PET Quantification for Assessing Tumour Response

by

Mike Sattarivand

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Medical Biophysics
University of Toronto

Copyright © Mike Sattarivand (2014)
Abstract

PET quantification for assessing tumour response

Mike Sattarivand, 2014

Doctor of Philosophy

Department of Medical Biophysics

University of Toronto

Treatment response assessment in advanced head and neck cancer patients using Positron Emission Tomography (PET) has potential to provide significant clinical benefit. PET quantification methods can be either static or dynamic. The static approach is simple and is widely used. The simplified dynamic PET quantification method is a promising approach as it provides a reasonable trade-off between accuracy and clinical practicality. This method requires a blood sample which makes it not ideal since the PET quantification accuracy may be compromised due to small activity and volume of the blood sample. The implementation of image-based simplified dynamic PET quantification in head and neck cancer patients requires partial volume correction due to small vessel sizes and limited PET spatial resolution. The objective of this thesis is to evaluate the accuracy of current PET quantification methods for response assessment in advanced head and neck cancer patients and to develop a novel and robust partial volume correction technique to improve PET quantification.

First, the static PET quantification method using fixed size ROI is evaluated. Significant variation in response assessment was observed suggesting that static PET quantification using a fixed-size ROI should be approached with caution in heterogeneous tumours.

Second, the accuracy of blood activity measurements and its effect on the accuracy of quantitative response assessment is evaluated. Significant inaccuracies in the blood sample based
simplified dynamic PET quantification method are identified. The results support a need to develop an image-based simplified dynamic PET quantification method with partial volume correction.

Finally, a novel partial volume correction technique was developed, validated, and its robustness was investigated. In comparison to previously published partial volume correction techniques, it performed better with noisy PET images and it was more robust for errors in PET-CT registration. The partial volume correction technique was also implemented and validated in sinogram space to provide additional advantages such as applicability to iterative reconstructions. The proposed partial volume correction technique enables the use of image-based simplified dynamic PET quantification in advanced head and neck cancer patients. Furthermore, the technique establishes a framework for future research to address the inherent low spatial resolution of PET.
To my parents
Acknowledgments

First and foremost, I am sincerely grateful to my supervisor Dr. Curtis Caldwell for his infallible guidance, availability, encouragement, and exceptional support. He has played a key role in fostering my scientific training and has taught me the importance of patience and persistence in research. With his gentle and kind reassurance, he has successfully navigated me around all obstacles and showed me that independent research can be a rewarding experience. It has been a great pleasure working with Curtis and I hope to always count on his support and friendship.

I would also like to thank the members of my supervisory committee: Dr. Ian Poon, Dr. Gregory Czarnota, and Dr. John Rowlands for their guidance and insightful comments on my thesis work. I would especially like to thank Dr. Ian Poon for the time he has spent helping me with the clinical aspects of my project. Guidance from my committee, especially Dr. Gregory Czarnota, has helped me stay on track in my PhD research.

I would like to thank the past and present members of Dr. Curtis Caldwell and Dr. Ian Poon’s research group for their scientific contribution and continuous help. Many special thanks goes to Maggie Kusano who helped me all along my PhD work and contributed to all chapters in this thesis.

Last but not least, I am eternally indebted to my family especially my wife for their unconditional love and encouragement, without which I would not have been able to make it through this challenging phase of my life.
## Contents

Chapter 1: Introduction .................................................................................................................. 1
  1.1. Overview of chapter .............................................................................................................. 1
  1.2. Why response assessment is needed in cancer therapy? .................................................. 2
  1.3. Current approaches for response assessment and limitations ........................................... 3
  1.4. Review of novel imaging methods for response assessment .............................................. 4
    1.4.1. Dynamic contrast enhanced - magnetic resonance imaging (DCE-MRI) ................... 5
    1.4.2. Magnetic resonance spectroscopy (MRS) ................................................................. 6
    1.4.3. Dynamic contrast enhanced - ultrasound (DCE-US) ............................................... 7
    1.4.4. Quantitative ultrasound (QUS) .................................................................................. 8
    1.4.5. Positron emission tomography (PET) ........................................................................ 10
  1.5. Clinical problem and its significance ................................................................................. 11
    1.5.1. Advanced head and neck cancer ............................................................................... 12
    1.5.2. Clinical significance ..................................................................................................... 13
  1.6. PET quantification methods ............................................................................................... 15
  1.7. Problems with current PET quantification methods ............................................................ 17
    1.7.1. Problem of ROI placement in PET quantification using SUV ................................... 18
    1.7.2. Problems of current PET quantification using blood sample based SKA-M ............. 19
    1.7.3. Problem of loss of PET quantification due to partial volume error ......................... 19
  1.8. Quantitative objective for PET tumour response assessment ........................................... 21
  1.9. Thesis outline ...................................................................................................................... 22

Chapter 2: Evaluation of static PET quantification using fixed size ROI for tumour response assessment ......................................................................................................................... 35
  2.1. Introduction ....................................................................................................................... 36
  2.2. Materials and methods ..................................................................................................... 37
    2.2.2. Population A ............................................................................................................... 39
    2.2.3. Population B ............................................................................................................. 41
  2.3. Results ............................................................................................................................... 43
    2.3.1. Population A ............................................................................................................... 43
    2.3.2. Population B ............................................................................................................. 44
  2.4. Discussion .......................................................................................................................... 45
    2.4.1. Effect of geometric changes ...................................................................................... 45
    2.4.2. Effect of ROI method ............................................................................................... 46
    2.4.3. Clinical significance ................................................................................................. 48
  2.5. Conclusion .......................................................................................................................... 50
Chapter 3: Accuracy of PET quantification using blood sample based simplified dynamic PET for response assessment ..............................................................60

3.1. Introduction .................................................................................................................61
3.2. Methods and materials ..............................................................................................63
  3.2.1. Thyroid probe technique (ProbePoint) .................................................................64
  3.2.2. Well counter techniques (WellPoint and WellCurve) ............................................65
  3.2.3. Precision of micro-pipette blood volume measurements ..........................................66
  3.2.4. Accuracy and precision of micro-pipette volume measurements for water ..........66
  3.2.5. Calculating SKA-M using the different blood sample measurement techniques ....67
3.3. Results ..........................................................................................................................68
3.4. Discussion ....................................................................................................................69
  3.4.1. Blood activity measurement ..................................................................................70
  3.4.2. Blood volume measurements .................................................................................71
  3.4.3. Effects of blood activity concentration measurements on SKA-M .......................71
  3.4.4. Clinical significance ...............................................................................................73
  3.4.5. Direct image-based measurement of input functions for SKA-M .........................74
3.5. Conclusion ....................................................................................................................75

Chapter 4: Improving PET quantification accuracy by developing a novel partial volume correction technique ..................................................................................83

4.1. Introduction ..................................................................................................................83
4.2. Theory ...........................................................................................................................84
  4.2.1. Analytical derivation ..............................................................................................87
  4.2.2. Physical interpretation ............................................................................................89
4.3. Materials and methods ...............................................................................................90
  4.3.1. Simulations ............................................................................................................90
  4.3.2. Physical sphere phantom .......................................................................................93
  4.3.3. Image analysis .......................................................................................................93
4.4. Results ..........................................................................................................................94
4.5. Discussion ....................................................................................................................96
  4.5.1. Accuracy ...............................................................................................................97
  4.5.2. Precision and noise characteristics .......................................................................97
  4.5.3. Robustness ..........................................................................................................98
  4.5.4. Clinical significance .............................................................................................100
4.6. Future directions .........................................................................................................101
4.7. Conclusion ....................................................................................................................101
4.8. Appendix .....................................................................................................................102

Chapter 5: Improving PET quantification accuracy by implementing the novel partial volume correction technique in sinogram space .........................................................112

5.1. Introduction ................................................................................................................112
5.2. Materials and methods .............................................................................................116
  5.2.1. Principles of region-based PVC methods ..............................................................116
  5.2.2. Simulations ..........................................................................................................119
5.2.3. Physical sphere phantom ...............................................................122
5.2.4. Image analysis ..............................................................................122
5.3. Results .............................................................................................124
5.4. Discussion .........................................................................................126
  5.4.1. Accuracy .......................................................................................127
  5.4.2. Noise characteristics .................................................................127
  5.4.3. Robustness to registration errors ................................................128
  5.4.4. The need for sinogram implementation ......................................129
5.5. Future directions ............................................................................130
5.6. Conclusion .......................................................................................131

Chapter 6: ............................................................................................139
Conclusions and future work ...............................................................139
  6.1. Thesis summary ...............................................................................140
  6.2. Future directions ............................................................................143
    6.2.1. Simplified dynamic PET quantification using an image-based input function ..........143
    6.2.2. Partial volume correction for iterative reconstructions ..................145
  6.3. Concluding remarks .....................................................................145

References ..........................................................................................150
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Functional versus anatomical approach for response assessment</td>
<td>25</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Basics of FDG PET imaging</td>
<td>26</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Management of advanced head and neck cancer patients</td>
<td>27</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Two PET scans to assess response in an advanced HNC patient</td>
<td>28</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>PET quantification methods</td>
<td>29</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Kinetic modeling</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Problem of ROI placement in PET quantification using SUV</td>
<td>31</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Image based SKA-M</td>
<td>32</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Partial volume error in PET imaging</td>
<td>33</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>The problem of image registration in current PVC methods</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Change in the distribution of FDG uptake during treatment</td>
<td>53</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Uptake profiles</td>
<td>54</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Therapy induce geometric changes</td>
<td>55</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Distance histogram</td>
<td>56</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Quantitative response assessment using two ROI method</td>
<td>57</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Uncertainties in ROIsame placement</td>
<td>58</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Quantification of therapy induce geometric changes</td>
<td>59</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Cross calibration</td>
<td>76</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Steps for calculating SKA-M</td>
<td>77</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>A sample calibration curve</td>
<td>78</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Accuracy of blood activity measurments and effects on SKA-M</td>
<td>79</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Precision of blood volume measurements</td>
<td>80</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Accuracy and precision of micro-pipette volume measurements for water</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Clinical impact of errors in blood activity concentration measurements</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Principals of GTM and sGTM partial volume correction techniques</td>
<td>104</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Simulated sphere phantom</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Simulated brain phantom</td>
<td>106</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Physical sphere phantom</td>
<td>107</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Accuracy and precision of RC</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>The effect of noise on the two PVC methods</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Sensitivity of the two PVC methods to mis-registration</td>
<td>110</td>
</tr>
</tbody>
</table>
Figure 4.8 RC values when PET PSF FWHM error is applied 111
Figure 5.1 Principals of GTMo and sGTMo partial volume correction techniques. 132
Figure 5.2 Simulated sphere phantom 133
Figure 5.3 Simulated brain phantom 134
Figure 5.4 Physical sphere phantom 135
Figure 5.5 Accuracy and precision of RC values using four PVC methods 136
Figure 5.6 Noise propagation characteristics of four PVC methods 137
Figure 5.7 The robustness of the PVC methods to registration errors. 138
Figure 6.1 Steps of calculating image-based SKA-M 147
List of Tables

Table 1.1  RECIST criteria for tumour response assessment  24
Table 2.1  Previous response assessment studies  51
Table 2.2  Patient Characteristics  52
Table 6.1  Estimates of clinical impact of response assessment in advanced HNC.  146
Table 6.2  Image-based vs blood sample based SKA-M for a sample patient from OCC trial  148
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>1-dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>2-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>3-dimensional</td>
</tr>
<tr>
<td>3D_RAMLA</td>
<td>3-dimensional row-action maximum likelihood algorithm</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>COM</td>
<td>center of mass</td>
</tr>
<tr>
<td>CR</td>
<td>complete response</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
</tr>
<tr>
<td>CT</td>
<td>computed tomography</td>
</tr>
<tr>
<td>DCE-MRI</td>
<td>dynamic contrast enhanced - magnetic resonance imaging</td>
</tr>
<tr>
<td>DCE-US</td>
<td>dynamic contrast enhanced - ultrasound</td>
</tr>
<tr>
<td>EORTC</td>
<td>European organization for research and treatment of cancer</td>
</tr>
<tr>
<td>FBP</td>
<td>filtered back projection</td>
</tr>
<tr>
<td>FDG</td>
<td>fluorodeoxyglucose</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FOV</td>
<td>field of view</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FWHM</td>
<td>full width half maximum</td>
</tr>
<tr>
<td>GTM</td>
<td>geometric transfer matrix</td>
</tr>
<tr>
<td>GTMo</td>
<td>GTM implemented in sinogram space</td>
</tr>
<tr>
<td>GTV</td>
<td>gross tumour volume</td>
</tr>
<tr>
<td>HNC</td>
<td>head and neck cancer</td>
</tr>
<tr>
<td>IDL</td>
<td>interactive data language</td>
</tr>
<tr>
<td>IMRT</td>
<td>intensity modulated radiation therapy</td>
</tr>
<tr>
<td>Intra-Tx</td>
<td>intra-treatment</td>
</tr>
<tr>
<td>IR</td>
<td>iterative reconstruction</td>
</tr>
<tr>
<td>LOR</td>
<td>line of responses</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MRS</td>
<td>magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NCI</td>
<td>national cancer institute</td>
</tr>
<tr>
<td>NMF</td>
<td>noise magnification factor</td>
</tr>
<tr>
<td>OCC</td>
<td>Odette cancer center</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>PD</td>
<td>progressive disease</td>
</tr>
<tr>
<td>PERCIST</td>
<td>PET response criteria in solid tumors</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PR</td>
<td>partial response</td>
</tr>
<tr>
<td>Pre-Tx</td>
<td>pre-treatment</td>
</tr>
<tr>
<td>PSA</td>
<td>prostate specific antigen</td>
</tr>
<tr>
<td>PSF</td>
<td>point spread function</td>
</tr>
<tr>
<td>PVC</td>
<td>partial volume correction</td>
</tr>
<tr>
<td>QA</td>
<td>quality assurance</td>
</tr>
<tr>
<td>QUS</td>
<td>quantitative ultrasound</td>
</tr>
<tr>
<td>RC</td>
<td>recovery coefficient</td>
</tr>
<tr>
<td>RECIST</td>
<td>response evaluation criteria in solid tumors</td>
</tr>
<tr>
<td>ROI</td>
<td>region of interest</td>
</tr>
<tr>
<td>RSF</td>
<td>regional spread function</td>
</tr>
<tr>
<td>SD</td>
<td>stable disease</td>
</tr>
<tr>
<td>sGTM</td>
<td>symmetric geometric transfer matrix</td>
</tr>
<tr>
<td>sGTMo</td>
<td>symmetric GTM implemented in sinogram space</td>
</tr>
<tr>
<td>SKA-M</td>
<td>simplified kinetic analysis - multiple time points</td>
</tr>
<tr>
<td>SPECT</td>
<td>single-photon emission computed tomography</td>
</tr>
<tr>
<td>STIR</td>
<td>software for tomographic image reconstruction</td>
</tr>
<tr>
<td>SUV</td>
<td>standardized uptake value</td>
</tr>
<tr>
<td>TAC</td>
<td>time activity curve</td>
</tr>
<tr>
<td>US</td>
<td>ultrasound</td>
</tr>
<tr>
<td>VOI</td>
<td>volume of interest</td>
</tr>
</tbody>
</table>
Chapter 1:

Introduction
1.1. Overview of chapter

This chapter reviews methods of assessing tumour response to cancer therapy. The current clinical standard for assessing response is discussed and its limitations are noted. A review of imaging methods which have been suggested for assessing tumour response is presented. A clinical problem which would benefit from an improved method of assessing response during treatment of advanced head and neck cancer is then presented. The use of positron emission tomography (PET) as a functional imaging modality for assessing tumour response during radiation treatment of advanced head and neck cancer is introduced. The need for accurate PET quantification is discussed and the thesis objectives to address this need are established.

1.2. Why response assessment is needed in cancer therapy?

Assessing response to cancer therapy has two important roles [1, 2]. First, it is used in clinical trials of new therapeutic approaches. The therapeutic development process is accelerated and the costs are reduced by providing an early indication to determine the efficacy of the new treatment method, whether a new pharmaceutical, a new method of delivering radiation therapy, a new surgical approach or some combination of these. Ideally, treatment response is assessed by direct measures such as years of disease-free survival in a randomized clinical trial comparing the new treatment approach with a standard treatment approach. However, it is of great benefit to patients, and would substantially reduce the cost of clinical trials, if other methods of assessing treatment response were available that were predictive of direct measures such as survival. Any technique for assessing response to therapy must be proven to be a powerful surrogate for direct indications of clinical outcome such as survival time. The second role for methods of assessing response is in personalized therapy by allowing physicians to assess the effectiveness of the therapy as early as possible during the course of that therapy. By signalling the need for therapeutic adjustments at
an early stage, patient management may be improved. Ineffective treatments could be abandoned for patients who are non-responders and alternative treatments that may be more effective for those patients could be selected. Thus, an effective response assessment may provide potential survival and/or quality of life benefits for the individual patient as well as cost benefits for the health care system.

1.3. Current approaches for response assessment and limitations

Currently, there are two primary approaches used clinically for assessing response to cancer treatments. The first is to measure specific markers secreted by cancer cells in blood samples. The second, and more common approach, is to measure changes in tumour size due to therapy using diagnostic imaging techniques.

Examples of serum markers are prostate specific antigen (PSA) for prostate cancer and thyroglobulin for well-differentiated thyroid cancer of follicular cell origin. The main advantage of using serum markers for assessing response are their relatively low cost and thus testing can be repeated frequently. However, suitable biomarkers that are secreted at sufficiently high levels to be used in assessing response have only been identified for a fraction of tumours. Moreover, these markers often are not specific to viable tumour cells. For example, PSA is specific to prostate but not to prostate cancer and its benefit for response assessment has been questioned [3].

The use of change in tumour size as a means of assessing treatment response is a common convention in current clinical practice. The response evaluation criteria in solid tumours (RECIST) rules have been developed [4] and updated [5] for this purpose. In the RECIST method, a tumour lesion is defined as being measurable if it is at least 10 mm in extent in any direction on computed tomography (CT) or magnetic resonance images (MRI). CT and MRI are
considered to be the best available and most reproducible imaging modalities for such assessment. When more than one measurable lesion is identified, all lesions (up to five) that represent involved organs are defined as target lesions. Lymph nodes are assessed differently since they are normal anatomical structures that could be visible by imaging even if they are not involved by disease. Any involved lymph node with a short axis greater than 15 mm is considered measurable and assessable as a target lesion. All other involved nodes with short axis between 10 mm and 15 mm are considered non-target lesions. Nodes having a short axis less than 10 mm are considered non-pathological and thus are not recorded.

A one-dimensional (1D) measurement on the longest diameter of each target lesion is performed and their sum is used to categorize treatment response into one of four groups: complete response (CR), partial response (PR), progressive disease (PD), and stable disease (SD). The RECIST rules based on changes in tumour size is summarized in table 1.1.

With 1D measurements of tumour lesions, RECIST evaluates response with minimal workload on routine clinical practice. However, since the approach is based on anatomical imaging, it has inherent limitations [6, 7]. RECIST does not address tumour heterogeneity. For example, tumour necrosis may not shrink due to therapy, causing an underestimate in the measure of response. Furthermore, RECIST may fail to appropriately assess response to therapies that are not cytocidal such as anti-angiogenesis drugs. Finally, inaccuracies may be introduced since assessing response using tumour size requires well defined tumour boundaries. Inter-observer variability in defining tumour boundaries has been reported to be as high as 100% in some cases [8].

1.4. Review of novel imaging methods for response assessment

In order to address some of the limitations of the RECIST criteria, a functional imaging approach may be used based on tumour biochemistry or cell biology. A functional approach has
the advantage of providing an earlier assessment of response since functional changes in tumours usually precede anatomical changes as illustrated in figure 1.1. Macroscopic changes in tumour size may take weeks to months to occur, while a functional imaging approach can measure change in tumour function from days to weeks after starting treatment [9].

Various image-based functional approaches have been applied to response assessment. These techniques include: dynamic contrast enhanced - magnetic resonance imaging (DCE-MRI), magnetic resonance spectroscopy (MRS), dynamic contrast enhanced - ultrasound (DCE-US), quantitative ultrasound (QUS), and positron emission tomography (PET). These techniques are briefly reviewed below.

1.4.1. Dynamic contrast enhanced - magnetic resonance imaging (DCE-MRI)

Magnetic resonance imaging involves detection of nuclear spin reorientation in a high magnetic field after the application of a radiofrequency pulse at the resonance frequency. Imaging is usually restricted to water protons since the interaction is weak and water is abundant in tissue. For typical magnetic field strengths used clinically (i.e., 1.5 or 3 T), image resolution is in the range of 2-3 mm which is comparable to that of CT. However, soft tissue contrast is higher in MR than in CT. The acquisition time, in part, depends on the resolution required and typically is several minutes. Similar to CT, conventional MRI provides an anatomical image of the subject. However, different techniques may be used to acquire functional images suitable for response assessment.

One such functional imaging technique is DCE-MRI. DCE-MRI images reflect microvasculature parameters. Angiogenesis is an important mechanism in malignant tumour growth and metastasis in which new capillaries are formed from existing blood vessels. Anti-angiogenesis drugs are not cytocidal to tumour cells directly but rather inhibit angiogenesis and
response to these drugs may be assessed using the DCE-MRI technique. In this technique, dynamic T1-weighted MR images are acquired immediately after an intravenous bolus injection of a contrast agent (e.g. gadolinium). The transit of the contrast agent through the tumor vasculature is recorded over time through rapid, dynamic MR image acquisition. The MR image intensity change is converted into contrast-agent concentration data and kinetic modeling is applied to extract vasculature parameters related to blood flow, capillary leakage and other physiological parameters [10]. These parameters reflect the tumor vascular micro-environment and changes in these parameters during therapy may be used as surrogate biomarkers for early assessment of treatment response. DCE-MRI parameters have been demonstrated to correlate with histopathological or clinical outcome data in patients treated with anti-angiogenesis drugs [11, 12]. DCE-MRI techniques can vary widely among different studies due to technical complexity, different modeling approaches, nonlinear relationships between the vasculature parameters, physiologic processes, and intrinsic tumor heterogeneity. Consensus recommendations to address the issue have been outlined and are being used for developing and validating the techniques for response assessment [13, 14]. However, DCE-MRI methods are still in their infancy and are actively being investigated.

1.4.2. Magnetic resonance spectroscopy (MRS)

The resonance frequency for the same MR active nucleus may vary slightly (on the order of a few kilohertz) depending on the surrounding molecular environment. For example, the resonance frequency of protons in a water molecule (64 MHz at 1.5 T) is slightly different than that of protons in choline. This phenomenon, which is known as chemical shift, is exploited in MRS to evaluate the molecular components of tissue. It provides a spectrum of frequencies where the height of the peaks is proportional to the metabolite concentration in the sampled tissue.
Proton MRS is used to investigate several metabolites such as choline containing compounds to study phospholipid metabolism. Choline compounds are involved in cell membrane synthesis and phospholipid metabolism. Therefore, tumours often have high levels of choline compared to normal tissues [15]. Decrease in the levels of choline compounds in tumours has been demonstrated to predict response to therapy in proton MRS studies of breast [16, 17], brain [18], and prostate cancer [18, 19]. Methodological differences among MRS studies calls for a need for standardization [14]. The major limitation of MRS for response assessment is its low sensitivity in tumours [20]. As the lesion shrinks during the course of therapy, partial volume error with surrounding normal tissue causes a significant challenge in detecting choline in the tumour. Higher magnetic field strengths may be used to improve sensitivity [21, 22].

1.4.3. Dynamic contrast enhanced - ultrasound (DCE-US)

Ultrasound (US) images are formed by transmitting high frequency sound waves through tissue and recording the reflected waves from the underlying tissue structures. Higher frequency sound waves have shorter wavelengths and are thus capable of recording higher resolution details. Unfortunately, the attenuation of US is higher at higher frequencies, limiting the depth in tissue where high resolution imaging can be used. Clinical imaging is performed at frequencies between 2 and 15 MHz to image depth of up to 2 -15 cm with a spatial resolution of 0.2 - 1.2 mm. Higher resolutions (with higher frequencies) are possible with limited depths in small animals. US is an attractive imaging technique since it is relatively inexpensive, widely available, portable, involves no ionizing radiation exposure to either the patient or the operator, and provides near real time imaging. The major limitations of conventional US are that it is limited to abdomen or superficial organs and it cannot provide whole body imaging (similar to CT or MRI). US waves have difficulty penetrating bone and the extreme differences in acoustic impedance
between air and soft tissue makes is not suitable to image lungs. Similar to MRI, conventional US provides anatomical imaging. However, different techniques are available to provide functional imaging applicable to tumour response assessment.

One approach to create functional imaging is DCE-US to image blood perfusion. Blood perfusion in the tumour is a characteristic of tumour microvasculature and it is directly associated with therapeutic outcome in radiotherapy [23], hyperthermia [24], and drug therapy [25]. In the DCE-US technique, a microbubble contrast agent is injected intravenously. The characteristics of tumour microvasculature can be obtained directly from the shape of the time intensity curves following a bolus injection of microbubbles, providing a semi-quantitative and model free approach [26, 27]. Alternatively, microbubbles are injected continuously and blood perfusion is quantified by recording replenishment kinetics of microbubbles after their local destruction by US [28, 29]. The tumour replenishment kinetics of microbubbles is then used to extract parameters such as blood volume, flow velocity and perfusion parameters useful to characterize microvasculature for response assessment purposes.

DCE-US has been used clinically for assessing response to hepatic and renal radiofrequency ablation therapy [30-32] as well as for assessing response to radiotherapy of liver metastases [33]. Another important clinical application of DCE-US is the assessment of response to anti-angiogenesis drugs by quantifying solid tumour perfusion and detecting early microvasculature changes [34]. Changes in microvasculature can be identified as early as 1-2 weeks post-therapy and these changes are correlated with patient survival in hepatocellular carcinoma and renal cell carcinoma [35].

1.4.4. Quantitative ultrasound (QUS)
The raw signal received by an US transducer contains information about tissue echogenicity that is related to structural characteristics of cells. However, conventional US imaging ignores this information and the raw data envelope is simply detected, log amplified and used to generate a B-mode image. QUS exploits more of the information contained in the raw data and performs spectral analysis to obtain backscatter parameters that are related to cell morphology. Using high frequency (20–60 MHz) QUS, morphological differences between living cells, necrotic cells, and apoptotic cells can be detected [36]. Several studies have confirmed the technique’s ability to monitor cell structural changes to chemotherapy [37] and radiotherapy [38] in-vitro and to detect cell death ex-vivo in liver tissue being preserved for transplantation [39]. The technique is robust and reproducible and can be used to evaluate the extent of cell death since nuclear condensation and fragmentation during cell death cause a large boost in the backscatter signal [40].

The in-vivo application of high frequency QUS to monitor response has been reported in the xenograft mouse models used to assess tumour response to photodynamic therapy [41] and radiotherapy [42]. In these studies, response to therapy was detected by QUS as early as 24 hours after the therapy. Due to the limited depth of penetration at high frequencies, the application of QUS to response assessment in patients is only possible if clinical frequencies (2-15 MHz) are used. In this frequency range, studies have demonstrated that QUS can detect as little as 10% apoptosis, paralleling changes observed at high frequency US [41, 43]. Results from an in-vivo mouse model experiment demonstrated that QUS using conventional clinical frequencies can predict histopathologically-confirmed tumour response to therapy [44]. QUS was also investigated to assess response in advanced breast cancer treated with chemotherapy [45]. The results indicated that QUS parameters, detected after 1-2 cycles of chemotherapy (a few weeks), were different between responders and non-responders while the clinical response in the tumour was observed many months later.
1.4.5. Positron emission tomography (PET)

PET is a functional imaging modality that provides images of an in-vivo biodistribution of a biologically active molecule analog (tracer) that is labelled with a positron emitting radionuclide. The basics of PET imaging are described in figure 1.2. The tracer is injected into the patient through a venous catheter. The most common PET tracer is fluorodeoxyglucose (FDG) which is a glucose analog and it is used to study glucose metabolism. FDG is created by chemically replacing a hydroxyl group in glucose with the radionuclide F-18. The FDG is distributed throughout the body by the blood stream and accumulates markedly in many tumour cells due to their higher glucose metabolism than normal cells. This phenomenon is known as the Warburg effect [46] (i.e. most cancer cells mainly produce energy by glycolysis rather than by oxidation of pyruvate like most normal cells).

The F-18 undergoes a $\beta^+$ decay by emitting a positron which travels for a short distance. The *positron range* is on the order of 1 mm [47]. When the positron slows down, the annihilation of the positron with an electron gives rise to two 511 keV photons emitted at $180^\circ \pm 0.25^\circ$ [48]. The PET detector system consists of crystals, typically 4 to 6 mm in cross sectional dimension, with a depth of 20 to 30 mm, attached to photo-multiplier tubes [47]. The annihilation events from all the coincident detector pairs are recorded to create the raw PET data or *sinogram*. A PET sinogram is analogous to a CT projection image. However, many fewer events are recorded in PET compared to CT and thus the PET data are inherently much noisier.

The sinograms are reconstructed to create the final PET image. Historically, the conventional reconstruction algorithm was filtered back projection (FBP) algorithm. This algorithm projects back the measured counts along the angles that form the line of response (LOR). Since back
projection also smoothes the image, each projection is first filtered before the back projection, hence the name filtered back projection. The choice of the filter is a trade-off between noise amplification and restoring the resolution. The FBP algorithm is an analytical approach with a single back projection step. In modern clinical PET systems, iterative reconstruction (IR) algorithms are used virtually exclusively. Unlike the FBP, the IR algorithms consist of repetitive forward and back projection steps starting with an initial estimate of the image. After each forward projection step, the resulting sinogram is compared to that of the measured sinogram. The steps are iterated until a desired convergence is achieved. Although FBP is faster, the IR approach can improve image quality by incorporating accurate physical and statistical modeling of photon production and detection processes during forward projection steps [49]. The reconstructed PET image provides the spatial distribution of the FDG throughout the patient's body and thus may be used to study glucose metabolism. Changes in glucose metabolism may reflect a metabolic response to treatment.

FDG PET has been demonstrated to have prognostic value in different cancer sites treated with radiation or chemotherapy [50-60]. These studies concluded that changes in tumour metabolic activity measured by FDG uptake is significantly associated with tumour response and patient survival. New methods such as the PET response criteria in solid tumours (PERCIST) are being developed [61]. The promising results of these studies triggered a pilot study with 10 advanced head and neck cancer patients in the Odette cancer centre (OCC) in 2007 and a peer reviewed research grant was obtained to conduct a clinical trial with 100 patients under Dr. Ian Poon's supervision. By 2012, the patient PET scans were complete and patient survival data are currently being collected.

1.5. Clinical problem and its significance
1.5.1. Advanced head and neck cancer

Head and neck cancer (HNC) is a group of malignancies that originates in areas of the head and neck such as the oral cavity, nasal cavity, pharynx, and larynx. About 90% of HNCs are squamous cell carcinomas, i.e. epithelial malignancies of the mucous membranes [62]. HNC often spreads to the lymph nodes in the neck before it is first diagnosed. Common risk factors include smoking, alcohol consumption, and certain strains of viruses such as the human papilloma virus. Every year about 40,000 new cases of HNC are diagnosed in North America [63]. More than 50% of all HNC patients have locally advanced (stage III and IV) disease at diagnosis [64].

The current management of advanced HNC patients is illustrated in figure 1.3. Standard treatment with curative intent is radiotherapy with or without the addition of chemotherapy depending on the patient's tolerance and surgery is reserved as a salvage treatment option due to its potentially high morbidity. Currently, response to therapy is not usually evaluated until 8-12 weeks after radiotherapy is completed, as anatomic changes reflective of response are unlikely to be detectable before that time. The current standard therapy evaluation methods used are based on change in tumour size and include clinical examination, endoscopy, and anatomical imaging such as CT and MRI. After therapy evaluation, salvage surgery may be offered to patients who are non-responders to the initial therapy. For these patients, the initial therapy will have failed to achieve loco-regional control, will have delayed a potentially more effective treatment (such as surgery), and may have caused significant morbidity.

Salvage surgery after the completion of unsuccessful initial therapy has increased surgical risks and a more radical surgical procedure may be required. Patients may require more tissue volume to be surgically removed. More extensive surgery also means longer post-surgery recovery time.
The likelihood of post-surgery infection and wound breakdown is higher for these patients. Moreover, radiotherapy induces an abnormal appearance in tissues that might compromise imaging that is required to plan surgery. Surgery must be delayed a few months in order for the normal tissue to recover. Radiation therapy may cause patients to lose weight and become fatigued. Some patients may not be ready for surgery and some may miss the window of curative surgical treatment. If, however, patients whose initial treatment is ineffective could be identified within the first few weeks of treatment, salvage surgery would not need to be delayed.

A therapeutic approach that is response adaptive would personalize treatment and reduce treatment toxicity. For responders, one would continue the effective treatment, while for non-responders one could stop the ineffective treatment and offer earlier surgical treatment. Currently, there is no reliable method to predict which patients will respond during therapy.

1.5.2. Clinical significance

An accurate response assessment to identify non-responders during chemo-radiation treatment could have great clinical benefit in terms of (a) improving survival, and (b) limiting morbidity. Theoretical estimates of these benefits are provided below.

Earlier trials demonstrated no significant difference in survival rates between chemo-radiation and surgery [65]. However, due to organ preservation of chemo-radiation compared to surgery, established practice is to keep surgery as a salvage treatment option [66]. The 2-year survival rate for advanced HNC patients treated with radiotherapy has been estimated to be 50%, with the addition of concurrent chemotherapy improving the survival rate to 54% [67]. Residual neck disease is reported in 30–60% of the patients after the completion of chemo-radiation treatment [68]. Salvage surgery is currently offered to these non-responders as demonstrated in figure 1.3. If non-responders could be accurately identified early, overall survival rates could potentially
improve. Surgery would then be offered only to non-responders who stop the ineffective chemo-radiation shortly after its start. We make the assumption that the “responders” identified correspond to the 50% of patients who are expected to survive for a minimum of 2 years post-chemo-radiation (i.e., 50% of the initial cohort would be expected to have 100% 2 year survival). Non-responders would then correspond to the remaining 50% not expected to survive 2 years following chemo-radiation. If we assume that failure to respond to chemo-radiation is independent of the probability of successful surgical management, then this group of non-responders to chemo-radiation may then have a 50% 2-year survival following surgery. Thus, assuming that chemo-radiation and surgery affect survival independently, the 2-year survival rate could be improved theoretically to approximately \( (0.5 \times 100\%) + (0.5 \times 50\%) = 75\% \). Here, the first bracket is for responders (having only chemo-radiation), and the second bracket if for non-responders (having early surgery). While the assumption that the “survival benefit” of surgery would be independent of whether surgery was offered to all patients in the population up-front or offered just to the sub-group who would fail to respond to chemo-radiation is likely optimistic, the calculation suggests that there is a potential for significant survival benefit of early, accurate response assessment, even if the 2-year survival rate were much less than 50% in this group.

Morbidities associated with the current management of advanced HNC patients with radiation includes xerostomia (dry mouth), acute mucositis, and dysphagia which occur in up to 33%, 50%, and 46% of the patients respectively [69]. Since these morbidities are deterministic effects (rather than stochastic), an early accurate response assessment could potentially eliminate these morbidities by stopping the ineffective treatment for non-responders, i.e. in almost half of the patients. Thus, xerostomia, acute mucositis, and dysphagia could theoretically decrease to 17%, 25% and, 23% of the patients respectively. Table 6.1 summarizes the theoretical estimates of clinical significance in terms of survival and morbidity.
The clinical trial at the OCC was designed to use PET to address the limitations of the current standard approach to assess response in HNC. For this purpose, two PET scans were performed as illustrated in figure 1.4. The pre-treatment (Pre-Tx) scan was performed before the treatment and the intra-treatment (Intra-Tx) scan was performed two weeks after the start of the treatment. Figure 1.4 reveals that the spatial distribution of tumour uptake has changed and the uptake in the tumour is relatively reduced because of the treatment. However, it is not clear how one would decide whether or not this patient is a responder or non-responder based on subjective evaluation of changes in tumour uptake. As has been shown with the RESIST and PERCIST criteria, a method of assessing response is only likely to be of clinical benefit if it can provide consistent, quantitative values. The answer to this key question of how to determine if a patient is a responder or not thus lies in PET quantification methods. An accurate response assessment using PET requires an accurate PET quantification method.

1.6. PET quantification methods

Different PET quantification methods have been proposed for the assessment of tumour response to treatment [70]. These methods in general can be divided into two broad categories of single time-point, static imaging or multiple time-point, dynamic imaging approaches (figure 1.5). In general a full dynamic approach, such as the Patlak method [71] is considered to be more accurate as an estimate of glucose metabolism than the static approach. However, this comes at the cost of a longer acquisition time (~60 minutes) and a number of arterial blood samples, which makes it clinically not practical. Puncturing an artery is generally more uncomfortable than puncture of a vein, because arteries are deeper than veins, have thicker walls, and have more nerves. Patients often find arterial sticks uncomfortable and pain may continue for some time
even after the needle is withdrawn. The risks of excessive bleeding, bruising and infection are greater with arterial sampling than with venous sampling.

Steps of calculating Patlak can be explained as following. Time activity concentrations of blood \( C_p(t) \) and tumour \( A(t) \) are measured post FDG injection for 60 minutes. The data are then fitted into a two-compartment kinetic model that includes both free FDG, \( C_F(t) \), and metabolized FDG, \( C_M(t) \), as illustrated in figure 1.6(a). The model assumes irreversible kinetics by ignoring un-phosphorylation of metabolized FDG (i.e. \( K_4=0 \)). Assuming a steady state is reached for the \( C_F(t) \) compartment, Patlak fits the data graphically into a linear model to obtain the tumour metabolic uptake rate [72]:

\[
\frac{A(t)}{C_p(t)} = K_i \frac{1}{C_p(t)} \int_0^t C_p(u)du + V_0,
\]

here \( V_0 \) is the initial volume of distribution and \( K_i = (K_1K_3)/(K_2+K_3) \) is the tumour uptake time rate constant that is used for response assessment.

On the other hand, static PET quantification, usually expressed as standardized uptake value (SUV), requires only 5 minutes of data acquisition (for a single bed position) and no blood sampling. The SUV is calculated by normalizing the measured tissue activity to the injected FDG activity and the patient's weight. The use of SUV is much less demanding than the Patlak method which accounts for its widespread clinical use. In fact, all commercial PET scanners automatically calculate SUV images when scanning a patient leaving minimal data analysis workload for the research team. Despite its extensive use, SUV has been criticized for its assumptions and over-simplifications [73, 74]. The consensus recommendation from the national cancer institute (NCI) is the use of the Patlak method in Phase I studies [75].

Efforts have been made to make quantitative FDG uptake measurement more accurate than SUV while keeping it clinically practical. Simplified kinetic analysis - multiple time points
(SKA-M) is one of these approaches (figure 1.5). The SKA-M method is mathematically identical to Patlak in terms of the kinetic modeling. However, rather than multiple arterial sampling as in Patlak method, the SKA-M method uses a single venous blood sample to scale a population average input function (figure 1.6(b)). PET image acquisition starts at 30 minutes post-injection, instead of immediately after injection as per Patlak. The total image acquisition time is about 30 minutes, which lies somewhere between the SUV and Patlak methods in terms of complexity.

Most PET response monitoring studies in head and neck cancer reported to date have used SUV for PET quantification [76-79]. However, a previous study that used both SUV and dynamic PET, suggested that dynamic PET quantification may be of greater value than SUV for response assessment [80]. In that study, an initial correlation was found between SUV and dynamic PET quantification for lower values of SUV. However, the correlation quickly diminished for higher SUV values. Moreover, the SUV method resulted in a poorer association with survival than dynamic PET quantification, favouring dynamic PET to have a greater prognostic value than SUV.

The clinical trial at the OCC uses the SKA-M method for PET quantification. With this approach both SUV and dynamic PET quantifications are possible, which is an additional advantage of dynamic PET quantification other than those discussed above.

1.7. Problems with current PET quantification methods

An accurate tumour response assessment using PET needs an accurate PET quantification method. In the following sections different problems of the current PET quantification methods (both SUV and SKA-M) will be identified and the thesis objective and specific aims will be built upon addressing these problems.
1.7.1. Problem of ROI placement in PET quantification using SUV

Response assessment using static PET quantification (SUV) requires a region of interest (ROI) to quantify tumour uptake. One of the simplest and most common methods of quantifying tumour uptake is to use the single voxel containing the maximum SUV [61, 81] as the ROI. Unfortunately, this approach is highly sensitive to image noise [82, 83] which leads to uncertainties and poor reproducibility. As a more robust alternative, an average SUV within a small fixed size ROI has been recommended by PERCIST to reduce uncertainties in PET quantification using SUV [61]. With this approach, the fixed-size ROI is placed at the maximum uptake point in the pre-treatment study. For intra-treatment, the ROI could be placed either at the maximum uptake point or at the same anatomic location as the pre-treatment ROI. The latter choice has been recommended by the European organization for research and treatment of cancer (EORTC) [84].

The distribution of uptake within the tumour may change in response to therapy such that the maximum uptake point in the Intra-Tx study is found at an anatomically different location than it was prior to treatment. This is illustrated in figure 1.7 for a sample advanced head and neck cancer patient from the OCC trial. The change in uptake pattern is due to a change in tumour heterogeneity during the course of therapy. This has implication on tumour response assessment as the maximum uptake point that corresponds to the most metabolic active part of the tumour shows a change in location within the tumour resulting in two different response assessment that use small ROI to quantify tumour uptake.

Uncertainty in the placement of the ROI could significantly affect the accuracy of quantitative response assessment. Uncertainties in quantitative response assessment could have significant impact on treatment decisions and clinical outcome. The first research project of this thesis is
designed to address this issue with the specific aim of evaluating static PET quantification method using fixed size ROI for assessing tumour response.

**1.7.2. Problems of current PET quantification using blood sample based SKA-M**

Quantitative PET response assessment using SKA-M is a promising method since it provides a reasonable trade-off between accuracy and clinical practicality, as discussed above. However, since the SKA-M method relies on a single blood sample from the patient, it is reasonable to postulate that the accuracy of a quantitative PET response assessment using the SKA-M technique strongly depends on the accuracy of blood activity concentration measurements. In order to obtain an accurate blood activity concentration, accurate measurements of both blood activity and blood volume are essential. The accuracy in measuring the blood activity concentration may be compromised due to the small activity and volume of the blood sample. One of the specific aims of this thesis is to evaluate the accuracy of measuring blood activity concentration and to assess how this accuracy will translate to the accuracy of quantitative response assessment using SKA-M.

**1.7.3. Problem of loss of PET quantification due to partial volume error**

In order to address the problems of the current blood sample based SKA-M, one approach would be to take advantage of PET-CT registration and measure blood activity concentration directly from the acquired dynamic PET images. This approach is referred to as deriving an image-based input function [85]. In this approach if the heart is in the field of view (e.g. whole body imaging), a region of interest is normally placed at the left ventricle in the PET image and the arterial blood activity concentration is read directly from the PET images.
For head and neck cancer patients, since the heart is not in the field of view of the PET scanner, an image-based input function may be possible if the blood activity concentration is read from major blood vessels in the head and neck area such as the carotid artery and jugular vein. This procedure is described in figure 1.8. Here, the CT images are used to identify the carotid artery and jugular vein. The contours for these blood vessels are then transferred to the co-registered PET images so that the blood activity concentration can be directly measured. This is an attractive approach since it avoids the blood sample problems of SKA-M and it can establish a non-invasive SKA-M PET quantification technique. A more accurate estimate of the patient's own input function can be obtained with this approach without any loss in clinical practicality of the SKA-M method. However, the partial volume error in PET images needs to be addressed before this approach is feasible.

Partial volume error in PET quantification stems from the fact that PET images are inherently limited in spatial resolution due to different PET image formation processes. Figure 1.9 explains this effect in relation to our attempt to determine the blood activity concentration from PET images of the vessels. For a typical 10 mm diameter vessel in the head and neck area there is 40% underestimation in estimating blood vessel uptake if no partial volume correction (PVC) is applied. For an accurate response assessment, 40% error is significant in PET quantification that could have substantial impact on treatment decisions and clinical outcome.

Many PVC methods have been proposed to recover the loss of accuracy in PET quantification due to limited PET spatial resolution [86]. The geometric transfer matrix (GTM) method [87] unarguably remains the most common PVC method [88] and it is often considered as the reference PVC method [89, 90]. In this method regions from the CT images are used as *a priori* information. In our case, for example, major blood vessels in the head and neck area may be used as regions for PVC. Although the GTM method is widely used, it has two key limitations. The
first limitation is its lack of robustness to mis-registration. Mis-registration between CT and PET images may be caused due to patient motion between the two scans. This problem is illustrated in figure 1.10. Mis-registration has been regarded as the major source of error [89] with the strongest impact on the accuracy of PET quantification [90]. The second limitation is the noise amplification problem. As mentioned in section 1.4.5, PET images are inherently noisy and the implementation of the GTM method propagates the noise leading to loss of precision in PET quantification.

In order to address these two limitations, PVC methods are needed that are robust to PET-CT mis-registration and image noise. One of the specific aims of this thesis is to address this issue by developing a novel partial volume correction technique.

1.8. Quantitative objective for PET tumour response assessment

An accurate PET quantification is a key requirement for an accurate PET-based tumour response assessment. A quantitative objective target for the accuracy is estimated by noting that inaccuracies in PET quantification may be due to either physical or biological variations. The physical variation may be evaluated by noting the statistical variation in phantom measurements without including patient scans. For example, variation due to PET calibration in a phantom has been reported to be on the order of 5% [91]. Studies performing test-retest PET uptake measurements of the same tumour within 3 days (while receiving no therapy) reported intra-patient biological variation to be 10% [92, 93]. Studies performing early PET response assessment, using PET uptake measurements of the tumour, reported inter-patient biological variations (within both responding and non-responding populations) to be ~15-20% [59, 94, 95].

A reasonable PET quantification metric should have accuracy less than inter-patient biological variation to be useful for response assessment. In this thesis, a quantitative objective target for the
accuracy of PET quantification is set to be 10%. Note that this accuracy target is a combination of both intra-patient biological variation and physical variability due to scanner effects. This choice may be justified by noting that below 10% is not practically achievable due to intra-patient biological variation as well as PET scanner calibration limitation. On the other hand, above 10% would in some cases significantly restrict ability to determine response. The choice of 10% is potentially attainable but it is below the inter-patient biological variations in most cases.

1.9. Thesis outline

This thesis is designed to address the problems of current PET quantification methods that were outlined in section 1.7. The objective of this thesis is to evaluate the accuracy of current PET quantification methods for response assessment in advanced head and neck cancer patients and to develop a novel partial volume correction technique to improve PET quantification. It is expected that the accuracy of PET quantification can be improved by development of an image-based simplified dynamic PET quantification method with partial volume correction.

Chapter 2 addresses the problem outlined in section 1.7.1. This chapter aims to evaluate static PET quantification method using fixed size ROI for assessing tumour response in advanced head and neck cancer patients. Patient data from the OCC trial have been used in chapter 2.

Chapter 3 investigates the problem outlined in section 1.7.2 regarding the current blood sample based SKA-M method. The aim of this chapter is to evaluate the accuracy of blood activity measurements and its effect on the accuracy of quantitative response assessment using the current simplified dynamic PET method (blood sample based SKA-M). Advanced head and neck cancer patient data from the current OCC trial have been used in this chapter. The results of chapter 3 support a need to develop an image-based simplified dynamic PET quantification method with partial volume correction.
Chapters 4 and 5 address the problem described in section 1.7.3 with regard to the loss of PET quantification due to partial volume error. Chapter 4 develops a novel and robust PET partial volume correction technique and chapter 5 extends this technique to the iterative PET reconstruction algorithms that are commonly used in the clinic.

Finally, chapter 6 summarizes the key results and implements the novel partial volume correction technique in an advanced head and neck cancer patient from the OCC trial. This chapter lays out the feasibility of an image-based simplified dynamic PET quantification method and identifies future research opportunities.
### Table 1.1 RECIST criteria for tumour response assessment.

<table>
<thead>
<tr>
<th>Response</th>
<th>RECIST change in sum of longest diameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>Disappearance of all target lesions. Any involved lymph nodes (whether target or non-target) must have reduction in short axis to &lt;10 mm.</td>
</tr>
<tr>
<td>PR</td>
<td>At least a 30% decrease in the sum of diameters of target lesions</td>
</tr>
<tr>
<td>PD</td>
<td>At least a 20% increase in the sum of diameters of target lesions or appearance of one or more new lesions.</td>
</tr>
<tr>
<td>SD</td>
<td>Neither sufficient shrinkage to qualify as PR nor sufficient increase to qualify as PD.</td>
</tr>
</tbody>
</table>

CR = complete response, PR = partial response, PD = progressive disease, SD = stable disease
Figure 1.1. **Functional versus anatomical approach for response assessment.** Functional changes in the tumour usually precede anatomical changes. This implies that functional imaging modalities (such as PET) can potentially predict the tumour response evaluated in a later time point by anatomical imaging modalities (such as CT).
Figure 1.2. Basics of FDG PET imaging. PET image formation processes consist of injecting FDG into the patient (a). During an uptake phase, the FDG distributes throughout the patient’s body and accumulates in areas that would normally use glucose. The F-18 in FDG is radioactive and decays by emitting a positron (b). The positron travels for a short distance, then it finds an electron and they annihilate to produce two photons emitted in opposite directions. PET raw data (or sinogram) are acquired from all PET coincident detectors at different angles (c). The sinogram is reconstructed to create a PET image (d) that presents the spatial distribution of FDG in the patient. FDG = fluorodeoxyglucose, $^{18}$F = fluorine-18, $e^+$ = positron, $e^-$ = electron, $\gamma$ = gamma ray.
Figure 1.3. Management of advanced head and neck cancer patients. A simplified chart illustrating the current management of advanced head and neck cancer patients.
Figure 1.4. Two PET scans to assess response in an advanced HNC patient. The Pre-Tx is the baseline scan performed before the treatment starts. The Intra-Tx scan is performed after two weeks of treatment. The PET images are superimposed on corresponding CT images and the primary tumour is identified by the red arrows. The pattern of tumour uptake has changed and the uptake in the tumour is relatively reduced. HNC = head and neck cancer, Pre-Tx = pre-treatment, Intra-Tx = intra-treatment.
Figure 1.5. PET quantification methods. Comparison of SUV, Patlak, and SKA-M analysis. SUV is a simple static acquisition, whereas Patlak and SKA-M are dynamic acquisitions. The SKA-M method is a promising method since it is clinically practical and yet its accuracy is close to that of that of the Patlak method.
Figure 1.6. Kinetic modeling. (a) Kinetic modeling using Patlak (and SKA-M) assumes two compartment kinetics, i.e. $C_F(t)$ and $C_M(t)$. In Patlak method, the input function $C_p(t)$ is required for 60 minutes post injection. (b) In SKA-M method, a single venous blood sample from the patient is used to scale a population-averaged time activity curve and to obtain the input function for the patient. $K_1$ = rate constant of FDG transport into the cell, $K_2$ = rate constant of FDG transport out of cell, $K_3$ = rate constant of phosphorylation of FDG, $K_4$ = rate constant of un-phosphorylation, SKA-M = simplified kinetic analysis-multiple time points.
Figure 1.7. Problem of ROI placement in PET quantification using SUV.
PET/CT images acquired Pre-Tx (a) and Intra-Tx (b) for an advanced head and neck cancer patient are illustrated. Two fixed size ROIs (15 mm in diameter) are centered at the maximum uptake points on both Pre-Tx (green) and Intra-Tx (red) images denoted by "M". An additional ROI is placed on the Intra-Tx image (blue) in a position judged to correspond to the same anatomical location as the ROI in the Pre-Tx. The uptake profiles along the black lines connecting the two ROIs are plotted in (c). The distribution of uptake within the tumour has changed during the therapy such that the maximum uptake point along the profile in Pre-Tx corresponds to a local minimum uptake point in the Intra-Tx. The maximum uptake point along the profile is now in a different location of the tumour. The choice of ROI placement affects quantitative tumour response assessment which could have significant impact on treatment decisions and clinical outcome. ROI = region of interest, SUV = standardized uptake value, HNC = head and neck cancer, Pre-Tx = pre-treatment, Intra-Tx = intra-treatment, SUV = standardized uptake value.
Figure 1.8. Image based SKA-M. With the availability of PET/CT (rather than PET alone), the current blood sample based SKA-M can be improved by directly measuring the patient's own blood activity from PET images. First, major blood vessels in the head and neck region in the CT image must be identified. Here, the carotid artery (red arrow) and jugular vein (blue arrow) are contoured in the CT image (a). These contours are then transferred to the PET image (b). An image based SKA-M is possible without a need for blood sample if partial volume correction is applied.
Figure 1.9. Partial volume error in PET imaging. The loss of resolution in PET imaging is characterized by point spread function (PSF). An ideal image of an F-18 point source should be similar to the image in (a). However, the real PET image is a blurry version of it as illustrated in (b) which is referred to as the PSF. Profiles of the ideal point image (black line) and the PSF (red curve) are plotted in (c). The PSF is characterized in terms of its full width half maximum (FWHM) which is about 7 mm. An ideal vessel image with 10 mm diameter is illustrated in (d) and a real PET image of it in (e). The profiles of the ideal vessel (black line) and the real vessel image (red curve) are plotted in (f). Partial volume error is quantified in terms of RC, i.e. ratio of measured to true uptake. For a typical vessel in the head and neck area (10 mm diameter), RC=60%, i.e. 40% underestimation in vessel uptake due to partial volume error. PSF = point spread function, FWHM = full width half maximum, RC = recovery coefficient.
Figure 1.10. The problem of image registration in current PVC methods. The major blood vessels are contoured for an advanced head and neck cancer patient on CT (cyan). If the patient moves between the PET and CT scans, these contours are sampling different parts of the PET image (red) that do not correspond to the vessel location on CT.
Chapter 2:

Evaluation of static PET quantification using fixed size ROI for tumour response assessment

This chapter investigates the efficacy of static PET quantification methods for response assessment in advanced head and neck cancer. In particular, it evaluates static PET quantification methods employing fixed size regions of interest (ROI) on standardized uptake value (SUV) images. Static SUV images of advanced head and neck cancer patients from an OCC-based clinical trial were used in this chapter. The degree of uncertainty for assessing treatment response using conventional static PET quantification methods was investigated and the potential clinical significance in terms of treatment management, survival, and morbidity are estimated.

The work presented in this chapter has been published in:

2.1. Introduction

As a powerful molecular imaging tool, positron emission tomography (PET) is increasingly being used for early assessment of tumour response to therapy [96-98]. Typically two sequential PET studies are performed and the tumour standardized uptake value (SUV) in the pre-treatment (Pre-Tx) study is compared to that of the intra-treatment (Intra-Tx) study.

Response assessment using SUVs requires the selection of either a representative tumour voxel or a region of interest (ROI) for quantification. One of the simplest and most common methods of quantifying tumour uptake is to use the single voxel containing the maximum SUV (SUVmax) [61, 81]. Unfortunately, SUVmax values are highly sensitive to image noise and voxel size [82, 83] which leads to uncertainties in quantitative response assessment. Moreover, Krak, et al.[83] reported that SUVmax has poor reproducibility compared to estimates of SUV made using ROI methods. As a more robust alternative, an average SUV within a small fixed size ROI has been recommended to provide adequate statistical quality in SUV measurements and to reduce uncertainties in quantitative response assessment [61].

Table 2.1 lists representative studies [51-60, 80, 99] that have used the fixed-size ROI method for early tumour response assessment. The Pre-Tx ROI is usually centred on the SUVmax voxel. However, there are two distinct approaches to the placement of the Intra-Tx ROI. Some studies have centred the Intra-Tx ROI on the SUVmax voxel (ROIpeak), whereas others have placed it at the same location as it was in the Pre-Tx image using anatomical landmarks (ROIsame). The distribution of uptake within the tumour may change in response to therapy such that the maximum uptake point in the Intra-Tx study is found at an anatomically different location than it was prior to treatment. This is illustrated in figures 2.1 and 2.2 for a sample advanced head and neck cancer (HNC) patient. Figure 2.2 illustrates two quantitative response assessments based on the two different choices of Intra-Tx ROI placement.
Using the ROIsame method is reasonable if the goal is to evaluate the change in uptake in the same area of the tumour. This method has been recommended by the European organization for research and treatment of cancer (EORTC) [84]. However, unlike ROIpeak, tumour response measured by ROIsame is prone to uncertainty due to the difficulty in positioning a ROI in the Intra-Tx scan in the exact anatomic location as it was in the Pre-Tx scan. Geometric changes of both tumour and normal tissue may occur during the therapy making it difficult to place a ROI at exactly the same location as it was in the Pre-Tx scan using anatomical landmarks. Figure 2.3 illustrates PET/CT images of a sample HNC patient illustrating the magnitude of typical geometric changes in terms of volume losses and shifts.

Uncertainty in the placement of the Intra-Tx ROI could significantly affect the accuracy of quantitative response assessment. Uncertainties in quantitative response assessment could have significant impact on treatment decisions and clinical outcome. Consequently, we investigated the effects of fixed-size ROI placement on quantitative response assessment. The purpose of this study was twofold: (1) to evaluate quantitative response assessment when Intra-Tx PET images are measured using the ROIpeak and ROIsame methods; (2) to quantify the geometric changes of both tumour and normal tissue and their impact on quantitative response assessment using the ROIsame method.

2.2. Materials and methods

2.2.1. Design of the study:

Two independent populations (A and B) were used. Population A consisted of 15 patients with a total of 38 gross tumour volumes (GTV) identified by experienced radiation oncologists. Population A was used to compare two quantitative tumour response assessments based on using the ROIpeak and ROIsame methods. Population B consisted of 10 patients with a total of 33
GTVs identified by experienced radiation oncologists and was used to quantify geometric changes of both tumour and normal tissues during therapy. The impact of these geometric changes on quantitative tumour response assessment was evaluated in population A. Both populations A and B were part of a clinical trial at Sunnybrook Health Sciences Centre (Toronto, Canada) to assess tumour response in patients with advanced HNC. Population B consisted of patients entered in the pilot study which proceeded the main trial, while population A consisted of patients entered in the clinical trial itself.

While populations A and B were very similar, there were some slight differences, primarily in the CT-voxel size used and the average time between the Pre-Tx and Intra-Tx scans. All patients in both groups had locally advanced HNC (stage III or IV) and underwent 6.6 weeks of radical radiotherapy with concurrent chemotherapy. Patients received intensity modulated radiation therapy (IMRT) of 70Gy in 33 fractions to all GTVs for both primary (GTVp) and involved lymph nodes (GTVn). All patients also received concurrent bolus platinum chemotherapy as tolerated by intravenously injecting 100 mg/m² Cisplatin on days 1, 22, and 43. Patients underwent two sequential ¹⁸FDG-PET/CT scans, one Pre-Tx and one Intra-Tx, both supine in the same position using a thermo-plastic radiotherapy immobilization mask. One 18 cm axial field of view (FOV) that covered the head and neck area was used. The PET/CT scanner was the GEMINI System (Philips medical system, Cleveland, Ohio). Prior to the PET/CT scans, patients were injected with 5 MBq of FDG per kg. Patients heavier than 75 kg were injected with a fixed dose of 370 MBq of FDG. PET images were reconstructed using a 3-dimensional row-action maximum likelihood algorithm (3D-RAMLA) and corrected for attenuation using CT. The PET images are already registered to the CT images as the scanners share the same bed. In order to register the Intra-Tx CT to the Pre-Tx CT images, a Chamfer matching algorithm [100] based on bony structures was implemented in house using the interactive data language (IDL) Ver. 6.4
(Research systems inc., Boulder, CO). The algorithm used 3D rigid body with rotation and translation but no scaling. Note that a rigid body (and not nonlinear) registration technique was chosen to obtain the geometric changes of the tumours (and other tissues). This choice was due to the fact that bone anatomy based registration was needed since bones are least likely to change due to therapy. Once the Intra-Tx bones are registered to Pre-Tx, the COM shift of tumour (and other tissues) can readily be calculated. To calculate changes in the volumes of tumour (and other tissues) no registration is required since they are contoured separately in both Pre-Tx and Intra-Tx. A manual registration could also be useful in cases where a patient is very loose to the immobilization mask and the Intra-Tx bony anatomy does not match Pre-Tx bones due to patient movement within the mask. In such cases, we still used chamfer matching but limited axial slices to those close to the C2 vertebral body. An IDL program was developed in house to simultaneously display the registered Pre-Tx and Intra-Tx PET/CT images, to contour ROIs, and to read SUV values. PET images were interpolated to match the voxel sizes of CT images. All statistical analyses were performed using the public domain package "R" (www.r-project.org).

2.2.2. Population A (study patients)

Pre-Tx FDG PET/CT scans were performed 14 ± 4 days (range, 8-22) prior to the start of the treatment. Intra-Tx FDG PET/CT scans were performed 16 ± 2 days (range, 11-20) after the first treatment day. The CT voxel size was 0.59 x 0.59 x 1.60 mm³ and the CT FOV was 300 x 300 x 210 mm³ in lateral, anterior-posterior, and superior-inferior directions respectively. The PET voxel size was 2 x 2 x 2 mm³ and the PET FOV was 576 x 576 x 180 mm³ in lateral, anterior-posterior, and superior-inferior directions respectively. PET images were acquired 50 minutes post-injection for 2.5 minutes. The Pre-Tx and Intra-Tx PET post-injection acquisition times were matched within 5 minutes.
The SUVs were normalized to the patients' body weight. For each GTV, ROIpeak (a circular ROI of 15 mm diameter) was placed on a single transaxial slice centered at the maximum FDG uptake point in both Pre-Tx and Intra-Tx images. For each GTV, ROIsame (a circular ROI of 15 mm diameter) was also placed on a single transaxial slice at the location of the Intra-Tx image that corresponded to the same physical location as the Pre-Tx max-point ROI. A dual-board certified, nuclear medicine/radiology physician positioned ROIsame based on anatomical landmarks. Thus, each GTV had two Intra-Tx ROIs. The distance between the centers of these two ROIs was measured in 3D geometry.

On the same transaxial slice where ROIsame was located, the Intra-Tx GTV size was measured by averaging the anterior-posterior and lateral extents of an oncologist drawn GTV. In order to reduce errors in FDG uptake from partial volume effects, only Intra-Tx GTVs larger than 15 mm were subsequently analyzed, reducing the total number of GTVs available for analysis from 38 to 26.

Tumour response assessments were obtained using two different methods, called $\Delta SUV_{peak}$ and $\Delta SUV_{same}$, by calculating the relative change in tumour uptake:

$$\Delta SUV_{peak} = 1 - \frac{\text{Intra}_T \text{X} \ SUV_{peak}}{\text{Pre}_T \text{X} \ SUV_{peak}}, \quad \Delta SUV_{same} = 1 - \frac{\text{Intra}_T \text{X} \ SUV_{same}}{\text{Pre}_T \text{X} \ SUV_{peak}}$$

Where $SUV_{peak}$ is the mean SUV within ROIpeak in either Pre-Tx or Intra-Tx PET images. $SUV_{same}$ is the mean SUV within the ROIsame in the Intra-Tx PET image. A positive value for $\Delta SUV_{peak}$ or $\Delta SUV_{same}$ indicates a decrease in uptake and a negative value indicates an increase in uptake.

In order to determine how uncertainties in positioning ROIsame due to geometric changes may impact $\Delta SUV_{same}$ values, the original ROIsame was systematically shifted in a 3D grid geometry up to 25 mm in three orthogonal directions. The sampling spaces of the grid were 1.17, 1.17, and
1.60 mm in the lateral, anterior-posterior, and superior-inferior directions respectively. For each point in the grid $SUV_{same}$ was determined. This data set was sorted based on the distance of the shifted ROIsame to the original ROIsame. For each GTV, $SUV_{same}$ was calculated and plotted as a function of this distance (i.e., positioning error). Each plot was normalized to the $SUV_{same}$ of the original ROIsame. This normalization makes the y-axes represented also normalized $(1 - \Delta SUV_{same})$. Plots of normalized $SUV_{same}$ were averaged over the 16 GTVs with Intra-Tx size smaller than 30 mm or the 10 GTVs with Intra-Tx size larger than 30 mm. The arbitrary 30 mm threshold (twice the ROI size) was chosen to emphasize the effects due primarily to tumour uptake heterogeneity versus the effects due primarily to the partial volume effect. Tumour uptake heterogeneity was expected to have greater impact in large GTVs (>30 mm) and the partial volume effect was expected to have greater impact in small GTVs (<30 mm).

2.2.3. Population B (pilot patients)

Pre-Tx FDG PET/CT scans were performed 17 ± 5 days (range, 13-28) prior to the start of the treatment. Intra-Tx FDG PET/CT scans were performed 33 ± 4 days (range, 28-40) after the first treatment day. The CT voxel size was 1.17 x 1.17 x 6.5 mm$^3$ and the CT FOV was 600 x 600 x 208 mm$^3$ in the lateral, anterior-posterior, and superior-inferior directions respectively. The PET voxel size was 2 x 2 x 2 mm$^3$ and the PET FOV was 576 x 576 x 180 mm$^3$ in the lateral, anterior-posterior, and superior-inferior directions respectively.

GTVs were contoured manually by an oncologist experienced in treatment of HNC. All the GTVs were contoured on CT images guided by co-registered PET images. Non-co-registered diagnostic MR images were available to aid contouring in all patients except one where no MRI
study was performed. Radiology reports on both PET/CT and MRI studies were also used to aid in contouring.

Geometric changes of the GTVs and normal tissues during treatment were thought to be potentially important in influencing the accuracy of placement of a ROI for quantitative tumour response assessment. In addition to GTVs, geometric changes of some normal tissues were also quantified. While the geometric shifts in tumours, not normal tissues, was of primary interest, the uncertainty in estimating geometric shifts in GTVs was greater than the uncertainty in estimating shifts in other structures, simply due to the difficulty in accurately delineating the GTV after treatment. Thus, the geometric shifts in normal tissues were used as surrogate measures of possible shifts in GTVs. Ten normal tissues were contoured on both Pre-Tx and Intra-Tx CT images for each patient. These normal tissues included the C2 vertebral body, mandible, hyoid, spinal cord, right and left sternocleidomastoid muscle, right and left parotid gland, and right and left submandibular gland. All normal tissues were contoured using consistent window and level settings under the guidance of an experienced oncologist. The most inferior extent for contouring the spinal cord and the sternocleidomastoid muscles was the most superior aspect of the apex of the lung. The most superior extent of the spinal cord was chosen to correspond to the most superior extent of the C2 vertebral body. Mandible and parotid contours were excluded from one patient since the scan did not include the entire organs in the superior direction.

Using both Pre-Tx and Intra-Tx contours for normal tissues and GTVs, an IDL program was developed in house to quantify the geometric changes by calculating:

(1) Percentage volume changes, i.e. Intra-Tx volume relative to Pre-Tx volume, and

(2) Shift of center of mass (COM), i.e. Intra-Tx COM relative to Pre-Tx COM. The shifts were calculated as a shift vector in a 3D geometry and the reported values are the absolute values of these vectors.
2.3. Results

Patient characteristics for both populations are listed in table 2.2.

2.3.1. Population A

The mean Intra-Tx GTV size (i.e., average of the anterior-posterior and lateral extents) was 25.7 ± 8.9 mm (range, 15.1-46.5). A histogram of the distances between the centers of the two Intra-Tx ROIs for each GTV is illustrated in figure 2.4. This histogram demonstrates that the Intra-Tx maximum uptake point does not normally correspond to the same physical location as that for Pre-Tx. The median distance between the centers of the two Intra-Tx ROIs was 7.4 mm. The two Intra-Tx ROIs were on the same transaxial slice in only 8% of the cases (in 2 out of 26 GTVs).

Figure 2.5A illustrates a scatter plot comparing quantitative tumour response assessments using the ROIpeak and ROIsame approaches. A high two-sided Pearson correlation coefficient was found between $\Delta{SUV}_{peak}$ and $\Delta{SUV}_{same}$ ($r = 0.93, p = 7e-12$) for all GTVs. Similarly, the $r$ value between Intra-Tx $SUV_{peak}$ and $SUV_{same}$ was 0.92, $p = 5e-11$ for all GTVs.

As expected, Intra-Tx $SUV_{peak}$ had a higher value than $SUV_{same}$ for most GTVs, resulting in a lower value for $\Delta{SUV}_{peak}$ compared to $\Delta{SUV}_{same}$ as seen in figure 2.5A. On average, $SUV_{peak}$ was 13.4% higher than $SUV_{same}$ (range -14% to 38%) and $\Delta{SUV}_{peak}$ was 7.9% lower than $\Delta{SUV}_{same}$ (range -5% to 36%). One unusual case, identified by the oblique arrow in figure 2.5A, is an example where the $\Delta{SUV}_{peak}$ was 5.3% higher than $\Delta{SUV}_{same}$. In this case, the ROIpeak region placed centred on the peak voxel in the Intra-Tx scan actually had a lower average uptake than the ROIsame region. GTVs in figure 2.5 are coded for primary versus nodal
mass as well as for large (>30 mm) versus small (<30 mm) GTVs. No statistically significance difference was found between $\Delta SUV_{peak}$ and $\Delta SUV_{same}$ on the basis of GTV size (large vs. small) or type (primary vs. node).

Figure 2.5B demonstrates classification of individual tumours based on the PET response criteria in solid tumours (PERCIST) [61] using either $\Delta SUV_{peak}$ or $\Delta SUV_{same}$. The PERCIST thresholds of ± 30% were applied to classify individual tumours into three categories of partial response, stable disease, and progressive disease. In 19% (5 out of 26) of the tumours this resulted in ambiguous tumour classification depending on the ROI method as indicated by the arrows.

Figure 2.6A illustrates an example plot for a tumour, demonstrating how uncertainties in positioning ROIsame may impact $\Delta SUV_{same}$ values. This plot indicates that positioning the ROIsame a few millimeters away may decrease or increase $\Delta SUV_{same}$ depending on if ROIsame is moving towards the maximum uptake point or is moving away from it. However, by moving a few centimeters away, the points eventually start to drop since the ROIsame is sampling the background normal tissue uptake. Individual plots such as figure 2.6A were averaged for all GTVs. The results are plotted as two curves in figure 2.6B based on Intra-Tx GTV size larger or smaller than 30 mm. A statistically significant difference between the two curves at the 95% confidence level was found to be between 5.4 mm and 16.2 mm

2.3.2. Population B

A total of 97 normal tissue regions were contoured in 10 patients in both Pre-Tx and Intra-Tx. Figure 2.7 demonstrates geometric changes due to therapy characterized in terms of percentage volume changes and COM shifts. Figure 2.7A illustrates the percentage volume changes for all
GTVs and normal tissues. Negative volume changes indicate a loss of volume during therapy. Both GTVp and GTVn demonstrated significant volume losses with median values of 76.1% and 60.1%, respectively. The median volume loss for all GTVs was 67.2% (range, 8.4-96.9%).

For normal tissues, significant volume losses were only found for the salivary glands. Median volume losses were 28.1% (range, 7.3-45.6%) for all parotid glands and 31.0% (range, 13.3-48.7%) for all submandibular glands. Other soft tissues (i.e. sternocleidomastoid muscles and spinal cord) and bones did not indicate significant volume losses.

Figure 2.7B illustrates COM shifts in GTVs and normal tissues. The median shift for all GTVs was 5.9 mm and the 95% CI range was 4.4-7.6 mm. The C2 vertebral body had the smallest shift with a median of 1.0 mm. Right and Left parotid glands had median shift values of 3.7 mm and 2.8 mm respectively. The median shifts in medial directions for right and left parotid glands were 1.4 mm and 2.5 mm respectively.

2.4. Discussion

2.4.1. Effect of geometric changes

Geometric changes during therapy can be expected to influence the accuracy with which an expert can place the tumour ROI and thus could affect tumour response assessment using the ROIsame approach. Our results in figure 2.7 are similar to those reported earlier [101, 102]. We found that the 95% CI for GTV COM shift to be between 4.4 and 7.6 mm. This range may represent the upper range of uncertainty in placing the Intra-Tx tumour ROI at the same location as the Pre-Tx tumour ROI. However in practice, attempts are made to correct for the geometric changes to some extent using anatomical landmarks. Moreover, in our study, population A had earlier Intra-Tx scans than population B. Due to these two factors, we expect that the uncertainties in placing the Intra-Tx tumour ROI to have a smaller range than the 95% CI shift,
possibly in the 0-5 mm range. Based on figure 2.6B, the impact of this uncertainty can be expected to be less than 10% on the measure of tumour response.

2.4.2. Effect of ROI method

The placement of the fixed size ROI could have a significant effect on PET quantification for tumour response assessment. In this study, we found that $\Delta S UV_{peak}$ was 7.9% lower than $\Delta S UV_{same}$ on average and difference was up to 36%. This degree of difference leads to different response assessment using PERCIST [61], resulting in overall 19% (5 out of 26) ambiguous tumour response assessment (figure 2.5B). This finding underscores the need for an optimized PET quantification method in individual patients using a consistent and standard ROI for an accurate response assessment. A small fixed size ROI placed on a single slice is a simplistic approach to sample tumour uptake. Figure 2.2 demonstrates that the change in heterogeneity within the tumour due to treatment could be significant. This indicates the disadvantage of PET quantification for response assessment using a small fixed size ROI [103].

With $\Delta S UV_{peak}$ one may risk underestimating response to treatment compare to $\Delta S UV_{same}$. This difference directly results from the fact that $S UV_{peak}$ was on average 13.4% higher than $S UV_{same}$ since it was centered at the maximum uptake point. Occasionally, $S UV_{peak}$ may be smaller than $S UV_{same}$. The outlier in figure 2.5A corresponds to a situation where the central pixel of ROIpeak has a high uptake but its surrounding pixels have a lower uptake than the pixels within ROIsame. Noisy PET images or high intra-tumour uptake heterogeneity might cause such a situation.
Considering the typical response thresholds that have been used to separate responding patients from non-responding patients (last column in table 2.1), the difference of 7.9% (and up to 36%) between the two ROI methods could be clinically significant.

Many recent studies on early tumour response assessment have used the single-voxel based, SUVmax method, while the new recommendation favours a fixed size ROI as a more robust alternative to reduce uncertainties due to noise [61]. The placement of the fixed size ROI in Intra-Tx, whether ROIpeak or ROIsame as per EORTC recommendations [84], could lead to significant uncertainties in response assessment. Thus, more studies are required to determine if either of these simple, fixed size ROI approaches are useful in assessing treatment response.

We found that the two ROI methods gave rise to highly correlated ($r = 0.93$) response assessments (figure 2.5A). This high correlation is a direct result of high correlation ($r = 0.92$) between the SUV values of the two Intra-Tx ROI methods. This suggests that the higher uptake in ROIpeak also means potentially higher uptake in ROIsame. ROIsame in general was sampling a different part of the tumour at some distance away from ROIpeak (figure 2.4). Part (but not all) of this correlation can also be explained by overlap of the two Intra-Tx ROIs (both 15 mm in diameter). In our patients, the two ROIs were in the same slice in only 8% of the GTVs. It is unsurprising that the pattern of tumour uptake could be considerably changed in response to therapy. Even without therapy, the pattern of uptake over time may alter as the tumour grows.

The EORTC [84] recommends placing the Intra-Tx ROI at the same anatomical location as the Pre-Tx ROI in order to sample the same area. This is a reasonable approach, that is, to evaluate the same location before and after some therapeutic intervention. It does not appear as intuitively reasonable to use the ROIpeak approach, which could mean comparing two anatomically distinct parts of the tumour pre- and intra-treatment. However, in a limited number of patients, we found that the two ROI methods were highly correlated ($r = 0.93$). This suggests that the two response
assessment methods would likely have a similar accuracy in terms of differentiating responders versus non-responders, although with different optimal response threshold values. In order to determine if a simple fixed ROI-based method has true utility for assessing response, substantial clinical trial data including patient outcomes is required. Such trial data could also be used to establish if there are threshold levels for ROI-based techniques that could reliably separate responders from non-responders for each disease site and given treatment type.

2.4.3. Clinical significance

The clinical significance of choices of ROI methods can be estimated following similar approach to that described in section 1.5.2 (chapter 1). First, we assume that ROIpeak provides the "true" classification between responders and non-responders using the PERCIST [61] threshold criteria, i.e. 30% decrease in uptake. With this assumption, approximately half of the patients in figure 2.5B with ROIpeak are responders and the other half are non-responders which is in agreement with previous data [68]. When ROIpeak is considered “truth”, the ROIsame method would mis-classify 15% (4 out of 26) of the patients as responders. For these patients the ineffective chemo-radiation would be completed followed by the salvage surgery. The clinical significance of this mis-classification could affect (a) patient survival, and (b) morbidity.

As per section 1.5.2, the theoretical estimates of patient survival using an accurate response assessment method (assumed ROIpeak here) is (0.5x100%) + (0.5x50%) = 75%. We assume that salvage surgery after completion of the ineffective chemo-radiation has only 25% survival rate (compared to 50% survival rate if early surgery is offered without completing of chemo-radiation). Thus with 15% error in identifying non-responders, the response assessment using ROIsame will have a theoretical survival estimates of [0.5x100%] + ((0.5-0.15)x50% +
0.15x25% = 71%. Here, the first square bracket is for responders and the second one is for non-responders.

In terms of morbidities, mis-classification of 15% of the patients as responders will translate to unnecessary chemo-radiation treatment for these patients. Therefore, rather than 50% of the patients (only responders) undergoing chemo-radiation, 65% will be treated with chemo-radiation. Thus, xerostomia, acute mucositis, and dysphagia which theoretically occur with “assumed ideal” ROIpeak method in 17%, 25% and, 23% of the patients respectively, will occur with ROIsame in 21%, 33%, and 30% of the patients respectively.

Conversely, one may assume that ROIsame provides the "true" response assessment. In this case, the ROIpeak method would mis-classify 15% of the patients as non-responders, suggesting that the effective chemo-radiation treatment should be stopped and early surgery should be offered to these patients. Again this mis-classification will have consequences on (a) patient survival and (b) morbidity.

We use the same assumptions as above with the additional assumption that responding patients to chemo-radiation will have a 50% survival rate if early surgery is offered to them. Thus, the theoretical survival estimates using ROIpeak will be [(0.5-0.15)x100% + 0.15x50%] + [0.5x50%] = 68%. Again, here the first and the second square brackets are for responders and non-responders to chemo-radiation treatment respectively.

Based on inaccurate response assessment using ROIpeak, 15% of the patients were offered early surgery without completion of chemo-radiation. Thus, only 35% of the patients will be treated with chemo-radiation (rather than 50% for an accurate response assessment). This will reduce the associated morbidities of xerostomia, acute mucositis, and dysphagia to 12%, 18%, and 16% respectively. Note that additional morbidities due to surgery exist that were not included.
in our discussion. However, accurate response assessment would identify only those patients who need surgery, thus would minimize overall surgical morbidity.

Table 6.1 summarizes the clinical impact of ROI method on response assessment in terms of treatment management, survival, and morbidity. As noted in this table the ROIpeak method has better morbidity at the expense of lower survival compared to ROIsame. This illustrates the relative effects of calling patients “responders” erroneously or calling patients “non-responders” erroneously.

2.5. Conclusion

PET quantification for assessing treatment response using a fixed size ROI is sensitive to the placement of the ROI within the tumour. The difference between the current recommendations favoring ROIpeak (over ROImax) and earlier recommendations using ROIsame could be substantial (36%) resulting in ambiguous treatment response assessment (19%). Methods making use of such small ROIs have the advantage of being relatively simple to implement while still providing improved statistical properties versus the SUVmax single voxel method. However, simplicity is not always an advantage and the use of a small fixed size ROI for tumour response assessment should be approached with caution in heterogeneous tumours. Clinical trials are necessary to compare the efficacy of a fixed size ROI over ROImax and establish a reliable threshold in a given cancer site.
<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Site</th>
<th>Pre-Tx ROI</th>
<th>Intra-Tx ROI</th>
<th>Res. Thr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schelling (2000)</td>
<td>22</td>
<td>breast</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>55%</td>
</tr>
<tr>
<td>Weber (2003)</td>
<td>57</td>
<td>lung</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>20%</td>
</tr>
<tr>
<td>Avril (2005)</td>
<td>33</td>
<td>ovarian</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>20%</td>
</tr>
<tr>
<td>Brun (2002)</td>
<td>47</td>
<td>head and neck</td>
<td>Fixed-size, square (4 or 9 pixels) at max</td>
<td>Fixed-size, square (4 or 9 pixels) at max</td>
<td>Median</td>
</tr>
<tr>
<td>Rousseau (2006)</td>
<td>64</td>
<td>breast</td>
<td>Fixed-size, 5 to 10 mm at max</td>
<td>Fixed-size, 5 to 10 mm at max</td>
<td>40%</td>
</tr>
<tr>
<td>Ott (2006)</td>
<td>65</td>
<td>EB junction</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max, using landmark</td>
<td>35%</td>
</tr>
<tr>
<td>Weber (2001)</td>
<td>40</td>
<td>EB junction</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max, using landmark</td>
<td>35%</td>
</tr>
<tr>
<td>Wieder (2007)</td>
<td>24</td>
<td>EB junction</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max, using landmark</td>
<td>35%</td>
</tr>
<tr>
<td>Ott (2003)</td>
<td>44</td>
<td>Gastric</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>Fixed-size, circular 15 mm at max, using landmark</td>
<td>35%</td>
</tr>
<tr>
<td>Wieder (2004)</td>
<td>38</td>
<td>esophagus</td>
<td>Fixed-size, circular 15 mm at max</td>
<td>N/Sp</td>
<td>30%</td>
</tr>
<tr>
<td>Schwarz (2005)</td>
<td>11</td>
<td>Breast</td>
<td>Fixed-size, circular (size N/Sp) manually placed</td>
<td>Fixed-size, circular (size N/Sp) manually placed</td>
<td>20%</td>
</tr>
</tbody>
</table>

n = number of patients, Pre-Tx = pre-treatment, Intra-Tx = intra-treatment, ROI = region of interest, Res Thr. = response Threshold, EG = esophagogastric, N/Sp = not specified

**Table 2.1. Previous response assessment studies.** A summary of previous Intra-Tx tumour response assessment studies that used the fixed-size ROI method.
<table>
<thead>
<tr>
<th></th>
<th>Population A</th>
<th>Population B</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Sex (F, M)</td>
<td>3F, 12M</td>
<td>1F, 9M</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean age ± SD</td>
<td>58.3 ± 5.7 yr</td>
<td>58.7 ± 11.6 yr</td>
</tr>
<tr>
<td>(age range)</td>
<td>(49 - 68) yr</td>
<td>(42 - 79) yr</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Stage IV</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Total no. of GTV</td>
<td>38</td>
<td>33</td>
</tr>
<tr>
<td>GTVp</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>GTVn</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tongue</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>tonsil</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>hypopharynx</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>larynx</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>paranasal sinus</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

F = female, M = male, SD = standard deviation, GTV = gross tumour volume, GTVp = primary tumour, GTVn = involved lymph node

Table 2.2. Patient Characteristics.
Figure 2.1. Change in the distribution of FDG uptake during treatment. The PET/CT images of Pre-Tx (A) and Intra-Tx (B) are of a patient with a base of tongue primary tumour. Two circular ROIs of 15 mm diameter are centered at the maximum uptake points on both Pre-Tx (green) and Intra-Tx (red) images denoted by "M". An additional 15 mm diameter circular ROI is placed on the Intra-Tx image (blue) in a position judged to correspond to the same anatomical location as the ROI as in the Pre-Tx. The FDG uptake profiles along the black lines connecting the two Intra-Tx ROIs are illustrated in figure 2.2.
Figure 2.2. Uptake profiles. The uptake profile from figure 2.1 normalized to Pre-Tx maximum SUV. The distribution of uptake within the tumour has changed during the therapy such that the maximum uptake point along the profile in Pre-Tx corresponds to a local minimum uptake point in the Intra-Tx. The maximum uptake point along the profile in Intra-Tx is now in a different location of the tumour.
Figure 2.3. Therapy induce geometric changes. Both tumours and normal tissues may shrink and shift during the treatment. The co-registered PET/CT images of Pre-Tx (A) and Intra-Tx (B) are cross-sectional images of a patient with a primary tumour of the tonsil. The patient’s left parotid gland in Intra-Tx (yellow contour) illustrates volume loss and shift relative to that in Pre-Tx (red contour). Similarly, the gross tumour volume for one nodal disease site in Intra-Tx (dotted green contour) illustrates volume loss and shift relative to that in Pre-Tx tumour (dotted blue contour).
Figure 2.4. **Distance histogram.** Histogram of distances between the centers of the two Intra-Tx ROIs.
Figure 2.5. Quantitative response assessment using two ROI method. Comparison between the two quantitative tumour response measurements when two different Intra-Tx ROI methods were used. Plot (A) is a scatter plot of the two methods. The solid line in this graph is the unity line where \( \Delta S_{\text{UV,peak}} = \Delta S_{\text{UV,same}} \). For most tumours \( \Delta S_{\text{UV,peak}} \leq \Delta S_{\text{UV,same}} \). An outlier is identified by the oblique arrow above the unity line where \( \Delta S_{\text{UV,peak}} > \Delta S_{\text{UV,same}} \). In plot (B), the tumour response on ordinate is plotted for all 26 tumours on abscissa. Thresholds of \( \pm 30\% \) as defined by PERCIST were applied to separate individual tumours to different categories using either \( \Delta S_{\text{UV,peak}} \) (red) or \( \Delta S_{\text{UV,same}} \) (blue). 19\% of the tumours (5 out of 26) were ambiguously classified as indicated by vertical arrows.
Figure 2.6. Uncertainties in ROIsame placement. Plots illustrating how uncertainties in positioning ROIsame impact tumour response assessment measured by $\Delta SUV_{same}$. Data for a sample tumour (A) and the average data for all tumours (B) is plotted. The error bars represent standard errors.
Figure 2.7. Quantification of therapy induced geometric changes. Geometric changes due to therapy for GTV and normal tissues characterized by percentage volume changes (A) and shifts (B). The bars are the median values and the error bars are the standard errors. □ = tumour, □ = salivary glands, □ = other soft tissues, □ = bones, GTVp = gross tumour volume (primary), GTVn = involved lymph node, LT = left, RT = right, Sman = submandibular, sMuscle = sternocleidomastoid muscle.
Chapter 3:

Accuracy of PET quantification using blood sample based simplified dynamic PET for response assessment

In the previous chapter, significant uncertainties in response assessment were identified when standard static PET imaging with fixed ROI analysis was used. It was demonstrated that these uncertainties could translate to significant clinical impact in terms of patients' treatment management, survival, and morbidity. This chapter evaluates the accuracy of a current simplified dynamic PET quantification method for assessing tumour response that is based on a patient's blood sample. Data from patients with advanced head and neck cancer enrolled in an OCC clinical trial were used in this chapter to evaluate the accuracy of blood activity measurements and its effect on the accuracy of a simplified dynamic PET quantification method.

The work presented in this chapter has been submitted for publication in:

- M. Sattarivand et al., Uncertainty in measurements of F-18 blood concentration and its effect on dynamic PET analysis, Journal of Nuclear Medicine Technology, 2013
3.1. Introduction

Nuclear Medicine technologists are sometimes asked to measure the concentration of activity in blood samples. Such data may be required, for example, for pharmacokinetic studies of new radiopharmaceuticals [104], to aid in calculating expected radiation dose to the bone marrow per unit administered activity in radionuclide therapy [105], and as a method of deriving an input function for kinetic analysis exploiting data from dynamic positron emission tomography (PET).

In the current case, blood sample concentration measurement was required due to a clinical trial that sought to assess the ability of fluorodeoxyglucose (FDG) PET scanning to predict treatment response for advanced head and neck cancer patients (Poon et al 2008). As part of this trial, dynamic PET scans were acquired and used to generate measures of activity versus time for volumes of interest thought to contain tumour.

Kinetic analysis of FDG PET studies normally requires absolute blood F-18 activity concentration measurements [71]. For full kinetic analysis, the patient's arterial blood activity concentration versus time, starting at the point of tracer administration and extending to about 1 hour after the injection of FDG, is the input function needed for kinetic analysis. Since taking arterial blood samples continuously for 1 hour is usually not clinically practical, dynamic PET studies commonly rely on a smaller number of blood samples [70]. The technique chosen in the trial in question, simplified kinetic analysis - multiple time points (SKA-M) relies on a single venous blood sample [106]. This single blood activity concentration is used to scale a population averaged blood activity curve and thus obtain the input function. Since the SKA-M method relies on a single blood sample from the patient, it is reasonable to postulate that the accuracy of any quantitative metric derived using the SKA-M technique strongly depends on the accuracy of blood activity concentration measurements. In order to obtain an accurate blood activity concentration, accurate measurements of both blood activity and blood volume are essential.
In many Nuclear Medicine facilities, the available detector systems for measuring blood activity are either a thyroid probe or a well counter. Direct measurement of blood activity in the dose calibrator is not practical due to the low activities involved. A cross calibration against a dose calibrator is necessary to convert the thyroid probe or well counter readings to activity (figure 3.1). A single point calibration factor using a standard solution [107] can be applied to either the thyroid probe (ProbePoint) or the well counter (WellPoint). Alternately, a calibration curve can be obtained for the well counter (WellCurve) in a full range of expected blood activity. The WellCurve method is more labor intensive, but it would be expected to provide the most accurate results.

Blood volume measurements (i.e. whole blood) are typically performed using micro-pipettes. Laboratory micro-pipettes are designed to measure the volume of liquids such as water precisely. Blood is more viscous than water and the precision of blood volume measurements may be compromised due to its viscosity. Slightly different pipetting techniques are recommended by manufacturers for pipetting blood. As the need to pipette any liquid may be rare in some Nuclear Medicine departments, it is unlikely that specialized techniques for pipetting blood will be familiar to all Nuclear Medicine technologists. For example, neither the Nuclear Medicine technologists who performed the initial pipetting work for the clinical study nor the researchers requesting the pipetting work were aware of the need to employ such methods.

For the purposes of the clinical trial, blood activity was originally measured using the ProbePoint method. The well counter was not used initially in the clinical trial, as cross-calibration of the well counter was considered less convenient. A standard solution of sufficient activity to be accurately measured in the dose calibrator contains too much activity to be used without decay with a standard well counter due to detector dead time effects. However, as the trial progressed, data emerged suggesting that there might be significant variation in the
measurements obtained using the ProbePoint technique. In response, potentially more accurate techniques such as the WellPoint and WellCurve methods were investigated.

The objective of this study was to quantitatively assess the accuracy of absolute blood activity measurements using the ProbePoint, WellPoint and WellCurve techniques. Moreover, the precision of blood volume measurements using a typical laboratory micro-pipette was also investigated. Finally, the effects of the accuracy of the blood activity concentration measurements on the accuracy of a SKA-M metric were evaluated.

3.2. Methods and materials

Thirty five patients with advanced head and neck cancer underwent dynamic PET scans using a Gemini PET-CT scanner (Philips medical systems, Cleveland, Ohio). Patients were injected with 5 MBq of FDG per kg of patient's weight. Patients heavier than 75 kg were injected with a fixed dose of 370 MBq (10 mCi) of FDG. The PET images were acquired at a single bed position and consisted of 12 frames of 2.5 minutes each for a total of 30 minutes (from 30 to 60 minutes post injection), which were similar to those of Sundaram’s (from 25 to 55 minutes post injection). The PET voxels were isotropic, 2 mm on each side. The reconstructed field of view (FOV) was 576 mm. Each patient received a low dose CT scan prior to their PET scan for PET attenuation correction purposes and a diagnostic, contrast enhanced CT scan post PET scan for radiation treatment planning. The treatment for the patients was 7 weeks of radiotherapy. If a patient could tolerate it, chemotherapy was also administered.

Blood samples of approximately 3 ml in volume were obtained immediately following the PET-CT scans. The average blood draw time was 77.8 ± 12.4 minutes post injection. Note that our blood drawn time was different than that of Sundaram’s (45 minutes) to make it more convenient to the patients while they are being scanned. From each gross blood sample, a testing
sample of 0.9 ml was extracted using a calibrated Corning Lambda micro-pipette (Corning life sciences, Amsterdam, The Netherlands). The pipette was a “air displacement” model with a continuously adjustable volume setting. The absolute activity of each testing blood sample was measured with three different techniques: “ProbePoint”, “WellPoint”, and “WellCurve”.

3.2.1. Thyroid probe technique (ProbePoint)

A thyroid probe was cross calibrated against a dose calibrator to convert counts to activities using a single point calibration factor. The thyroid probe used was an Atomlab 950 and the dose calibrator was an Atomlab 500 (both from Biodex medical systems, Shirley, NY). Cross calibration was performed on the day of the patient PET scan immediately after obtaining the blood sample from the patient. A standard solution for cross calibration was obtained by diluting FDG with saline such that about 370 kBq (10 µCi) in a 0.9 ml volume could be read from the dose calibrator. The test tube and the volume of the standard solution (0.9 ml) were the same as those of the blood sample to avoid geometric effects. Three one-minute count measurements of blood sample and standard solution were made using the thyroid probe and the averages of the three counts were used in the calculations. Two one-minute count measurements of background were also acquired. The geometrical orientation of the probe for all count measurements (i.e. standard solution, blood sample, and the background) was kept the same. Blood samples were corrected first for background counts and then for F-18 decay. With the known activity of the standard solution from the dose calibrator and the counts of the standard solution in the probe, a single point calibration factor was obtained. This calibration factor and the blood volume (0.9 ml) were used to obtain the activity concentrations (kBq/ml) of the blood samples. The ProbePoint technique required about 20 minutes to measure the activity of each blood sample.
3.2.2. Well counter techniques (WellPoint and WellCurve)

A well counter was cross calibrated against a dose calibrator to convert counts to activities using a standard F-18 solution. The well counter used was an optional accessory to the Atomlab 950 thyroid probe and the dose calibrator was again the Atomlab 500 (both from Biodex medical systems, Shirley, NY). On the day of the patient's PET scan, the same blood sample used in the ProbePoint method was counted in the well counter three times one-minute each. The background was also counted twice. Moreover, on the same day of the patient scan, a 0.9 ml standard solution of F-18 was prepared in the same kind of test tube as the blood sample. The activity of this standard solution was approximately 130 MBq (3.5 mCi) such that its activity during the 8 hour working time on the following day was within the 1 to 20 kBq range (0.03 to 0.5 µCi) due to decay. This range was chosen to cover the whole expected range of the activities of all blood samples. On the day following the PET scan, the counts of the standard solution were measured in the well counter 8 times separated by 1 hour each. From these 8 measurements, a calibration curve was obtained for the well counter by calculating the known activity at each of the 8 measurement times after accounting for decay. The calibration curve was a linear fit to all 8 points. This calibration curve and the blood volume (0.9 ml) were used to calculate the activity concentration of the blood sample in kBq/ml. The WellCurve technique requires a full day of measurements to obtain the calibration curve and apply it to a blood sample. The WellPoint method used a single point randomly selected from the WellCurve calibration curve. Similar to the ProbePoint method, only 20 minutes were required for the WellPoint method, with the provision that the F-18 calibration source used in the WellPoint method had to be decayed for ~ 24 hours. The 24 hour waiting time is necessary to allow the decay of standard solution to match to the small activity of the blood samples. The WellCurve technique was assumed to represent the ground truth and the accuracies of the ProbePoint and WellPoint techniques were compared.
against the WellCurve. This is because it incorporates both the more sensitive detector and multiple calibration points.

3.2.3. Precision of micro-pipette blood volume measurements

Six experienced nuclear medicine technologists were involved with blood pipetting for all 35 patients that were a mix of Pre-Tx and Intra-Tx. For 10 of the 35 patients, multiple blood samples (3-6 samples) were obtained and the precision of blood volume measurements was evaluated. All the blood samples for a given patient were measured by the same nuclear medicine technologist. The same, calibrated Corning Lambda micro-pipette described above was used to draw 0.9 ml for each blood sample. The activities of all blood samples were measured using the WellCurve technique. The precision of the blood volume measurements was evaluated by calculating the percent difference in activity of each blood sample relative to each individual patient's sample mean blood activity.

3.2.4. Accuracy and precision of micro-pipette volume measurements for water

A quality assurance (QA) test was performed to evaluate the accuracy and precision of volume measurements of water using the micro-pipette and resulting values were compared with those stated by the manufacturer. Pipetting QA with water was chosen to ensure the accuracy and precision of volume measurements as stated in the pipetting manual. Five trained nuclear medicine technologists participated in the QA test. Using the micro-pipette, each technologist drew four samples of 0.9 ml of water into test tubes. The test tubes were the same as those used for the blood samples. The test tubes were weighed with and without the water using a micro-balance and the volume of the water for each test tube was calculated assuming that the density of water is 1.0 g/ml. The micro-balance used was a Sartorius CPA2P (Precision weighing balances,
Bradford, MA). The accuracy and precision (standard deviation) of the micro-balance as stated in the manual were both 1 μg.

3.2.5. Calculating SKA-M using the different blood sample measurement techniques

For FDG-PET imaging, the “SKA-M” parameter may be considered an estimate of the more well-known Patlak constant, $K_i$, which is linearly related to the glucose metabolic rate. The principles of SKA-M are described by Sundaram et al [106]. A single venous blood sample and a series of dynamic PET images are required. The three activity concentration measurements obtained using the ProbePoint, WellPoint and WellCurve activity measurement techniques of the same blood sample were used to calculate three values of SKA-M. Each blood activity concentration was used to scale a population-averaged blood activity curve obtained from previously published data [106]. The scaled blood activity curve was used to estimate a patient-specific blood activity curve $c_p(t)$. An activity concentration versus time curve for every PET voxel, $A(t)$, was extracted from the 12 frames of the PET images. In order to suppress noise in $A(t)$, the average voxel value in a 3x3x3 voxel window centered on each PET voxel was used as that voxel’s value for subsequent calculations. For each voxel, 12 points were obtained by plotting $\frac{\int c_p(t)}{c_p(t)}$ on the abscissa vs $\frac{A(t)}{c_p(t)}$ on the ordinate for each PET frame as illustrated in figure 3.2. A line was fitted to these 12 points and the slope of this line was taken as the SKA-M uptake rate for that voxel. This procedure was repeated for all PET voxels and a SKA-M image was obtained for every PET study and each blood activity measurement technique. From these SKA-M images, mean SKA-M was calculated for the primary tumour of all 35 patients for all the voxels within a physician's manually contoured gross tumour volume.
All the SKA-M calculations were performed in interactive data language (IDL) version 8.1 (Research systems inc., Boulder, CO).

3.3. Results

Background and decay corrected blood sample counts ranged from 110 to 1516 for the thyroid probe technique and from 18595 to 177927 for the well counter technique. Based on Poisson statistics alone, the percentage uncertainty in the thyroid probe counts thus ranged from 2.6% to 9.5% for the thyroid probe technique and from 0.24% to 0.73% for the well counter technique. A sample calibration curve that converts sample counts in the well counter to sample activities is illustrated in figure 3.3. All 8 data points are close to the fit line as reflected in the high $R^2$ value (0.9998). Similar calibration curves were obtained for all blood activity measurements acquired using the well counter. All calibration curves had good linear fits with $R^2 >= 0.9998$.

Figures 3.4A and 3.4B demonstrate the accuracy of the blood activity measurements using the ProbePoint and WellPoint techniques respectively relative to the WellCurve technique. The sample numbers (on abscissa) are sorted based on the date of blood sampling and no trend over time (i.e., a period of one year) was observed. The ordinate is labeled as the percentage error in blood activity measured using the ProbePoint or WellPoint technique relative to the WellCurve technique, since the WellCurve technique was assumed to represent the true blood activity. The average value of the percent difference between the ProbePoint and WellCurve techniques was -0.5% and the standard deviation was 6.2%. A paired t-test with 95% confidence interval indicated no statistically significant differences between the average sample activities measured using either the ProbePoint or WellCurve techniques. Thus, there was no systematic error in blood activity measurements using the ProbePoint technique, but only random errors. Errors in
ProbePoint technique ranged from about -9.5% to 7.6% as indicated in figure 3.4A. Errors in the WellPoint technique were significantly smaller, ranging from -1.3% to 0.9%.

Figures 3.4C and 3.4D illustrate the accuracy of SKA-M calculated using the ProbePoint and WellPoint techniques, respectively, relative to the WellCurve technique. Since the WellCurve technique was assumed to provide the true blood activities, the SKA-M values calculated using WellCurve-based measurements were taken as the true SKA-M values. Similar to the errors of blood activity data in figure 3.4A, the errors of SKA-M in figure 3.4C were also found to be random and in the same range. The errors in SKA-M in figure 3.4C are reciprocal to the errors in blood activities in figure 3.4A. The same reciprocal relationship applies in the WellPoint techniques between figures 3.4B and 3.4D.

Figure 3.5 illustrates micro-pipette volume precision with multiple blood sample volumes from the same patient. Figure 3.5A is the absolute blood activity concentration measurements in 10 patients labeled A to J. Each patient had multiple blood samples that are numbered following the letters A to J. Figure 3.5B was obtained by calculating the percentage difference of each blood sample from the mean of each patient and thus it represents the precision of the micro-pipette in measuring blood volume. As seen in figure 3.5B, blood sample volume estimates varied widely, with differences from the mean value from -6% to 12%.

Figure 3.6 demonstrates the results of the QA test to confirm the accuracy and precision of micro-pipette volume measurements for water. The range of accuracy and precision (standard deviation) for all technologists were (-1% to 0.7%) and (0.2% to 0.4%) respectively. These values were in reasonable agreement with the corresponding values in the micro-pipette manual, i.e. ± 0.6% and 0.2% for accuracy and precision respectively.

3.4. Discussion
In order to measure accurate blood activity concentration values, accurate measurements of both blood activity and blood volume are essential.

3.4.1. Blood activity measurement

In this study, the WellCurve method was taken as the best available estimate of the truth since it incorporated both the more sensitive detector and multiple calibration points. The well counter has a fixed geometry and is less likely to be affected by setup errors. Moreover, the geometric efficiency of the well counter is higher than the thyroid probe due to its near 4-pi geometry. Thus, the well counter is less likely to be affected by transient variation in background which is particularly important for measurements of low blood activity after biological and physical decay.

Unlike the ProbePoint or WellPoint techniques which rely on a single point calibration factor, the WellCurve technique used 8 data points to obtain calibration over the full range of the expected blood activities, thus reducing the overall uncertainties. The paired t-test indicated that there are only random errors and not systematic errors in the ProbePoint technique. This can be explained by the fact that thyroid probes are more susceptible to variation in background levels and setup errors as mentioned above, and both these errors have a stochastic nature. In addition, it is clear from the low number of counts acquired using the thyroid probe technique that longer counting times could have been used to reduce the percentage uncertainties in the results. However, long counting times are not convenient in a busy Nuclear Medicine department and can present problems with the potential for variation in background levels as injected patients move about the department.

The disadvantage of the WellCurve technique is that it requires a full working day of measurements, while the ProbePoint technique needs only 20 minutes. The ProbePoint or WellPoint techniques also automatically account for any drift in response over time. However, in
the WellCurve technique, regular calibration of the well counter (e.g. every two weeks) is required if a drift is observed over time [107]. Thus, the WellPoint technique provides a compromise between clinical practicality and accuracy (within 1.3%).

3.4.2. Blood volume measurements

The results in figure 3.5B demonstrate that errors of up to 12% may be encountered if one relies on a single blood sample from the patient. Large errors such as these in reproducibility of volume measurements with lab micro-pipettes were not expected. The most common variety of micro-pipettes, the “air cushion” or “air displacement” type, are designed to measure volumes of liquids such as water precisely. Whole blood, however, has different physical properties from water. Blood has both higher viscosity and density than water, contains proteins which can interact with the micro-pipette tip and cause “foaming”, and is not just a simple fluid, but a suspension of blood cells. Part of the observed error could be due to these differences between blood and water. This error can be minimized by using pipetting techniques more suitable for blood, such as “reverse pipetting” with pre-wetting of the pipette tip. Alternatively, the use of a different type of pipette, a positive displacement pipette, may be considered [108]. In addition, it is reasonable to have multiple blood samples and average the results. For example, 5 to 6 ml of blood could be collected from the patient in a syringe and then, using a micro-pipette, multiple 0.9 ml of blood samples could be drawn into test tubes.

3.4.3. Effects of blood activity concentration measurements on SKA-M

The SKA-M method relies on a patient’s single blood activity concentration measurement to estimate the input function. Inaccuracies in blood activity or volume measurement will directly impact the accuracy of SKA-M measurement. The relationship between the blood scale factor
and SKA-M is mathematically reciprocal and no other factors contribute. For example, if the measured blood activity concentration is 0.9 of its true value (i.e., 10% underestimation), then the calculated SKA-M is $1/0.9=1.11$ of its true value (i.e. 11% overestimation). The reason for the reciprocal relationship lies in the mathematical calculation of SKA-M as illustrated in figure 3.2B. A relative error in the estimate of blood activity concentration by a constant factor $\alpha$ will not affect the abscissa of figure 3.2B, but will translate to an error of $\frac{1}{\alpha}$ in the estimate of the ordinate. Thus, the SKA-M will also be inaccurate by a factor of $\frac{1}{\alpha}$. This reciprocal relationship is reflected in the results demonstrated in figure 3.4, that is, every data point in figure 3.4A is translated to a reciprocal point in figure 3.4C. The range in error of blood activity in figures 3.4A and 4B is also translated to the same range of errors of SKA-M in figures 3.4C and 4D respectively. A similar argument applies to figure 3.5B in terms of precision of blood volume and its effect on the precision of SKA-M. Therefore, errors in precision of blood volume measurements up to 12% will be translated to errors in precision of SKA-M approximately up to 12%.

Paired t-tests with 95% confidence interval for both blood activity measurements (figure 3.4A) and SKA-M (figure 3.4C) indicated that the difference between the average blood activity measured by the ProbePoint and WellCurve techniques was not statistically significant. Therefore, the errors in SKA-M using the ProbePoint technique are random. The fact that the nature of the error is random and not systematic is of special concern. In order to quantify PET uptake in a tumour using SKA-M, typically pre-treatment (Pre-Tx) and intra-treatment (Intra-Tx) PET scans are obtained and the relative change in uptake is used for tumour response assessment:

$$
\Delta SKA - M = \left( \frac{SKA - M_{\text{intra}}}{SKA - M_{\text{pre}}} - 1 \right) \times 100\% 
$$
Systematic errors in SKA-M in both Pre-Tx and Intra-Tx will cancel out while the random errors will combine to cause a larger error in $\Delta$SKA-M. Theoretically, the combined error in $\Delta$SKA-M will be on average a factor of $\sqrt{2}$ greater than the error in SKA-M. For example, errors in both Pre-Tx and Intra-Tx of 10% in both activity and volume measurements will cause an error of $\sqrt{2} \times 10\%$ in blood activity concentration estimates and approximately an error of $\sqrt{2} \times \sqrt{2} \times 10\% = 20\%$ in $\Delta SKA_M$.

3.4.4. Clinical significance

Inaccurate response assessment up to 20% error in $\Delta$SKA-M could have significant clinical impact. An estimate of the clinical significance is provided here similar to discussions in sections 1.5.2 (chapter 1) and 2.4.3 (chapter 2). Since SKA-M depends not only on the blood activity concentration, but also on tumour uptake, we assumed that a "true" $\Delta$SKA-M will have a distribution similar to figure 2.5B and a PERCIST threshold criteria [61] is used to classify the patients to responders and non-responders. Subsequently, we apply a random 20% error to the data to identify the mis-classified patients as demonstrated in figure 3.7. As illustrated, errors in blood activity concentration measurement, caused 15% of patients to be mis-classified as responders. These patients would complete the ineffective chemo-radiation and still require salvage surgery (figure 1.3). On the other hand 8% of the patients are mis-classified as non-responders which would cause them to stop the effective chemo-radiation and have an early surgery. These mis-classifications will have effects on (a) survival and (b) morbidity.

Theoretical survival rates can be estimated using the two assumptions in section 2.4.3. The first assumption is that salvage surgery after an ineffective chemo-radiation has only a 25% survival rate. The second assumption is that responding patients to chemo-radiation will have a
50% survival if they are treated with early surgery. Thus, mis-classifying 15% patients as responders and 8% as non-responders could translate to a theoretical survival of \([0.5-0.08)\times100\% + 0.08\times50\%] + [(0.50-0.15)\times50\% + 0.15\times25\%]= 67\%\). Here the first square bracket is for responders and the second one is for non-responders.

As discussed in section 1.5.2, an accurate response assessment will allow 50% of the patients (only responders) to complete chemo-radiation thus reducing morbidities as per table 6.1. Mis-classification errors in figure 3.7 will allow additional 8% (4-2=2 out of 26) of patients to complete chemo-radiation. Thus, xerostomia, acute mucositis, and dysphagia morbidities will increase to 19%, 29%, and 27% of the patients respectively. Table 6.1 summarizes the clinical impact of blood activity concentration measurement errors on response assessment in terms of treatment management, survivals, and morbidity.

3.4.5. Direct image-based measurement of input functions for SKA-M

To avoid loss of accuracy in blood activity concentration measurements and the resulting loss in accuracy of SKA-M, an alternative approach is to measure the blood activity concentration directly from the PET images. A blood activity concentration curve obtained with this approach is referred to as an image-based input function [85]. Figure 3.2 indicates that any systematic error in the PET image will not affect the accuracy of the SKA-M uptake rate: a systematic error in the PET image will affect both \(c_p(t)\) and \(A(t)\), thus the error will cancel out on both axes in figure 3.2B without affecting the SKA-M uptake rate.

A region of interest should ideally be placed at the center of the left ventricle to measure blood activity concentration from PET images. For cases such as head and neck patients, where the heart is not in the field of view, major blood vessels such as the carotid artery may be used from
the CT images. However, this approach is only feasible if corrections to the images can be made to obtain accurate measures of absolute activity concentrations in the vessels (e.g., partial volume correction).

3.5. Conclusion

While one might think that measuring the concentration of activity in a blood sample would be a simple task, in practice care must be taken to select an appropriate technique and to properly apply that technique. Pipetting techniques suitable for blood should be applied to minimize errors in blood volume measurements. If possible, techniques that use the thyroid probe should be avoided in favour of those that use a well counter. The WellPoint technique provided a compromise between clinical practicality and accuracy (within 1.3%). Random errors in blood activity and volume measurements may accumulate and compromise the SKA-M estimates of tumour uptake rate. Such errors would also be of concern in other situations where blood activity concentration must be assessed, such as the measurement of blood concentration time-activity curves for radionuclide treatment planning applications. Finally, the results provide support for the development of direct image-based measurements of input functions for kinetic analysis.
Figure 3.1. Cross calibration. A cross calibration against a dose calibrator (A) is needed to convert counts readings from the thyroid probe (B) and well counter (C) to the units of activity (kBq or μCi) for blood samples.
Figure 3.2. Steps for calculating SKA-M. (A) A single patient's blood sample is used to scale a population average blood activity curve and estimate the patient's blood activity curve $C_p(t)$. The $A(t)$ is a representative tumour uptake curve and is measured during PET acquisition between 30-60 minutes post-injection in 12 frames. The two time curves, i.e. $C_p(t)$ and $A(t)$, are used to generate plot (B) with data points corresponding to each PET frame. The slope of the line fitted to these 12 points is SKA-M. (C) The relationship between the blood scale factor and SKA-M.
Figure 3.3. A sample calibration curve. The calibration is obtained for the well counter in a range of typical blood activities for SKA-M.
Figure 3.4. Accuracy of blood activity measurements and effects on SKA-M. Errors in blood activity measurements using the ProbePoint (A) or WellPoint techniques (B) in 35 patients. Errors are relative to the WellCurve technique which is assumed to be the ground truth. The accuracy of blood sampling techniques in (A) and (B) is reflected on the accuracy of SKA-M in (C) and (D) respectively. The red lines are the average values for each plot.
Figure 3.5. Precision of blood volume measurements. Micro-pipette volume precision is estimated for 10 patients. (B) Absolute blood activity in 10 patients having 3-6 blood samples. The letters A to J correspond to 10 patients. The numbers following each letter are sample numbers for each patient. (B) Percentage difference of each blood sample from the mean for the patient.
Figure 3.6. Accuracy and precision of micro-pipette volume measurements for water. The pipetting 0.9 ml of water measured by 5 nuclear medicine technologists.
Figure 3.7. Clinical impact of errors in blood activity concentration measurements. The applied 20% random errors caused 15% (4 out of 26) of patients mis-classified as responders (green arrows) and 8% (2 out of 26) of patients are mis-classified as non-responders (black arrows).
Chapter 4: Improving PET quantification accuracy by developing a novel partial volume correction technique

In the previous chapter, the accuracy of current blood sample based simplified dynamic PET quantification method was evaluated for tumour response assessment. Significant errors were identified in the blood sample based dynamic PET quantification method. It was demonstrated that corresponding errors in assessing response to treatment could have significant clinical impact in terms of patients' treatment management, survival, and morbidity. The results in chapter 3 support a need for a direct image based dynamic PET quantification method. However, as mentioned in chapter 1, due to limited PET spatial resolution, this is only possible if partial volume correction (PVC) is employed to recover the loss in PET quantification accuracy in small vessels in head and neck area. Standard existing PVC techniques have limitations in their application for small vessels, particularly for noisy PET images. In this chapter, a novel partial volume correction technique is presented to improve PET quantification accuracy.

The work presented in this chapter has been published in:

4.1. Introduction

Positron emission tomography (PET) is inherently a low spatial resolution imaging modality in comparison to anatomical imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI). Without partial volume correction (PVC), accuracy in quantification of PET imaging can be significantly affected [109-112]. Two distinct phenomena both arising from limited resolution contribute to the partial volume errors in PET imaging [113]. The first is referred to as the “tissue fraction” effect. With typical PET image digital sampling, a single voxel may contain more than one tissue type, each of which may exhibit different physiological uptake of tracer. This effect can be accounted for using additional a priori information from co-registered high resolution anatomical images such as CT or MRI. The second effect is the “system response” or “point spread function (PSF)” effect, i.e. spill over between regions due to physical 3-dimensional (3D) PET image formation processes. To correct for this effect, the PET PSF is usually characterized with a 3D Gaussian function. Without PVC, quantification of tracer uptake in objects smaller than 2 to 3 times the full width half maximum (FWHM) of the PSF is affected [113].

PVC can be applied either during or post reconstruction. During reconstruction PVC methods incorporate the system response into the PET reconstruction process [114-118]. The drawbacks of this approach are limited access to proprietary reconstruction algorithms as well as the wide variety of algorithms and PET scanners used in the clinical setting.

Post reconstruction methods can be divided into two broad categories, those that apply PVC at the voxel level and those that apply PVC at the region level. Some PVC methods in the first category perform PVC without a need for co-registered anatomic images. They rely on an iterative deconvolution using maximum likelihood [119], expectation maximization [120, 121], or other iterative deconvolution techniques [122]. Those that need a co-registered anatomic image
include those that perform wavelet transforms [88, 123] or Bayesian iterative deconvolution [124]. Although these voxel-based PVC methods do not need CT or MR segmentation, noise amplification is the main concern in iterative deconvolution based methods [113], since deconvolution is an ill-posed problem.

A class of voxel-based PVC methods that need tissue segmentation, assume homogeneous and known activity for all regions except the one to be corrected (i.e. the target region). Regions in these methods are referred to as “compartments” and such methods most commonly used in neurological studies. Different variants of these methods are referred to as two- [125], three- [126], or four- [127] compartment models. In these methods, an accurate knowledge of the activities in other regions is generally required in order to have an accurate PVC for the target region.

A generalization of voxel-based PVC that uses MR segmentation exists that does not require any regional activity to be known, but assumes homogeneous uptake within every region [128]. In this method, spill over is modeled at the voxel level using a least squares approach. The method requires calculations involving matrices of large dimension, since a linear equation for each voxel is considered in the least squares fitting. This method was originally described and implemented for a 2-dimensional (2D) case. However, the matrix size for a 3D implementation would be substantially larger. Thus, its clinical implementation in 3D may not be feasible due to large storage and computational requirements [129]. This restriction may be mitigated by using algorithms such as tensor operations, although with a restrictive separability assumption [130].

Post reconstruction region-based PVC methods have drawn special attention in PET imaging as they are more feasible to implement than voxel-based PVC. Non-overlapping regions representing different tissue types are usually segmented in MR (or CT) and are used as \textit{a priori} information. With knowledge of the PET system response, PVC attempts to recover mean
activity within each region, accounting for both the tissue fraction effect and the PSF effect. One of the first studies using region-based PVC [87] derived an analytical linear equation for the uptake in each region and calculated the effect of PSF on segmented anatomical images. These equations form a geometric transfer matrix (GTM) describing spill over between regions. This method was originally described in 2D with an image based convolution of the regions with the PSF. Subsequently, the GTM method was described in 3D with a sinogram implementation without the need for PSF estimation [89]. The 3D implementation accounts for the 3D acquisition mode of current PET scanners. Another study [130] presented the GTM method in a weighted least squares framework with a 3D implementation. This study presented another PVC method based on weighted least squares fitting but with an added explicit noise model. The GTM method was also extended to iterative reconstructions where the linearity assumption of the system response does not hold true [131].

Noise characteristics and robustness are important issues for any PVC method. Real PET data are inherently noisy and the assumptions of perfect PET-CT registration or exact knowledge of the PSF may not hold true. Limited studies have been published in this regard for the 3D implementation of the GTM or Labbe's methods in PET. Earlier studies [129, 132] compared the GTM to Labbe's method in 2D. A later 3D study [89] compared the robustness of image space versus sinogram space implementations of the GTM method. The effects of mis-registration errors were investigated but not the effects of PSF errors.

In this study, a new 3D region-based PVC method is presented with an analytical derivation. It uses a least squares fitting technique analogous to Labbe's approach. It is demonstrated that the new method is mathematically equivalent to Labbe's method, however, since the new method is region-based rather than voxel-based, it avoids handling large matrices. The method yields a geometric transfer matrix similar to that of the GTM method, while providing meaningful
physical interpretations for spill over between regions in a symmetric GTM framework. The method was validated using two 3D simulated phantoms and a measured physical phantom. Noise characteristics and robustness of the new PVC method were assessed and compared to those of the 3D GTM method in terms of noise propagation and accuracy lost due to mis-registration errors and errors in PSF measurements.

4.2. Theory

4.2.1. Analytical derivation

The measured PET image is represented by \( I(r) \), where \( r \) is a spatial location in three dimensional (3D) image space, \( r \in FOV \). Assuming PET imaging to be a linear operation, in the absence of noise, the expected image \( \bar{I}(r) \) is:

\[
\bar{I}(r) = I_{true}(r) \ast h(r),
\]

where \( I_{true}(r) \) is the ideal PET image with no resolution lost, \( \ast \) is the 3D convolution operator, and \( h(r) \) is the imaging system point spread function (PSF) which is assumed to be spatially invariant. Similar to GTM, it can be assumed that \( I_{true}(r) \) is composed of \( N \) tissue volumes, each with a homogeneous uptake \( T_i \), i.e.:

\[
I_{true}(r) = \sum_{i=1}^{N} VOI_i(r) \cdot T_i, \quad i=1,2,\cdots N,
\]

where, \( VOI_i(r) \) is the mask of the volume of interest for the \( i \)th tissue, i.e. \( VOI_i(r) = 1 \) for voxels inside and zeros outside. The VOIs can be obtained from a co-registered high spatial resolution imaging modality such as CT or MRI. Assuming that the VOIs have homogeneous uptake, there is no overlap between them, and they add up to unity image.
\[
\tilde{I}(r) = \sum_{i=1}^{N} VOI_i(r) \cdot h(r) \cdot T_i = \sum_{i=1}^{N} RSF_i(r) \cdot T_i,
\]

where \( RSF_i(r) \) is the regional spread function for the corresponding VOI and is defined [87] as:

\[
RSF_i(r) = VOI_i(r) \cdot h(r).
\]

A cost function is defined based on a least squares fit of the expected image to the measured image:

\[
E = \int_{FOV} \left[ I(r) - \tilde{I}(r) \right]^2 dr = \int_{FOV} \left[ I(r) - \sum_{i=1}^{N} RSF_i(r) \cdot T_i \right]^2 dr,
\]

The minimization of this cost function implements linear square fitting "at the voxel level". This technique is analogous to Labbe's approach. This equation appears to differ from Labbe's equation set (equation (4.A1) in the appendix), where a linear equation is presented for every voxel. In the appendix however, it is demonstrated that the results of the two implementations are mathematically equivalent. Nonetheless, as illustrated below, the cost function implementation is converted to an equation that applies PVC "at the region level", thus avoiding handling the excessively large matrices required in Labbe's method.

In order to minimize the cost function with respect to \( T_j \):

\[
\frac{\partial E}{\partial T_j} = -2 \int_{FOV} \left[ I(r) - \tilde{I}(r) \right] RSF_j(r)dr = 0, \quad j = 1, 2, \ldots N
\]

\[
\int_{FOV} RSF_j(r) \sum_{i=1}^{N} RSF_i(r) \cdot T_i dr = \int_{FOV} I(r) RSF_j(r)dr
\]

In matrix form, this equation is written as:
\[ i = 1, \quad 2, \quad \cdots \quad N \]
\[
\begin{bmatrix}
\omega_{11} & \omega_{12} & \cdots & \omega_{1N} \\
\omega_{21} & \omega_{22} & \cdots & \omega_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\omega_{N1} & \omega_{N2} & \cdots & \omega_{NN}
\end{bmatrix}
\begin{bmatrix}
T_1 \\
T_2 \\
\vdots \\
T_N
\end{bmatrix}
= \begin{bmatrix}
t_1 \\
t_2 \\
\vdots \\
t_N
\end{bmatrix},
\]

Eq.(4.1)

where
\[
\omega_j = \int_{FOV} \text{RSF}_i(r) \cdot \text{RSF}_j(r) dr \quad \text{and} \quad t_j = \int_{FOV} I(r) \cdot \text{RSF}_j(r) dr.
\]

By taking the inverse of the \( \omega_{ij} \) matrix, equation (4.1) can be solved to calculate \( T_i \) values thus obtaining the true PET image. Unlike the GTM method, here the geometric transfer matrix is symmetric (thus named sGTM). Equation (4.1) has the same appearance as equation (8) in the original GTM paper [87], but with different elements in the matrix and t-values.

4.2.2. Physical interpretation

In the GTM method, \( \omega_j = \int_{FOV} \text{RSF}_i(r) \cdot \text{VOI}_j(r) dr \) and \( t_j = \int_{FOV} I(r) \cdot \text{VOI}_j(r) dr \). The weighting factor \( \omega_j \) represents spill in \((i \neq j)\) or spill out \((i = j)\) from \( \text{RSF}_i \) to \( \text{VOI}_j \). Note that unlike the original GTM paper, normalization of both \( \omega_j \) and \( t_j \) by the number of voxels in the \( \text{VOI}_j \) mask \( (n_{\text{px}}) \) is omitted here in order to facilitate comparison, without changing the concept. In the sGTM method, \( \omega_j \) represents spill in \((i \neq j)\) or spill out \((i = j)\) from \( \text{RSF}_i \) to \( \text{RSF}_j \). In order to get the t-values, the PET image is sampled by \( \text{RSF}_j \) in the sGTM method rather than by the \( \text{VOI}_j \) mask as in the GTM method. Thus, rather than treating every voxel as either belonging to a given \( \text{VOI} \) (voxel value =1), or not (voxel value =0), the voxel is weighted to have a value between 0 and 1 as defined by the corresponding \( \text{RSF} \). For example, a voxel within \( \text{VOI}_l \) and close to its
boundary may belong 70% to \( RSF_1 \) and belong 30% to \( RSF_2 \), if this voxel is adjacent to \( VOI_2 \) and far enough from other \( VOI \) masks. Both GTM and sGTM methods assume that all \( VOI \) masks add up to a unity image, i.e.:

\[
\sum_{i=1}^{N} VOI_i(r) = 1
\]

Due to the linear nature of convolution, this assumption extends to the \( RSF \), i.e.:

\[
\sum_{i=1}^{N} RSF_i(r) = \left[ \sum_{i=1}^{N} VOI_i(r) * h(r) \right] = \left[ \sum_{i=1}^{N} VOI_i(r) \right] * h(r) = 1
\]

Figure 4.1 demonstrates a comparison of GTM and sGTM for an image of a vessel.

4.3. Materials and methods

Two 3D simulated phantoms (a sphere phantom and a brain phantom) and one physical sphere phantom were used to validate the sGTM method and compare it against the conventional GTM method. The accuracy, precision, and noise propagation characteristics of the two PVC methods were evaluated. Moreover, the robustness of the sGTM method was compared to that of the GTM method in regard to (a) mis-registration error between PET and CT images, and (b) error in PSF measurement. These two sources of error are important in clinical implementation of any PVC method. For example, patient motion may cause mis-registration errors. On the other hand, errors in estimation of PET PSF may be caused by inaccurate PSF measurement or use of literature values or vendor data rather than actual PSF measurements. Since the values in the weighting matrices used in the GTM and sGTM methods rely on accurate estimates of PSF, this error will translate into inaccuracies in PVC calculations.

4.3.1. Simulations
4.3.1.1. Sphere phantom simulation

An ideal 3D PET image was created from CT images of a physical phantom having spheres in a cylindrical tank filled with a sphere to background (tank) activity concentration ratio of 3 to 1. Six spheres were simulated with inner diameters ranging from 5 to 30 mm and the wall thickness of 0.6 mm for all spheres. The isotropic voxel size was 0.6 mm and 2 mm for CT and PET respectively. The ideal 3D PET images were created by auto-contouring spheres and the tank seen in CT images using an analytical least squares fitting algorithm. This gives an ideal PET image in CT space, which was then down-sampled using a tri-linear interpolation to PET voxel size. This image was then convolved with a 3D Gaussian PSF with a FWHM of 8.7, 8.3, and 7.8 mm in X, Y, Z directions respectively. This PSF was chosen to match exactly the same PSF used in the physical sphere phantom below. After the convolution, uncorrelated Gaussian noise of 27% was added uniformly across the image and a total of 100 simulated PET images were created each different in statistical noise. The 27% noise corresponds to the voxel noise level of the physical sphere phantom. The noise added was spatially uniform for the non-uniform sphere phantom. This choice was based on a previous study [134] where it was demonstrated that noise is more or less uniform and independent of the activity level for FBP-reconstructed PET images due to back projection of the sinogram data. Figure 4.2 illustrates the ideal PET image (in CT space, i.e. before down-sampling to PET voxel size) and a simulated PET image. PVC using GTM or sGTM was applied using 3x3 weighting matrices. For each sphere size, the 3 VOIs were inner sphere, sphere wall, and the background volume. The background volume was a cylinder encompassing the sphere and its walls such that the cylinder walls were at least 20 mm away from the sphere walls in any direction. Thus, the size of the background cylinder was different for different sphere sizes. For example, the diameter of the background cylinder for the sphere with inner diameter of 13 mm was $13+2*(0.6+20)=54.2$ mm. Here 0.6 mm is the thickness of the
sphere wall. The height and the axial position of the background cylinder was such that it was 20 mm larger than the sphere in the axial directions. Since the simulated phantom was created from the CT images of the physical phantom, the weighting matrices of the GTM (or the sGTM), which reflect the geometrical relationship of the VOIs, were identical for both simulated and physical sphere phantoms. This approach along with identical PSFs and noise levels provided an opportunity to directly compare the simulation results with those of the physical phantom experiments.

4.3.1.2. Brain phantom simulation

A 3D MRI brain phantom based on the “BrainWeb” phantom [133] was used to simulate a 3D PET image. The isotropic voxel sizes were 1 mm and 2mm for MRI and PET respectively. Three VOIs were used: cerebrospinal fluid (CSF), grey matter, and white matter. The rest of the brain was assigned to be the background volume, thus a total of 4 VOIs were defined. An ideal PET image was created by assigning standardized uptake values (SUV) of 1, 9, 3, and zero to CSF, grey matter, white matter, and background respectively. This provided an ideal PET image in CT space, which was then down-sampled using tri-linear interpolation to PET voxel size. This image then was convolved with an isotropic 3D Gaussian PSF with a FWHM of 8.0 mm. The PSF was chosen to be close to that of the physical sphere phantom below (8.7x8.3x7.8 mm). However, since the geometry of the brain phantom is different than that of the physical phantom, no exact match of the PSF was implemented. Finally, a uniform voxel noise of 27% was added to create a simulated PET image. A total of 100 simulated 3D brain images were created each with a unique statistical noise pattern similar to the simulated sphere phantom.
Figure 4.3 illustrates the 4 mask images, the ideal PET image (in CT space), and a simulated PET image (in PET space). GTM and sGTM methods were used to obtain corrected uptake values \( T_f \) for all 4 VOIs by calculating the corresponding 4x4 weighting matrices.

4.3.2. Physical sphere phantom

A 20 cm diameter cylindrical plastic phantom was constructed containing 6 fillable spheres ranging in diameter from 5 to 30 mm. The wall thickness was 0.6 mm for all spheres. Figure 4.4 illustrates the physical phantom and its CT and PET images. A total activity of 1.11 mCi F-18 was injected into the water in the phantom to create a 3 to 1 activity concentration ratio between spheres and the tank. The phantom was scanned using a PET/CT GEMINI System (Philips medical system, Cleveland, Ohio). The CT and PET voxel sizes were 0.6 and 2 mm isotropic respectively. A total of 18 PET frames were scanned with a scan time of 5 minutes for each frame. The PET reconstruction algorithm was 3D filtered back projection (FBP). The PET PSF was measured using five F-18 point sources placed 3.5 cm apart along the axial FOV and at 7.5 cm off the center of transaxial FOV. The average fitted Gaussian FWHMs to PSF were 8.7, 8.3, and 7.8 mm in the X, Y, Z directions respectively. For each sphere size, the GTM or sGTM 3x3 weighting matrices were calculated for the 3 VOIs similar to the simulated sphere phantom above.

4.3.3. Image analysis

A computer program written in-house in interactive data language (IDL) Ver. 8.1 (Research systems inc., Boulder, CO) was developed to implement the GTM and sGTM methods. The software was used to display the 3D PET and CT images, draw the VOIs, calculate RSFs and
weighting factors \( (\omega_i) \), obtain PVC corrected uptake values \( (T_i) \), and plot the results. To compare the accuracy and precision of sGTM versus GTM method, recovery coefficients (RC) were calculated as the figure of merit [47]:

\[
RC = \frac{\text{measured activity within VOI}}{\text{true activity within VOI}}
\]

The figure of merit to evaluate noise propagation characteristics of the two PVC methods was the Noise magnification factor (NMF) (i.e. ratio of coefficient of variance after PVC to that before PVC). The NMFs were calculated using [87]:

\[
NMF = \frac{dT_i / T_i}{dt_i / t_i}
\]

The ideal values for RC and NMF are both unity, i.e. no loss of accuracy due to partial volume effect and no amplification of noise due to PVC. The RC values before and after PVC using GTM or sGTM were calculated with or without errors in PET-CT registration or errors in PET PSF measurements. To evaluate the effect of mis-registration on PVC, the PET image was shifted with respect to the mask images separately in X, Y, and Z directions before PVC. Mis-registration ranged from 0 to 10 mm. To assess the effects of PET PSF errors, PSFs with different FWHMs than those of the true PSFs were used during PVC. The true PSF was assumed to be either the measured PSF (for the physical phantom) or the PSF used to simulate PET images (for the simulated phantoms). Error factors in the range of 0.5 to 1.5, (i.e. +/-50 % error) in FWHM were applied for this purpose.

4.4. Results

Figure 4.5 demonstrates the accuracy and precision of RC with and without PVC using the GTM and sGTM methods. The RC results for the simulated sphere and brain phantoms are
illustrated in figures 5(a) and 5(b) respectively and the results of the physical sphere phantom is in 5(c). The graphs for the simulated and physical sphere phantoms (5(a), 5(c)) are for the inner sphere volume. For both the GTM and sGTM methods, the accuracy was within 5% for all three phantoms and for all sphere sizes. The only exception was the 5 mm sphere size for the physical phantom, where accuracy within 10% was obtained for the GTM method due to small object size. The error bars in figure 4.5 are significantly smaller for the simulated brain phantom than for the sphere phantoms. This result is due to the fact that VOIs of the brain have many more voxels than the spheres. Smaller objects are expected to be affected more by PET image noise. For all graphs in figure 4.5 no errors in PET-CT registration or in PET PSF measurements were applied.

Figure 4.6 compares the NMF values for the GTM and sGTM methods. The NMF results for the simulated sphere and brain phantoms are illustrated in figures 4.6(a) and 4.6(b) respectively and for the physical sphere phantom in 4.6(c). As expected, application of PVC magnifies noise and the NMF values are all larger than unity. However, in both simulation phantoms and in the physical phantom, noise propagation is smaller with the sGTM method compared to that of the GTM method, particularly for smaller objects. All plots in figure 4.6 are obtained without applying errors in PET-CT registration or in PET PSF measurements.

Figure 4.7 illustrates normalized RC values when PET-CT mis-registration was applied to the simulated phantoms (4.7(a) and 4.7(b)) and for the physical phantom (4.7(c)). The RC values were normalized to the RC values with no mis-registration error. The mis-registration was applied by shifting the VOI mask in the lateral (X) direction. Similar results were found for shifts in the other directions (Y and Z). For the simulated sphere phantom and for the physical sphere phantom, results for two spheres (13 and 30 mm) are demonstrated in figures 4.7(a) and 4.7(c). The results for other sphere sizes were similar. The results in figure 4.7 indicates that the sGTM method is more robust than the GTM method for mis-registration errors.
Figure 4.8 demonstrates normalized RC values when PET PSF FWHM error was applied to the simulated phantoms (4.8(a) and 4.8(b)) and for the physical phantom (4.8(c), 4.8(d)). The RC values were normalized to the case with no FWHM error. For the simulated sphere phantom and for the physical sphere phantom, results are plotted for two spheres (13 and 30 mm) in figures 4.8(a) and 4.8(c). Similar results were observed for other sphere sizes. Figure 4.8(d) illustrates normalized RC values for the physical sphere phantom for the case with 50% overestimation in FWHM as a function of sphere size. The results indicates that the sGTM method is more robust than the GTM method for errors in FWHM.

4.5. Discussion

A new, region-based PVC method was derived through analytical least squares fitting equations. Although the least squares approach is analogous to that of Labbe, the new method is region-based rather than voxel-based and thus avoids computational problems imposed by handling excessively large matrices for 3D PVC. While Labbe's method requires a linear equation for every voxel, the sGTM method only requires a linear equation for every region, similar to the conventional GTM method. The basic assumptions of both the GTM and sGTM methods are the same. Both methods require the measured PET PSF as well as segmented, co-registered images from CT (or MRI) to segment the PET image into different non-overlapping tissue types or VOIs. The impacts of region size and shape are automatically taken into account by the 3D convolutions to obtain the RSFs in both GTM and sGTM methods. The number of regions are not expected to affect these PVC methods for limited regions. However, if many regions are chosen (e.g. every voxel as a region), then noise amplification could be problematic similar to voxel based PVC methods. The linear equations in the GTM method solve the spill-in and spill-out from one VOI to another, while the sGTM method solves the spill-in and spill-out
from one RSF to another. This physical interpretation is interesting since RSFs add up to unity image (just as VOIs do) to satisfy the non-overlapping assumption and to fulfill no spill loss requirement. The data presented validate the sGTM method and provide a comparison of it to the GTM method with regard to accuracy, precision, noise characteristics, and robustness.

4.5.1. Accuracy

In the absence of errors in PET-CT registration or PSF measurement, the accuracy of the two methods is similar as indicated in figure 4.5. As expected, with no PVC, the accuracy of the recovery coefficient is worse for smaller objects than for larger objects, e.g. CSF versus white matter in figure 4.5(b).

4.5.2. Precision and noise characteristics

The results in figures 4.5 and 4.6 indicate that the precision and noise characteristics are improved with the sGTM algorithm compared to the GTM method for noisy PET images, particularly for small objects. The smaller RC error bars of the sGTM method relative to the GTM method in figure 4.5 translates to smaller NMF values in figure 4.6. This improvement is more pronounced for smaller object sizes (< 13 mm sphere diameter or CSF). The improvement in precision may be explained by the different ways that the sGTM and GTM methods handle each PET voxel value. The sGTM method uses a least squares fitting algorithm, thus penalizing voxels with random noise more than other voxels by taking the square difference of the expected voxel value and the measured voxel value. The GTM method, on the other hand, treats all the voxels the same by taking an average within a VOI. The superior noise characteristics of the least squares approach was also reported previously [129, 132] when the 2D GTM method was compared to the 2D Labbe's voxel-based PVC method.
The added noise level was 27% for the simulated phantoms, a value that was approximately the same as the voxel noise from the physical sphere phantom in test acquisitions. This noise level corresponds to noisy PET images where acquisition time is short (such as dynamic PET) or where little smoothing is applied during reconstruction.

4.5.3. Robustness

Application of a PVC method in clinical situations requires a robust method since clinical data are not ideal and accuracy may be compromised due to (a) registration errors or (b) errors in estimation of the PSF. The results demonstrated that sGTM method is more robust than GTM method in both situations.

4.5.3.1 Robustness to mis-registration errors

PET-CT (or PET-MRI) mis-registration errors may be caused by patient motion during the scan. Any mismatch between the CT (or MR) VOI that the transfer matrix weighting factors are based on and the PET VOI where the measured PET data are sampled, will translate into a loss of RC accuracy. For the GTM method, this effect has already been pointed out as a major source of error [89]. The results plotted in figure 4.7 illustrate that the sGTM method is more robust than the GTM method in this regard. This can be explained by noting how each method samples PET image data for a given VOI. Note that mis-registration error will not affect the weighting matrix which was calculated from the CT (or MR) image alone but rather the t-values where the PET image is sampled. The GTM method uses a VOI mask with sharp boundaries to sample the PET image. However, the sGTM method uses the RSF with fuzzy boundaries where the boundaries are blurred by the PET PSF. In effect, the sGTM method takes advantage of resolution loss in PET. As expected, for a given mis-registration error, the loss in RC accuracy is more significant.
for smaller objects compared to bigger objects as indicated in figure 4.7. For this reason, the advantage of the sGTM method over the GTM method is more pronounced for smaller objects (e.g. CSF more than white or grey matters). Similarly, for a given object size, this advantage is more pronounced for a wider PSF compared to narrower PSF (data not shown). This means that the poorer the PET resolution, the greater the advantage in using the sGTM method over the GTM method for PVC.

4.5.3.2 Robustness to PSF measurement errors

As mentioned earlier, inaccuracies in the PSF used in the PVC method will affect the accuracy of the RC values by means of changes in weighting matrices in both GTM and sGTM equations. Figure 4.8 demonstrates that the sGTM method is more robust than the GTM method given an error in estimating the PSF. An explanation for this observation may be realized by reflecting on equation (4.1). An error in estimation of the PSF will translate to an error in the RSF for a given VOI and not to the VOI mask itself. The sGTM method applies the RSF to both sides of equation (4.1), i.e. to both weighting factors and t-values. However, the corresponding equation for the GTM method [87] applies the RSF to one side of the equation (the weighting factors) and the VOI mask to the other side of the equation (t-values). Thus, one expects that the sGTM method will be more robust for errors in estimation of the PSF. This effect is seen for both simulated and physical phantoms in figure 4.8. Similar to the robustness issue for PET-CT mis-registration, this effect is more pronounced for smaller than for larger objects (13 mm vs. 30 mm spheres, or CSF vs. white and grey matters). Obviously, the selection of the tissue VOI depends on the clinical application. For instance, striatal PET imaging [89] is an example where small VOIs have to be included in the PVC method. As expected, the improvement in robustness is highlighted more for larger errors in PSF estimation than for smaller PSF estimation errors as seen in figures 4.8(a)-
(c). In other words, the bigger the error in estimation of PSF, the greater the advantage in using the sGTM method over the GTM method. Resolution lost (and thus RC values) due to the partial volume effect is well known to be object size dependant [113]. The results in figure 4.8(d) for the measured sphere phantom reveal an interesting observation: for a given error in estimation of PSF, the RC values for GTM are object size dependant as expected. However, the RC values for sGTM are more or less independent of the object size, although some dependency still exists. A similar observation was found for the simulated sphere phantom (data not shown).

4.5.4. Clinical significance

PET quantification accuracy loss due to partial volume error could have significant clinical impact. Following the discussions in previous chapters (sections 1.5.2, 2.4.3, and 3.4.4), here we provide an estimate on how such an error could translate to response assessment in advanced head and neck cancer patients. As per section 1.7.3 and figure 4.5, for typical head and neck vessel sizes (10 mm diameter), quantifying PET vessel uptake will result in accuracy loss of approximately 40% underestimation due to partial volume error. As per figure 3.2C this underestimation translates to approximately $1/0.6=1.67=67\%$ overestimation in an image-based SKA-M without a PVC. The clinical significance of errors of this magnitude can be assessed by generating a plot similar to figure 3.7. In this case, 67% overestimation will cause all the patients to be identified as responders resulting in a treatment management that is the same as the current treatment (figure 1.3), i.e. all the patients are assumed responders and undergo chemo-radiation and salvage surgery is offered to non-responders after chemo-radiation. Thus, without a PVC, the theoretical estimates of survival rate and morbidity are the same as those of current treatment (table 6.1), i.e. 50% survival rates, 33% xerostomia, 50% acute mucositis, and 46% dysphagia. Due to significant loss in PET quantification accuracy and its clinical impact, it is unlikely that
image-based SKA-M without a PVC would be attempted. Results presented in this chapter based on phantom data, demonstrated that PVC can potentially recover PET accuracy and thus provide an accurate response assessment with substantial clinical impact, i.e. 75% survival, 17% xerostomia, 25% acute mucositis, and 23% dysphagia. These theoretical estimates need to be evaluated in patient studies. Table 6.1 summarizes the estimated potential clinical impact of PVC on response assessment.

4.6. Future directions

The GTM method can be implemented in image space or in sinogram space [89, 111]. Here the sGTM method was implemented in image space due to ease of implementation. It is expected that a sinogram implementation of sGTM method will yield similar results to this study and this issue will be investigated in the next chapter. The FBP reconstruction algorithm was used for the current study. Iterative methods of reconstruction of PET images are used more commonly today and an extension of the sGTM method to iterative reconstruction algorithms is being investigated.

4.7. Conclusion

A new analytic PVC method for PET was proposed, validated, and its robustness was investigated. The sGTM method is mathematically equivalent to Labbe's method, however it avoids the computational penalty of handling large matrices. The sGTM method provides a weighting matrix similar to the conventional GTM method but it is based on a least squares fitting algorithm and relies on RSFs rather than VOI masks. The accuracy of the sGTM method is similar to that of the GTM method, however, the sGTM method has a better performance with noisy PET images. It is also more robust in practical situations where errors in PET-CT (or PET-
MRI) registration or errors in estimation of PET PSF exist. All three advantages of sGTM over GTM are more pronounced for smaller objects.

4.8. Appendix

In this appendix it is demonstrated that the sGTM method is mathematically equivalent to Labbe's PVC method [128], however, the sGTM does not require handling large matrices. Labbe's method has the same assumptions as the GTM or sGTM methods. Assuming that the measured PET image $I(r)$ having $M$ voxels is segmented to $N$ non-overlapping homogenous regions, applying the linearity property of convolution results in the following equation:

$$[A]_{M \times N} [x]_{N \times 1} = [b]_{M \times 1},$$

where $b$ is a vector containing $I(r)$ values for every voxel $r=1,2,\cdots M$, and vector $x$ contains unknown uptake values $T_j$ for each region $j=1,2,\cdots N$. Matrix $A$ is a large matrix and is composed of $RSF$ values for every region and every voxel. Each column of this matrix is an $RSF$ for a given region, Thus:

$$\begin{bmatrix}
    RSF_1(1) & RSF_2(1) & \cdots & RSF_N(1) \\
    RSF_1(2) & RSF_2(2) & \cdots & RSF_N(2) \\
    \vdots & \vdots & \ddots & \vdots \\
    RSF_1(M) & RSF_2(M) & \cdots & RSF_N(M)
\end{bmatrix}
\begin{bmatrix}
    T_1 \\
    T_2 \\
    \vdots \\
    T_N
\end{bmatrix}
= \begin{bmatrix}
    I(1) \\
    I(2) \\
    \vdots \\
    I(M)
\end{bmatrix}$$

Eq.(4.A1)

Since there are more equations ($M$) than unknowns ($N$), the best solution is the one that satisfies all the equations in a least squares sense. Labbe's approach was to use singular value decomposition [135] to compute $x = A^{-1}b$. The classical solution is

$$x = (A^T A)^{-1} A^T \cdot b,$$

Eq.(4.A2)

where $A^T$ is the transpose of matrix $A$. 

102
$A^T = \begin{bmatrix}
RSF_1(1) & RSF_1(2) & \cdots & RSF_1(M) \\
RSF_2(1) & RSF_2(2) & \cdots & RSF_2(M) \\
\vdots & \vdots & \ddots & \vdots \\
RSF_N(1) & RSF_N(2) & \cdots & RSF_N(M)
\end{bmatrix}$

If one denotes $[t] = A^T \cdot b$ and $[\omega] = A^T A$, then one notes that equation (4.A2) is equivalent to equation (4.1) due to the definition of matrix multiplication. Thus sGTM and Labbe’s methods are equivalent. However the implementation of the sGTM method does not require the formation of matrix $A$. 
Figure 4.1. Principals of GTM and sGTM partial volume correction techniques. (a) VOIs, corresponding RSFs, and a measured PET image of a vessel, (b) weighting matrix and t-values calculated using the GTM method, and (c) weighting matrix and t-values in equation (4.1) calculated using the sGTM method.
Figure 4.2. Simulated sphere phantom. A transverse slice of the ideal sphere phantom before down-sampling to PET resolution (a) and a simulated PET image of it (b). The ideal PET image is based on CT images of a physical phantom in figure 4.4 (b).
Figure 4.3. Simulated brain phantom. Different activities were assigned to 4 VOI masks in (a) to create an ideal PET image (b). The ideal PET image (in CT space) was then down-sampled to PET voxel size, convolved with the PET PSF, and noise added to create a simulated PET image (c).
Figure 4.4. Physical sphere phantom (a), a transverse slice of its CT (b) and PET (c) images.
Figure 4.5. Accuracy and precision of RC. While both methods produced similar average RC, note the larger standard deviations (see error bars) in the GTM method. (a) simulated sphere phantom, (b) simulated brain phantom, (c) physical sphere phantom. The green lines are the ideal RC. RC = recovery coefficient, PVC = partial volume correction, GTM = geometric transfer matrix, sGTM = symmetric GTM, csf = cerebrospinal fluid, gry = grey matter, wht = white matter, VOI = volume of interest.
Figure 4.6. The effect of noise on the two PVC methods. The noise magnification factor (NMF) was consistently smaller for the sGTM method, with the most pronounced effect for small objects. NMF for (a) simulated sphere phantom, (b) simulated brain phantom, and (c) physical sphere phantom. The green lines are the ideal NMF. NMF = noise magnification factor, GTM = geometric transfer matrix, sGTM = symmetric GTM, csf = cerebrospinal fluid, gry = grey matter, wht = white matter, VOI = volume of interest.
Figure 4.7. Sensitivity of the two PVC methods to mis-registration. Normalized RC values are plotted when PET-CT mis-registration is applied in the lateral direction to (a) simulated sphere phantom, (b) simulated brain phantom, (c) physical sphere phantom. The green lines are the ideal RC and the error bars are standard deviations. RC = recovery coefficient, GTM = geometric transfer matrix, sGTM = symmetric GTM, csf = cerebrospinal fluid, gry = grey matter, wht = white matter.
Figure 4.8. RC values when PET PSF FWHM error is applied. (a) simulated sphere phantom, (b) simulated brain phantom, (c) and (d) physical sphere phantom. The green lines are ideal RC and the error bars are standard deviations. RC =recovery coefficient, PVC =partial volume correction, GTM =geometric transfer matrix, sGTM =symmetric GTM, csf =cerebrospinal fluid, gry =grey matter, wht =white matter.
Chapter 5:

Improving PET quantification accuracy by implementing the novel partial volume correction technique in sinogram space

In the previous chapter, a novel partial volume correction technique was presented to improve PET quantification accuracy. The implementation in chapter 4 was performed in PET image space and using conventional filtered back projection algorithms. This chapter extends the partial volume correction in chapter 4 to PET sinogram space making it applicable to iterative reconstruction algorithms which are used virtually exclusively in modern clinical PET systems.

The work presented in this chapter has been submitted for publication in:

- M Sattarivand et al., Region-based partial volume correction techniques for PET imaging: sinogram implementation and robustness, Physics in Medicine and Biology
5.1. Introduction

In spite of continuous improvement in the instrumentation of positron emission tomography (PET), its spatial resolution still remains relatively low compared to anatomical imaging modalities such as magnetic resonance (MR) or computed tomography (CT). Failure to implement a partial volume correction (PVC) in quantitative PET imaging may result in significant bias in the estimate of regional radioactivity uptake [109, 110, 112]. The limited spatial resolution of PET is due to several factors that influence the image formation processes, including positron range, non-collinearity, detector width, and reconstruction filtering [136]. Two distinct effects are usually associated with the partial volume effect [113]. The first is the point response effect, which causes spill-over between different regions. This effect can be accounted for with a knowledge of the three-dimensional (3D) PET image formation processes or a measurement of the global PET point spread function (PSF). Usually, a fitted 3D Gaussian curve characterized by its full width half maximums (FWHMs) in the x, y, and z directions is used to estimate the global PET PSF. The second effect is the tissue fraction effect due to the coarse spatial sampling of the PET images, which may cause a single PET voxel to contain more than one tissue type with different tracer uptakes. With the availability of anatomical images in current PET-CT (or PET-MR) systems, this effect can be accounted for by segmented co-registered CT (or MR) images.

Two general categories for PVC implementations are voxel-based and region-based approaches [113]. Some voxel-based techniques require anatomical CT (or MR) co-registered images [119-122] and others in this category do not [88, 123, 124]. Voxel-based techniques most often do not require CT (or MR) image segmentation, which could be considered appealing in some applications. Nevertheless, since the correction is performed at the voxel level, the main disadvantage of these techniques is the noise amplification [113] which could be a limiting factor.
A region containing many voxels is less likely to be affected by image noise compared to a single voxel.

Region-based PVC approaches are mainly based on a key study by [87], in what is commonly referred as the geometric transfer matrix (GTM) method. In this method, the CT (or MR) images are segmented to different non-overlapping regions representing different tissue types. Using these regions, the GTM method corrects for both the PET tissue fraction effect and the point response effect in an analytical approach. The GTM method is attractive since it is straightforward to implement and it provides meaningful physical interpretations for spill-over between all regions. The GTM method indeed remains the most widely used PVC method [88] and is usually considered as the reference PVC method [89, 90]. Often new PVC methods, even those that are non-region-based, are evaluated in performance against the GTM method [88, 118, 124, 137, 138]. The GTM method was first implemented in two dimensional (2D) form [87]. However, an extension to 3D implementation was soon introduced [89]. The 3D implementation is important to account for the 3D acquisition mode used by most current PET systems. These GTM implementations require the convolution of the segmented regions in "image-space" with the PET PSF to obtain what are known as regional spread functions (RSFs).

The study by [89] also introduced another 3D implementation of the GTM method. This implementation was performed in "sinogram-space" rather than in image-space. Throughout this paper, the sinogram implementation of the GTM method is referred to as the GTMo method. The "o" here refers to the letter "o" in "sinogram". Rather than convolving regions as in the GTM method, the GTMo method requires a forward projection followed by the reconstruction of each region to obtain the RSFs. Although the GTMo method is computationally slower than the GTM method, the GTMo method more closely simulates the 3D image formation and acquisition processes of a physical PET system compared to the GTM method which uses a global 3D PSF.
Thus, the GTMo method automatically accounts for spatial variations in the PET PSF [113]. Moreover, only a sinogram-based method can be readily extended to the iterative reconstruction algorithms [131] which are now more commonly used in the clinic. The 3D implementation of GTMo by Frouin was adapted by some studies for clinical applications [139].

While accurate regional recovery is the main goal of any PVC, precision and noise propagation are important considerations in evaluating the performance of any PVC method. Performance can be further evaluated in terms of robustness to PET-CT (or PET-MR) mis-registration, as well as with regard to errors in PET PSF measurements. The study by [89] demonstrated that the performance of the GTMo method is similar to that of the GTM method in terms of accuracy, precision, and robustness with regard to registration errors. Robustness to PET-CT (or PET-MR) registration is of special concern if patient motion exists between the two scans. Both the GTM and GTMo methods are especially vulnerable to registration errors [89]. Mis-registration has been reported [90] to be the factor with the largest impact on the accuracy and precision of the GTM method.

Recently, a new region-based PVC has been reported [140] which is referred to as the symmetric geometric transfer matrix (sGTM) method. The accuracy of the sGTM method is similar to that of GTM, while it has better characteristics for noisy PET images in terms of precision and noise propagation. The sGTM method was also reported to be more robust than the GTM method, both in terms of registration errors and errors in PSF measurements. Similar to GTM, the sGTM method is attractive in a sense that it provides an analytical equation with meaningful physical interpretations for spill-over between all regions. The implementation of sGTM does not incur any additional computational cost compared to the GTM method.

The sGTM method referred to above was implemented in image-space, similar to the GTM method. No previous study has implemented the sGTM method in sinogram-space. As mentioned
previously, a sinogram implementation is of interest since it more closely simulates 3D PET image formation processes compared to the image-space implementation and it provides the other advantages noted above. The objective of this study was to implement and validate the sGTM method for 3D PET in sinogram-space. In this study we refer to this implementation as the sGTMo method. This study also compares the performance of the sGTMo method to previous region-based PVC methods, i.e. GTMo and sGTM. Two hypothesis are tested in this comparison: (a) just as GTMo was reported to be similar in performance to the GTM method [89], the first hypothesis is that the sGTMo method performs similarly to the sGTM method; and (b) since the sGTM was demonstrated to have performance advantages over the GTM method [140], the second hypothesis is that the sGTMo method would have similar performance advantages over the GTMo method.

5.2. Materials and methods

5.2.1. Principles of region-based PVC methods

The principles of region-based PVC methods are described in the literature [87, 140]. In this section, the implementation of four different region-based PVC methods are described, i.e. GTM, GTMo, sGTM, and sGTMo, that were used to obtain the results in this paper.

In short, high resolution CT (or MR) images are segmented into \( N \) tissue types with non-overlapping volumes of interest \( VOI_i \), \((i = 1 \ldots N)\). Uptake for each tissue type is assumed to be homogenous and the goal is to obtain this uptake value with a knowledge of the PET 3D PSF or the PET 3D image formation process. The mask image for each VOI contains binary voxel values of unity (if the voxel belongs to the tissue) or zero (if the voxel is outside the tissue). The \( N \) VOI masks are each interpolated from the CT (or MR) voxel size to the coarser voxel size of PET to
account for the tissue fraction effect and obtain PET-space VOI images. This makes the voxels near the boundary region to have values between 0 and 1.

In the GTM and sGTM methods, each PET-space VOI image is convolved in 3D with the PET PSF to obtain the RSFs, which are blurred versions of the corresponding tissue masks. In the GTMo and sGTMo methods, the global PET PSF is not used, and in order to obtain the RSFs three steps are required. First, the PET-space VOI images are forward projected with knowledge of the 3D PET scanner acquisition geometry. The resulting projection sinograms containing 3D line of responses (LORs) are then blurred. Finally, the projection sinogram for each VOI is reconstructed with a 3D reconstruction algorithm to obtain the corresponding RSF. These three steps, in effect, simulate the PET image formation process for each tissue volume separately. We implemented the first and last steps using the open source software for tomographic image reconstruction (STIR), details of which are described in section 2.4.2 below. The middle step (blurring) may be performed by convolving each LOR with the intrinsic PSF of 2 crystals [89]. We implemented this step in the Fourier domain by multiplying the Fourier transform (FT) of the sinograms with a Gaussian function such that the resulting 3D PSF from a simulated point source matched the measured PSF of the PET scanner.

Note that the RSF images are different for the image-based PVC methods (GTM and sGTM) versus the sinogram-based PVC methods (GTMo or sGTMo) due to essentially different methods of calculation that account for resolution loss in PET.

To perform a PVC, all four methods use the following equation:

\[
\omega_j [T_j]_{N \times 1} = [t_j]_{N \times 1}, \quad \text{Eq.}(5.1)
\]

where in GTM and GTMo methods [87]:

\[
\omega_j = \int_{\text{FOV}} \text{RSF}_i(r) \cdot \text{VOI}_j(r) dr \quad \text{and} \quad t_j = \int_{\text{FOV}} I(r) \cdot \text{VOI}_j(r) dr,
\]
and in sGTM and sGTMo methods [140]:

$$\omega_{ij} = \int_{\text{FOV}} \text{RSF}_i(r) \cdot \text{RSF}_j(r) \, dr \quad \text{and} \quad t_j = \int_{\text{FOV}} I(r) \cdot \text{RSF}_j(r) \, dr.$$ 

Here, $I(r)$ is the measured PET image with a given field of view (FOV) and $[\omega_{ij}]_{N \times N}$ is commonly referred to as the geometric transfer matrix, which contains weighing factors with a physical interpretation for each element. In the GTM and GTMo methods, $\omega_{ij}$ describes spill-over from one VOI mask to another. In the sGTM and sGTMo methods, on the other hand, this matrix is symmetric and $\omega_{ij}$ describes spill-over from one RSF to another. $[t_j]_{N \times 1}$ on the right side of the equation is where we sample the measured PET image for each region. In GTM and GTMo methods, this is performed by multiplying the PET image by the corresponding VOI mask, while in the sGTM and sGTMo methods, this task is performed by multiplying the PET image by the corresponding RSF image. $[T_j]_{N \times 1}$ is a vector containing estimates of true uptake values for each tissue type and in order to apply PVC, equation (5.1) can be solved by multiplying both sides of the equation by the inverse of $[\omega_{ij}]_{N \times N}$.

Figure 5.1 illustrates the steps of implementing PVC in sinogram-space (GTMo and sGTMo) for an image of a vessel having only two tissue types ($N=2$). Note that unlike RSFs obtained by convolution in GTM and sGTM, where voxel values are only between zero and one, RSFs obtained for GTMo and sGTMo may contain negative voxel values and streak artifacts as indicated in figure 5.1(e). This effect is the result of the reconstruction process, and may translate into the weighting matrices as illustrated in figure 5.1(g) and 5.1(h). However, this effect does not pose a problem to the calculation of weighting factors. This can be confirmed by adding all the RSF images and verifying that all the voxels are close to unity [140]:
\[
\sum_{i=1}^{N} RSF_i(r) = 1
\]

For the data presented in this paper, the deviations from unity were less than 0.5% for all voxels and thus we ignored this effect.

The performances of the four region-based PVC methods were evaluated by implementing them for two 3D simulated phantoms (a sphere phantom and a brain phantom) and one physical sphere phantom.

5.2.2. Simulations

5.2.2.1. Simulation of 3D sphere phantom

A 3D sphere PET phantom was simulated using the geometry obtained from a CT scan of the physical sphere phantom illustrated in figure 5.4(a). The spheres were fixed in a cylindrical tank and their sizes ranged in inner diameter from 5 to 30 mm and the wall thickness for all spheres was 0.6 mm. The CT voxels were isotropic, 0.6 mm on each side. The CT images of the phantom were contoured automatically using an in-house analytic sphere segmentation algorithm to obtain the location of the spheres and the outer walls relative to the tank. Using the CT geometry and voxel size, an ideal PET image was obtained by assigning the relative uptake of the sphere-to-background ratio of 3-to-1 and the wall uptake to zero as illustrated in figure 5.2(a). This CT-space ideal PET image was then down-sampled using a tri-linear interpolation algorithm to the PET voxel size to account for the tissue fraction effect. The PET voxel size was 2x2x3.15 mm³ in the X, Y, and Z directions respectively, the same as that of the physical phantom PET scan. Note that voxel size here is different than that in chapter 4, since two different reconstruction software were used (Philips software in chapter 4 and STIR in chapter 5). For the GTM and sGTM PVC methods, the ideal PET image was then convolved in 3D with the PET PSF and noise was added.
to obtain the simulated PET image. The PSF was Gaussian with FWHMs of 7.23, 7.14, and 6.65 mm in the X, Y, and Z directions respectively, matching the PSF for the physical sphere phantom. For the GTMo and sGTMo PVC methods the ideal PET image was forward projected, filtered, reconstructed using the 3D filtered back projection (FBP) algorithm, and noise was added to create a simulated PET image as demonstrated in figure 5.2(b). The forward projection and reconstruction parameters were chosen to match those of the physical sphere phantom. The details of these parameters are described in section 5.2.4.2 below. The added noise was uniform across the phantom, uncorrelated for all four PVC methods and had a Gaussian distribution with a standard deviation of 25% relative to the mean tank uptake. This noise level was chosen to match the voxel noise obtained from the physical phantom. A total of 100 3D PET images were simulated, each with different stochastic noise and the four PVC methods were applied to each 3D PET image. For all PVC methods, three VOIs were chosen, i.e. inner sphere volume, sphere wall, and the background volume, which resulted in 3x3 weighting matrices. The background volume for each sphere size was a cylinder in the tank around the sphere such that its dimensions were at least 20 mm larger than the outer sphere walls in all directions. The geometry and the parameters (reconstruction, noise, etc) of the simulated PET sphere phantom were chosen to be identical to those of the physical PET phantom. This approach provides an opportunity to directly compare the results for the simulated and the physical sphere phantoms.

5.2.2.2. Simulation of 3D brain phantom

A 3D PET brain phantom was simulated from the segmented images of a dedicated MRI head phantom made available by techniques described in a previous study [141]. Five VOIs were chosen from the Zubal phantom: right and left putamen, right and left caudate, skin and skeletal muscle, grey matter, and white matter. The choice of tissue volumes is typically based on the
research question involved and the characteristics of the tracer used. The VOIs for the brain phantom in this study are those of a previous study [89] for striatal brain PET imaging using 18F-L-dopa where small VOIs (e.g. putamen and caudate) are involved in the PVC. For this reason we chose Zubal phantom and not the BrainWeb phantom used in chapter 4, as BrainWeb phantom did not have putamen and caudate. The five VOIs were assigned relative uptake values of 4.5, 4.0, 1.0, 2.5, and 2.0 respectively. These uptake values were taken from [89] for the 18F-L-dopa tracer. The rest of the image volume was assigned to be the background VOI with a relative uptake of zero. Thus a total of six VOIs were assigned to create 6x6 weighting matrices for all four PVC methods. These six VOIs are illustrated in figure 5.3(a).

The ideal PET image in MR-space after assigning the uptake values is illustrated in figure 5.3(b). The MR voxel size was 1.1x1.1x1.4 mm in the X, Y, and Z directions respectively. The ideal 3D MR-space image was then down-sampled to match the voxel size of the PET image to account for the tissue fraction effect using a tri-linear interpolation algorithm. The PET voxel size was 2x2x3.15 mm³ in the X, Y, and Z directions respectively. For the GTM and sGTM PVC methods, the down-sampled PET image was then convolved in 3D with the PET PSF and noise was added to create simulated 3D brain PET images. The Gaussian 3D PSF had FWHM values of 7.23, 7.14, and 6.65 mm in the X, Y, and Z directions respectively. For the GTMo and sGTMo PVC methods, the down-sampled ideal PET image was forward projected, filtered, and reconstructed using the 3D FBP algorithm and then noise was added to create simulated 3D brain PET images. Figure 5.3(c) demonstrates the simulated PET image for the GTMo and sGTMo PVC methods. Similar to the simulated sphere phantom, a 25% uncorrelated voxel noise was added uniformly for all four PVC methods. A total of 100 3D PET images were created each with different stochastic noise and the four PVC methods were applied to each 3D image.
5.2.3. Physical sphere phantom

A physical sphere phantom illustrated in figure 5.4(a) was constructed that had six fillable spheres with inner diameters ranging from 5 to 30 mm and a wall thickness of 0.6 mm for all the spheres. The spheres were fixed in a cylindrical tank with a diameter of 20 cm and a height of 20 cm. The spheres and the tank were filled with F-18 radionuclide solution with a total activity of 40.7 MBq (1.1 mCi) such that the uptake ratio of sphere to background was 3-to-1 for all spheres. The phantom was then scanned with a Gemini PET-CT scanner (Philips medical system, Cleveland, Ohio). The reconstructed CT voxels were isotropic, of size 0.6 mm. A slice of the CT image is demonstrated in figure 5.4(b). The PET FOV was 256 mm and a total of 18 frames were acquired each with an acquisition time of 5 minutes. The acquired PET sinograms were reconstructed in STIR using 3D FBP as described in section 5.2.4.2 below. The reconstructed PET voxel size was 2x2x3.15 mm³ in the X, Y, and Z directions respectively. A slice of the PET image is illustrated in figure 5.4(c). The PET PSF was measured in air using five F-18 point sources and the STIR parameters used to reconstruct the PET images of the point sources were the same as those used to reconstruct the PET images of the sphere phantom. Gaussian fits were obtained in three orthogonal profiles to measure the FWHMs of PSF. The average FWHMs of the PSF were 7.23, 7.14, and 6.65 mm in the X, Y, and Z directions respectively. Similar to the simulated sphere phantom, a total of 3 VOIs, i.e. inner sphere volume, sphere walls, and the background volume were created, resulting in 3x3 weighting matrices to apply for all four PVC methods.

5.2.4. Image analysis

5.2.4.1. In-house IDL software
An in-house image analysis software toolkit was developed in interactive data language (IDL) version 8.1 (Research systems inc., Boulder, CO). This software was used to create the simulated PET images, display the 3D images, create and display VOIs, calculate the RSFs, implement the four PVC methods, and plot the results. The STIR routines for forward projection or reconstruction were called from the IDL software as needed (see below). For all PVC methods, accuracy, precision, noise propagation characteristics, and the robustness in terms of PET-CT (or PET-MR) mis-registration were evaluated.

The accuracy and precision were evaluated by calculating the mean and standard deviation of the recovery coefficient (RC) of a given PVC method using [47]:

\[
RC = \frac{\text{measured activity within VOI}}{\text{true activity within VOI}}.
\]

The ideal value for RC is unity, however RC before PVC might be smaller or larger than unity if the VOI is hotter or colder than its surrounding background.

To evaluate the noise propagation characteristics of the PVC methods, noise magnification factors (NMFs) were calculated, i.e. ratio of the coefficient of variation after PVC to that before PVC [87]:

\[
NMF = \frac{dT_i / T_i}{dt_i / t_i},
\]

were the \(T_i\) and \(t_i\) are as defined above for equation (5.1). The ideal value for NMF is unity. However, since PVC usually amplifies the noise, the value of the NMF is often greater than unity.

To evaluate robustness to mis-registrations, the CT (or MR) mask images were shifted with respect to the PET images before performing PVC and the values of RC were calculated after the shifts. Mis-registrations up to 10 mm were applied in the X, Y, and Z directions separately.
5.2.4.2. *STIR forward projection and reconstruction*

In order to perform the 3D forward projection and 3D reconstruction needed to implement the PVC methods in sinogram-space, routines from STIR release 2 [142] were used. These tasks require knowledge of the 3D PET detector geometry and STIR contains this information for a number of commercially available PET scanners including the Philips Allegro scanner used in this study. In order to call STIR routines, both image and sinogram data were converted to the interfile format [143] that is compatible with STIR.

In order to calculate RSFs for GTMo or sGTMo PVC methods in all three phantoms, VOI masks were forward projected and the sinograms were reconstructed using the STIR "fwdtest" and "fbp3drp" routines respectively. For the GTMo and sGTMo PVC methods in all three phantoms, the PET images were reconstructed using the "fbp3drp" routine. This routine is based on a 3D FBP algorithm [144] for a given PET detector geometry. The measured sinogram data from the physical phantom scan needed to be corrected for detector gaps in the Philips Allegro scanner prior to reconstruction. Since STIR does not provide a routine for this task, an in-house gap filling algorithm based on a previous study [145] was implemented. The measured sinogram data was further corrected for attenuation and scatter before reconstruction. The attenuation correction was performed in STIR based on the measured CT images of the phantom. The scatter correction performed in STIR was based on the single scatter simulation algorithm [146].

5.3. Results

Figure 5.5 demonstrates the results for accuracy and precision of the RC values with different PVC methods and without using PVC. All RC values in figure 5.5 are for the case without PET-CT (or PET-MR) mis-registration. Figures 5.5(a), 5.5(b), and 5.5(c) are the results for simulated
sphere, simulated brain, and the physical sphere phantoms respectively. Figures 5.5(a) and 5.5(c) are for the inner sphere VOIs. In general, the corrected RC for all brain and sphere VOIs and for all four PVC methods were within 5% of the ideal value. The only exceptions were for the smallest sphere size (5 mm diameter) for 2 PVC methods (GTM and sGTM) of the physical sphere phantom, although accuracy within 10% was still obtained even for this small object size. The variation in RC with image noise for the sGTM and sGTMo methods was less than that for the GTM and GTMo methods especially for smaller objects, as indicated by the error bars for each method. Note that the error bars in figure 5.5 are somewhat different than those in figure 4.5 in chapter 4. This is due to the fact that two different reconstruction software were used (Philips software in chapter 4 and STIR in chapter 5). Moreover, slightly different voxel noise was used (27% in chapter 4 and 25% in chapter 5). No statistical tests of difference in standard deviations were used in either chapter. The error bars for the simulated brain phantom in figure 5.5(b) are significantly smaller than those for the sphere phantoms (figure 5.5(a) and 5.5(c)). This effect is due to the fact that the brain VOIs had more voxels than the small sphere VOIs and noise is expected to affect smaller VOIs more than bigger VOIs.

Figure 5.6 demonstrates the noise propagation plots characterized by the NMF values. The NMFs in figure 5.6 are for the case with no errors in registration. The plots in figures 5.6(a), 5.6(b), and 5.6(c) are the results for the simulated sphere phantom, simulated brain phantom, and the physical sphere phantom respectively. Figures 5.6(a) and 5.6(c) are for the inner sphere VOIs. As expected, the NMF values for GTM were similar to those for GTMo method, and the values for sGTM were similar to those for sGTMo method. Moreover, the values of NMF for sGTMo were smaller than those of GTMo method indicating an improvement in noise propagation when the PVC matrix is symmetric, even when the PVC method is performed in the sinogram-space. The improvement in NMF is more pronounced for smaller objects.
Figure 5.7 demonstrates the normalized RC values when mis-registration is applied between PET-CT (or between PET-MR) images. The plots in figures 5.7(a), 5.7(b), and 5.7(c) are the results for the simulated sphere phantom, simulated brain phantom, and the physical sphere phantom respectively. Figures 5.7(a) and 5.7(c) are for the inner sphere VOIs.

Errors in registration were applied by shifting the CT (or MR) VOIs relative to the PET image before applying the PVC. All curves in figure 5.7 are plotted as a function of mis-registration in the lateral (X) direction. The results of mis-registration in other two directions (Y and Z) were similar (data not shown). The RC values in the figure were normalized to the RC values with zero mis-registration. Data for one small (13 mm diameter) and one large (30 mm diameter) sphere is illustrated in figures 5.7(a) and 5.7(c) to make the plots less cluttered. The results for other sphere sizes were similar (data not shown). The sGTM curves for the brain phantom and the GTM curves for all phantoms are not shown in figure 5.7 in order to clarify other curves in the figure. However, as expected in all phantoms, the curve for GTM was close to that of GTMo and the curve for the sGTM was close to that of sGTMo. The results demonstrate that the symmetric PVC method is more robust than the non-symmetric PVC method even if the PVC method is performed in the sinogram-space.

5.4. Discussion

In this study, the sGTMo PVC method, a sinogram implementation of sGTM, was implemented, validated, and its performance was compared to previously established region-based PVC methods. In order to test our two hypotheses, all four region-based PVC methods were applied to images of three different phantoms and their relative performance was evaluated in terms of accuracy, noise characteristics, and robustness with regard to PET-CT (or PET-MR) mis-registrations. A discussion of how the results reflect upon the two hypotheses is presented
below. In addition, situations where the sinogram implementation could be of interest are also discussed below.

5.4.1. Accuracy

The accuracy of the new sGTMo method is similar to all other region-based PVC methods. As illustrated in figure 5.5, with no PVC on PET images, accuracy is lost as expected especially for smaller objects. However, using any of the four PVC methods and in the absence of registration errors, the accuracy of recovered uptake measurements will generally be within 5%.

5.4.2. Noise characteristics

The results presented in figure 5.6 demonstrate that the noise characteristic of the sGTMo method is similar to those of the sGTM method, while it is improved compared to the GTMo method. This improvement is more pronounced for smaller objects as expected [140]. The precision (standard deviations) of the sGTMo method illustrated in figure 5.5 is similar to that of the sGTM method while in general smaller than that of the GTMo method. Similarly, this is more notable for smaller objects. Better precision in the RC value translates to a better noise propagation characteristics when the weighting matrix is symmetric, regardless of its method of calculation, i.e. image-based or sinogram-based.

The added noise for the simulations was uniform spatially across the image even though the uptake was not uniform within the image. This decision was based on a previous study [134] which demonstrated that for the FBP algorithm the noise is almost uniform even if the uptake is spatially non-uniform. The 25% noise level used in the simulations was the same noise level as obtained from the physical phantom. This corresponds to noisy PET images where the acquisition time is short (as for dynamic PET) or light filtration is applied during reconstruction. Note that
the noise in the physical phantom experiments was inherently Poisson noise at the projection level, while, for simplicity, noise was added at the post-reconstruction level for the phantom simulation experiments. The results in terms of correction sensitivity to noise level was similar for both physical experiments and simulations, suggesting that the differences in noise spectrum characteristics (i.e., one would expect a correlated image noise spectrum for noise added at the projection level) did not significantly change the interpretation of the results.

5.4.3. Robustness to registration errors

A practical PVC method for clinical applications requires robustness to PET-CT (or PET-MR) mis-registration in order to preserve accuracy. Errors in registration may be due to patient movement during or between the two scans. The loss of accuracy due to mis-registration has been regarded as a major source of error affecting region-based PVC methods [89]. Compared to other sources of error such as mis-segmentation or errors in PSF measurements, the mis-registration has the strongest impact on accuracy and precision of uptake recovery [90]. The results in figure 5.7 demonstrate that the sGTMo method is more robust than the GTMo method. Moreover, the sGTMo and sGTM methods are equally robust. The improvement in the robustness of sGTMo over GTMo is greater for smaller objects or tissue volumes. This was also the case when sGTM was compared to GTM in a previous study [140]. Thus, a symmetric implementation of the region-based PVC method is more robust to mis-registration than the non-symmetric implementation regardless of the image-based or sinogram-based approach.

It is interesting to point out that the registration errors do not affect the weighting matrices, but only the t-values on the right side of equation (5.1). As discussed previously [140], the improvement in robustness to mis-registration is due to the fact that the symmetric implementation samples the PET image (t-values) by applying the RSFs with blurry boundaries.
rather than by utilizing the VOIs with sharp boundaries. Sampling the PET image with RSFs, in effect, takes advantage of resolution loss in PET. This implies that, for a given object size, the more resolution loss in PET, the more advantage in using the symmetric region-based PVC (sGTM or sGTMo) than the non-symmetric PVC (GTM or GTMo). Thus, the 3D implementation of the PVC (compared to 2D) is important to realize the advantage of improvement in robustness.

5.4.4. The need for sinogram implementation

5.4.4.1. Speed of implementation

In the image-based PVC methods (GTM and sGTM) every VOI is convolved in 3D with the PET PSF to obtain the RSFs and thus the spill-over information between the regions. On the other hand, in the sinogram-based PVC methods (GTMo and sGTMo) every VOI is forward projected and then reconstructed to get the same information on spill-over. Unlike image-based PVC, the sinogram-based approach intrinsically accounts for local variations in the spatial resolution in non shift-invariant systems [113]. If the PSF is shift-invariant, the convolution step for the image-based PVC can be performed in Fourier domain using the fast Fourier transform (FFT) [147] which significantly speeds up the PVC implementation. Thus, the fast implementation of the image-based PVC is its major advantage over the sinogram-based PVC [89].

However, if the PSF is non-shift-invariant, the FFT cannot be used for convolution. Although in this case in principle it is possible to perform the convolution in the image domain, in practice this approach is not used for various reasons. For example, one needs a fully measured and characterized PSF for all points in the space. Moreover, it requires an extensive computational cost in terms of memory and speed. Thus, the advantage of a fast implementation for image-based PVC approach may be lost in non-shift-invariant systems. Non-shift-invariant behavior
could be caused by detector parallax [48] which is more pronounced in small animal PET systems [115] due to small bore size. Another example is the head-dedicated PET systems in multi-slice configuration, where more than 40% difference exists between the axial PSF FWHM in the peripheral FOV compared to that of the central FOV [148].

Once the RSFs are calculated using either convolution or forward projection followed by reconstruction, the remaining implementation steps are based on equation (5.1) for all four PVC methods. This implies that the implementation of sGTM does not add any additional computational cost to the GTM method. Similarly, the implementation of sGTMo is just as fast as the GTMo method.

5.4.4.2. Extension to iterative reconstructions

Due to their advantage over conventional analytic FBP algorithms, iterative reconstructions are commonly used today in the clinic for PET systems. Through repetitive forward projection and reconstruction techniques, iterative reconstructions model physical and statistical processes of photon production and detection more accurately and thus improve the image quality [49]. In terms of region-based PVCs, only a sinogram-based PVC method is readily extendable to iterative reconstructions [131]. In this method, the transfer matrix weighting factors are calculated through a perturbed forward projection and reconstruction technique, an approach which is not feasible for image-based PVC using the convolution technique. This is an advantage of the sinogram-based PVC over the image-based PVC.

5.5. Future directions

The sinogram-space implementation of a region-based PVC method (GTMo) has been extended to iterative reconstruction algorithms [131]. These algorithms are now commonly used
in the clinic due to their advantages over conventional FBP algorithms. An extension of the sGTMo method to iterative reconstructions is of great interest and will be investigated as future work.

5.6. Conclusion

A sinogram-space implementation of the symmetric region-based PVC method (sGTMo) was implemented and validated using two 3D digital phantoms and one physical phantom. The performance of the sGTMo method was compared in terms of accuracy, noise characteristics, and robustness to registration errors to previously established region-based PVC methods, i.e. sGTM and GTMo. The results confirm our two hypotheses that sGTMo method is similar in performance to the sGTM method while its performance is improved compared to the GTMo method.
Figure 5.1. Principals of GTMo and sGTMo partial volume correction techniques. High resolution CT (or MR) image of a vessel is segmented to create two VOI masks (a). These VOI masks are then interpolated to the coarser PET voxel size to account for the tissue fraction effect (b). Note that unlike VOI masks in (a), voxel values in (b) may have values between zero and one in the boundary region. The PET-space VOI image is then forward projected using the 3D PET scanner geometry (c). The projection image is blurred in sinogram space (d). The sinograms are then reconstructed with a 3D PET reconstruction algorithm to obtain RSF images (e). With a given measured PET image (f) and calculated RSF images, parameters of equation (5.1) can be obtained for GTMo (g) or sGTMo (h) PVC methods.
Figure 5.2. Simulated sphere phantom. A transaxial slice of an ideal 3D PET image (in CT-space) is illustrated in (a) which was created from the physical sphere phantom in figure 5.4(a). A transaxial slice of the simulated 3D PET image is demonstrated in (b) which was created for the GTMo and sGTM PVC methods.
Figure 5.3. Simulated brain phantom. Five VOIs plus background (a) were chosen from the Zubal brain phantom. These VOIs were assigned different relative uptake values to create an ideal PET image in MR-space (b). The ideal PET image was then down-sampled, forward projected, filtered, and reconstructed to create a simulated PET image (c) for GTMo and sGTMo PVC methods.
Figure 5.4. Physical sphere phantom (a) and a transaxial slice of CT (b) and PET (c) images of it.
Figure 5.5. Accuracy and precision of RC values using four PVC methods. The plots are for the simulated sphere phantom (a), simulated brain phantom (b), and the physical sphere phantom (c). The error bars are standard deviations. The ideal values for RC are indicated as green lines. RC = recovery coefficient, PVC = partial volume correction, GTM = geometric transfer matrix, sGTM = symmetric GTM, GTMo = GTM implemented in sinogram-space, sGTMo = sGTM implemented in sinogram-space, VOI = volume of interest, puta = putamen, caud = caudate nucleus, gry = grey matter, wht = white matter
Figure 5.6. Noise propagation characteristics of four PVC methods. The ordinates are noise magnification factors (NMF) for the simulated sphere phantom (a), simulated brain phantom (b), and the physical sphere phantom (c). The ideal values for NMF are indicated as green lines. The NMF values are smaller for sGTM and sGTMo compared to those of GTM and GTMo methods specially for smaller objects. NMF = noise magnification factor, GTM = geometric transfer matrix, sGTM = symmetric GTM, GTMo = GTM implemented in sinogram-space, sGTMo = sGTM implemented in sinogram-space, VOI = volume of interest, puta = putamen, caud = caudate nucleus, gry = grey matter, wht = white matter
Figure 5.7. The robustness of the PVC methods to registration errors. Normalized RCs are plotted as a function of registration error. The misregistration was applied in the lateral directions to (a) the simulated sphere phantom, (b) the simulated brain phantom, and (c) the physical sphere phantom. The error bars are standard deviations. The ideal values for RC are indicated as green lines. RC = recovery coefficient, GTM = geometric transfer matrix, sGTM = symmetric GTM, GTMo = GTM implemented in sinogram-space, sGTMo = sGTM implemented in sinogram-space, VOI = volume of interest, puta = putamen, caud = caudate nucleus, gry = grey matter, wht = white matter.
Chapter 6:

Conclusions and future work
6.1. Thesis summary

The objective of this thesis was to evaluate the accuracy of current PET quantification methods for assessing tumour response in advanced head and neck cancer patients and to develop a novel partial volume correction technique to improve the accuracy of PET quantification. In section 1.5.2 (chapter 1), the clinical significance of an adaptive treatment approach based on an accurate response assessment was estimated in comparison to the current treatment approach (table 6.1). Clinical significance was evaluated in terms of patients' treatment management, survival, and morbidity. The required accuracy of PET quantification for response assessment was estimated to be 10% in this thesis. This choice was based on previous studies of PET uncertainties described in the introduction chapter (section 1.8). Three specific aims were identified in this thesis. The first specific aim was to evaluate the variability of a current static acquisition PET tracer uptake quantification method which uses fixed size ROIs for response assessment in advanced head and neck cancer patients. The second specific aim was to evaluate the accuracy of a current simplified dynamic PET quantification method for response assessment in advanced head and neck cancer patients, which depends on the accuracy of activity concentration measurements of a single blood sample. The third specific aim was to develop a novel partial volume correction technique to improve PET quantification accuracy. The three specific aims were addressed in different chapters of this thesis.

Chapter 2 addressed the first specific aim. Derivation of standardized uptake values from static PET acquisitions using a small fixed size ROI is a common PET quantification method for assessing tumour response. However, this chapter demonstrated that this approach is prone to significant variations depending on the ROI method selected. The earlier ROI method recommended by the European organization for research and treatment of cancer (EORTC) [84] was ROI\_same, while the current recommendations by PET response criteria in solid tumours
(PERCIST) [61] favours $\text{ROI}_{\text{peak}}$ (over $\text{ROI}_{\text{max}}$). The results in chapter 2 demonstrated that the difference between the current recommendations ($\text{ROI}_{\text{peak}}$) and earlier recommendation ($\text{ROI}_{\text{same}}$) could be substantial (e.g., more than 35% in some cases). The variations resulted in ambiguous tumour response assessment in 19% of the cases in the advanced head and neck cancer patients from the OCC trial. Section 2.4.3 in chapter 2 suggested that a relatively small difference in the ROI method used could result in mis-classification of 15% of the patients, in turn resulting in significant clinical impact in terms of patient's treatment management, survival, and morbidity (table 6.1). Quantification methods based on static PET acquisition using a fixed size ROI have the advantage of simplicity and improved statistical properties over $\text{ROI}_{\text{max}}$. However, simplicity is not always an advantage and static PET quantification using a small fixed size ROI should be approached with caution in heterogeneous tumours.

Chapter 3 addressed specific aim 2. It evaluated the effect of inaccuracies in blood sample concentration measurements on a simplified dynamic PET quantification (SKA-M) method in advanced head and neck cancer patients from the OCC trial. This chapter demonstrated that significant errors in both blood activity (10%) and blood volume measurements (12%) using the measurement techniques employed in the OCC trial exist, which were translated to significant errors in $\Delta$SKA-M estimate of tumour uptake rate change due to therapy. Section 3.4.4 in chapter 3 demonstrated that inaccuracies in blood sample concentration measurements could potentially lead to inaccuracies in treatment response assessments, in turn leading to mis-classifying 15% of the patients as responders and 8% as non-responders. These mis-classifications could have significant impact on patients' treatment management, survival, and morbidity, as summarized in table 6.1. The results in this chapter provided support for the development of a direct image-based simplified dynamic PET quantification method that does not rely on patient's blood sample. However, an image based approach requires a partial volume correction technique due to limited
PET spatial resolution. Without partial volume correction, significant errors in PET quantification accuracy (40% underestimation) are encountered for typical vessel sizes (10 mm in diameter) in the head and neck area.

Chapters 4 and 5 addressed specific aim 3. To recover the loss in PET quantification accuracy, in chapter 4 a novel partial volume correction technique was developed, validated, and its robustness was investigated. PET quantification accuracy of this novel partial volume correction technique was within 5% for objects greater than 5 mm in diameter. In comparison to previously published partial volume correction techniques, the new technique in chapter 4 performed better with noisy PET images and it was more robust in practical situations where errors in PET-CT registration or errors in estimation of PET point spread function exist. Section 4.5.4 in chapter 4 demonstrated that response assessment without a partial volume correction has major clinical impact as it would result in labeling all patients as responders (table 6.1). However, results of phantom data, demonstrated that partial volume correction can potentially recover PET quantification accuracy and thus provide an accurate response assessment with substantial clinical impact in terms of treatment management, survival, and morbidity.

The partial volume correction technique of chapter 4 was implemented (and validated) in sinogram space in chapter 5. The results in this chapter demonstrated that the sinogram implementation has the same advantages as the image-space implementation used in chapter 4 in terms of achieving accurate PET quantification (5%), better performance with noisy PET images, and improved robustness to PET-CT registration. However, the sinogram implementation provided additional advantages such as automatically accounting for local variations in PET point spread function and more importantly applicability to iterative reconstructions which are exclusively used in modern clinical PET systems.

In summary, the major original contributions of this thesis to knowledge are:
(1) Evaluation of the accuracy of current PET quantification methods for assessing tumour response in advanced head and neck cancer. Assessing response in patients from the OCC trial using current PET quantification methods (either static method or blood sample-based simplified dynamic method) resulted in significant errors or uncertainties (more than 10%).

(2) Improvement of PET quantification accuracy by implementing novel and robust partial volume correction techniques. Image-based simplified dynamic PET quantification requires partial volume correction to recover the loss of accuracy. With the addition of the presented partial volume correction, it appears feasible to achieve accuracy in PET quantification of 10% or better for vessels larger than 5 mm in diameter.

A limitation of this work is that the partial volume correction techniques were performed in simulated and physical phantoms only. A natural extension of this thesis is to apply these techniques to patients. Moreover, an accurate and robust method to segment blood vessels on CT needs to be employed. Several key considerations are outlined below for future work.

6.2. Future directions

6.2.1. Simplified dynamic PET quantification using an image-based input function

The current method of deriving a quantitative metric for assessing response based on simplified dynamic PET analysis requires a blood sample from the patient to estimate the input function. There are problems with this approach as discussed in section 1.7.2 of chapter 1. Data was also provided in chapter 3 to support some aspects of these problems. A non invasive, image-based approach to obtain such a metric is of great interest to address these problems. In order to apply such an approach to head and neck cancer patients, partial volume corrections need to be applied to restore the loss of accuracy in PET quantification due to small vessel sizes in the head
and neck area and limited PET spatial resolution. Thus, a non-invasive, image-based simplified
dynamic PET quantification method would be a natural extension to this thesis work.

The workflow that connects different chapters in this thesis needs to be extended as a future
work and to develop an image-based dynamic PET quantification with a robust PVC technique.
Then, the image-based simplified dynamic PET quantification needs to be applied for the patients
in the OCC clinical trial and to evaluate its clinical prognostic value. Other parameters also need
to be optimized. For instance, the method of ROI for the dynamic PET may not be the same ones
as those discussed in chapter 2. Moreover, a robust and accurate method of segmenting vessels in
CT needs to be implemented. The patient outcome from the OCC trial is the key to optimize the
parameters.

Image-based simplified dynamic PET quantification was implemented for a sample patient
from the OCC trial by applying the partial volume correction technique in chapter 4. The steps
required to obtain the image-based input function, and hence purely image-based response
metric, are described in figure 6.1. Note that the VOIs (i.e. carotid artery and jugular vein) needed
for the partial volume correction can readily be contoured in CT images especially if the CT
images are contrast based as per our patients. Note that both carotid artery and jugular vein can
be combined to increase the vessel VOI volume and improve the PVC accuracy. This is justified
since after 30 minutes post injection, the blood is mixed and similar activities in both vessels are
expected. The final results are summarized in table 6.2. No firm conclusions can be made at this
point with these results. However, they demonstrate that dynamic PET quantification using an
image-based input function provides results similar to those obtained from dynamic PET
quantification which used a blood sample based input function for this patient. Unlike the blood
sample derived input function based approach, the image-based input function method has a key
advantage that errors in PET scanner calibration will not affect the results of dynamic PET quantification as it cancels out in the ordinate in figure 6.1(b).

6.2.2. Partial volume correction for iterative reconstructions

The proposed partial volume correction techniques in this thesis enable the use of image-based input functions for simplified dynamic PET quantification in the current OCC trial. Furthermore, the techniques establish a framework for future research to address the inherent low spatial resolution of PET. One area of future research is to extend the partial volume correction technique of this thesis (which was implemented for filtered back projection algorithms) to iterative reconstruction algorithms.

Currently, iterative reconstructions are used virtually exclusively and have largely replaced filtered back projection algorithms. Although iterative reconstructions require more computational power, advancements in the speed of modern computers have led to their widespread clinical adoption. Through repetitive forward projection and back projection steps, the iterative reconstructions improve PET image quality by incorporating accurate physical and statistical modeling of photon production and detection processes [49].

The sinogram implementation of the partial volume correction in chapter 5 can be extended to iterative reconstructions with an approach similar to a previous study [131]. Quantitative imaging using both PET and SPECT could potentially have substantial benefit from this approach.

6.3. Concluding remarks

For early tumour response assessment, while functional imaging such as PET is preferred over anatomical imaging such as CT, the high resolution anatomical information in CT images was used in this thesis to identify regions in order to correct for inherently low spatial resolution of
PET. This is a clear example of using hybrid imaging where a complimentary strength from one modality is exploited to compensate for a limitation of the other modality.

Hybrid imaging techniques such as PET-CT and SPECT-CT have made advancements in fusion imaging for research and clinical applications. The recent introduction of PET-MR is opening new horizons of offering simultaneous morphologic, functional, and molecular imaging of a living system. The advancements in hybrid imaging along with progress in the field of biomolecules and particles to provide multi-modality agents, offer great promise in clinical oncology and basic medical biophysics research. Although significant advances have been made in molecular multi-modality imaging [149], many challenges still exist for future research [150].

Finally, while this thesis has not resulted in a definitive PET-based method of treatment response assessment, it has clarified that a technique of discriminating between responders and non-responders early in the treatment of advanced head and neck cancer, would have significant clinical benefit to patients, even if that technique were only moderately accurate. It is hoped that such a technique is soon developed and applied to increase the likelihood of survival for these patients as well as reduce treatment morbidity.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Adaptive treatment based on inaccurate response assessment due to ROI method:</th>
<th>Change in Tx management due to inaccurate PET quantification</th>
<th>Survival</th>
<th>Morbidity (due to radiation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5.2</td>
<td>Current treatment: CRT to all patients then salvage surgery to non-responders (figure 1.3)</td>
<td>50%</td>
<td>33%</td>
<td>50%</td>
</tr>
<tr>
<td>2</td>
<td>2.4.3</td>
<td>Adaptive treatment based on inaccurate response assessment: CRT to responders and early surgery to non-responders</td>
<td>75%</td>
<td>17%</td>
<td>25%</td>
</tr>
<tr>
<td>3</td>
<td>3.4.4</td>
<td>Adaptive treatment based on inaccurate response assessment due to ROI method: ROIpeak is true, but ROIsame is used, 15% mis-classified as responders</td>
<td>15% of patients complete ineffective CRT followed by salvage surgery</td>
<td>71%</td>
<td>21%</td>
</tr>
<tr>
<td>4, 5</td>
<td>4.5.4</td>
<td>Adaptive treatment based on inaccurate response assessment due to inaccurate blood activity concentration: 15% mis-classified as responders and 8% mis-classified as non-responders</td>
<td>15% of patients stop effective CRT and offered early surgery</td>
<td>68%</td>
<td>12%</td>
</tr>
<tr>
<td>4, 5</td>
<td>4.5.4</td>
<td>Adaptive treatment based on inaccurate response assessment due to lack of PVC: All patients are classified as responders</td>
<td>All patients receive CRT (same as current Tx)</td>
<td>50%</td>
<td>33%</td>
</tr>
<tr>
<td>4, 5</td>
<td>4.5.4</td>
<td>Adaptive treatment based on accurate response assessment: Potential improvements due to PVC</td>
<td>75%</td>
<td>17%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 6.1. Estimates of clinical impact of response assessment in advanced HNC. CRT = chemo-radiation treatment, Tx = treatment, ROI = region of interest, PVC = partial volume correction, HNC = head and neck cancer.
Figure 6.1. Steps of calculating image-based SKA-M. Plot (a) demonstrates the activity concentrations of tumour and blood vessels. The curve labeled as 'vessel' is the combined volume of carotid artery and jugular vein before partial volume correction. After partial volume correction, the time activity curve is labeled as 'vessel_PVC'. A scaled population average time activity curve to 'vessel_PVC' is plotted in dashed line. Plot (b) is the SKA-M graph obtained using tumour and blood time activity curves. PVC = partial volume correction, SKA-M = simplified kinetic analysis - multiple time points.
Table 6.2. Image-based vs blood sample based SKA-M for a sample patient from the OCC trial. The slope of the SKA-M plot in figure 6.1 (b) is Ki.
References


