW, Chung J-K, Kim EE, Lee DS: **Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis.** *J Nucl Med* 2014, **55:**884–890.
18. Chung MDHH, PD, Kwon MDHW, Kang MDKW, Park MDN-H, Song MDY-S, Chung MDJ-K, Kang MDS-B, Kim MDJW: **Prognostic value of preoperative metabolic tumor volume and total lesion glycolysis in patients with epithelial ovarian cancer.** *Ann Surg Oncol* 2012, **19:**1966–1972.
19. Hyun S, Ahn H, Kim H, Ahn M-J, Park K, Ahn Y, Kim J, Shim Y, Choi J: **Volume-based assessment by 18F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer.** *Eur J Nucl Med Mol Imaging* 2014, **41:**50–58.
20. Panin VY, Kehren F, Michel C, Casey M: **Fully 3-D PET reconstruction with system matrix derived from point source measurements.** *Med Imaging, IEEE Trans* 2006, **25:**907–921.
21. Alessio AM, Stearns CW, Shan T, Ross SG, Kohlmyer S, Ganin A, Kinahan PE: **Application and evaluation of a measured spatially variant system model for PET image reconstruction.** *Med Imaging IEEE Trans* 2010, **29:**938–949.
22. Conti M, Bendriem B, Casey M, Chen M, Kehren F, Michel C, Panin V: **First experimental results of time-of-flight reconstruction on an LSO PET scanner.** *Phys Med Biol* 2005, **50:**4507.
23. Kalemis A, Delattre BMA, Heinzer S: **Sequential whole-body PET/MR scanner: concept, clinical use, and optimisation after two years in the clinic. The manufacturer's perspective.** *Magn Reson Mater Phy* 2013, **26:**5–23.
24. Karp JS, Surti S, Daube-Witherspoon ME, Muehllehner G: **Benefit of time-of-flight in PET: experimental and clinical results.** *J Nucl Med* 2008, **49:**462–470.
25. Lois C, Jakoby BW, Long MJ, Hubner KF, Barker DW, Casey ME, Conti M, Panin VY, Kadrmas DJ, Townsend DW: **An assessment of the impact of incorporating time-of-flight information into clinical PET/CT imaging.** *J Nucl Med* 2010, **51:**237–245.
26. El Fakhri G, Surti S, Trott CM, Scheuermann J, Karp JS: **Improvement in lesion detection with whole-body oncologic time of flight PET.** *J Nucl Med* 2011, **52:**347–353.
27. Akamatsu G, Ishikawa K, Mitsumoto K, Taniguchi T, Ohya N, Baba S, Abe K, Sasaki M: **Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters.** *J Nucl Med* 2012, **53:**1716–1722.
28. Kadrmas DJ, Casey ME, Black NF, Hamill JJ, Panin VY, Conti M: **Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes.** *Med Imaging IEEE Trans* 2009, **28:**523–534.
29. Schaefferkoetter J, Casey ME, Townsend DW, El Fakhri G: **Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study.** *Phys Med Biol* 2013, **58:**1465–1478.
30. Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW: **Impact of time-of-flight on PET tumor detection.** *J Nucl Med* 2009, **50:**1315–1323.
31. Tong S, Alessio AM, Thielemans K, Stearns C, Ross S, Kinahan PE: **Properties and mitigation of edge artifacts in PSF-based PET reconstruction.** *Nucl Sci IEEE Trans* 2011, **58:**2264–2275.
32. Rahmim A, Qi J, Sossi V: **Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls.** *Med Phys* 2013, **40:**064301–064315.
33. Rapisarda E, Bettinardi V, Thielemans K, Gilardi MC: **Image-based point spread function implementation in a fully 3D OSEM reconstruction algorithm for PET.** *Phys Med Biol* 2010, **55:**4131–4151.
34. Watson CC: **Estimating effective model kernel widths for PSF reconstruction in PET**. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE; 23–29 Oct. 2011*; 2011:2368–2374.
35. Kotasidis FA, Matthews JC, Angelis GI, Noonan PJ, Jackson A, Price P, Lionheart WR, Reader AJ: **Single scan parameterization of space-variant point spread functions in image space via a printed array: the impact for two PET/CT scanners.** *Phys Med Biol* 2011, **56:**2917–2942.
36. Lasnon C, Hicks RJ, Beauregard J-M, Milner A, Paciencia M, Guizard A-V, Bardet S, Gervais R, Lemoel G, Zalcman G, Aide N: **Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non–small cell lung cancer.** *Clin Nucl Med* 2012, **37:**971–976.
37. Andersen FL, Klausen TL, Loft A, Beyer T, Holm S: **Clinical evaluation of PET image reconstruction using a spatial resolution model.** *Eur J Radiol* 2013, **82:**862–869.
38. Prieto E, Dominguez-Prado I, Garcia-Velloso MJ, Penuelas I, Richter JA, Marti-Climent JM: **Impact of time-of-flight and point-spread-function in SUV quantification for oncological PET.** *Clin Nucl Med* 2013, **38:**103–109.
39. Lasnon C, Desmonts C, Quak E, Gervais R, Do P, Dubos-Arvis C, Aide N: **Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients.** *Eur J Nucl Med Mol Imaging* 2013, **40:**985–996.
40. 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:**181–200.

41. Kelly MD, Declerck JM: **SUVref: reducing reconstruction-dependent variation in PET SUV.** *EJNMMI Res* 2011, **1:**16.
42. Jakoby BW, Bercier Y, Conti M, Bendriem B, Townsend D: **Physical and clinical performance of the mCT time-of-flight PET/CT scanner.** *Phys Med Biol* 2011, **56:**2375–2389.
43. Conti M, Bendriem B, Casey M, Mu C, Kehren F, Michel C, Panin V: **Implementation of time-of-flight on CPS HiRez PET scanner.** In *Nuclear Science Symposium Conference Record, 2004 IEEE; 16–22 Oct. 2004*; 2004:2796–2800.
44. National Electrical Manufacturers Association: *NEMA Standards Publication NU 2–2007: Performance Measurements of Positron Emission Tomographs*. NEMA 2007.
45. Rahmim A, Tang J: **Noise propagation in resolution modeled PET imaging and its impact on detectability.** *Phys Med Biol* 2013, **58:**6945–6968.
46. Hofheinz F, Dittrich S, Potzsch C, Hoff J: **Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold.** *Phys Med Biol* 2010, **55:**1099–1113.
47. Berthon B, Marshall C, Edwards A, Evans M, Spezi E: **Influence of cold walls on PET image quantification and volume segmentation: a phantom study.** *Med Phys* 2013, **40:**082505.
48. Budinger TF: **Time-of-flight positron emission tomography: status relative to conventional PET.** *J Nucl Med* 1983, **24:**73–78.

doi:10.1186/s40658-014-0099-3
Cite this article as: Armstrong *et al.*: Impact of point spread function modelling and time of flight on FDG uptake measurements in lung lesions using alternative filtering strategies. *EJNMMI Physics* 2014 **1**:99.

6. THIRD PAPER

The assessment of time-of-flight on image quality and quantification with reduced administered activity and scan times in ^{18}F-FDG PET

Article status:
Submitted to Nuclear Medicine Communications in December 2014
Accepted decision received January 2015

The paper included in this chapter is the published version:

Armstrong IS, James MD, Williams HA, Kelly MD, Matthews JC. *The assessment of time-of-flight on image quality and quantification with reduced administered activity and scan times in ^{18}F-FDG PET*. Nucl Med Commun 2015; **36**:728-737

The assessment of time-of-flight on image quality and quantification with reduced administered activity and scan times in ^{18}F-FDG PET

Ian S. Armstrong[a,b], Jackie M. James[a], Heather A. Williams[a], Matthew D. Kelly[c] and Julian C. Matthews[b]

Objectives **The last decade has seen considerable technological innovations in PET detectors with the availability, among other advances, of time-of-flight (TOF). TOF has been shown to increase the signal-to-noise ratio (SNR), which should allow for a reduction in acquired counts while maintaining image quality.**

Methods **Fifty-eight patients referred for routine ^{18}F-flurodeoxyglucose (^{18}F-FDG) oncology PET studies were included in this study. Patients with weight below or above 100 kg were prescribed 350 or 400 MBq of ^{18}F-FDG, respectively. Listmode data were acquired for 2.5 min per bed position and reconstructed with ordered-subset expectation maximization (OSEM) reconstruction. TOF reconstruction was performed on reduced-count data, with two levels of reduction (− 20 and − 40% for patients < 100 kg and − 16 and − 30% for patients > 100 kg) achieved by clipping the listmode data. Liver SNR, mediastinum mean standardized uptake value (SUV_{mean}), and lesion maximum standardized uptake value (SUV_{max}) were measured in all images. All images were visually assessed as adequate or suboptimal.**

Results **No significant difference was seen in mediastinum SUV_{mean} or lesion SUV_{max} when comparing reduced-count TOF with full-count OSEM images. Compared with the original OSEM images, liver SNR was higher for TOF images using the more conservative − 20% reduction of counts ($P < 0.001$, Wilcoxon's signed-rank test), whereas no significant statistical difference was seen with − 40% reductions.**

Conclusion **Incorporation of TOF allows for a reduction in acquired counts; this method has been implemented at our institution, with administered activity reduced for all patients to 280 MBq and a reduction in scan times for all but the largest patients. This has significantly reduced the patient radiation dose and improved scanner flexibility and throughput.** *Nucl Med Commun* **36:728–737**

Nuclear Medicine Communications 2015, **36**:728–737

Keywords: dose reduction, ^{18}F-flurodeoxyglucose, oncology, PET, time-of-flight

[a]Nuclear Medicine Centre, Central Manchester University Hospitals NHS Foundation Trust, [b]Institute of Population Health, MAHSC, University of Manchester, Manchester and [c]Siemens PLC, Healthcare Sector, Molecular Imaging, Oxford, UK

Correspondence to Ian S. Armstrong, MPhys, MSc, Nuclear Medicine Centre, Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester M13 9WL, UK
Tel: + 44 161 276 4785; fax: + 44 161 276 8023;
e-mail: ian.armstrong@cmft.nhs.uk

Received 10 December 2014 **Revised** 30 January 2015
Accepted 30 January 2015

Introduction

PET with computed tomography (PET/CT) using ^{18}F-fluorodeoxyglucose (^{18}F-FDG) has long been recognized as a key factor in patient management in oncology. The recommended provision in the UK in 2009 for PET-CT for oncology was between 57 000 and 61 500 scans per annum [1]. In addition, the role of ^{18}F-FDG PET-CT in nononcological pathology has also grown, with pyrexia of unknown origin, cardiac device infection, vasculitis and the differential diagnosis of dementia being among the recommended indications [2]. The emerging interest in nononcological indications for ^{18}F-FDG PET-CT has had considerable impact on work patterns at this institution, with increasing numbers of full-body and/or dual-time-point imaging studies being performed for pyrexia of unknown origin, vasculitis and suspected cardiac device infection [3]. This has resulted in scheduling difficulties within the department.

In the UK, the quantity of injected radioactivity for ^{18}F-FDG is specified by the Notes for Guidance published by the Administration of Radioactive Substance Advisory Committee (ARSAC) [4]. This guidance states that the Diagnostic Reference Level (DRL) for ^{18}F-FDG PET for tumour imaging is 400 MBq. This has not been changed since 1998, and is associated with a quoted effective dose to the patient of 8 mSv from the ^{18}F-FDG. The dose rate to nuclear medicine practitioners from patients administered 350 MBq of ^{18}F-FDG at a distance of 1 m immediately after injection has been reported as being ~45 μSv/h [5]. This is ~10 times that from patients injected with similar quantities of ^{99m}Tc tracers [6]. Consequently, staff doses from ^{18}F-FDG PET patients are considerably higher than those

DOI: 10.1097/MNM.0000000000000305

from patients administered comparable quantities of ^{99m}Tc radiopharmaceuticals [7].

Over the past decade, the two components of a PET-CT system – the PET detector and the CT system – have undergone significant technological advances with respect to both instrumentation and image reconstruction. Advances in PET technology include a shift from two-dimensional to three-dimensional (3D) systems, faster scintillation crystals such as lutetium orthosilicate, which have facilitated collection of time-of-flight (TOF) information, and more sophisticated image reconstruction algorithms that include point spread function resolution modelling. Advances in CT technology include the introduction of tube current modulation, automatic tube voltage selection and the availability of iterative reconstruction. These advances in CT have been widely demonstrated to bring about reductions in the effective dose to patients without a substantial impact on image quality [8–10]. Conversely, advances in PET detector technology appear to have resulted in very little change in the administered activity of ^{18}F-FDG, and hence patient effective dose. Typical nuclear medicine practice in the UK is to prescribe all patients with activity either equal to or close to the ARSAC DRL, with little change in practice following the transition to more advanced scanners. This failure to deviate from the existing ARSAC DRL, given technological advances, has recently been criticized [11]. Data collected by the UK National Cancer Research Institute PET Network from 18 sites in the UK show that the majority of sites administer between 350 and 400 MBq (personal communication – Lucy Pike, NCRI).

The use of TOF is known to increase the signal-to-noise ratio (SNR) in PET images [12]. In recent times, gains in SNR from TOF on current PET scanners have been demonstrated [13–18], with SNR gains of ~14–35% measured from clinical data [13,16,18]. In addition, the increases in SNR have been widely demonstrated to improve lesion detection [19–23]. Two studies have shown that performance of lesion detection with TOF was comparable to that of non-TOF lesion detection, where the image acquisition time for the TOF data had been reduced [20,21]. Specifically, the study by Kadrmas *et al.* [21] stated that lesion detection performance measured using the same scanner as in this study was comparable between non-TOF and TOF images, where the image acquisition time for the TOF data had been reduced by 40%. These studies on lesion detection performance were either based solely on phantoms with lesions *in situ* [19–21], or artificial lesions were imaged in air and inserted into patient images through data manipulation [22,23].

Assuming that the acquired data are Poisson in nature, the SNR is expected to be approximately proportional to the square-root of the noise-equivalent counts (NECs) that are acquired [24]. Therefore, reducing the acquired NEC by 40% would reduce the SNR by ~22%. On the basis of published data on phantoms discussed previously, it was hypothesized that, by applying TOF to images with 40% fewer counts, the magnitude of SNR gain from TOF should approximately compensate for the SNR reduction resulting from fewer counts.

The robustness of typical ^{18}F-FDG uptake measurements, maximum standardized uptake value (SUV_{max}) and mean standardized uptake value (SUV_{mean}), with decreasing levels of counts has been studied previously [25]. A range of images from clinical data were created using resampled listmode data. These images were reconstructed only with a TOF algorithm, and the study showed that the counts could be reduced by 50% while maintaining good agreement between uptake metrics. This is encouraging, as it suggests that quantification of lesion uptake should be relatively unaffected by the 40% reduction in acquired counts proposed in this study.

The aim of this study was to assess the use of TOF acquisition and reconstruction for routine clinical imaging with a reduction in acquired counts in our local practice. As mentioned, studies comparing lesion detection between TOF and non-TOF images have been widely reported in the literature. As such, the focus of this study is on the conservation of ^{18}F-FDG uptake quantification and image quality, with the latter being assessed visually and quantitatively. We aimed to investigate a selection of metrics that we considered clinically important so that, by demonstrating that metrics are consistent, we could show that the images have the same clinical utility in terms of these parameters. We believe that the comparison of reduced-count images with TOF against non-TOF full-count images has yet to be investigated in terms of clinical data. Reduction of counts was achieved by trimming listmode data acquired from patient images to shorter times per bed position. An alternative to reducing the acquisition time is reducing the administered ^{18}F-FDG activity, and therefore a phantom experiment was conducted to estimate the equivalence of these two approaches. Reducing activity levels will reduce the radiation exposure to both staff and patients, whereas a reduction in acquisition time will improve the flexibility of scanner scheduling and patient throughput to accommodate changing workflow patterns.

Methods

Patient study group

Retrospective data from 58 patients referred for routine PET/CT imaging were included in this study. Patient demographics are given in Table 1. Clinical indications for the patients were as follows: assessment of non-small-cell lung cancer (32 patients), lymphoma post-treatment assessment (10 patients), assessment of a single pulmonary nodule (four patients), restaging of colorectal cancer (four patients), staging of oesophageal cancer

Table 1 **Patient demographics of this study group**

	All	Male	Female
N	58	31	27
Median age (range)	69 (25–89)	69 (27–89)	68 (25–85)
Median weight (range) (kg)	78 (61–142)	80 (61–123)	70 (61–142)
Weight > 100 kg (*n*)	9	8	1
Median height (range) (m)	1.65 (1.48–1.91)	1.74 (1.57–1.91)	1.58 (1.48–1.83)
Median BMI (range) (kg/m^2)	27.3 (19.3–42.4)	27.0 (19.3–38.2)	27.5 (23.8–42.4)

(three patients), primary of unknown origin (two patients), initial lymphoma staging (one patient), and assessment for recurrent breast cancer (one patient) and cervical cancer (one patient). All data were fully anonymized before inclusion in this study. After seeking advice from the institutional research department, ethical approval was not deemed necessary for reprocessing of anonymized retrospective data.

^{18}F-FDG PET imaging

Patients with a body weight less than 100 kg (49/58) were prescribed 350 MBq of ^{18}F-FDG, whereas the remaining nine patients with a body weight greater than 100 kg were prescribed 400 MBq. This was in accordance with the local protocol used at this institution at the time these data were collected. PET image acquisition commenced at a median (interquartile range) time of 63 (61–66) min after injection of ^{18}F-FDG. Data were acquired on a Siemens Biograph mCT (Siemens Healthcare, Knoxville, Tennessee, USA) with an extended axial field of view, using a multibed listmode acquisition for 2.5 min per bed position for all patients. The axial scanning range was from skull base to mid thigh and comprised acquisitions over either six or seven bed positions. The scanner is capable of collecting TOF information and has a quoted timing resolution of 550 ps. Performance characteristics of this scanner have been published [15].

Data were reconstructed according to a local protocol consisting of 3D ordinary Poisson ordered-subset expectation maximization (OSEM) iterative reconstruction using three iterations and 24 subsets. A 200×200 image matrix was used, with a transaxial slice thickness of 4.0 mm, and a 5.0 mm full-width at half-maximum Gaussian postfilter was applied to all images. This resulted in isotropic voxel dimensions of $4.0 \times 4.0 \times 4.0$ mm.

For each patient, two additional images were reconstructed with two different levels of reductions in the number of acquired counts, implemented by trimming the listmode data to a shorter time per bed position and by reconstruction with TOF. In this study, conservation of lesion quantification was required and the number of reconstruction iterations was chosen to reflect this. TOF has been widely demonstrated to result in faster reconstruction convergence [14–16]. Work performed at this institution and elsewhere has shown that TOF reconstruction with two iterations, compared with three iterations for non-TOF, provides comparable quantification, whether this be contrast recovery or SUV [16,18,26, 27]. The degrees of count reduction were different between patients with weight less than 100 kg and those with weight greater than 100 kg. For the 49 patients with weight less than 100 kg, the time per bed position, and hence counts, was reduced by 20 and 40%, whereas for the remaining nine patients with weight greater than 100 kg the time per bed position was reduced by 16 and 30%. The choice of count reduction was based on the aim of this work, which was to reduce both the administered activity and the scanning time. However, the prioritized reduction was that of administered activity, and therefore, for patients with body weight less than 100 kg, the intended method of implementing a 20% count reduction would be to reduce only the administered activity by 20%, whereas a 40% count reduction would be achieved by reducing the scanning time in addition to a 20% reduction in activity. In addition to this, the intention was also to modify the current practice of increasing the prescribed activity to 400 from 350 MBq for patients with body weight above 100 kg. Considering this, it was hypothesized that a 30% reduction from 400 MBq would be approximately equal to a 20% reduction from 350 MBq, potentially allowing continuity of administered activity across the existing threshold of 100 kg. Regardless of patient weight, the two groups are referred to as 'TOF-1' and 'TOF-2', where TOF-1 is the group with more conservative count reduction compared with the TOF-2 group.

Image analysis

On comparing the reduced-count TOF images with the full-count non-TOF images, we had several desirable criteria for this study: (i) to maintain or potentially improve image quality, quantified by measurement of the SNR within an area of uniform ^{18}F-FDG uptake; (ii) to maintain ^{18}F-FDG uptake measurements, specifically SUV_{mean}, in normal tissue, which is increasingly used as a reference in lymphoma [28]; and (iii) to maintain ^{18}F-FDG uptake measurements, specifically SUV_{max}, in any lesion that was identified as being metabolically active.

All images were analysed with Siemens TrueD image software on a Siemens Syngo workstation (Siemens Healthcare). Volumes of interest (VOIs) were defined on the images reconstructed with non-TOF OSEM. A 3-cm-diameter spherical VOI was placed in a uniform area of uptake within the patient's liver. A manually defined ellipsoidal VOI was placed in the mediastinum corresponding to the pulmonary trunk, using the CT image for guidance. An ellipsoidal VOI was placed over each metabolically active lesion that was suspicious for malignancy, as identified by an experienced nuclear

medicine consultant physician (J.M.J.). SNR was measured in the liver VOI and calculated as the ratio of the mean to the SD of voxel values within the VOI, where the units of the voxel values were absolute activity concentration – that is, Bq/ml – and not SUV. This measure of SNR in the liver is a useful quantitative surrogate for the more subjective visual assessment of image quality [29] and has been used as a metric to compare image quality for different reconstructions [16]. SUV_{mean} was measured in the mediastinum VOI, and SUV_{max} was measured for each lesion. Both SUV_{mean} and SUV_{max} were normalized using injected activity and patient body weight. All images (OSEM, TOF-1 and TOF-2) were assessed visually by the same experienced observer and rated as adequate or suboptimal, with the observer being blinded to the image type.

Simulating reduced administered activity

The clinical images in this study with reduced counts would ideally reflect a reduction in the administered activity and/or scanning time. In this study, the reduced-count images were produced by reducing the time per bed position. It was therefore important to quantify the equivalence of reduced activity and reduced acquisition time. This is because the count-rate response of a PET scanner is known to be nonlinear because of the different dependences of true and random coincidences from inside and outside the PET field of view and detector dead time [30].

To measure the count-rate response under imaging conditions that are similar to those in humans, the background cavity of the National Electrical Manufacturers Association (NEMA) image quality phantom [31], representing a torso, was filled with ^{18}F-FDG at a concentration of 8 kBq/ml. The fillable spheres were present in the phantom but filled with nonactive water. This is typical of the activity concentration in an area of uniform ^{18}F-FDG distribution in a patient's liver that is seen at this institution. The total activity within the phantom was 73 MBq. A 70-cm-long NEMA scatter phantom was placed adjacent to the image quality phantom, with the internal scatter line filled with 177 MBq ^{18}F-FDG. This gave a total activity of 250 MBq, which is approximately the activity inside a patient injected with 370 MBq ^{18}F-FDG at 1 h after injection, ignoring any excretion. A 2-h dynamic acquisition of the phantoms was performed using the PET scanner. The image quality phantom was placed at the centre of the field of view and data were acquired as 48×2.5 min frames. For each image, the NEC was calculated as follows:

$$\mathrm{NEC} = \frac{T^2}{(T+S+R)},$$

where T, S and R are the number of true, scattered and random coincidences, respectively, that were detected and which were determined from sinogram header files for each frame.

Statistical testing

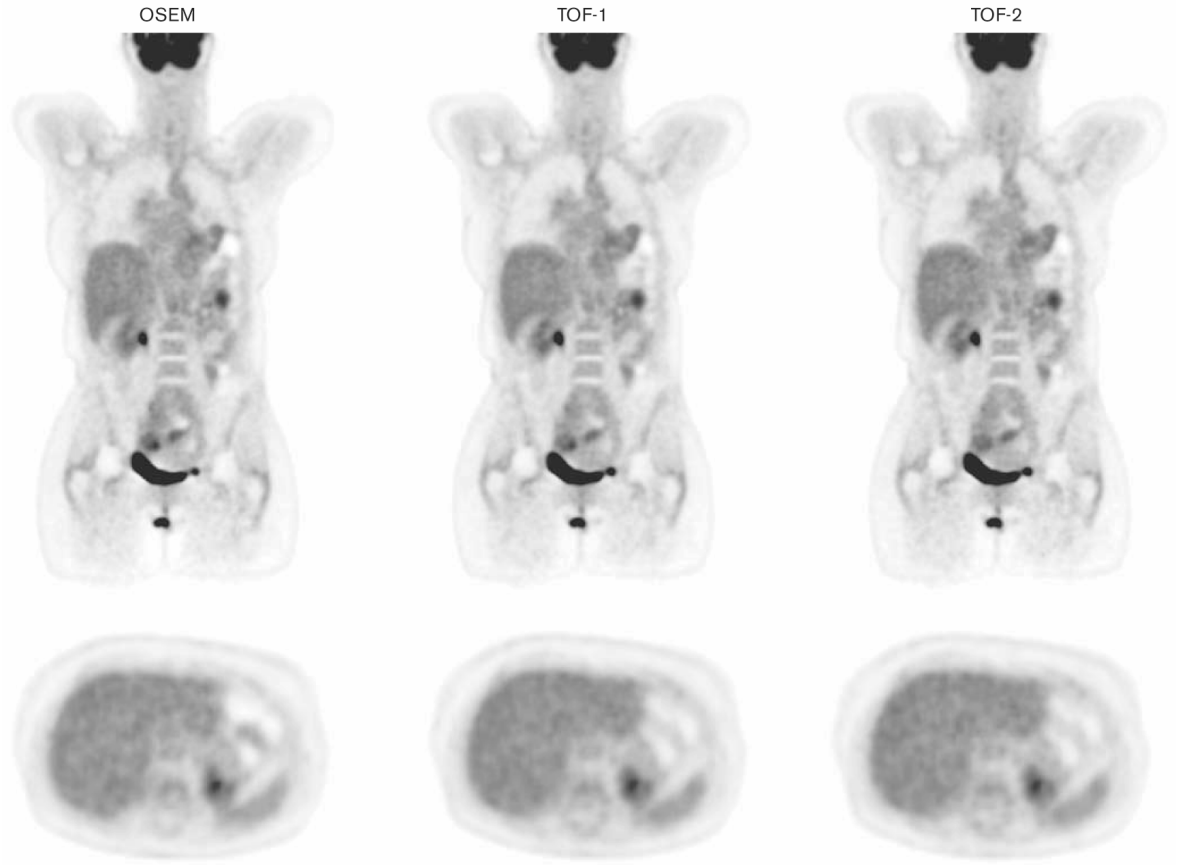
A Shapiro–Wilk test for normality was performed on the absolute differences of liver SNR, mediastinum SUV_{mean} and lesion SUV_{max} for each of the two reduced-count TOF images compared with the original OSEM images. The test determined that the absolute differences were not normally distributed for any of these data. As such, a nonparametric Wilcoxon's signed-rank test ($\alpha = 0.05$) for paired samples was performed for data from each of the TOF images compared with the original OSEM images. Following a Bonferroni multiple comparisons correction for the six tests, P-values of less than 0.008 were considered statistically significant. Relative differences in the metrics were also shown to exhibit a non-normal distribution, and hence these data were represented by medians and interquartile ranges. Finally, Spearman's rank coefficient tests were performed to establish agreement. All statistical analyses were carried out using StatsDirect v2.8.0 (StatsDirect Statistical Software, Altrincham, UK).

Results

Patient study group

SNR was measured in the liver for all patients, but the mediastinum SUV_{mean} was not measured in one patient because of widespread mediastinal disease. A total of 97 metabolically active lesions were identified, and SUV_{max} was measured for each of these. Example PET images are shown in Figs 1 and 2, with the images in Fig. 2 being those from a patient with high body weight and BMI. As can be seen from the images, the SNR in the liver for the TOF images with a 20% count reduction (TOF-1) is superior to that for OSEM images, whereas the TOF images with a 40% count reduction (TOF-2) result in an SNR that is comparable to that of OSEM images. The liver SNR, mediastinum SUV_{mean} and lesion SUV_{max}, measured for the patient group, are given in Table 2 for OSEM and both sets of reduced-count TOF images. Table 3 shows the relative differences in and the data from statistical analysis of the three measurements from the reduced-count TOF images compared with OSEM images. The liver SNR from the TOF-1 images was significantly greater ($P < 0.001$) than that from OSEM, with a median relative increase of 8.3%. No significant difference in the liver SNR ($P = 0.16$) was seen between TOF-2 images and OSEM images. The median relative difference for TOF-2 was -0.43% and the interquartile range was slightly biased towards a negative reduction. No significant difference was observed in mediastinum SUV_{mean} between TOF-1 or TOF-2 images and OSEM images ($P = 0.49$ and 0.42 for TOF-1 and TOF-2, respectively), nor in lesion SUV_{max} between TOF-1 or TOF-2 images and OSEM images ($P = 0.02$ in both cases). In addition, the small relative differences and the interquartile range for these differences for both of these metrics, as shown in Table 3, suggest that quantification

Fig. 1



Coronal and transaxial PET images for a female patient with weight 63 kg and BMI 26.6 kg/m^2. The signal-to-noise ratio, measured in the liver, for OSEM, TOF-1 (20% count reduction) and TOF-2 (40% count reduction) images was 12.4, 13.3 and 12.6, respectively. The upper window level has been set to an SUV of 5.0. OSEM, ordered-subset expectation maximization; SUV, standardized uptake value; TOF, time-of-flight.

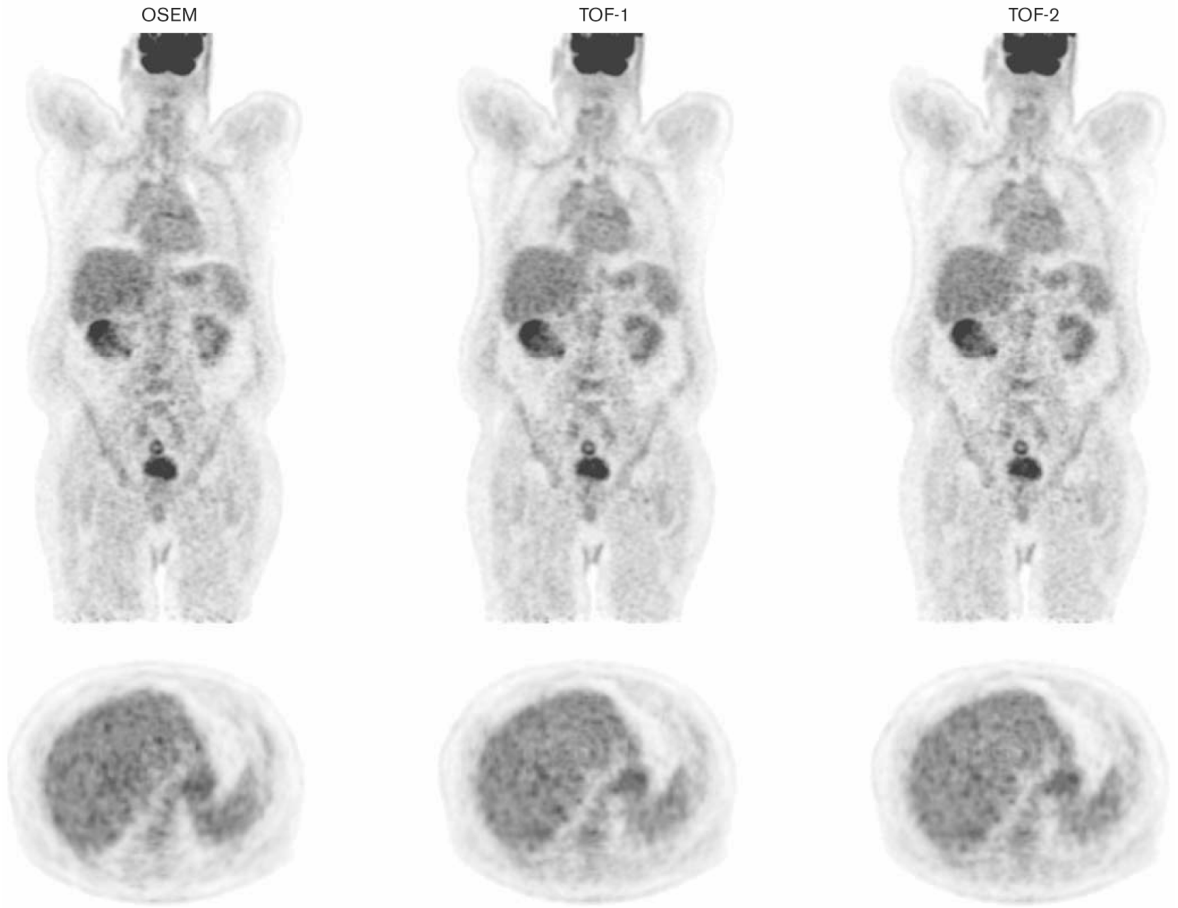
by either of the reduced-count TOF images is essentially equivalent to the data from OSEM images.

Figure 3 shows the dependency of liver SNR on patient weight for all three sets of images, with there being a clear reciprocal relationship for all reconstructions. Figure 4 shows the results from the visual assessment during which images were scored as adequate or suboptimal. Data are shown for OSEM and the greater level of count reduction (TOF-2 images): that is, 40% reduction for body weight less than 100 kg and 30% reduction for body weight greater than 100 kg. It can be seen that some patients were considered suboptimal for the OSEM images. Among the patients with body weight less than 100 kg who were prescribed 350 MBq ^{18}F-FDG, the TOF-2 images (40% reduction) from 39/49 patients were rated as adequate. All TOF-1 images (20% count reduction) were rated as adequate. From these results, two cutoff values were defined for which the 40% count reduction had to be applied. These were body weight less than 85 kg and BMI less than 28. These cutoff values are shown in the figure, and their implementation would result in images for 32/49 patients having counts reduced by 40%. Among patients with body weight greater than 100 kg who were prescribed 400 MBq ^{18}F-FDG, the TOF-2 images (30% count reduction) from five of nine patients were rated as adequate and the TOF-1 images (16% count reduction) in the remaining four patients were rated as adequate. From these limited data, a single cutoff of 115 kg for body weight was chosen.

Count-rate response verification

The relationship between the NEC rate and phantom activity from the 2-h dynamic acquisition is shown in Fig. 5. It can be seen from Fig. 5b that there is only a small deviation from the line of proportionality. This

Fig. 2



Coronal and transaxial PET images for a female patient with weight 142 kg and BMI 42.4 kg/m^2. The signal-to-noise ratio, measured in the liver, for OSEM, TOF-1 (16% count reduction) and TOF-2 (30% count reduction) images was 6.6, 7.3 and 6.7, respectively. The upper window level has been set to an SUV of 5.0. OSEM, ordered-subset expectation maximization; SUV, standardized uptake value; TOF, time-of-flight.

Table 2 Liver SNR, mediastinum SUV_{mean} and lesion SUV_{max} for the three sets of images

	OSEM	TOF-1	TOF-2
Liver SNR ($n=58$)			
Median	10.1	11.6	10.9
Interquartile range	8.9–11.8	9.9–13.1	9.1–12.0
MS SUV_{mean} ($n=57$)			
Median	1.78	1.80	1.81
Interquartile range	1.71–1.96	1.69–1.96	1.68–1.95
Lesion SUV_{max} ($n=97$)			
Median	8.3	8.5	8.6
Interquartile range	5.6–11.2	6.0–11.4	6.0–11.4

MS, mediastinum; OSEM, ordered-subset expectation maximization; SNR, signal-to-noise ratio; SUV, standardized uptake value; TOF, time-of-flight.

suggests that using a reduction in acquisition time as a surrogate for a reduction of administered activity leads to a slight overestimation of the reduction in acquired NEC, and hence SNR, that would occur. This is, however, considered advantageous, as it demonstrates the 'worst-case' image quality for a given reduction in administered activity.

Discussion

This study has investigated the effectiveness of TOF in facilitating a reduction in counts while obtaining adequate image quality and maintaining consistent quantification of ^{18}F-FDG uptake. The study has demonstrated that, from the perspective of the chosen quantitative metrics, the reduced-count TOF images maintain the same clinical utility as the full-count images without TOF. Other studies have demonstrated gains in SNR with PET, but we believe that this is the first study to demonstrate the usefulness of TOF in allowing

reductions of acquired counts in clinical data. The supplementary phantom experiment examined how reduction of scanning time may be used as a surrogate for reduction of the administered activity to obtain similar levels of count reduction.

The aim of implementing TOF for ^{18}F-FDG PET/CT at this institution was to enable a reduction in both the prescribed activity and the scanning time. Following this work, these reductions have been implemented locally since September 2014 as a combination of reduced activity and scanning time, as summarized in Table 4. Given that the count reductions applied range from 16 to 40%, we infer from published data [21] that lesion detection will not be compromised following implementation of the scheme.

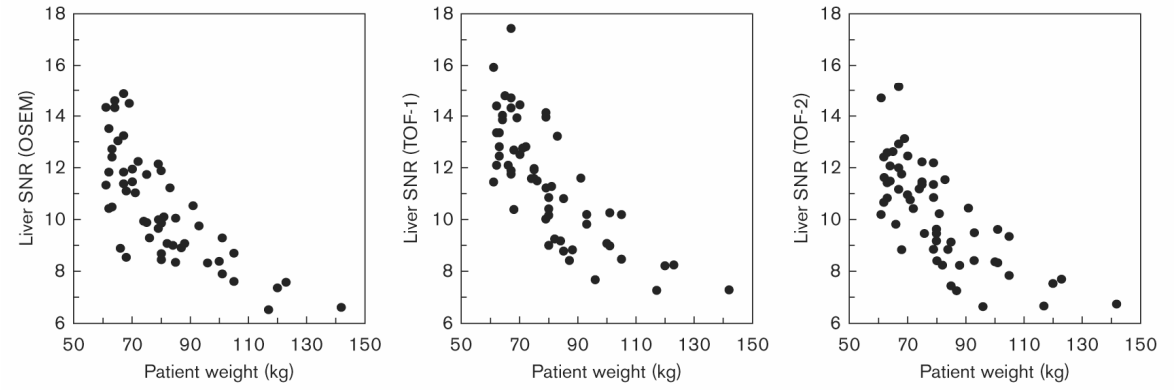
Table 3 The relative changes and statistical test results of the liver SNR, mediastinum SUV_{mean} and lesion SUV_{max} for the TOF-1 and TOF-2 images against images reconstructed with OSEM

	TOF-1	TOF-2
Liver SNR		
Relative difference to OSEM		
Median	+8.3%	−0.4%
Interquartile range	+1.0 to +16.1%	−9.7 to +2.8%
Wilcoxon's *P*-value	<0.001	0.16
Spearman's coefficient	0.83	0.84
MS SUV_{mean}		
Relative difference to OSEM		
Median	+0.6%	+0.6%
Interquartile range	−2.0 to +2.7%	−1.6 to +2.8%
Wilcoxon *P*-value	0.49	0.42
Spearman's coefficient	0.95	0.96
Lesion SUV_{max}		
Relative difference to OSEM		
Median	+1.4%	+1.6%
Interquartile range	−1.8 to 4.9%	−2.8 to 5.5%
Wilcoxon *P*-value	0.02	0.02
Spearman's coefficient	0.99	0.99

MS, mediastinum; OSEM, ordered-subset expectation maximization; SNR, signal-to-noise ratio; SUV, standardized uptake value; TOF, time-of-flight.

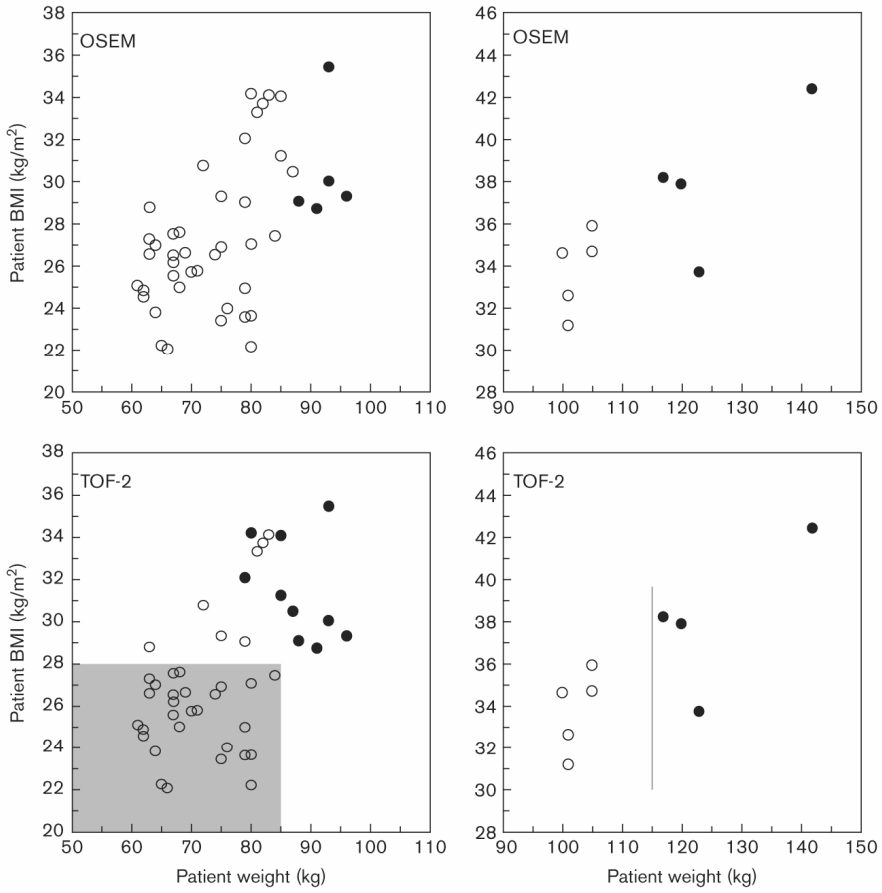
As can be seen from the table, the prescribed activity has been standardized to 280 MBq for all patients. For patients with body weight less than 100 kg, this 20% reduction from the previously prescribed dose of 350 MBq has provided scope for reducing the scanning time in smaller patients in whom the TOF-2 images were considered adequate. This was achieved by reducing the scan time per bed position by 20% – that is, to 2.0 min – and reducing the total time for a six bed-position scan from 15 to 12 min and that for a seven bed-position scan from 17.5 to 14 min. The 30% count reduction among patients prescribed 400 MBq approximates to reducing the prescribed activity by 30%, from 400 to 280 MBq, while maintaining the same scanning time. The 16% count reduction among these patients approximates to reducing the prescribed activity by 30%, from 400 to 280 MBq, but increasing the scanning time by 20%, from 2.5 min per bed position to 3.0 min, with the revised scheme in Table 4 reflecting this. The frequency of patients who require this increase in bed time is less than 5% and is not expected to have a significant impact on throughput. Importantly, the increase in prescribed activity, from 350 to 400 MBq, among patients with body weight greater than 100 kg is no longer applied. Instead, the prescribed activity is now 280 MBq for all patients, with scanning time increased to compensate. This leads to a significant reduction in radiation exposure to both patients and practitioners working with PET/CT. The reduction in radiation dose is particularly important for patients with certain types of cancer, such as lymphoma, who can be relatively young and frequently respond extremely well to treatment. These patients are expected to undergo multiple PET/CT examinations to assess the

Fig. 3



Plots of the SNR measured in the liver ($n = 58$) against patient weight for the OSEM and the two TOF images. OSEM, ordered-subset expectation maximization; SNR, signal-to-noise ratio; TOF, time-of-flight.

Fig. 4



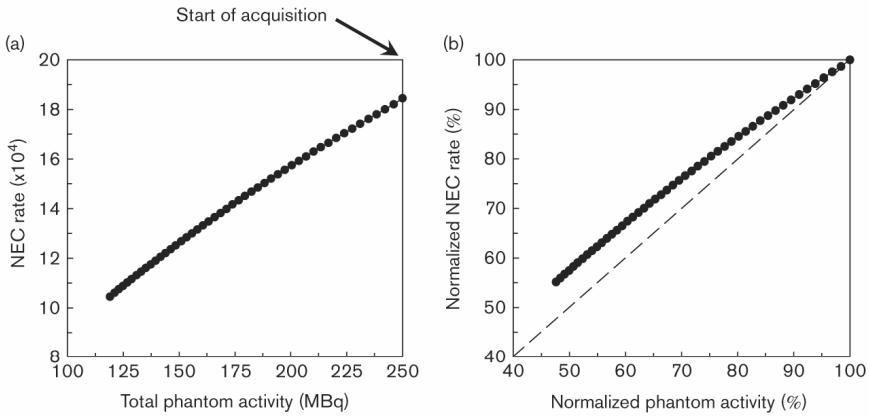
Data showing the categorical visual assessment results, plotted as patient body weight and BMI. Data shown in the upper plots are for OSEM reconstruction, whereas data shown in the lower plots are for the TOF-2 images with the greater level of count reduction. Left: data for patients with body weight less than 100 kg, and data for a 40% reduction in counts. Right: data for patients with body weight greater than 100 kg, and data for a 30% reduction in counts. Unfilled circles represent images considered 'adequate' and filled circles represent images considered 'suboptimal'. The shaded region in the lower left panel represents the chosen cutoff region in which TOF-2 images were to be used instead of TOF-1 images. All TOF-1 images were considered adequate, and hence the plots are not shown for these data. OSEM, ordered-subset expectation maximization; TOF, time-of-flight.

response to therapy. Following the implementation of the scheme, our experience suggests a marked improvement in the flexibility of scanner scheduling, meaning additional scans such as local views or separate head and neck scans had a reduced impact on the workflow for the session.

The reconstruction with TOF takes longer than that with OSEM, with time taken for a single bed position reconstruction on the current hardware being 75 and 45 s, respectively, using the parameters in this study. This is despite using one less iteration with TOF compared with OSEM, as TOF has been shown to provide faster convergence [16]. The scanner reconstructs during the acquisition phase, and, despite the increase in reconstruction time with TOF and decrease in acquisition times, there is still sufficient time to complete the reconstruction of the previously acquired bed position while acquiring the current bed position so that pile up of reconstruction jobs does not occur.

It is noted that the current European guidelines for use of ^{18}F-FDG PET/CT in oncology [32] recommend a linear weight-based protocol for the prescribed activity, which is dependent on the acquisition type of the PET scanner and on scan time. Applying the recommendations to the scanner used in this study, which acquires images in the 3D mode and has a bed overlap of 41%, the prescribed activity should be (6.9 × weight in kg)/(minutes per bed position). For a 70 kg patient, the prescribed activity

Fig. 5



Acquired noise-equivalent count (NEC) rate for each frame against the combined ^{18}F-FDG activity in the NEMA phantom and scatter phantom (a) over the course of a 2 h dynamic acquisition and (b) normalized to the respective values from the first frame of the acquisition. The dashed line on the normalized plot represents the line of identity. ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; NEMA, National Electrical Manufacturers Association.

Table 4 A summary of how reductions in prescribed activity and scanning time have been implemented locally following this work

	Prescribed activity (MBq)	Scan time per bed (min)
Body weight < 85 kg and BMI < 28	280	2.0
Body weight 85–115 kg or BMI > 28	280	2.5
Body weight >115 kg	280	3.0

would be 240 and 193 MBq for 2.0 and 2.5 min per bed position, respectively. The scheme implemented locally would prescribe 280 MBq to a 70 kg patient and acquire for 2.0 min per bed position, which still appears to be a conservative count reduction with respect to the recommendation set out in the European guidelines. However, the feasibility of using the levels of activity recommended by the European guidelines is dependent on how the image is reconstructed, which will influence how statistical noise is handled. The guidelines specify recovery values of SUV_{max} that should be achieved on measurement in the six spheres of an NEMA image quality phantom. These recovery values are strongly dependent on reconstruction parameters and postfilter. A study [33] using the same scanner as in this study demonstrated that it was necessary to apply an additional smoothing filter to a typical reconstruction for it to fall within the recovery criteria specified in the guidelines. This evidence suggests that PET images reconstructed to meet the recovery criteria specified in the guidelines will be considerably smoother than those described here, which would obviously allow for further reductions in administered activity.

The validity of increasing the administered activity linearly with body weight has been questioned recently, with an alternative nonlinear protocol [34]. In both linear and nonlinear models, the upper limit of the administered activity is either not clearly defined or is still relatively high. The study by De Groot *et al.* [34] does not mention an upper limit of administered activity. The European guidelines recommend an upper limit of 530 MBq but suggest increasing bed time rather than increasing activity to maintain image quality. This philosophy has been used in this study to maintain the image quality among larger patients and remove the increase from 350 to 400 MBq for patients with body weight greater than 100 kg. Although this study is based on patients referred for oncology studies, the implementation of the scheme has been extended to non-oncologic ^{18}F-FDG studies that are conducted at this institution, with successful results.

The scheme implemented locally following this study involves a reduction in the prescribed activity to all patients to a fixed level that is considerably less than the UK DRL specified by ARSAC. However, the ideal outcome from the optimization of activity and scanning protocol would be the achievement of consistent image quality, quantified by SNR in an area of uniform uptake, such as the liver. As can be seen from the data presented in this work, this is still not the case with this implementation. The image quality in larger patients has been maintained by increasing the bed time, but the lower limit of 2 min per bed position results in an increasing, and arguably unnecessary, SNR in smaller patients. At our centre, there is no real value in further decreasing the scanning time in smaller patients because of the current provision of uptake facilities placing a limit on the throughput of ^{18}F-FDG patients. However, other centres with a provision for high-throughput scanning may

benefit from reducing scanning times. This lower limit on scan times at our centre highlights the obvious shortcomings of a fixed activity for all patients. A solution may be to implement a weight-based protocol for smaller patients, with an upper limit of 280 MBq. Despite this, we consider the reductions in both administered activity and scanning time to be a significant step in the right direction for clinical practice. This study encourages the continuation of optimization work in the future, particularly the evaluation of a weight-based protocol for small patients.

Conclusion

This study has shown that TOF information can allow for a reduction in acquired counts while maintaining image quality and commonly used quantification metrics. Locally, this has been implemented as a combination of reduced administered activity and reduced scanning time in ^{18}F-FDG PET studies. After optimizing the choice of scanning time based on patient weight and/or BMI, the scanning time was reduced for the majority of patients, improving the flexibility of scanner scheduling to more easily accommodate changing workflow patterns.

Acknowledgements

Conflicts of interest

This work is part of the corresponding author's PhD project that is registered with the University of Manchester. The PhD study receives financial support (university course fees) that is paid by Siemens Healthcare to the University of Manchester. Ian Armstrong receives no direct financial support for PhD study. Matt Kelly is an employee of Siemens Healthcare. For the remaining authors, there are no conflicts of interest.

References

1 UK PET CT Advisory Board. Oncology FDG PETCT scan referral criteria. Available at: *http://www.bnms.org.uk/images/stories/downloads/documents/referral_indications_pet-ct_board_2009.pdf*. [Accessed 16 September 2014].

2 Royal College of Physicians and Royal College of Radiologists. *Evidence-based indications for the use of PET-CT in the UK*. London: Royal College of Physicians and Royal College of Radiologists; 2013.

3 James J, Ahmed F, Armstrong IS, Zaidi A, Arumugam P. Use of ^{18}F-FDG PET/CT in suspected Cardiac Implantable Electronic Device (CIED) related infection. *Nucl Med Commun* 2014; **35**:A16.

4 Administration of Radioactive Substances Advisory Committee. *Notes for Guidance on the Clinical Administration of Radiopharmaceuticals and Use of Sealed Radioactive Sources – March 2006 (updated 2014)*. Available at: *www.gov.uk/government/publications/arsac-notes-for-guidance*. [Accessed 26 February 2015].

5 Benatar NA, Cronin BF, O'Doherty MJ. Radiation dose rates from patients undergoing PET: implications for technologists and waiting areas. *Eur J Nucl Med* 2000; **27**:583–589.

6 Harding LK, Mostafa AB, Roden L, Williams N. Dose rates from patients having nuclear medicine investigations. *Nucl Med Commun* 1985; **6**:191–194.

7 Chiesa C, De Sanctis V, Crippa F, Schiavini M, Fraigola CE, Bogni A, *et al.* Radiation dose to technicians per nuclear medicine procedure: comparison between technetium-99m, gallium-67, and iodine-131 radiotracers and fluorine-18 fluorodeoxyglucose. *Eur J Nucl Med* 1997; **24**:1380–1389.

8 Lee TY, Chhem RK. Impact of new technologies on dose reduction in CT. *Eur J Radiol* 2010; **76**:28–35.

9 Prakash P, Kalra MK, Kambadakone AK, Pien H, Hsieh J, Blake MA, Sahani DV. Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique. *Invest Radiol* 2010; **45**:202–210.

10 McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for reducing radiation dose in CT. *Radiol Clin North Am* 2009; **47**:27–40.

11 McMeekin H, Wagner T, Burniston M, McCool D. 400 MBq of ^{18}F-FDG: one size no longer fits all? *Nucl Med Commun* 2014; **35**:781–782.

12 Budinger TF. Time-of-flight positron emission tomography: status relative to conventional PET. *J Nucl Med* 1983; **24**:73–78.

13 Karp JS, Surti S, Daube-Witherspoon ME, Muehllehner G. Benefit of time-of-flight in PET: experimental and clinical results. *J Nucl Med* 2008; **49**:462–470.

14 Lois C, Jakoby BW, Long MJ, Hubner KF, Barker DW, Casey ME, *et al.* An assessment of the impact of incorporating time-of-flight information into clinical PET/CT imaging. *J Nucl Med* 2010; **51**:237–245.

15 Jakoby BW, Bercier Y, Conti M, Casey ME, Bendriem B, Townsend DW. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. *Phys Med Biol* 2011; **56**:2375–2389.

16 Akamatsu G, Ishikawa K, Mitsumoto K, Taniguchi T, Ohya N, Baba S, *et al.* Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters. *J Nucl Med* 2012; **53**:1716–1722.

17 Akamatsu G, Mitsumoto K, Ishikawa K, Taniguchi T, Ohya N, Baba S, *et al.* Benefits of point-spread function and time of flight for PET/CT image quality in relation to the body mass index and injected dose. *Clin Nucl Med* 2013; **38**:407–412.

18 Akamatsu G, Mitsumoto K, Taniguchi T, Tsutsui Y, Baba S, Sasaki M. Influences of point-spread function and time-of-flight reconstructions on standardized uptake value of lymph node metastases in FDG-PET. *Eur J Radiol* 2014; **83**:226–230.

19 Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. *J Nucl Med* 2009; **50**:1315–1323.

20 Surti S, Karp JS. Experimental evaluation of a simple lesion detection task with time-of-flight PET. *Phys Med Biol* 2009; **54**:373–384.

21 Kadrmas DJ, Oktay MB, Casey ME, Hamill JJ. Effect of scan time on oncologic lesion detection in whole-body PET. *IEEE Trans Nucl Sci* 2012; **59**:1940–1947.

22 El Fakhri G, Surti S, Trott CM, Scheuermann J, Karp JS. Improvement in lesion detection with whole-body oncologic time-of-flight PET. *J Nucl Med* 2011; **52**:347–353.

23 Schaefferkoetter J, Casey M, Townsend D, El Fakhri G. Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study. *Phys Med Biol* 2013; **58**:1465–1478.

24 Watson CC, Casey ME, Bendriem B, Carney JP, Townsend DW, Eberl S, *et al.* Optimizing injected dose in clinical PET by accurately modeling the counting-rate response functions specific to individual patient scans. *J Nucl Med* 2005; **46**:1825–1834.

25 Murray I, Kalemis A, Glennon J, Hasan S, Quraishi S, Beyer T, Avril N. Time-of-flight PET/CT using low-activity protocols: potential implications for cancer therapy monitoring. *Eur J Nucl Med Mol Imaging* 2010; **37**:1643–1653.

26 Armstrong IS, Kelly MD, Williams HA, Matthews JC. Impact of point spread function modelling and time-of-flight on FDG uptake measurements in lung lesions using alternative filtering strategies. *EJNMMI Physics* 2014; **1**:99.

27 Conti M, Bendriem B, Casey M, Chen M, Kehren F, Michel C, Panin V. First experimental results of time-of-flight reconstruction on an LSO PET scanner. *Phys Med Biol* 2005; **50**:4507–4526.

28 Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leuk Lymphoma* 2009; **50**:1257–1260.

29 McDermott GM, Chowdhury FU, Scarsbrook AF. Evaluation of noise equivalent count parameters as indicators of adult whole-body FDG-PET image quality. *Ann Nucl Med* 2013; **27**:855–861.

30 Tarantola G, Zito F, Gerundini P. PET instrumentation and reconstruction algorithms in whole-body applications. *J Nucl Med* 2003; **44**:756–769.

31 National Electrical Manufacturers Association. 2007*NEMA Standards Publication NU 2-2007: Performance Measurements of Positron Emission Tomographs*. Rosslyn, VA: National Electrical Manufacturers Association.

32 Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, *et al.* FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* 2010; **37**:181–200.

33 Kelly MD, Declerck JM. SUVref: reducing reconstruction-dependent variation in PET SUV. *EJNMMI Res* 2011; **1**:16.

34 De Groot EH, Post N, Boellaard R, Wagenaar NR, Willemsen AT, van Dalen JA. Optimized dose regimen for whole-body FDG-PET imaging. *EJNMMI Res* 2013; **3**:63.

7. FOURTH PAPER

Evaluation of the utility of estimated covariance kernels for predicting regional ensemble variance

Article status:

This paper is a follow-on to the IEEE Conference Proceedings that was accepted for the Molecular Imaging Conference in October 2015

Submitted to IEEE Transactions on Medical Imaging in March 2016
Rejection decision received April 2016 following review

The paper will be revised and submitted to an alterative journal in the future taking account of the detailed comments received from the three reviewers.

The paper included in this chapter was the version submitted to IEEE Transactions on Medical Imaging.

Evaluation of the utility of estimated covariance kernels for predicting regional ensemble variance

Ian S. Armstrong, Heather A. Williams, Matthew D. Kelly, and Julian C. Matthews *Member IEEE*

***Abstract*—This work aims to develop and evaluate methods to estimate voxel covariance kernels and evaluate their utility in predicting regional ensemble variance (EV). The goal is to enable tools for comparing EV from reconstruction algorithms with very different correlation structures.**

50 (uniform Ge-68 phantom A), 200 (uniform Ge-68 phantom B) and 24 (NEMA image quality) replicate images were acquired on a Siemens Biograph mCT. Images were reconstructed: with and without resolution modelling (RM); with 3 and 10 iterations; and with 2mm isotropic voxels. Cubic samples of 40mm $\times$ 40mm $\times$ 40mm were extracted from various locations and used to estimate covariance kernels. EV was measured in spherical, ellipsoidal and hollow spherical shell regions of different diameters (40mm max). This was compared against estimates of EV derived from the covariance kernels. Bootstrap resampling was use to evaluate the precision of kernel derived EV. Additionally, covariance kernels were generated from multiple locations within the liver of 4 routine oncology patients.

Covariance kernels were broader when RM was used. Good agreement between the measured and kernel estimated EV was observed, being mostly within 95% confidence interval expected from measured EV. Good repeatability and precision of kernel derived estimates was observed, with small differences with position generally less than that seen for different reconstruction algorithms. Changes in EV within the hot NEMA spheres with reconstruction algorithms generally reflected observed changes in background EV regions. The kernels shapes and EV values derived from patients reflected that observed with phantom data.

***Index Terms*—Nuclear imaging; Image reconstruction - iterative methods, quantification and estimation, optimization.**

I. Introduction

PET reconstruction with resolution modelling (RM) has been available for several years [1], [2], [3] and has been shown to improve contrast recovery and lesion detection [1], [4]. The addition of RM reduces the magnitude of the voxel variance but increases the width of the voxel covariance kernel, resulting in increased near-neighboring voxel correlation [5], [6], [7], [8], [9]. This reduction of voxel variance has been shown to reduce variance on single-voxel measurements across multiple image replicates [10]. It has been suggested that this increased voxel correlation with RM will increase the variance of region-based measurements which have been referred to as ensemble variance (EV) in previous publications [8]. This is particularly relevant for alternative uptake metrics such as SUV_{peak} [11]. While Tong *et. al.* showed EV in background regions of a NEMA appeared comparable with and without RM at low numbers of reconstruction iterations [7], we believe that this potential degradation of EV with RM has not been fully evaluated. Methods to determine variance maps have been reported [12], [13] but these approaches are quite involved and do not address the effect of covariance on EV. EV may be estimated by bootstrap re-sampling of projection data [14] but such approaches are very demanding computationally. As such, we believe there is a need for a simple and pragmatic approach.

The aim of this work was to estimate covariance kernels and evaluate their utility in predicting EV. Specifically we hypothesise that although voxel covariance will vary spatially, within a relativity small homogeneous volume such variations will be sufficiently small to enable estimation and that this estimate can predict regional EV. This hypothesis was evaluated by comparing such estimates with those derived from replicate images of uniform Ge-68 phantoms and a NEMA image quality phantom. Finally, the feasibility of estimating a covariance kernel from clinical data using a single image was assessed.

I. Armstrong and H. Williams are with Nuclear Medicine, Central Manchester University Hospitals, Oxford Road, Manchester, UK

I. Armstrong and J. Matthews are with Imaging Sciences, University of Manchester, UK

M. Kelly is with Siemens Molecular Imaging, Oxford, UK

II. Theory

A. Estimation of covariance matrix

Given a homogeneous sub-volume of an image, $\mathcal{S}$, and assuming a spatially invariant covariance structure, the covariance kernel can be estimated in image space using:

$$k_{i-i',j-j',k-k'} = \sum_{(i',j',k')} (x_{i,j,k} - \mu_{\mathcal{S}})\,(x_{i',j',k'} - \mu_{\mathcal{S}}) \quad (1)$$

where $x_{i,j,k}$ represents the voxel concentration in 3 dimensions indexed by i, j, k, $\mu_{\mathcal{S}}$ is the mean concentration in $\mathcal{S}$ and such that $(i', j', k') \in \mathcal{S}$ and $(i, j, k) \in \mathcal{S}$. Alternatively the covariance kernel can be estimated in the Fourier domain using

$$K_{i,j,k} = E\left[\left|X_{i,j,k}\right|^2\right] \quad (2)$$

where $K_{i,j,k}$ and $X_{i,j,k}$ are the three dimensional discrete Fourier transforms of $k_{i,j,k}$ and $x_{i,j,k}$ and $E[\cdot]$ is the expectation [15]. The implementation using a 3-dimensional fast Fourier transform will result in the incorporation erroneous wrapped around voxels, which in reality will have little if any correlation. If it is assumed that the true correlation of such wrapped around voxels is zero then the kernel estimated using equation 2 can be corrected using:

$$k_{i,j,k} = \frac{n_x n_y n_z}{(n_x - |i|)\,(n_y - |j|)\,(n_z - |k|)} \mathcal{F}^{-1}\left(K_{i,j,k}\right) \quad (3)$$

where n_x, n_y, and n_z are the number of voxels in the x, y, and z directions of a cuboid sub-volume $\mathcal{S}$, $|\cdot|$ is the absolute value, and $\mathcal{F}^{-1}$ is the discrete inverse 3D Fourier transform.

B. Estimation of regional ensemble variance

Let $\mathbf{x}$ be a vector of image voxel intensities which is distributed as a multi-variate distribution with a mean vector of μ_x and a covariance matrix of $\mathbf{\Sigma}$. Let $\mathbf{b}$ be a vector of weights used to sample the voxels such that $y = \mathbf{b}^T\mathbf{x}$. These weights are typically equal to $\frac{1}{N}$ or zeros where N is the number of voxels in a region of interest resulting in y being the mean concentration within this region. With this it follows that

$$E\left[\sigma_{en}^2\right] = E\left[(y-\mu_y)^2\right] = \mathbf{b}^T\mathbf{\Sigma}\mathbf{b} \tag{4}$$

where σ_{en}^2 is the ensemble noise or variance, the variance of y, $\mu_y = \mathbf{b}^T\mu_x$, and $E[\cdot]$ is the expectation. This is a generalisation of equation 10 in [8].

If voxel covariance is spatially invariant, $\mathbf{\Sigma}$ will be a Töplitz matrix with coefficients defined by the covariance kernel and with multiplication equivalent to convolution of this kernel. Consequently estimation of the EV equates to (figure 1):

1) create an image of zeros and $\frac{1}{N}$ values defining the shape of the region for which to estimate the EV;
2) convolve this image with an estimated covariance kernel;
3) and calculate the mean value of this convolved image within the region.

Thus the EV of a single voxel region is just the central value of covariance kernel, the voxel variance, whereas a large contiguous region will tend to the average values of the covariance kernel over the number of region voxels. We refer to the estimated EV as the Fourier derived EV (FDEV) when the kernel is estimated using equations 2 and 3. We choose to perform the kernel estimation using this Fourier method due to greater computational efficiency.

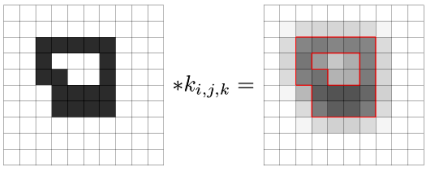


Fig. 1. Graphical representation of the estimation of EV for a spatially invariant covariance kernel and an irregular region. The initial image intensity is $\frac{1}{N}$ where N is the number of non-zero voxels with EV the mean value within the red region.

III. Methods

A. Simulations

Simulations were conducted in order to test the accuracy of kernel estimation (Matlab, The MathsWorks). Gaussian white noise of amplitude 4 times the mean value was applied to images of uniform concentration. These images were subsequently convolved with a 3D Gaussian kernel with a full width half maximum (FWHM) of 2 voxels in x, y and z directions in order to create a covariance structure. This resulted in images with a voxel variance of approximately 43% of the image value.

Cuboid sections of 21×21×21 voxels within the image were extracted, and the covariance kernels estimated using equations 1, 2 and 3 and compared to the true covariance. In addition these covariance kernels were used to estimate EV using equation 4 and the procedure described in section II-B. To examine how the precision of the kernels and estimated EV was dependent on the number of image volumes used these calculations were performed for 2, 5, 10, 20, 30, 50, and 100 volumes. The whole process was then repeated 50 times from which mean and standard deviations of the estimate kernels and EV values were derived.

B. Ge-68 phantom data

Two Ge-68 phantoms were used: one with an activity concentration of approximately 2.6 kBq/mL (Ge-68 phantom A) and the other with a concentration of approximately 10.1 kBq/mL (Ge-68 phantom B). For both phantoms, data were acquired for 250 minutes. Data from phantom A were split into fifty sequential 5 minute frames. Data from phantom B were split into 200 sequential 75 second frames. For both phantom datasets, the number of acquired true coincidences in each frame was approximately 35×10^6. Data were acquired using a Siemens Biograph mCT (Siemens Healthcare, Knoxville, Tennessee, USA) with a 21.6 cm axial field of view (TrueV). Images were reconstructed: with and without RM (Siemens HD-PET); with 3 and 10 iterations, 24 subsets and with isotropic voxel dimensions of 2 mm. No reconstruction post-filtering was applied. Initial findings using data acquired with phantom A have been presented previously [9] and inclusion in this is study is primarily to assess repeatability.

Cuboid regions of 40 mm × 40 mm × 40 mm were extracted from the center of images and used to estimate the covariance kernel using equations 2 and 3. EV was measured within spherical regions of increasing diameter (max 40 mm) across the 50 or 200 replicates for phantoms A and B respectively. These measurements were then compared against the FDEV estimates using the method described in section II-B.

Further to this, four other region shapes were defined: an ellipsoid with dimensions along one axis being half that of the other two axes and three hollow spherical shells with one, two and three-voxel thick shell wall. The latter shape was an attempt to produce results akin to those that may be obtained from a relatively uniform hot rim around a large necrotic tumour. In all regions, EV was measured across the 50 and 200 replicates from phantoms A and B respectively and compared with FDEV estimates as described previously. Illustrative examples of the region shapes can be seen in figure 2.

Fig. 2. Sections through examples of various 3-D region shapes that were evaluated comparing FDEV estimates with EV determined from replications. The example of the shell region shows a two-voxel thick shell.

The repeatability of kernel estimation and resulting FDEV estimates was assessed by comparing the data acquired from phantoms A and B. A difference in voxel variance was expected due to the different activity concentrations of the two phantoms but it was hypothesized that the correlation structure may be consistent, resulting in a similar relationship of FDEV with region size.

As with the simulations, the impact of the number of images used for the kernel on the precision of the FDEV was assessed by generating covariance kernels from 2, 5, 10 and 20 replicates that were randomly sampled with replacement from the 200 replicate dataset from phantom B. Each kernel was convolved with a sphere of increasing diameter to calculate the FDEV as previously described. The sampling was repeated 200 times and, for each sphere size, the coefficient of variation of the FDEV across the 200 samples was calculated.

To investigate the spatial variation of the covariance kernel, an additional 26 cuboid regions, surrounding the central sample volume were extracted from the phantom creating a total of 27 sampling regions. These 27 regions formed a 3×3×3 array centered at the image center and offset by -40 mm to +40 mm in x, y and z planes. At each sample location, a covariance kernel was produced using the 50 and 200 replicate images and the FDEV estimated for the spherical regions.

C. F-18 NEMA image quality phantom

The ultimate goal of this work is to try and infer region variance in a tumor volume based upon covariance data that can be extracted from an area of uniform tracer uptake, such as a patient liver.

To assess the feasibility of this, analysis was performed on a NEMA image quality phantom (PTW, Freiburg, Germany). The phantom was filled with a background activity concentration of approximately 5.2 kBq/ml and all six spheres had approximately 4:1 contrast to the background. The phantom was acquired for 90 minutes on the scanner with an ECG gating input provided by an ECG simulator. A gated reconstruction consisting of 24 gating bins was performed to produce 24 replicate images. Images were reconstructed: with and without RM; with 3 and 10 iterations; 24 subsets; with and without a 5.0 mm FWHM Gaussian post-filter and with isotropic voxel dimensions of 2 mm.

Six sets of spherical regions ranging from 10 mm to 37 mm (corresponding to the hot sphere diameters) were placed in the background area in 12 locations. Regions were centered on the transaxial plane intersecting the center of the spheres. At each location and for each sphere size, the EV was measured across the 24 replicate images. The mean and standard deviation was calculated across the 12 locations. At the same 12 locations, a covariance kernel was calculated with a cuboid region of 40 mm × 40 mm × 40 mm and the FDEV was calculated for spherical regions as performed for the Ge-68 experiments.

The EV across the 24 replicates was then measured in all six hot spheres. If the FDEV estimates predict background EV then these changes in background EV with reconstruction parameters (+/-RM, number of iterations, post-filter), may or may not reflect changes in EV within the hot spheres. To assess this, the ratio of measured EV changes within background and hot sphere regions were calculated when reconstruction parameters were altered to assess how relevant the background predicted FDEV estimates are to non-uniform areas.

D. Clinical data

Covariance kernels were generated from cubic regions placed within the relatively uniform uptake of the liver for four routine oncology PET patients. Each clinical image was reconstructed with isotropic 2 mm voxels with and without RM using 3 iterations, 24 subsets and no reconstruction post-filter. The kernels were generated for 15-18 positions, using 13 × 13 × 13 voxel regions within the liver. Region location was performed manually to ensure no region fell outside the boundaries of the liver.

For each patient image, using the average kernel across the multiple positions, the same process of convolving the covariance kernel with a spherical index image of increasing diameter was performed to calculate the FDEV.

IV. Results

A. Simulations

Coefficient of variation (COV) of estimated EV values from the simulations are shown in figure 3. Further data on the accuracy and precision of the estimated kernels can be found in preliminary work by the authors [9].

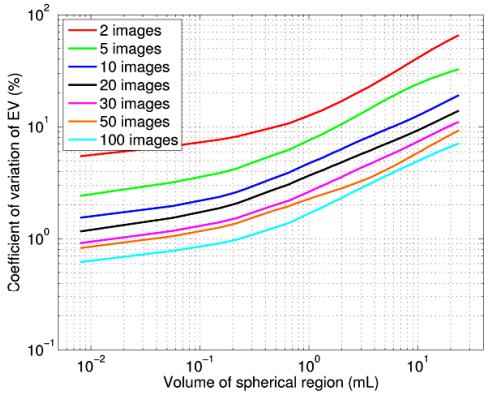


Fig. 3. Coefficient of variation of EV values of spherical regions of varying size and calculated using simulated data. The different lines refer to the number of images used in the esimation of the covariance kernel.

Small covariance underestimations were observed using equation 2, however this underestimation was not observed when using equations 1 or 3. This strongly suggests that this underestimation is due to voxel wrap around of the discrete Fourier transform but that the proposed correction adequately corrects for this.

The precision of the estimated EV values was approximately proportional to $\frac{1}{\sqrt{n}}$ where n are the number of image volumes used for the kernel estimation. Relatively the precision deteriorates as the spherical region volume increases. However such deterioration can be compensated for by increasing the number of image volumes used.

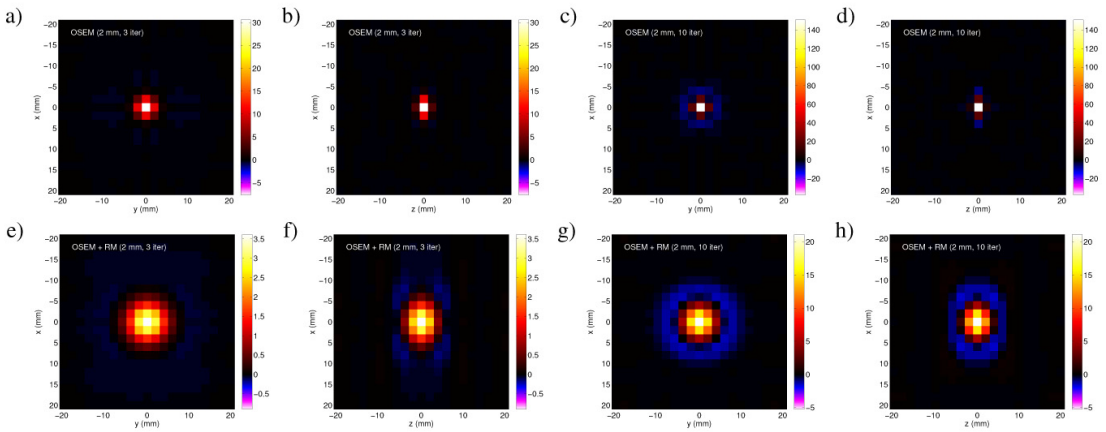


Fig. 4. Estimated covariance kernels estimated using equations 2 and 3 and data from 200 replicate images of a uniform Ge-68 phantom reconstructed without RM (a-d) and with RM (e-h), and with 3 iterations (a,b,e,f) and 10 iterations (c,d,g,h). The images show slices through the central section in the x-y plane (a,c,e,g) and the x-z plane (b,d,f,h). The units are $(kBq/mL)^2$ for a phantom with a mean concentration of approximately $10.1kBq/mL$.

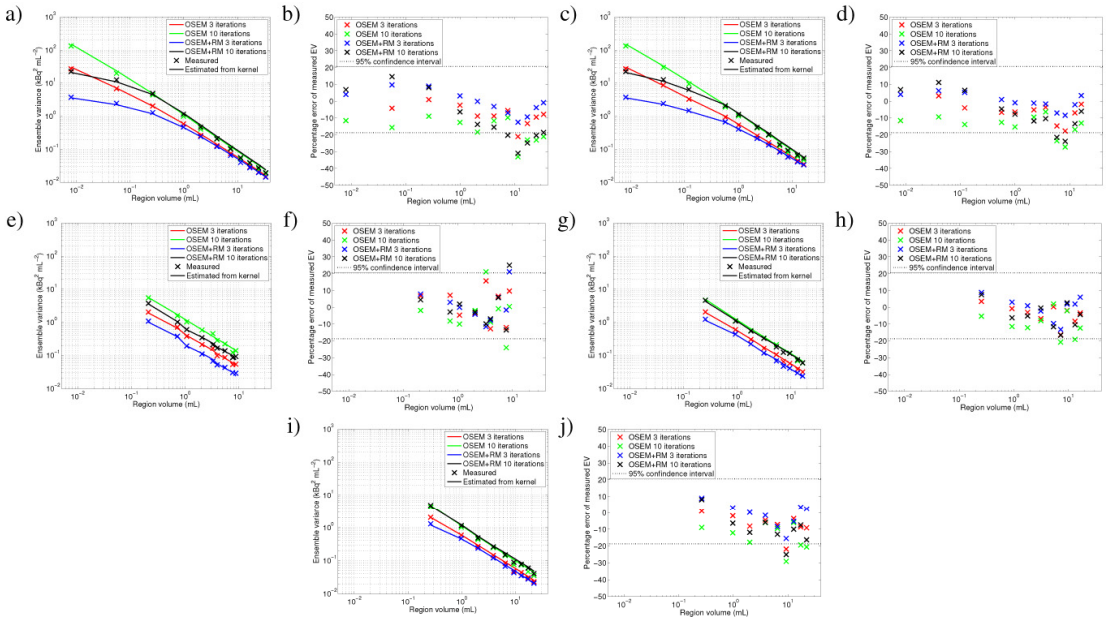


Fig. 5. Comparison of estimated and measured regional EV (a,c,e,g,i) and relative differences (b,d,f,h,j) for spherical regions of varying diameter (a,b), ellipsoids of varying sizes (c,d), and shell regions of various diameter with a thickness of 1 voxel (e,f), 2 voxels (g,h), and 3 voxels (i,j). The 95% confidence intervals on the difference plots were calculated assuming that $m-1$ times the ratio of the measured EV over the true EV is distributed as a χ^2 distribution with $m-1$ degrees of freedom.

B. Ge-68 phantom data

Estimated covariance kernels are shown in figure 4. Kernels shown were derived from phantom B and show similar features to data previously reported from phantom A [9]. Notably: greater correlation is observed with RM; kernel width decreases both with and without RM as the iterations are increased from 3 to 10; increased negative correlation is seen with 10 iterations and greater correlation is seen in the x-y plane compared with the axial plane, particularly without RM.

A comparison of measured and estimated FDEV for the various region shapes is shown in figure 5. General agreement was obtained with most differences for phantom B within the 95% confidence intervals (-19% to +21%) of measured EV values, with similar trends observed with region size and reconstruction algorithms. RM resulted in reduced EV values for small spherical regions at 3 and 10 iterations. The EV for large regions decrease at $\frac{1}{V}$ where V is region volume. With the exception of the one-voxel thick shell, very

small differences in EV were observed between reconstruction algorithms with and without RM for large diameter regions. In the one-voxel thick shell a consistent reduction of EV is seen when RM is used for both 3 and 10 iterations. Increases in EV are observed when increasing the number of iterations to 10 from 3 for all shapes and sizes but particularly for the smallest regions.

Data showing the repeatability and precision of FDEV are shown in figure 6. Repeatability was found to be good for the two phantoms. A systematic offset was observed between each phantom due to the different voxel values, consistent with the different activity concentrations, but the slope and shape of the FDEV curves was comparable for phantoms A and B for all reconstruction parameters investigated. As with simulation data, the precision of the FDEV was seen to follow a $\frac{1}{\sqrt{n}}$ relation. The relationship between the COV of EV and region size is also very similar for simulations and phantom data. The increased COV with RM is due to reduced mean EV values with similar standard deviation values observed.

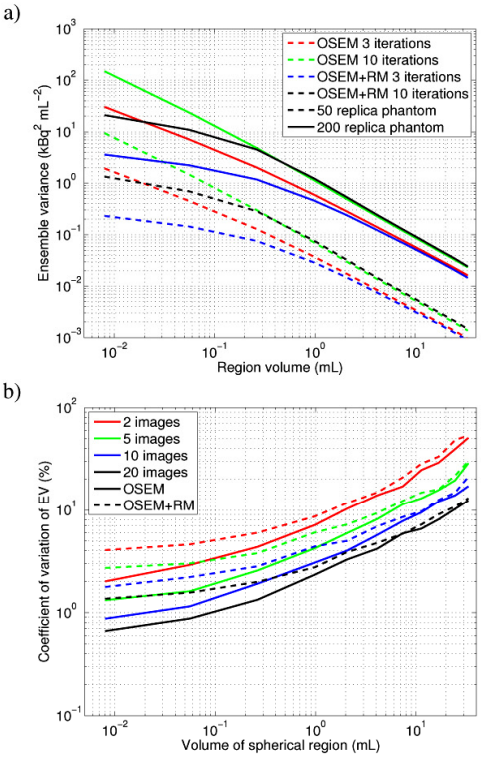


Fig. 6. a) Repeatability: EV values estimated for spherical regions of increasing size using data from two separate Ge-68 phantoms and using 50 and 200 replicates and b) Precision: coefficient of variation values of EV values estimated for spherical regions of increasing size using re-sampling of the 200 replicates of Ge-68 phantom B and using different numbers of images.

The covariance kernels derived from various locations within the Ge-68 phantom are shown in figure 7. For images with and without RM, there is a subtle tangential elongation of the x-y plane kernel sections that are radially offset. The variance is seen to increase slightly for axially offset regions. The variation of the FDEV using these 27 kernel locations is show in figure 8, with the differences with position typically smaller that the differences with reconstruction parameters, particularly for small regions.

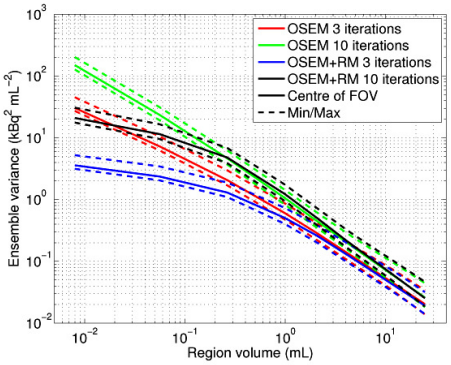


Fig. 8. EV values calculated using covariance kernels derived from different locations with the Ge-68 phantom. The solid line represents the value in the center of the field of view. The dashed lines are the minimum and maximum values across all 27 covariance kernels.

C. F-18 NEMA image quality phantom

The calculated FDEV for spherical regions using kernels derived from the background section of the NEMA phantom are shown in figure 9. Also shown in the figure are the measured EV for the six spheres corresponding to the dimensions of the hot spheres.

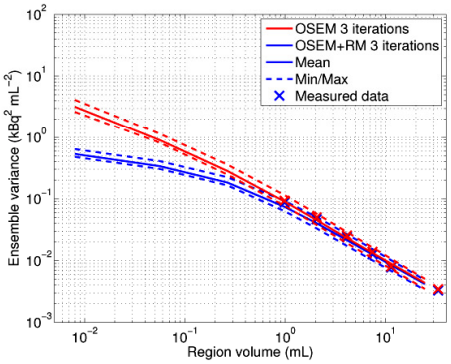


Fig. 9. Comparison of estimated EV using covariance kernels estimated from background regions of a NEMA image quality phantom. Data is shown for the mean (solid lines) and for the minimum and maximum values across 12 locations (dashed lines) with data reconstructed with (blue) and without (red) RM and using 3 iterations. Red and blue crosses are also plotted showing the measured EV for 10, 13, 17, 22, 28, 37 mm diameter spherical regions across the 24 replicates.

Ratios of the measured EV for images with varying reconstruction parameters that are expected to alter the correlation structure (RM, post-filter, iteration) are shown in figure 10. The ratios show good agreement when calculated for the hot

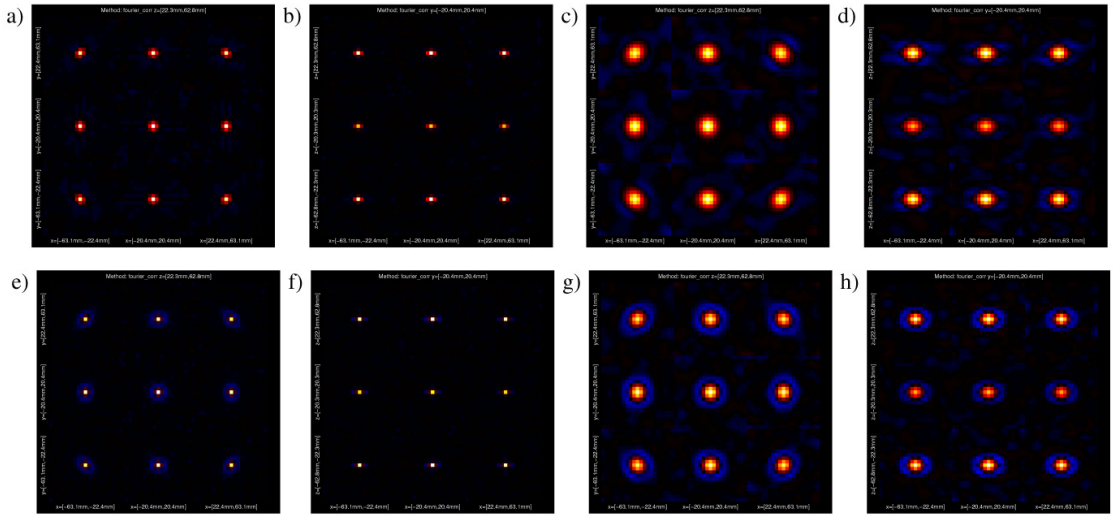


Fig. 7. Covariance kernels derived from regions of a continuous Ge-68 phantom forming a 3 × 3 × 3 grid of contiguous 40 mm × 40 mm × 40 mm volumes. The title, x and y axis labels define the position of the region relative to the phantom center. Data shown for: 3 (a,b,c,d) and 10 (e,f,g,h) iterations; without (a,b,e,f) and with (c,d,g,h) RM; and x-y (a,c,e,g) and x-z (b,d,f,h) sections through the covariance kernel center. Data scaled to the maximum value across all 27 spatial locations for each panel.

spheres and the background region when looking at the impact of RM and post-filter on EV. The ratios show less agreement when the number of iterations is increased, with a greater magnitude of increase of EV with 10 iterations observed in the colder background region compared with the hot spheres. Nevertheless although smaller, similar directions in change are observed.

D. Clinical data

The four sets of covariance kernels derived from the clinical data with and without RM are shown in figure 11. The general appearance and extent of the covariance kernels in the figure is similar to those observed from the phantom data in figure 4 with some elongation which is likely to be a consequence of transaxially offset location of the regions altering the kernel shape.

Plots of FDEV for spherical regions are shown in figure 12. As with the two Ge-68 phantom data, there are subtle systematic offsets in the data which could be due in part to variations in activity concentration and patient size but the slope and shape of FDEV against region volume is very consistent. The differences in FDEV with and without RM for spherical regions of varying volume in figure 12 is very similar to that observed for phantom data in figure 5a.

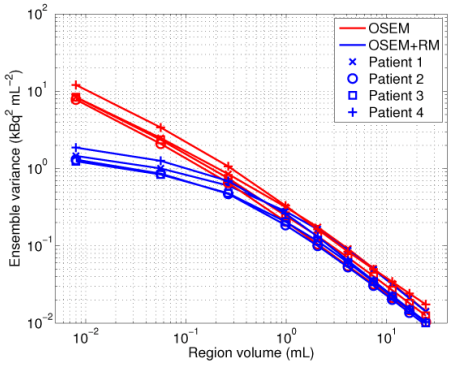


Fig. 12. Comparison of estimated EV values using kernels derived from the livers of 4 patients. Data shown for reconstructions with (blue) and without RM (red), 3 iterations, and no post reconstruction smoothing.

V. Discussion

From the evaluations conducted, the use of voxel covariance kernels to estimate regional EV appears to be valuable in the assessment of how EV changes with region shape, region size, and image reconstruction parameters.

The simulations demonstrate that the kernels could be estimated with acceptable accuracy and precision both with image domain calculations using equation 1 or with Fourier domain calculations using equation 2. Furthermore for the phantom data, the precision can be increased through the use of replicate acquisitions. Our implementation of the Fourier domain calculations uses the discrete Fourier transform and results in an underestimation of covariance values particularly for voxels within close proximity. This is because the distances between voxels can be wrapped around with two voxels on opposing edges being inferred as being close to each other. However if the extracted image section is large enough that the real correlation of such cases is negligible and effective correction is possible using equation 3. In practice there is little difference in Fourier or image derived kernels apart from computation. This computational advantage of the Fourier derived kernel we have exploited in applying the method to multiple replications and re-sampled data.

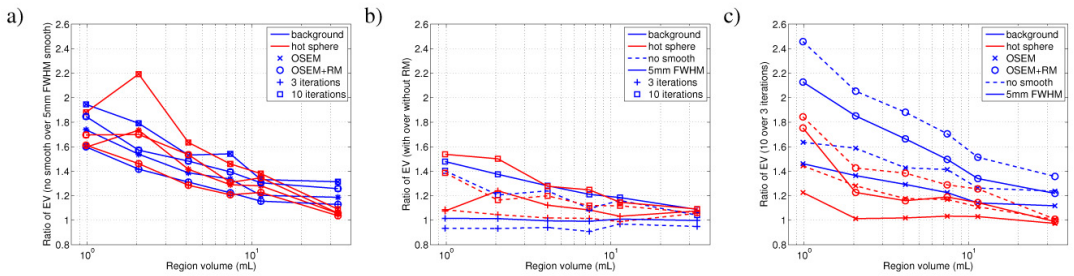


Fig. 10. Ratios of measured EV values across 24 replications with reconstruction parameters altered: a) Post reconstruction smoothing (none over 5mm FWHM); b) Resolution modelling reconstruction (with over without RM); and c) Iterations (10 over 3 iterations). Data is shown for 10, 13, 17, 22, 28, 37 mm diameter spherical regions derived from background (blue) and hot (red) regions and for 8 reconstructions combinations: (+RM (circle), -RM (cross)) $\times$ (no smooth (dashed), 5mm FWHM Gaussian (solid)) $\times$ (3 iterations (plus), 10 iterations (square)).

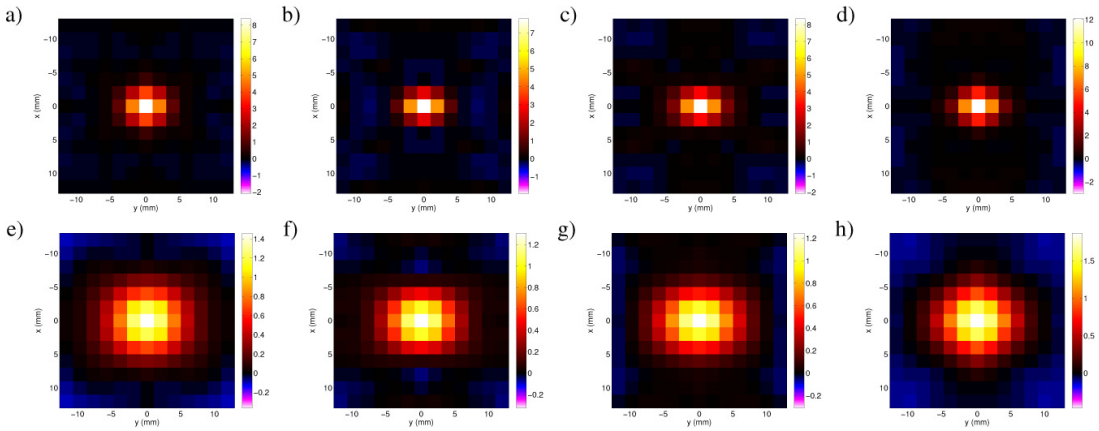


Fig. 11. Covariance kernels estimated from the livers of 4 patients (columns) and for reconstruction without RM (a,b,c,d) and with RM (e,f,g,h), and reconstructed using 3 iterations with no post reconstruction smoothing. The images show slices through the central section in the x-y plane.

The estimated FDEV values show smooth changes with region sizes, decreasing from the voxel variance value to a common rate of decline approximately inversely proportional to the volume size. In contrast the measured EV values have more fluctuations which probably reflect random error. Assuming independence and normality the precision of the measured EV precision was estimated with the observed differences consistent with this lack of precision. The similar values in these measured EV to FDEV differences with similar region size probably reflects the use of common portions of the image used when calculating the measured EV estimation, with significant overlap of the evaluated regions. This interpretation is consistent with the observation that the similarities in measured EV values was less for the shell regions, and particularly with a one voxel width.

Evaluating the benefit of different reconstruction algorithms and parameters typically requires the trade-off of bias and variance. Metrics such as image roughness have been proposed for evaluating variance (Tong et al), and which for large homogeneous regions is a reasonable approximation of voxel coefficient of variance. However for quantitative use of PET images, region of interest assessment is typical and consequently EV is a more appropriate metric particularly when voxel correlations change. Measured EV lacks precision, particularly for a low number of replicates. In contrast, FDEV has greater precision as demonstrated in figures 3 and 6b, can be estimated from just a few replicates and consequently has value in performing reconstruction optimization.

When comparing the impact of RM on both the estimated FDEV and measured EV values we note that, from figures 5 and figures 10b, with 3 iterations, there is no degradation of the EV when RM is incorporated. In fact in these calculations reductions of EV in small regions were observed with large regions resulting in very similar values with and without RM. However, with 10 iterations, we note that there is a worsening of EV when RM is used in the reconstruction. As illustrated in figure 10b, this continues to occur with the addition of a post-filter.

In a clinical setting, one would likely expect the minimum volume used in region-based measures to be 1.0 ml as is specified by the PERCIST protocol when defining SUV_{peak} [11]. Based on the results from this study using clinically relevant reconstruction parameters (3 iterations and a 5 mm post-filter in our center), one would expect comparable degree of variance on this metric for reconstructions with and without RM.

The findings in the NEMA phantom imply that FDEV derived from a uniform area of warm uptake has some value in assessing the impact of reconstruction parameters on hot regions. This work does not attempt to derive absolute values of EV for the hot spheres from measurements in the background but demonstrates that relative changes of EV in background regions, which can be accurately estimated by FDEV, are predictive of EV changes within hot regions. Figure 10 showed that this is the case when RM is or is not used, and when a post reconstruction filter is applied. When the iterations are increased from 3 to 10 iterations, although the magnitude of the EV changes do not correspond the general pattern and direction are consistent.

The kernels derived from clinical data demonstrate that there is a consistent appearance in the covariance kernels derived from areas of uniform uptake within the liver. The difference in kernel appearance with and without RM is very similar to that observed for the phantom data. In addition the shape of the estimated FDEV values with region size in figure 12 were both similar between patients and similar to values observed with the phantom data (figures 5a and 9). Consequently the use of phantom data to predict these changes in FDEV in patients seems to have validity.

We note that, even though several geometries of region were used for convolution in this work, all of them assumed spatial stationarity of the covariance kernel which is unlikely within heterogeneous regions. This may be a valid assumption for very small regions on clinical data placed either within a small lesion or a particular area of a larger tumor with a heterogeneous uptake pattern. However, if one wanted to define a region to encompass an entire tumor volume, this assumption is likely to break down. Nevertheless we believe that the proposed method still enables evaluation of how the EV is likely to change with region size, region shape and reconstruction parameters. In addition, the measurement of EV within regions of patient data is typically not possible due to a lack of acquisition replications, and even with limited replications lacks precision. Consequently the proposed FDEV estimate has potential utility.

VI. Conclusion

The use of voxel covariance kernels to estimate regional EV appears to be valuable in the assessment of how EV changes with region shape, region size, and image reconstruction parameters. The derivation of a covariance kernel from a single image allows for an estimation of EV without the need for multiple image replications, which is convenient for routine clinical images where multiple images are rarely available. Good agreement was observed between measured and predicted EV in a range of region shapes and the technique appears to be repeatable and precise. The NEMA phantom suggest that kernels derived from areas of uniform uptake may be useful to infer relative changes of EV in focal areas of greater uptake as reconstruction parameters are adjusted.

VII. Acknowledgments

This work was conducted as part of a Siemens funded Ph.D. studentship.

References

[1] V. Y. Panin, F. Kehren, C. Michel, and M. Casey, "Fully 3-d pet reconstruction with system matrix derived from point source measurements," *Medical Imaging, IEEE Transactions on*, vol. 25, pp. 907–21, 2006.

[2] E. De Bernardi, M. Mazzoli, F. Zito, and G. Baselli, "Resolution recovery in pet during awosem reconstruction: A performance evaluation study," *Nuclear Science, IEEE Transactions on*, vol. 54, pp. 1626–38, 2007.

[3] E. Rapisarda, V. Bettinardi, K. Thielemans, and M. C. Gilardi, "Image-based point spread function implementation in a fully 3d osem reconstruction algorithm for pet," *Physics in Medicine and Biology*, vol. 55, pp. 4131–51, 2010.

[4] D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois, and D. W. Townsend, "Impact of time-of-flight on pet tumor detection," *J Nucl Med*, vol. 50, pp. 1315–23, 2009.

[5] F. C. Sureau, A. J. Reader, C. Comtat, C. Leroy, M.-J. Ribeiro, I. n. Buvat, and R. g. Trbossen, "Impact of image-space resolution modeling for studies with the high-resolution research tomograph," *Journal of Nuclear Medicine*, vol. 49, pp. 1000–8, 2008.

[6] C. W. Alessio, A. M.and Stearns, S. Tong, S. G. Ross, S. Kohlmyer, A. Ganin, and P. E. Kinahan, "Application and evaluation of a measured spatially variant system model for pet image reconstruction," *Medical Imaging, IEEE Transactions on*, vol. 29, pp. 938–49, 2010.

[7] 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*, vol. 55, pp. 1453–73, 2010.

[8] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in pet imaging: Theory, practice, benefits, and pitfalls," *Medical Physics*, vol. 40, no. 6, pp. 064 301–15, 2013.

[9] I. S. Armstrong, H. A. Williams, and J. C. Matthews, "Evaluation of the utility of estimated covariance kernels for predicting regional ensemble variance," in *Nuclear Science Symposium Conference Record, 2015*, 2015.

[10] ——, "Accuracy and variability of quantitative measurements using pet with time-of-flight information and resolution modelling," in *Nuclear Science Symposium Conference Record, 2011*, 2011, pp. 4167–70.

[11] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From recist to percist: Evolving considerations for pet response criteria in solid tumors," *J Nucl Med*, vol. 50, pp. 122S–50, 2009.

[12] H. H. Barrett, D. W. Wilson, and B. M. Tsui, "Noise properties of the em algorithm: I. theory," *Phys Med Biol*, vol. 39, pp. 833–46, 1994.

[13] J. Qi and R. M. Leahy, "Resolution and noise properties of map reconstruction for fully 3-d pet," *Medical Imaging, IEEE Transactions on*, vol. 19, pp. 493–506, 2000.

[14] P. J. Markiewicz, A. J. Reader, and J. C. Matthews, "Assessment of bootstrap resampling performance for pet data," *Physics in Medicine and Biology*, vol. 60, pp. 279–99, 2015.

[15] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Wiley-Interscience, 2004.

8. SUMMARY AND CONCLUSIONS

The four pieces of work in this thesis have studied the impact of three advanced iterative reconstruction algorithms, OSEM with resolution modelling, OSEM with TOF and OSEM with combined resolution modelling and TOF. These three algorithms are commercially available on the Siemens Biograph mCT PET/CT system at CMUH. The performance of these three algorithms, from a quantitative accuracy and image quality perspective was compared against the standard OSEM algorithm that was used clinically on the scanner until mid-2014. From mid-2014, the department switched to OSEM with TOF as the default reconstruction algorithm for FDG oncology scanning based on outcomes from the work in this thesis. Reconstruction with FBP is available on the system but was not included as part of this project because it has never been used nor plans to be used clinically. As discussed in the introductory chapter of this thesis, the primary aim of the project was to evaluate the impact on quantification but the aims evolved to also take advantage of the observed gains in image quality to increase throughput on the PET/CT scanner. The findings of the work are summarised in the following two sections accordingly.

8.1 Impact on quantification

Due to the inability to apply TOF without resolution modelling at the time, the primary focus of the first paper was the use of resolution modelling. The work demonstrated the benefits of resolution modelling, particularly in small objects, with a notable increase of activity concentration recovery for the smaller spheres in the NEMA phantom. The change in noise structure due to increased voxel correlation was also evident, and resulted in a reduced voxel variance in the images. This reduced the variability of activity concentration metrics, with the most notable impact seen for SUV_{max} methods. It is also the likely cause for a reduction in the positive bias, when compared with OSEM, with SUV_{max} methods observed in the larger spheres and particularly at low contrast. The work illustrated the necessity of applying a moderate degree of post-reconstruction smoothing to images reconstructed with standard OSEM algorithms to avoid substantial positive bias and high variability of measurements.

The first paper also clearly demonstrated that SUV_{max} rarely provides an accurate measure of the true activity concentration and is particularly susceptible to positive bias, particularly at low contrast levels in images reconstructed with standard OSEM algorithms. The paper showed the benefits of alternative metrics such as SUV_{peak} that provide notably greater accuracy in larger lesions and reduced variability. However, a substantial negative bias was still observed with SUV_{peak} in the smaller objects. The calculation of SUV_{peak} can be considered as a convolution of the image data with a uniform spherical region of 1.0 ml in volume, which results in variability that shows very little dependence on reconstruction methods. Therefore it is likely that SUV_{peak} will be a more appropriate metric to use in place of SUV_{max}. The overall findings from the paper suggested that the benefits of SUV_{peak} will be greatest when combined with advanced reconstruction algorithms as the reduction in the degree of post-filtering may aid lesion visualisation combined with the reduced variability inherent with the metric. The adoption of alternative metrics may present a challenge due the ubiquitous use of SUV_{max}.

The second piece of work that followed on from this work applied a subset of reconstruction parameters in clinical oncology imaging for lung cancer. The parameters for the reference reconstruction were those that were used routinely in the local imaging protocol for oncology. Parameters for the advanced reconstructions were selected to allow the two matching criteria to be achieved using pragmatic levels of post-filtering. For example, a very high degree of post-filtering would have been required to align quantification for resolution modelling images if high numbers of iterations were used, and hence findings would be of limited clinical relevance. Three uptake metrics, which have been reported for use in routine imaging, were evaluated under the two matching schemes. The use of post-filtering to match either image noise or SUV_{max} was seen to translate very well from the NEMA phantom to patient imaging. The notion of smoothing images that have been reconstructed with resolution modelling may seem counter-intuitive as the rationale for using this algorithm is to improve spatial resolution. The translation was slightly less successful with TOF image when trying to match SUV_{max} with a small positive bias seen in patient studies that was not observed in the phantom. As with the first paper, this second paper illustrated the significant impact that algorithms with resolution modelling had on quantification.

The main clinical implication from these two papers is that the locally adopted SUV_{max} thresholds, used in our department, for discriminating benign and malignant lesions cannot be applied when using reconstructions that include resolution modelling.

The next paper aimed to exploit the observed reductions in image noise described in the second paper to allow reductions in administered activity or imaging time. The second paper demonstrated that the 4.8 mm FWHM post-filter applied to the TOF images aligned SUV_{max} with the non-TOF images, which had a 5.0 mm post-filter applied. When evaluated in the patient studies, this resulted in a slight positive bias of SUV_{max} measured in the TOF images compared with the non-TOF images. Therefore, for the third paper, the post-filter FWHM for the TOF images was increased to 5.0 mm. This resulted in no significant difference observed in lesion SUV_{max} for the TOF and non-TOF images. This conservation of lesion SUV_{max} allowed for a seamless transition to TOF reconstruction for the routine oncology imaging.

The other factor to assess from the work is the impact of the advanced algorithms on the precision of uptake metrics. The first paper suggested that variability of SUV_{max} was reduced with resolution modelling, due to reduced voxel variance, while there seemed to be very little change on SUV_{peak} variability. The final paper aimed to focus in more detail on the impact of resolution modelling on the variability of region-based measurements – quantified by EV. The findings in the fourth paper provided encouraging results on estimating EV from a clinical image using covariance kernels derived from a uniform area of uptake. The notable finding of this work was that resolution modelling does not appear to introduce an undesirable increase of EV using reconstruction parameters that are used clinically. This is particularly relevant for SUV_{peak}, which may achieve more widespread adoption in the future.

The use of reconstruction with resolution modelling raises a big challenge to PET/CT centres. The NHS PET/CT service contract [1] requires centres to perform a clinical audit that requires a minimum of 10% of scans to be re-reported by an independent clinician. Disagreement may occur between the two reports, particularly on lesions

where the SUV is considered borderline for malignancy. The views of the two clinicians may differ on the interpretation of SUVs from these lesions. The significant impact of resolution modelling on SUV_{max} demonstrated, particularly in small lesions, will exaggerate the problem as one clinician may be unfamiliar with the impact of the reconstruction. There are currently no studies that have aimed to reassess the use of FDG uptake for the purposes of discriminating benign from malignant disease. An additional risk in implementing resolution modelling is the requirement, also stated in the NHS contract, that all images must be transferred electronically to the referring clinician. These images may be open to mis-interpretation if the referring clinician is not familiar with the resolution methods.

8.2 Improvements in image quality

There are increasing pressures on NHS hospitals, both to meet waiting time targets while also needing to save money and work more efficiently. Throughout 2016, negotiations have been on-going regarding the re-tender of the second phase of the national PET/CT scanning provision. The clinical service at CMUH, as part of the North West FDG Oncology Service, falls within the remit of these negotiations. As such there has been a growing need to demonstrate excellence of service.

The work performed in the third paper of this thesis aimed to assess whether TOF could be used to refine the imaging protocol for FDG oncology at CMUH. The aims were to reduce either or both the administered radioactivity of the FDG and the scanning time on the scanner. Reducing the scanning time would hopefully allow greater throughput on the scanner. More scanning throughput would make it easier to comply with the existing service contract of a seven-day turnaround from PET/CT referral to a completed clinical report. Additional benefits would be increased revenue from performing a greater number of PET/CT scans and the ability to provide additional service to expanding potential clinical indications.

The outcome of the work saw the adoption of OSEM with TOF reconstruction and a change to the local protocol, with all patients being prescribed 280 MBq instead of the 350 MBq or 400 MBq for patients with a body weight below and above 100 kg respectively. The scanning time per bed position for patients below 85 kg was

reduced from 2.5 minutes to 2.0 minutes. This translated to a total reduction in imaging time of 3.0 to 3.5 minutes for approximately half of all patients scanned. Since publication of the third paper in this thesis, follow-on work has been performed on the low-weight FDG patients. The theory of TOF suggests that the benefits of TOF, defined by an increase in image SNR, would diminish for smaller patients [2]. However, gains in SNR with TOF measured at our department suggest that there is very little relation with SNR gain and patient size, as seen in Figure 8.1, and this also includes two paediatric patients.

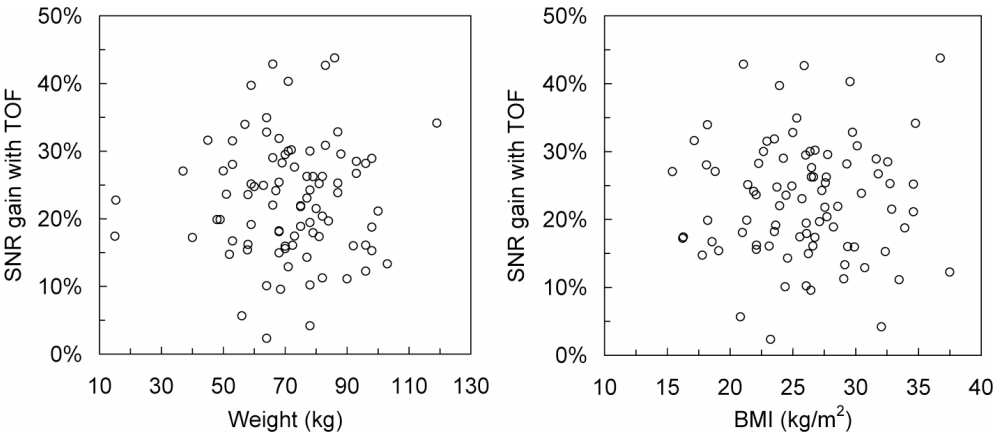


Figure 8.1 Gains in image SNR measured in the livers of 88 patients that underwent standard FDG PET/CT oncology imaging at CMUH.

Based on these measurements, the local imaging protocol was refined further, with the implementation of a weight-based scheme for the administered activity of FDG for patients up to a body weight of 80 kg. The work demonstrated that SUV_{max} was conserved with this new protocol compared with the initial protocol with non-TOF reconstruction. This work is currently unpublished but was presented at the 2015 British Nuclear Medicine Society Annual Meeting [3].

The combination of the OSEM with TOF imaging protocol and the weight-based scheme was introduced at the start of 2015 and resulted in an increase of 100 additional patients being scanned at CMUH each year while also reducing the median administered activity by 30% [4].

8.3 Conclusions

Advanced iterative reconstruction algorithms that include resolution modelling and TOF provide clear substantial benefits over standard iterative reconstruction algorithms in oncology PET/CT imaging.

Resolution modelling improves visualisation of small objects and results in significant increases of uptake measurements, alongside a likely shift from the monotonic relation of uptake measurement and object size that clinicians have become used to over many years. This may pose a challenge when interpreting established uptake metrics and further work is required to determine whether these algorithms can improve diagnostic accuracy of lesion characterisation based on uptake measurements.

The increased voxel correlation of resolution modelling reduces the voxel variance, which reduces the variability of SUV_{max}. The work here also suggests that it does not increase variability of region-based measures such as SUV_{peak}. Both of these outcomes are particularly beneficial when changes in uptake are used for assessing the response of lesions to cancer therapy. The reduced voxel variance may also facilitate either a reduction in administered activity and/or the scanning time.

The implementation of TOF evaluated here has been shown to result in reductions of image noise that are relatively consistent across the patient population that has been evaluated. In addition the conservation of FDG uptake measurements, relative to non-TOF algorithms, is trivial to achieve. This makes the strong case that incorporation of TOF information in PET images should become standard practice where available. This has been the case since mid-2014 at the CMUH department.

Overall, the work in this thesis demonstrates that high accuracy of uptake metrics, particularly of SUV_{max}, is rarely achieved across a range of object sizes with any of the reconstruction algorithms. In clinical imaging, the importance of a truly accurate uptake metric is debatable. It may be better to suggest that one should strive for a combination of reconstruction parameters and appropriate uptake measures that achieve two aims:

- to provide the highest ability to discriminate benign and malignant tissue with the most consistency;
- to confidently identify a change in uptake when assessing response to therapy.

The reduced variability achieved from both resolution modelling and TOF, as well as improved SNR, argue the case that both should be used routinely in clinical FDG imaging. The remaining challenge of implementing resolution modelling is to establish an understanding of uptake metrics. This is likely to require either a revalidation of existing metrics, when using both algorithms combined, against histological confirmation of malignancy or evaluation of alternative metrics such as SUV_{peak} or TLG that are not yet fully available for routine use.

8.4 References

1, NHS England. *2013/14 NHS Standard Contract for Positron Emission Tomograph – Computed Tomograph (PET_CT) Scanning (All Ages)*. B02/s/a. NHS England, 2013

2, Budinger TF. *Time-of-Flight Positron Emission Tomography: Status Relative to Conventional PET*. J Nucl Med 1983; **24**:73-78

3, Armstrong IS, James JM, Muthu S. *Combining a weight-based dosing scheme with Time-of-Flight for further reductions in patient and staff FDG doses*. Nuc Med Commun 2015; **36**:A41

4, CMFT Research and Innovation. *One hundred cancer patients a year in Manchester benefit from scan technology*. CMFT Press Release 2015.

https://research.cmft.nhs.uk/news-events/one-hundred-cancer-patients-a-year-in-manchester-benefit-from-scan-technology. Accessed 5th September 2016

9. FUTURE WORK AFTER PHD

There are three pieces of work that naturally follow the progression of the work of this thesis.

9.1 Correlation with lung nodule and lymph node histology

To implement resolution modelling on a routine basis and maintain using FDG uptake metrics, it is essential to understand the clinical implication of changes to quantification.

The work in this thesis from phantom and patient lung cancer studies has illustrated the impact of the algorithms with resolution modelling and TOF on SUV. The work on clinical data has used the standard technique as a reference and hence ground truth of lesion uptake was unknown. The aim of this follow-on work would be to deduce whether these algorithms can be used to improve the diagnostic accuracy of SUV_{max} (or other metrics) when differentiating benign and malignant lesions.

The current NICE Guideline CG121 [1] for the diagnosis and management of lung cancer states the following with regards to the use of PET-CT in the management of lung cancer:

- "Ensure all patients potentially suitable for treatment with curative intent are offered PET-CT before treatment."

It also states the following with regards to assessment of mediastinal lymph nodes:

- "Evaluate PET-CT-positive mediastinal nodes by mediastinal sampling (except when there is definite distant metastatic disease or a high probability that N2/N3 disease is metastatic [for example, if there is a chain of lymph nodes with high 18F-deoxyglucose uptake])."

The key issue here is in the involvement of PET/CT in determining positive malignant lymph nodes. While this is performed mainly as a visual interpretation, there is a risk that small lesions that do not appear particularly FDG-avid, due to

partial volume effects, may be missed. As discussed, SUV_{max} is commonly used to support characterisation of lymph nodes but thresholds to suggest malignancy are based on historical data. In the early 1990s, one study demonstrated that no benign pulmonary lesions had a SUV_{max} of greater than 2.5 [2]. The use of this value became frequently used as a discriminator for malignancy [3, 4] and is still used in practice today. This is despite technological advances in PET/CT systems where changes in image quality and quantification may not necessarily mean that this is still appropriate, which has attracted criticism [5]. A later study comparing lymph node histology with FDG PET showed that the median SUV_{max} was greater in all lymph node locations when malignant disease was present but there was considerable crossover in SUV_{max} for malignant and non-malignant nodes [6]. A second study from the same institute [7] showed that SUV_{max} was significantly greater in malignant lymph nodes than non-malignant nodes and using ROC analysis demonstrated an optimal cut-off value of lymph node SUV_{max} of 5.3 (sensitivity 91%; specificity 88%). A recent study with data acquired on a GE PET/CT system has demonstrated that the use of advanced reconstruction algorithms are likely to require higher thresholds for discrimination [8] compared with traditional iterative algorithms. The study compared images reconstructed with OSEM and TOF against *Q.Clear*, an algorithm with resolution modelling and regularisation. To date, there are no studies that have evaluated the need for revised thresholds for data acquired with resolution modelling and TOF on the Biograph mCT.

The studies in this thesis have demonstrated the significant impact of the advanced reconstruction algorithms on quantification when compared with traditional algorithms. It follows that, if SUV_{max} is to continue being used in FDG PET, there is a need to establish new thresholds for differentiating benign and malignant lesions. Using standard algorithms without resolution modelling results in a monotonic relation of measured SUV and object size [9]. While a criticism of resolution modelling has been the impact on this monotonic response curve, the first paper in this thesis demonstrated that reconstruction parameters can be selected that result in a monotonic response albeit with a substantially reduced dependence on object size.

During this PhD, a pilot audit was undertaken to establish the feasibility of correlating findings from histo-pathology for patients that had undergone a PET/CT

scan to evaluate the ease of correlating lesions and lymph nodes identified on PET scan with the histo-pathology findings. This work was performed retrospectively and illustrated the many challenges of such an investigation. Histo-pathology information was obtained from 21 primary tumours and 43 lymph nodes from 26 patients. For all primary tumours, the question of malignancy was not necessary and other factors such as cell type, such as adenocarcinoma or squamous cell carcinoma, were of interest. Unfortunately, such information was only available in samples from nine primary tumours. Only six lymph nodes showed positive malignancy and of these five could be correlated with the PET/CT image. The time between PET imaging and biopsy was typically at least one to two months and so disease progression could not be ruled out. The correlation of lymph nodes was particularly difficult due to the different conventions in the PET report and the surgical convention of defining lymph node location. This was especially problematic for nodes without disease, with no appreciable FDG uptake on the PET images, making it essentially impossible to locate the nodes. Due to the low dose CT acquisition and lack of CT contrast, locating nodes on the CT was also incredibly difficult.

At Central Manchester University Hospitals, a service is being discussed to provide a "one-stop shop" solution for patients undergoing PET/CT for diagnosis and staging of lung cancer that are likely to go to surgery to have a biopsy guided by endobronchial ultrasound bronchoscopy. If the service is to be established, it offers an excellent opportunity to compare uptake FDG measurements from the PET with prompt histo-pathology findings. Images could be reconstructed with a range of reconstruction parameters with lesion uptake measured by a number of typical metrics such as SUV_{max}, SUV_{peak} and total lesion glycolysis with ROC analysis being performed to deduce the performance of each combination of reconstruction and uptake metric using the optimal cut-offs derived for each method.

9.2 Continuation of development of extended NEMA phantom

The first piece of work in this thesis evaluated the impact of resolution modelling and TOF on the quantification of the six spheres within the NEMA image quality phantom. The phantom consists of six spheres with diameters 10, 13, 17, 22, 28 and 37 mm. The work demonstrated that, even with very low 2:1 contrast, the smallest

10 mm sphere could be visualised with combined RM and TOF reconstruction as shown in Figure 9.1.

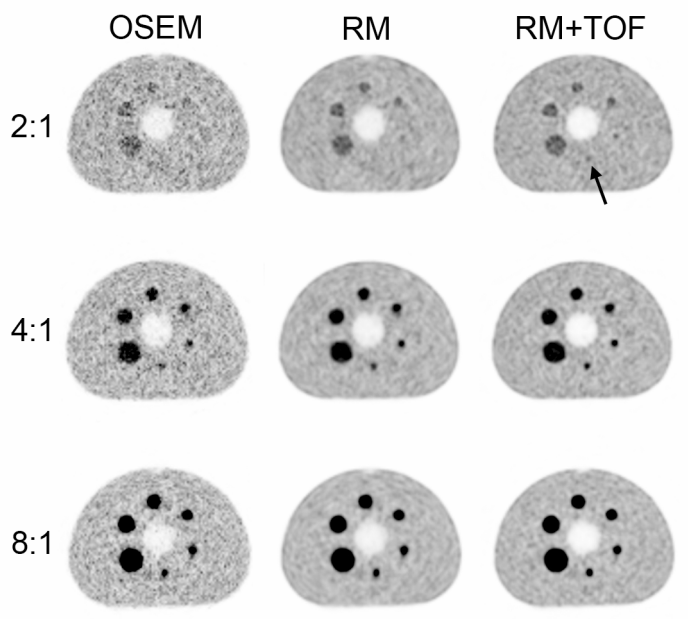


Figure 9.1 Images showing visualisation of spheres with various contrast in the NEMA image quality phantom using three different reconstructions algorithms.

While the standardised NEMA-specification phantom is clearly an essential tool for comparison and standardisation of images across different PET scanners and sites, there is a need to explore the limitations of current and next generation PET scanners. During this PhD thesis a prototype phantom has been developed by Ian Armstrong and Heather Williams in collaboration with Leeds Test Objects. The phantom can be configured according to the NEMA specification but alternative configurations are possible. Firstly, two additional compartments can be placed around the standard NEMA phantom to provide an additional thickness of 5 cm. Secondly, the standard spheres can be replaced by smaller spheres of diameter 7, 10, 13, 17, 22 and 28 mm. Preliminary data on a prototype showed that, when also using the phantom extension compartments, the 7 mm was just visible using combined RM and TOF reconstruction when filled with 8:1 contrast.

The primary aim of this phantom is to explore the possible limits of PET scanners, focussing on improvements in scanner spatial resolution and the impact of TOF on larger objects. There may also be an opportunity to try and evaluate the feasibility of

harmonising quantification with advanced reconstructions across multiple vendors. This is particularly relevant as more and more new scanners are installed and centres wish to utilize the benefits of resolution modelling and TOF.

In the UK, the National Cancer Research Institute (NCRI) provides a means of ensuring standardisation of recovery measurements for sites wishing to be enrolled in clinical trials within the UK. For accreditation, a centre is required to fill a NEMA phantom with F-18. The NCRI acceptance criteria for activity concentration recovery in each of the six spheres closely follow the specification set out by EARL, the research branch of the European Association of Nuclear Medicine [10]. A significant implication of this is that iterative reconstruction that uses resolution modelling falls outside the acceptable range as the recovery for the smaller spheres is greater than the upper bounds that are specified as acceptable. Given the clear benefits of advanced algorithms, there is an argument that standards should try to move towards incorporating these algorithms. A pilot study was performed during this PhD using standard NEMA data acquired on a GE PET/CT at Oxford University Hospitals and the Biograph mCT at CMUH. A range of reconstructions were performed on each system, all with resolution modelling [11]. The study suggested that the two manufacturers' algorithms gave notably different performance in terms of recovery of activity concentration measurements in the six spheres of the NEMA phantom.

Given that multiple copies of the prototype NEMA phantom are available in the department at CMUH, a potentially useful project may be to fill either all of one phantom or simply just the spheres with a long-lived isotope such as Ge-68, with a 271 day half-life, in a resin form. This would provide a phantom that could be transported to various sites and compare performance of new PET scanners.

9.3 Assessment of image quality with next-generation PET scanners

9.3.1 Solid state PET detectors

In late 2016, a GE SIGNA PET/MR will be installed at CMUH a part of the Dementias Platform UK grant. The PET detector uses Silicone Photomultipliers (SiPM) instead of traditional photomultiplier tubes and its performance represents a substantial change compared with the currently available scanners. There are two reasons for this: firstly the intrinsic sensitivity of the detector is approximately

23 cps/kBq [12], compared with the 9.7 cps/kBq sensitivity of the Biograph mCT [13] that is currently in the department; secondly the TOF timing resolution of the SiPM detector is quoted at approximately 390 ps [14] compared with the quoted 530 ps of the Biograph mCT. These two factors suggest the SNR in the images from the SIGNA should be greater than those in the mCT and characterising this would be a useful exercise for protocol development. The role of PET/MR within the NHS is yet to be established and in the short-term, it is unlikely to replace PET/CT as a high-throughput solution for routine oncology scanning based upon the currently common clinical indications. However, in summer 2016, at the Society of Nuclear Medicine and Molecular Imaging Annual Meeting, GE launched a PET/CT system with the same SiPM-based detector [15] and other manufacturers will likely follow suit as has been the case historically. Consequently, the characterisation of image quality gains from the SIGNA PET/MR at Manchester should be transferable to future-generation SiPM-based systems. The Biograph mCT is currently the most sensitive TOF-capable PET/CT that is available commercially and so this comparison should indicate the progress due to technological advances.

9.3.2 The evaluation of new reconstruction algorithms

The GE SIGNA PET/MR system will also be available with the *Q.Clear*, which has been available since 2014, and consists of OSEM reconstruction with MAP regularisation to encourage image smoothness and suppress the noise propagation during the iterative reconstruction process. This has been shown to increase SNR for lesions by increasing the lesion SUV while reducing background noise [8, 16]. Currently, no comparison has been performed against the current best performance of the Biograph mCT. The first logical investigation should be a side-by-side assessment of the NEMA phantom on the two systems, taking a similar method as the first study in this thesis.

9.4 References

1, National Institute for Health and Care Excellence. *Lung cancer: diagnosis and management*. NICE Clinical Guideline 121; 2011

2, Patz EF, Lowe VJ, Hoffman JM, Paine SS, Burrowes P, Coleman RE, et al. *Focal pulmonary abnormalities: evaluation with F-18 fluorodeoxyglucose PET scanning*. Radiology 1993; **188**:487-490

3, Hain SF, Curran KM, Beggs AD, Fogelman I, O'Doherty MJ, Maisey MN. *FDG-PET as a "metabolic biopsy" tool in thoracic lesions with indeterminate biopsy*. Eur J Nucl Med 2001; **28**:1336-1340

4, Beggs A, Hain S, Curran K, O'Doherty M. *FDG-PET as a "metabolic biopsy" tool in non-lung lesions with indeterminate biopsy*. Eur J Nucl Med Mol Imaging 2002; **29**:542-546

5, Kwee T, Cheng G, Lam MEH, Basu S, Alavi A. *SUVmax of 2.5 should not be embraced as a magic threshold for separating benign from malignant lesions*. Eur J Nucl Med Mol Imaging 2013; **40**:1475-1477

6, Cerfolio RJ, Bryant AS, Ohja B, Bartolucci AA. *The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival*. J Thorac Cardiovasc Surg 2005; **130**:151-159

7, Bryant AS, Cerfolio RJ, Klemm KM, Ojha B. *Maximum Standard Uptake Value of Mediastinal Lymph Nodes on Integrated FDG-PET-CT Predicts Pathology in Patients with Non-Small Cell Lung Cancer*. Ann Thorac Surg 2006; **82**:417-422

8, Teoh EJ, McGowan DR, Bradley KM, Belcher E, Black E, Gleeson FV. *Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules*. Eur Radiol 2016; **26**:576-584

9, Hoffman EJ, Huang SC, Phelps ME. *Quantitation in positron emission computed tomography: 1. Effect of object size*. J Comput Assist Tomogr 1979; **3**:299-308

10, EANM Research Ltd (EARL). New EANM FDG PET/CT accreditation specifications for SUV recovery coefficients http://earl.eanm.org/cms/website.php?id=/en/projects/fdg_pet_ct_accreditation/accreditation_specifications.htm. Accessed 01 Sept 2016

11, Armstrong IS, Kenny K, Rowley L, McGowan D. *Comparison of PET reconstructions from two manufacturers with point spread function modelling and time-of-flight*. Nuc Med Commun 36(5):P58, 2015

12, Grant AM, Deller TW, Khalighi MM, Maramraju SH, Delso G, Levin CS. *NEMA NU 2-2012 performance studies for the SiPM-based ToF-PET component of the GE SIGNA PET/MR system*. Med Phys 2016; **43**:2334-2343

13, Jakoby BW, Bercier Y, Conti M, Bendriem B, Townsend D. *Physical and clinical performance of the mCT time-of-flight PET/CT scanner*. Phys Med Biol 2011; **56**:2375-2389

14, Levin CS, Maramraju SH, Khalighi MM, Deller TW, Delso G, Jansen F. *Design Features and Mutual Compatibility Studies of the Time-of-Flight PET Capable GE SIGNA PET/MR System*. IEEE Trans Med Imaging 2016; **PP**:1

15, GE Healthcare. *Going Beyond Analog*. GE Healthcare Press Release 2016. http://www.genewsroom.com/press-releases/going-beyond-analog-ge-healthcare-launches-digital-next-generation-molecular-imaging. Accesses 5th September 2016.

16, Parvizi N, Franklin JM, McGowan DR, Teoh EJ, Bradley KM, Gleeson FV. *Does a novel penalized likelihood reconstruction of 18F-FDG PET-CT improve signal-to-background in colorectal liver metastases?* Eur J Radiol 2015; **84**:1873-1878