2010

DEVELOPMENT OF MULTI-DIMENSIONAL X-RAY COMPUTED TOMOGRAPHY MEASUREMENTS OF LUNG TUMOURS

Laura Claire Rebecca Wilson

DEVELOPMENT OF MULTI-DIMENSIONAL X-RAY COMPUTED TOMOGRAPHY MEASUREMENTS OF LUNG TUMOURS

(Spine Title: Volumetric Tumour Response Measurements)

(Thesis Format: Integrated Article)

By

Laura Claire Rebecca Wilson

Graduate Program in Medical Biophysics

Submitted in partial fulfillment
of the requirements of the degree of
Master of Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Certificate of Examination

Supervisor

Grace Parraga, Ph.D.

Supervisory Committee

Glenn Bauman, M.D., FRCPC

Aaron Fenster, Ph.D., FCCPM

Elaine O'Riordan, M.D., FRCPC

Examiners

Shou Li, Ph.D.

Richard Malthaner, M.D., M.Sc., FRCSC

Eugene Wong, Ph.D.

The thesis by

Laura Claire Rebecca Wilson

entitled

Development of Multi-Dimensional X-Ray Computed Tomography Measurements of Lung Tumours

is accepted in partial fulfilment of the
requirements for the degree of
Master of Science

Date: _____

Chair of Thesis Examination Board

Abstract

Metastatic disease is the most common cause of cancer-related deaths. Two established image analysis tools, the World Health Organization Handbook for Reporting Results for Cancer Treatment (2D measurement), and Response Evaluation Criteria in Solid Tumours (1D measurement), have been used to quantify metastatic tumour burden in vivo. Limitations of the 1D and 2D measurements may be addressed using a 3D technique. The overall objective of this thesis was to determine the accuracy and reproducibility of a 3D measurement technique to be used as a potential imaging biomarker to quantify pulmonary metastases in vivo, using x-ray CT. We compared the accuracy and reproducibility of our 3D technique to the 1D and 2D measurements using lung tumour phantoms of known dimensions and seven subjects with pulmonary metastases. Three-dimensional measurements accurately quantified spherical and irregularly-shaped tumour phantoms ($p<0.05$), and most observers measurement patient metastases with high intra- and inter-observer reproducibility ICC>0.900.

Keywords: Response Evaluation Criteria in Solid Tumours; World Health Organization; pulmonary metastases; three-dimensional tumour measurements; x-ray computed tomography

Co-Authorship

The following thesis contains a manuscript prepared for submission for publication to a peer-reviewed scientific journal. Chapters 2 and 3 form the basis of an original manuscript in preparation for submission to the journal *Medical Physics* entitled, "Reproducibility of Multi-Dimensional Measurements of Pulmonary Metastases from X-Ray Computed Tomography" which is coauthored by L. Wilson, J. McCallum, E. O'Riordan, R. Etemad-Rezai, M. MacKenzie, H. Keller, G. Bauman and G. Parraga. Laura Wilson, the Master's candidate performed all the analyses on CT images. She analyzed all results, performed all statistical analysis, and prepared first drafts of the manuscript. Dr. Grace Parraga, the candidate's supervisor, conceptualized the project, the experimental approach and manuscripts, provided guidance, reviewed the results and statistical analysis, provided mentorship and editorial assistance. Jessica McCallum performed analyses and assisted in the preparation of manuscript drafts. Dr. Elaine O'Riordan and Dr. Roya Etemad-Rezai performed analyses, provided clinical expertise, and helped in the interpretation of the results. Dr. Mary MacKenzie provided the database of subjects with renal cell carcinoma and lung metastases used in Chapter 3 from Victoria Hospital, and performed analyses. Dr. Harald Keller provided the chest phantom images used in Chapter 2, acquired at Princess Margaret Hospital, and provided all ground truth measurements. Dr. Glenn Bauman provided guidance with the conceptualization of the manuscript, provided clinical expertise and helped in the interpretation of the results.

To my mother, Claire Wilson, my father, Lascelles Wilson, Sr., my sister, Louise Wilson, and my brothers, Lascelles Wilson, Jr., and Luke Wilson; and most importantly, to my nieces, Mekayla and Deja, and my nephew, Ryan.

"I do not feel obliged to believe that the same God who has endowed us with sense, reason, and intellect has intended us to forgo their use." – Galileo Galilei

Acknowledgements

I would like to thank my supervisor, Dr. Grace Parraga, for her guidance, support, and mentorship throughout my graduate studies. Her enthusiasm and commitment were vital as she encouraged me to think logically and critically, and perform to the best of my ability. I would also like to thank my advisory committee and collaborators, Dr. Glenn Bauman, Dr. Elaine O'Riordan, Dr. Aaron Fenster, Dr. Roya Etemad-Rezai, and Dr. David McCormack for their scientific contributions and clinical insights, support and enthusiasm, and keen interest.

I would also like to acknowledge the trained and expert observers who spent many hours and great effort performing analyses of the datasets: Andrew Wheatley, Lindsay Mathew, Jessica McCallum, Fred Seale, Sarah Svenningsen, Dr. Mary Mackenzie, Dr. Roya Etemad-Rezai, and Dr. Elaine O'Riordan. Thank you to Dr. Harald Keller and Brandon Driscoll at Princess Margaret Hospital in Toronto, ON, for providing the phantom database in Chapter 2.

Many thanks are due to the clinical staff and lab support at Robarts and at Western. In the Parraga Lab, I would like to thank Sandra Halko for her help regarding ethics and protocols, Andrew Wheatley for technical assistance, and Chris Piechowicz, and Shayna McKay for their assistance will all lab-related issues and inquiries. Special thanks are owed to Igor Gyacskov and Joseph Awad for their technical support. Financial support is greatly appreciated from the Western Graduate Research Stipend (WGRS) from the Schulich School of Medicine and Dentistry.

There were many people at Western who have made my two years in London enjoyable, and have contributed to my success. Most importantly, I would like to thank my fellow lab members, Hassaan Ahmed and Miranda Kirby, for sharing in every triumph as well as their commiseration in the darkroom and in class; Lindsay Mathew, for her probing questions and encouragement; Amir Owrangi for his correspondence while at PMH, as well as his sense of humour; and former Parraga Lab members, Adam Krasinski and Christiane Mallett for providing tips and tricks necessary to successfully complete this thesis. I would also like to thank the rest of my friends in the Department of Medical Biophysics. Many thanks to Mary Owusu, Ashley MacGregor, and Yordanos Woldemariam, for always showing me a good time when I was in London on the weekends.

I would like to express my sincere gratitude to those who have provided support from afar: many thanks to Jane Marum, Peta-Gay Williams, Janine Chin, and Stephanie Jones in Toronto, and Crystal Ward, Natasha Crenshaw, and Cassia Carson, in Texas and Alabama, who provided daily words of encouragement and good laughs, and helped me stay focused. Last, but certainly not least, I would like to thank my family, for their unconditional support and endless motivation: to my parents, Lascelles Sr. and Claire, my siblings, Lascelles Jr., Louise, and Luke, and my greatest sources of inspiration, Mekayla, Deja, and Ryan, my nieces and nephew.

Table of Contents

Chapter 2: Accuracy and Reproducibility of Multi-Dimensional X-Ray Computed Tomography Measurements of Pulmonary Tumour Phantoms

Chapter 3: Multi-Dimensional X-Ray Computed Tomography Measurements of Pulmonary Metastases in Patients

Chapter 4: Conclusions and Future Work

List of Tables

List of Figures

Appendix C:

Appendix E:

List of Appendices

List of Abbreviations

ANOVA	Analysis of variance
CR	Complete response
CT	Computed tomography
CV	Coefficient of variance
DICOM	Digital Imaging and Communications in Medicine
DR	Delayed scan region of interest
ER	Early scan region of interest
HASTE	Half-Fourier acquisition single-shot turbo spin echo
ICC	Intraclass correlation coefficient
keV	kiloelectronvolts
kVp	kilovolt peak
mA	milliamperes
MDCT	Multidetector row computed tomography
MRI	Magnetic Resonance Imaging
NC	No change
ns	nanosecond
PD	Progressive Disease
PR	Partial Response
PET	Positron Emission Tomography
RCC	Renal Cell Carcinoma
RECIST	Response Evaluation Criteria in Solid Tumours
RI	Retention index
ROI	Region of interest
SD	Stable Disease
SPECT	Single Photon Emission Computed Tomography
T/N	Tumour-to-normal ratio
TSE	Turbo spin echo
VIBE	Volumetric interpolated breath hold examination
WHO	World Health Organization
18-F-FDG	18-F-fluorodeoxyglucose
1D	One-dimensional
201-Tl	201-Thallium
2D	Two-dimensional
3D	Three-dimensional
99m-Tc	99m-Technecium

Chapter 1: Introduction

1.1 Overview

The work in this thesis uses image analysis tools for x-ray computed tomography (CT) to quantify pulmonary metastatic tumour size in one-, two-, and three-dimensions. These methods are also used to quantify longitudinal changes in tumour size. We evaluate intra-observer, inter-observer, and inter-scan precision and measurement accuracy.

In this introductory chapter, the motivation, imaging principles, and research objectives are described as a foundation for understanding the methods used, results and discussion in subsequent chapters. The use of x-ray CT as the primary method for imaging pulmonary metastases is explained and compared to other methods. Comprehensive descriptions of the image analysis tools are provided: the one-dimensional (1D) Response Evaluation Criteria in Solid Tumours (RECIST) method, the two-dimensional (2D) World Health Organization (WHO) method, and the three-dimensional (3D) volumetric analysis. Finally, the research questions, objectives, and hypotheses that we tested are described.

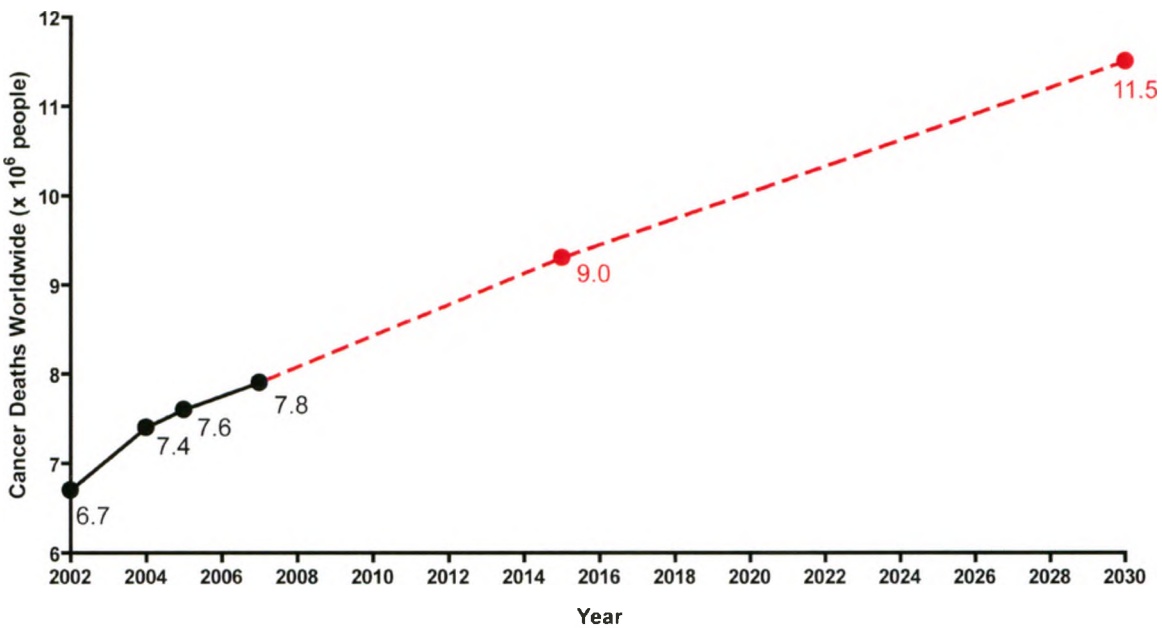
1.2 Cancer

1.2.1 Global Impact and Research Motivation

Cancer is the leading cause of death among Canadians (1;2). It is responsible for approximately 29% of all deaths in Canada, and in 2009, there were an estimated 171,000 new cases of cancer diagnosed (1). The economic burden that cancer creates for the

Canadian economy is an estimated $2 billion CAD/yr in direct costs and $12 billion CAD/yr in indirect costs, annually (3;4). Cancer is the second leading cause of deaths worldwide, despite it being one of the most preventable non-communicable chronic diseases (5); Figure 1-1 illustrated the predicted increase in worldwide cancer mortalities. The World Heath Organization reports that healthy lifestyle changes, including a decrease in tobacco use and alcohol consumption, increased physical activity and improved diets would prevent approximately 40% of all cancers (5). In 2005, cancer accounted for 7.6 million deaths worldwide; this mortality statistic is expected to increase to 9 and 11.5 million in 2015 and 2030, respectively (5). Accordingly, there has been an international research effort to improve not only cancer diagnosis and care, but also the detection of treatment effects.

Figure 1-1: World Heath Organization Worldwide Mortality Rate Projections
World Health Organization (WHO) projects an increase in cancer-related deaths worldwide (5-7). The reported values are labelled in black, and the projected values are labelled in red.



Currently, diagnostic imaging modalities such as CT, magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT), and positron emission tomography (PET) are used to detect the presence of metastatic lesions and monitor their changes longitudinally. Since 1979, two major criteria have been developed and implemented to standardize the reproducible quantification of cancer metastases *in vivo*. In 1979, the World Health Organization (WHO) introduced a 2D measurement of metastases from x-ray or x-ray CT images. This 2D measurement was defined as the product of the maximal dimension and the longest perpendicular bisector of each tumour (8). The sum of this measurement for each tumour was used to assess the change in tumour size. The WHO criteria defined four distinct categories of tumour response: stable disease (SD), progressive disease (PD), partial response (PR), and complete response (CR). Physicians used these classifications, to select treatment options for the patient (9), some options including surgical resection, radiotherapy, or chemotherapy (9;10). There were limitations with the WHO criteria: the minimum tumour size, guidelines for selecting target lesions, and number of target lesions to be selected were not stated in the original WHO guidelines. These issues led to investigator-specific criteria modifications, and as a result, response criteria were no longer comparable between research groups (11-14).

To provide a simple solution to some of the limitations of the WHO criteria, and re-establish a simple, standardized method of tumour quantification (14), the Response Evaluation Criteria in Solid Tumours (RECIST) was introduced in 2000. RECIST defined a 1D measurement of the maximal dimension of each tumour, based on x-ray CT and MRI (13). It addressed shortcomings in the WHO criteria by defining the minimum

lesion size allowable as no less than double the slice thickness of x-ray CT or MRI (13); it also restricted the maximum of 10 target lesions to be selected (maximum of five lesions per organ), with the criteria for selecting these lesions expressed clearly and defined prospectively (13). The sum of the longest axis of each tumour for all tumours measured was used to assess response. The four response classifications outlined in WHO criteria (SD, PD, PR, and CR) were redefined using RECIST, and were also used to manage patient care (13).

The RECIST criteria was widely adopted by many within the clinical and scientific communities to assess treatment outcomes (15). However, investigators questioned the validity of the number of target lesions necessary to determine response or progression (9;15;16), as well as the necessity of the four-week period in which to confirm tumour response (15). To address these among other issues, the RECIST working group published the revised RECIST guideline, version 1.1 (RECIST 1.1) in 2009 (15). RECIST 1.1 was developed after the prospective analysis of over 6500 patients with over 18,000 target lesions, and two major modifications were the reduction of the number of target lesions from 10 to five (maximum of 2 lesions per organ), and the reclassification of PD to prevent over-classifying SD as PD (15). Although revisions were made to RECIST to improve its clinical applications, the fundamental principle remained unchanged: RECIST 1.1 implemented a 1D measurement.

unknown primary cancer site. Although the cancerous cells had also metastasized to the liver, bone, and intestines, we focus on the lung because it is a common site of metastases for renal cell carcinoma (29). Furthermore, the inherent contrast of the tumour boundary and the lung parenchyma makes the analysis of lung metastases ideal with CT as the imaging modality. This CT property is discussed in section 1.3.3.2.

1.3.2 Diagnostic Imaging of Pulmonary Metastases

In this thesis, we focus on quantifying pulmonary metastases using CT; however, it is important to address the various methods used for the diagnosis of pulmonary metastases. Cahan et al. (27) showed that when a solitary lung lesion was detected in over 800 patients with previously diagnosed cancer, 25% were solid metastatic lesions. Gross et al. (30) showed that when multiple lung lesions were detected in 137 patients, where 123 patients were previously diagnosed with cancer, 73% had metastatic disease. Medical imaging is used to detect the presence of suspicious lesions; however, a number of other more invasive methods are used in conjunction with these imaging methods. A number of invasive or minimally invasive approaches have been used to determine malignancy: percutaneous fine needle aspiration biopsy (31-34), bronchoscopy (35-38), and open lung biopsy (39). The modalities used for non-invasive diagnostic imaging include CT, magnetic resonance imaging (MRI), chest radiography, positron emission tomography (PET), and single photon emission computed tomography (SPECT).

1.3.3 Diagnostic Imaging of Pulmonary Metastases

Medical imaging is commonly used to detect pulmonary metastases. This is accomplished by an array of diagnostic imaging modalities described in sections 1.3.3.1 to 1.3.3.4. Medical imaging uses energies along the electromagnetic spectrum to penetrate the tissues within the human body; as the energies are absorbed, attenuated, or scattered within the tissues, the detected resultant energies are used to construct an anatomical or functional image (40). Anatomical imaging allows for visualization of the internal anatomy, and functional imaging provides information regarding the physiology or metabolism of the patient (40).

1.3.3.1 Chest X-Ray

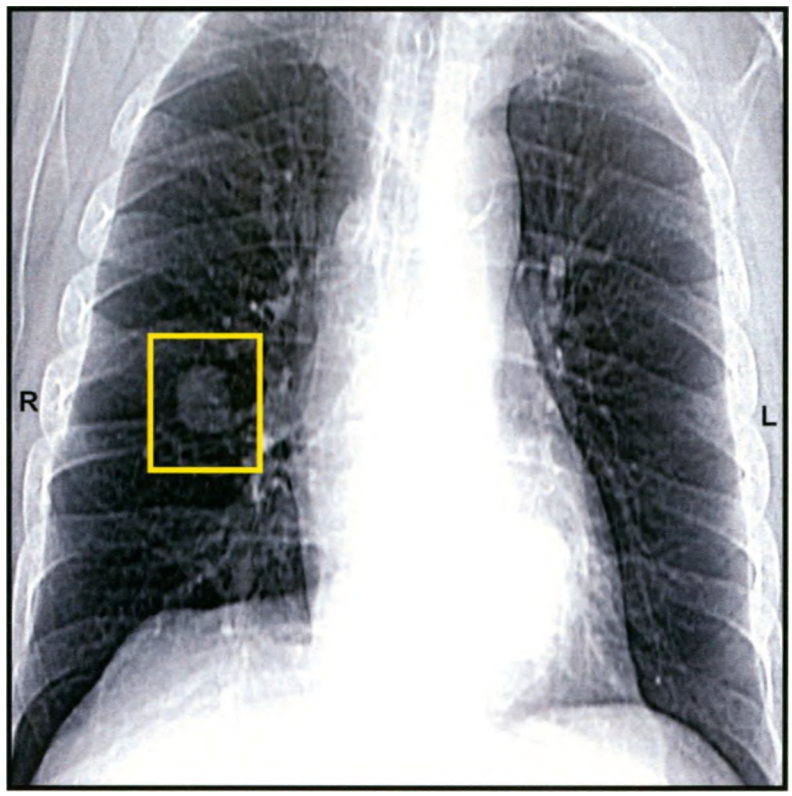
X-ray radiography was the first medical imaging technology, made possible when the physicist Wilhelm Roentgen discovered x-rays in 1895 (40). Chest radiography is the most frequently performed diagnostic X-ray examination for a variety of pulmonary diseases (40). Lesions on a chest radiograph may be considered measurable if they are clearly defined and surrounded by aerated lung (15), and radiographs have been used to detect malignant nodules in the lung (31;32;39). An example of a chest radiograph is shown in Figure 1-3.

When a radiograph is performed, an x-ray source and x-ray detector are on either opposing sides of the patient. Within a fraction of one second, the x-ray tube emits a pulse of x-rays which are transmitted through the patient to the detector, where the radiographic image is generated. The homogeneous distribution of x-rays that enter the patient is modified by the degree to which the x-rays are removed from the beam by

scattering and absorption within the tissues. The attenuation properties of tissues such as bone, soft tissue, and air inside the patient are very different, resulting in heterogeneous distribution of x-rays that emerges from the patient and onto the detector.

Some advantages of chest radiography over other 3D and functional methods are lower cost, lower dose, and speed of acquisition and diagnosis (40). One disadvantage of using a chest radiograph is that any out-of-plane information, the dimension parallel to the x-ray beam, is unattainable (40). Although traditional assessment of pulmonary nodules was performed on chest radiographs (34), CT is viewed as the best currently available and reproducible method to measure tumour response (9;15).

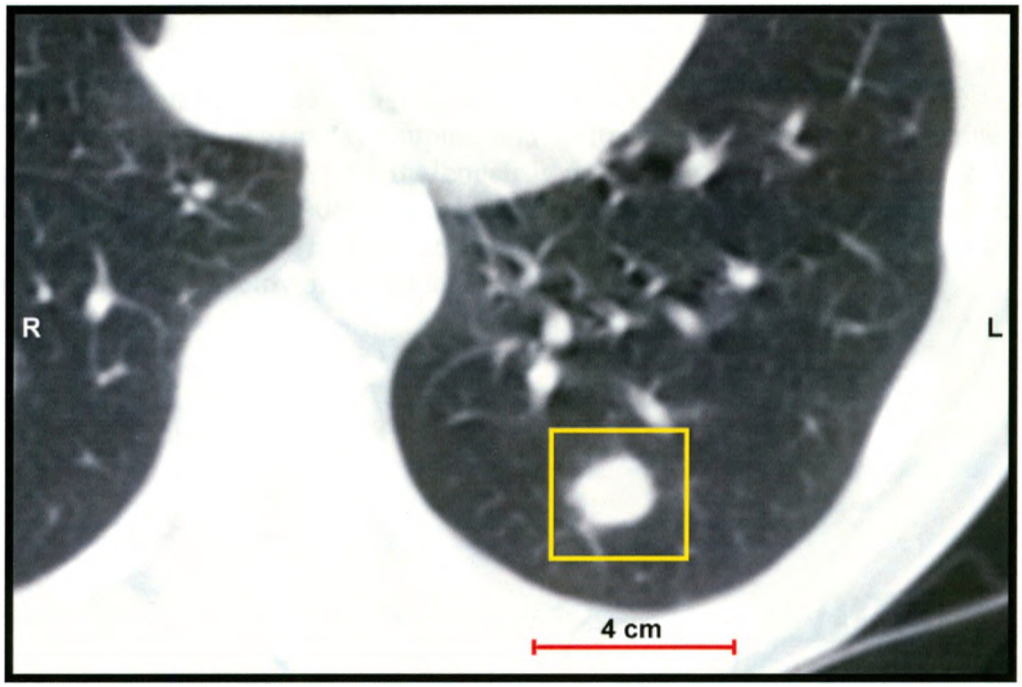
Figure 1-3: Chest Radiograph of Subject 2
This chest x-ray shows a posterior-anterior view of the lungs. The tumour is located in the right lung, enclosed in the yellow box.



1.3.3.2 X-Ray Computed Tomography

X-ray CT addresses the flat plane limitation of the x-ray radiograph, with the introduction of the z-plane which provides the opportunity for 3D views. CT is frequently used to evaluate tumour response, as conventional response criteria are CT based (41;42). It is common to incidentally find pulmonary nodules on a chest CT (20;31;32;39;43); CT will disclose approximately twice as many nodules as a chest radiograph (44;45). Since the introduction of multidetector row CT (MDCT) in the early 1990s, there has been a substantial increase in the number of incidentally detected small pulmonary nodules (46). One advantage of CT over chest radiography is the ability to display anatomical structures in a slice of tissue without interference from over- or underlying structures (40). CT changed the practice of medicine by considerably reducing the need for exploratory surgery (40). An example of a chest x-ray CT is shown in Figure 1-4.

Figure 1-4: X-Ray Computed Tomography Image of Subject 7
An x-ray CT image of one tumour from Subject 7 is shown. This subpleural tumour located in the right lung, lower lobe lateral basal segment, enclosed in the yellow box.



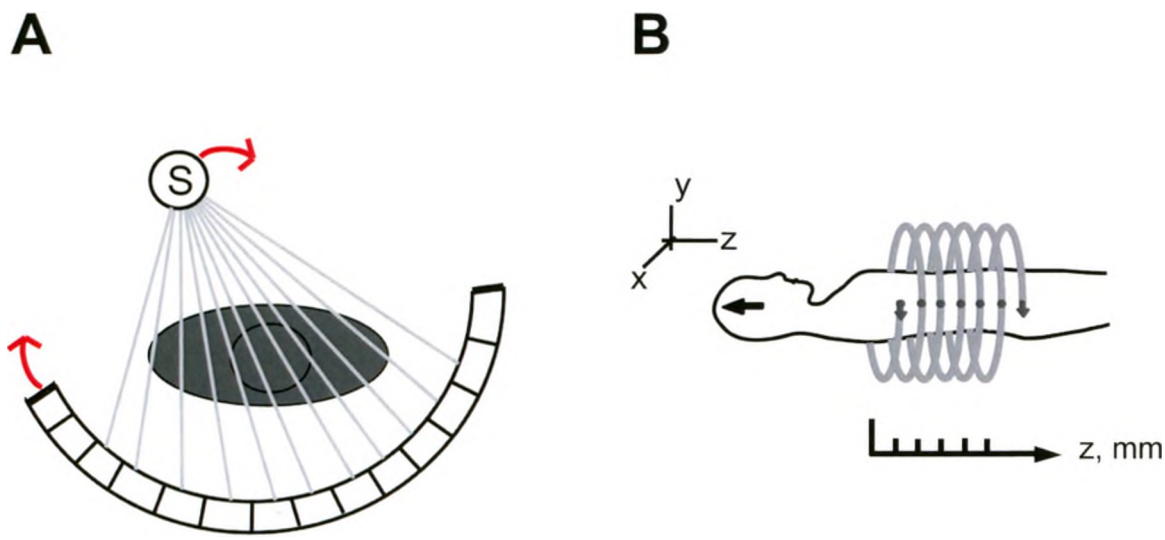
In CT, x-rays are transmitted through and absorbed by tissues within the body. The detectors measure the attenuated x-ray intensity I_t. With the use of a reference detector, the intensity of the unattenuated x-ray beam, I_o, is also known (40). The following equation expresses the relationship between the two x-ray beam intensities, where t is the thickness of the patient, and μ is the average linear attenuation coefficient:

$$I_t = I_o e^{-\mu t} \tag{1}$$

Once all of the projections have been acquired, the raw data is reconstructed using filtered backprojection, and the CT image is created (40). During backprojection, the image is developed by assigning the same linear attenuation coefficient value along a known path of an x-ray beam. This information reinforces the raw data that was collected, and the

computer programming uses the collection of the total data to reinforce areas of high and low attenuation (40). A schematic of the image acquisition process is shown in Figure 1-5.

Figure 1-5: Schematic of CT image acquisition
With helical CT, the patient lays supine and is translated through the machine in the cranial-caudial direction. (A) As the couch is translating, the X-ray source (S) rotates within the enclosed circle of detector elements and transmission measurements are being acquired at all angles. (B) The translational couch results in a series of contiguous slices within the selected anatomy of the patient (40;47).



In a CT image, darker regions represent tissues of low x-ray absorption, and lighter images represent tissues of high absorption. Each pixel in the image is represented by a CT number and is calculated by:

$$CT(x,y) = 1000\frac{\mu(x,y) - \mu_{water}}{\mu_{water}} \quad (2)$$

The units of the CT number is Hounsfield units (HU), named after Godfrey Hounsfield, one of the inventors of CT. The CT number of water is 0 HU, and the value of air is -

1000HU. The CT values for soft tissues range from -300 to -100 (40). Lung nodules usually have higher attenuation than do the surrounding parenchyma (42). The low-density lung provides natural contrast for the dense pulmonary nodules (9). The lungs are unique in that they provide high-contrast images for computer analysis where the solid structures such as airways, blood vessels and nodules have a much higher intensity that the surrounding lung parenchyma (34). A disadvantage of CT is small pulmonary lesions often have similar attenuation levels of blood vessels, and may not be easily distinguished from vessels of similar size (48). Although serial examinations avoid unnecessary surgery when benign lesions are present, they also result delayed diagnosis and treatment when malignancy is present (49). Another major disadvantage of serial examinations is the inherent exposure to ionizing radiation (48;50).

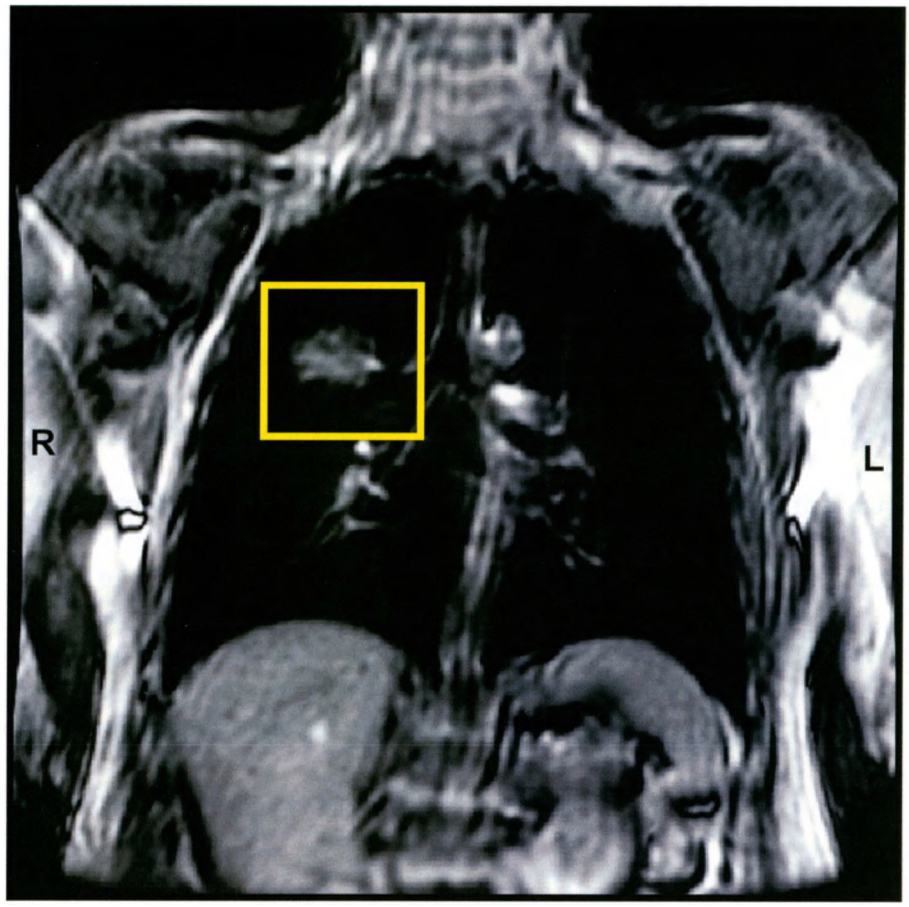
CT contrast agents can be used to differentiate between malignancies and juxtaposed tissues, and to differentiate pulmonary metastases from primary lung cancer (51;52). These are direct agents in that they contain an atom such as iodine (53) that attenuates or scatters the incident x-ray beam differently from the surrounding tissue. This scattering permits direct visualization of the agent itself regardless of its location (54).

1.3.3.3 Magnetic Resonance Imaging

Another imaging method that is used to detect cancer is magnetic resonance imaging (MRI). Traditionally, MRI has not been a primary tool for evaluating pulmonary lesions because of inherent low proton density (48;55-57), limited spatial resolution (48;55;56), high susceptibility differences between air spaces and the pulmonary interstitium (48;55-57), and the presence of respiratory and cardiac motion (40;48;50;55;56). Motion

artefacts may preclude the detection of a small pulmonary nodule (50). A conventional proton MRI of the lungs is shown in Figure 1-6. However, with the implementation of specific pulse sequences, such as 2D T2-weighted turbo-spin echo (TSE), 2D half-Fourier acquisition single-shot turbo spin-echo (HASTE), and T1-weighted 3D volumetric interpolated breath-hold examination (VIBE), MRI can provide good anatomical resolution (55-58), high lesion contrast (55-58), and can detect pulmonary lesions greater than 5mm in diameter (48;55). Vogt et al. (55) showed that when MDCT was used as the standard of reference, MRI was significantly more accurate than conventional chest radiography.

Figure 1-6: Proton MRI of the Lungs
This is a Coronal view of the lungs of a subject with non-small cell lung cancer. The tumour is located in the right lung, enclosed in the yellow box. *Image courtesy of Lindsay Mathew.*



The T2-weighted triple inversion black blood TSE sequence detects small pulmonary nodule by suppressing blood signal intensity; as a result, vascular structures become invisible in the peripheral aspect of the lung and thus improves detection of small pulmonary nodules (56). HASTE images are characterized by high signal intensity in water-rich tissues; thus, lung parenchymal lesions and vessels appear bright whereas surrounding air-filled lung parenchyma display low signal intensity, and the black blood preparation assures flow voids with no apparent signal in pulmonary vessels, thereby facilitating the differentiation of small lung nodules from arteries and veins (55). VIBE

produces near isotropic resolution with fat saturation, and lung vessels appear hyper-intense due to inflow artefacts (59;60). Single breath-hold acquisition of VIBE sequences allows for thin-sliced contiguous images with no interslice gap, and no cardiac motion artefacts (60).

In image acquisition, MRI utilizes the nuclear magnetic resonance properties of the proton, which is the nucleus of the hydrogen atom, and is very abundant in biologic tissues (40). The patient is placed in a magnetic field inside a radio wave producing coil, and protons within the patient's tissues absorb and reemit the radio waves, which are detected by the surrounding coils. Pulse sequences are used to vary the magnetic field strength with the use of magnetic field gradients. The variation of the magnetic field also causes the relaxation time of the protons in various tissues. These relaxation times, T1 and T2, are used to image various anatomies (48).

Although MRI has lower spatial resolution than CT (48;61), advantages of using MRI over CT to detect lesions in the lung are its lack of ionizing radiation (48;50;55;56), and use of contrast agents. Contrary to the direct contrast agents using in CT, MRI contrast agents are indirect agents which affect the relaxation times of the water protons in the nearby image (54). Because of the low concentration and dosage for MRI contrast agents there have been a lower occurrence of adverse reactions (54). Within three to four days, MRI contrast agents are excreted through the renal system. Contrast agents are usually categorized as T1 or T2 agents based on their primary effect of shortening the T1 or T2 relaxation times, respectively (54). Contrast agents commonly contain gadolinium as the metal ion (62), with a half-life typically in the order of 90 minutes (54).

1.3.3.4 Functional Imaging

Functional imaging may also be used for the detection of pulmonary metastases. Contrary to anatomical imaging, which assesses lesions based on their geometrical properties, functional imaging uses metabolic properties to determine malignancy. In nuclear medicine, a radioisotope is administered to the patient orally, intravenously, or via inhalation (40;63) which localizes in the suspicious nodules throughout the body upon administration. Functional imaging methods include positron emission tomography (PET) and single photon emission computed tomography (SPECT).

Cancer cells metabolize glucose at higher rates than normal cells (49;64-66) and PET exploits this characteristic by utilizing radiotracers which are glucose analogues (49;63;64) to quantify the metabolic activities of viable cancer cells; a common radiotracer used in oncology is 18-F-fluorodeoxyglucose (FDG) (64;67). A variety of metrics have been developed to detect the presence of cancerous cells based on 18-F-FDG uptake, but the method most commonly used to evaluate pulmonary nodules is standard uptake value (SUV) (66;68;69). To determine SUV, the concentration of 18-F-FDG within the tumour is normalized to the volume of distribution within the patient; SUV may be calculated in terms of body weight, lean body mass, or body surface area (66;70). The ability of PET to detect pulmonary metastases is also dependent on tumor size (69).

Standard uptake value is increased when the cells are malignant, because of the increased glycolysis, and it has been shown that glucose metabolism correlates with tumour doubling time (63;65). With respect to tumour size, a larger tumour will be detected as

cancerous with greater accuracy than one of smaller dimensions. Gould et al. (49) showed that PET could accurately detect pulmonary malignancies greater than 1cm in diameter, and Khalaf et al. (68) showed that the average SUV of pulmonary nodules less than 1cm in diameter are equal for benign and malignant nodules.

During PET imaging, gamma rays pairs are simultaneously emitted from within the patient. These gamma rays are annihilation radiation, as they are the product of an annihilation event between a positron and an electron. When 18-F-FDG (or any other radiotracer) decays it produces a positron, which travels a few millimetres in tissue while depositing kinetic energy (71). When the positron comes in contact with an electron, mutual annihilation occurs: from conservation of energy, two 511 keV annihilation photons are formed; from conservation of momentum, the annihilation photons are emitted 180^{o} in opposing directions from each other (64;71). Coincidence logic is used within the PET image processing for the analysis of signals from the opposing detectors. A coincident event is the when two photons contact detectors 180^{o} apart, and with a precision of a few nanoseconds, the electronics can detect a coincidence event (40;72). The timing window of a coincidence event is typically 6-12ns (72). It is possible to simultaneously acquire data at all projection angles, because a stationary ring or geometric array of detectors completely surrounds the patient (72). This information is used to mathematically compute the 3D distribution of the PET agent, resulting in a series of tomographic emission images (40). An advantage of using PET is the radionuclides used have short half-lives (11-C, 20min; 13-N, 10min; 15-O, 2min; 18-F, 110min), which allow for same day rescans without interference from background activity from prior injections (71).

Single photon emission computed tomography (SPECT) uses cationic or receptor-specific radiotracers that are metabolized by cancer cells to detect malignancies. Common radionuclides used in SPECT are 201-Tl and 99m-Tc based agents (72-75). Radiotracer uptake by malignant cells is evaluated by the tumour-to-normal lung (T/N) activity ratio and retention index (76-79). The T/N ratio is calculated from the average photon count per pixel within two selected regions of interest (ROIs): the early scan ROI (ER), for example 15 mins post-radiotracer administration (80), and the delayed scan ROI (DR), for example, 3h post-radiotracer administration (80). These values are used to calculate the retention of radiotracer within the lesion (76;79):

$$RI = \frac{DR - ER}{ER} \times 100\% \quad (3)$$

The RI can be used to distinguish benign from malignant nodules, as two groups showed that there was no difference between early and delayed scans of benign lesions (79;80).

In SPECT, a gamma camera on a rotating gantry is used to record radioactive emissions from the patient tissues at numerous angles. These 1D projection data are used to reconstruct a series of cross-sectional, tomographic emission images (40). An advantage of SPECT is it more widely available and less expensive than FDG-PET (74). A disadvantage is that the radiopharmaceuticals fail to delineate the structures involved in pathological processes or in describing the biological meaning of some uptakes (73). The rotation of the detector head is a disadvantage of SPECT, causing the image acquisition process to occur on the order of minutes (40;72). In an attempt to increase sensitivitiy, manufacturers are now producing multicamera SPECT systems. These incorporate two or three cameras to increase sensitivity (71;72).

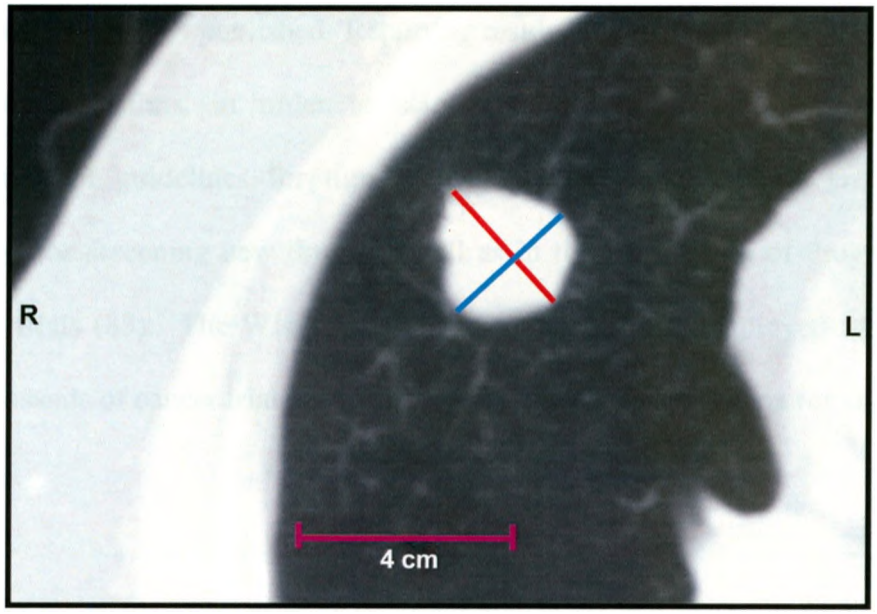
1.4 Measurements of Pulmonary Metastases

Beyond the detection of pulmonary metastases, standardization in reporting tumour response is necessary to compare two or more trials or treatment regimen (81). Two measurement criteria have been developed and applied to classify tumour progression and response to treatment. These guidelines are described in sections 1.4.1 and 1.4.2.

1.4.1 World Health Organization Criteria for the Measurement of Tumours (WHO)

In 1979, WHO introduced standard criteria by which clinicians and researchers could categorize the change in total tumour burden (8). WHO implemented a two-dimensional measurement that was defined as the product of the longest axis of the tumour and its longest perpendicular bisector (8), as measured from x-ray film or x-ray CT. An example of WHO measurements are shown in Figure 1-7.

Figure 1-7: World Health Organization Tumour Measurement
The World Health Organization (WHO) measurement on an x-ray CT image is calculated as the product the longest axis of the tumour (red line) and its longest perpendicular bisector (blue line). The lengths of the red and blue lines are 2.8 cm and 2.6 cm, respectively; the WHO measurement for this tumour is 7.28 cm2.



Four treatment response classifications were defined in the WHO guidelines, to assess tumour progression or response: 1) complete response (CR), 2) partial response (PR), 3) no change (NC), and 4) progressive disease (PD) (8).

Interpretation of the four response categories varied over time; for example, some investigators classified disease progression as a 25% increase in the sum of the products, rather than on the basis of change in a solitary lesion (10;82), and defined partial response as a less than 50% decrease in tumour size (10). Clinicians became interested in changed from the measurement of each individual tumour, to the tumour load. When the reproducibility of WHO measurements were evaluated, it was determined that there

would have to be a drastic change in the longest axis for the two dimensional measurement to change significantly.

In 1981, Miller et al. (10) published 'Reporting results of cancer treatment', based on the WHO recommendations, in order to standardize the response assessment. Such standardization of guidelines for tumour response evaluation helped greatly in the methodology for screening new drugs as well as in the comparison of drug efficacy in randomized trials (83). The WHO recommendations had been employed for almost 20 years in thousands of cancer trials and still appeared as reference criteria for any historical comparison.

1.4.2 Response Evaluation Criteria in Solid Tumours (RECIST)

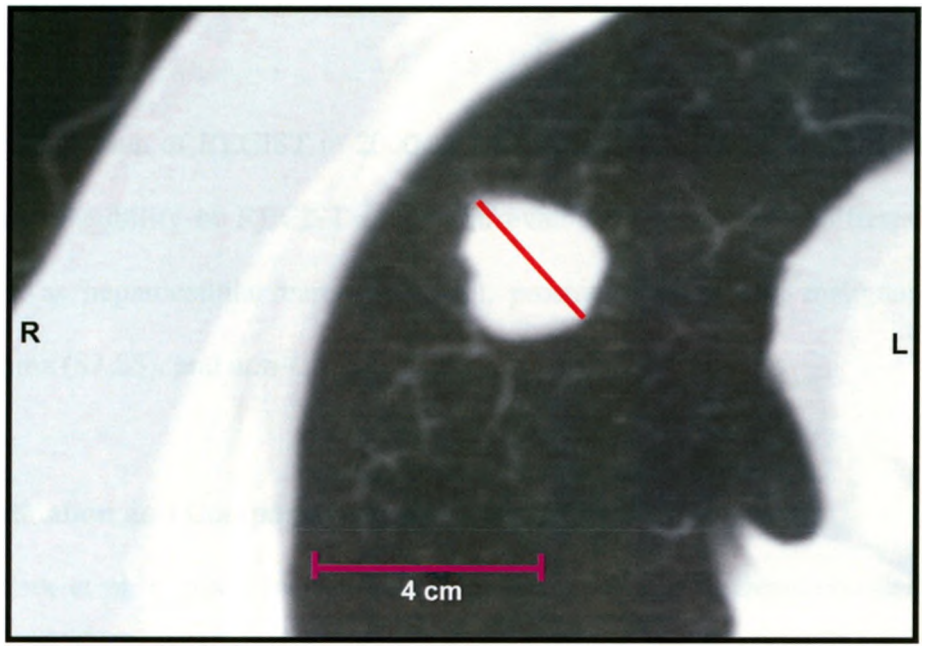
Response Evaluation Criteria in Solid Tumours is widely accepted as the new standard (14). The first study to propose the use of a unidimensional measurement was in 1966 by Gurland and Johnson (84). The authors reported good correlations between the maximum diameter and its longest perpendicular bisector (R=0.79-0.99); the surface areas of tumours with various shapes (R=0.85-0.99); and also with the perimeter of the tumour boundary (R=0.98-0.99). In 1984 Spears (85) reported that the measurement of the tumour diameter alone was only inaccurate as a surrogate of tumour size when it exceeded twice the tumour width; this suggested an assumption of tumour sphericity. In 1999, James et al (82) demonstrated good agreement between WHO response criteria and a one-dimensional measurement in 1999. The theoretical reasoning supporting this study was that the change in tumour size should be linearly proportional to the logarithm of the number of cells killed by a fixed dose of a cytotoxic agent, under the assumption of

spherical tumours (82). Because the absolute number of cells killed by a given dose of drug depends on the number of cells actually present at the time of drug exposure, attempts to measure the degree of lethality should relate to proportional reductions in tumour volume, which is assumed to be the logarithm of the number of cells killed (82). Using the Kappa statistic for concordance, agreement of overall response between both measurement techniques was 0.95 (82). In 2000, two co-authors from James et al joined a collaborative effort with their professional colleagues and developed, validated, and implemented the Response Evaluation Criteria in Solid Tumours (RECIST). Contributors of this RECIST paradigm were members of the European Organization for Research and Treatment of Cancer, National Cancer Institute and National Cancer Institute of Canada Clinical Trials Group, who became known as the RECIST Working Group (13).

The first publication of the RECIST guidelines was a retrospective analysis of 14 different studies demonstrating that bidimensional or unidimensional measurement of tumour lesions did not change the response rate in each individual study. The RECIST Working Group proposed this unidimensional measurement in order to minimize the risk of measurement error as compared to WHO criteria, and to prevent overestimation of response rates (13). In 2009, the Group modified these criteria in the RECIST guideline version 1.1 (15). An example of the RECIST measurement is shown in Figure 1-8. As expected, the theoretical validity and practical feasibility of measuring in only one dimension were questioned by a number of research groups (15). The criteria for partial response between RECIST and WHO are almost equivalent under the assumption that tumours are round and shrink uniformly (22). Table 1-1 outlines the response

classifications for RECIST and WHO, as well as 3D classifications derived from the two existing guidelines.

Figure 1-8: Response Evaluation Criteria in Solid Tumours Measurement
The Response Evaluation Criteria in Solid Tumours (RECIST) measurement on an x-ray CT image is the longest axis of the tumour (red line). The length of the red line is 2.8 cm; therefore, the RECIST measurement for this tumour is 2.8 cm.



	RECIST (cm)	WHO (cm^2)	Volume (cm^3)
Stable Disease (SD)	▲< PD threshold **OR** ▼< PR threshold		
Progressive Disease (PD)	▲20% + 5mm absolute increase	▲25%	▲73% ▲40%
Partial Response (PR)	▼30%	▼50%	▼65%
Complete Response (CR)	Disappearance of all measurable lesions		

Table 1-1: Response Classification for RECIST, WHO, and 3D Measurements
RECIST and WHO criteria define four distinct tumour response classifications in each respective guideline. Volumetric response classifications have been derived from the established criteria (8;13;15)

Even if extensive tumour necrosis is achieved, this may not be paralleled by a reduction in the greatest dimension of the lesion (11). It seems reasonable to evaluate response by measuring the extent of tumour necrosis instead of the mere measurement of size by 1D or 2D systems (11). The assessment of necrosis could evaluate a potential tumour response more accurately, and it would prevent the rejection of a promising treatment because of an underestimation of its real anti-tumour activity (11).

Since the publication of RECIST in 2000, several reports have been published regarding the low reproducibility of RECIST criteria in evaluating response in different tumour types, such as hepatocellular carcinoma (11), prostate cancer (86), malignant pleural mesothelioma (87;88), and non-small cell lung cancer (14;89).

1.4.2.1 Validation and Comparative Studies of RECIST

Joon Oh Park et al, a group in Korea collected data from 79 patients enrolled in eight prospective phase II trials at Samsung Medical Center, which were retrospectively re-analyzed to determine concordance between RECIST and WHO criteria.(12) The studies conducted were for gastric cancer, breast cancer, non-small cell lung cancer (NSCLC), and hepatocellular carcinoma (12). There were 25 of 79 (31.6%) partial responses (PR) to the WHO criteria and 24 of 79 (30.4%) to the RECIST criteria. Only one of 79 (1.3%) patients had PR according to the WHO criteria but not according to the RECIST criteria. In addition, two of 79 patients were assess as PR according to the unidimensional criteria but not according to the bidimensional criteria. The overall response rate according to WHO was 31.6% and for RECIST was 30.4%. In total, nine patients were reclassified; six were switched from progressive disease (PD) to stable disease (SD), one from the SD

to the PR group, and two from the PR to the SD group. The number of SD group patients was higher, and the number in the PD group was lower in those assessed using the RECIST criteria. Using the Kappa statistic to test concordance for overall response rate, the group found that there was excellent agreement between the unidimensional and bidimensional criteria in 23 of 25 responses (92%). The Kappa statistic for concordance in overall response was 0.91. However, when the responses were subdivided into PR, SD, and PD, the concordance for the response rate was 0.83. This phenomenon was the result of the largest re-categorization of patients from PD to SD. The re-categorization of patients from PD to SD would have an impact on a time to progression, but was not further analyzed as patients were taken off study or treatment when they met the WHO criteria (12).

Sohaib et al. (22) performed a study of 16 patients with 20 lymph node masses pre- and post- chemotherapy for germ cell cancer or lymphoma. When comparing the response classifications between 1D, 2D, and 3D measurements, 2 of 20 (10%) showed a difference between 1D and 2D assessment criteria of response. Volumetric tumour measurements were used as the gold standard for this phantom study. Watanabe et al. (14) excluded patients who were treated in daily clinical practice as tumour response evaluation in the daily clinical practice of oncology.

RECIST guidelines presume that tumours are spherical and change in a uniform symmetric manner. In actuality, tumours do not necessarily grow symmetrically; different portions may grow at different rates (21). Significant variability in RECIST measurements exists among different observers (46;90;91).

1.4.3 Volumetric Measurements

There are several limitations of 1D and 2D tumour measurements, such as difficulty in estimating size of irregular or confluent lesions, discrepancies in scan planes and patient positioning leading to error in measurement (92). Also, measuring in two dimensions followed by the calculation of products and sums is laborious and may have the risk of variability (12). Two-dimensional measurements may misrepresent change in tumour size by disregarding alteration to the third dimension. More accurate determination of tumour size have may be obtained using three-dimensional volume measurements, with consequent implications for assessment of response (22).

Several theoretical advantages to using 3D volumetric measurements are that is allows for better quantification of total tumour bulk by incorporating multiple tumour sites into one tumour volume measurement, more accurate assessment of tumour change by adding a third dimension of measurement, and better measurement of irregular masses. Volume measurements potentially offer a more accurate representation of change in size of a tumour following treatment (22).

Direct volume measurements are theoretically preferable to diameter or perimeter measurements because nodules are seldom perfectly spherical and often have irregular or difficult-to-define margins (20). If nodules were perfectly spherical, a change in diameter would accurately reflect overall changes in volume (20). However, lesions are frequently lobular, so judgments based on long- and short- axis measurements are necessarily subjective (20). Tumours such as mesothelioma, or those located within hollow viscera,

may grow circumferentially, requiring special modifications to the longest diameter rule in standardized reporting (41;82).

With the introduction of multi-detector scanners, thinner slices are more routinely available and better z-axis resolution renders nodule volume measurements more accurate (20). New imaging technologies and progress in the development of new classes of anti-cancer agents required the establishment of a new methodology, which in turn led to a number of different modifications of WHO criteria (12). For similar reasons, WHO criteria was modified and later replaced with RECIST criteria, and RECIST 1.0 replaced by version 1.1. Technological and pharmaceutical advances suggest the development of volumetric measurements should be explored to compensate for the limitations of RECIST 1.1.

Currently, the standard method of measuring change is the sequential segmentation of the tumour in interval examinations followed by subtraction of the value of the tumour volume of the previous examination from that derived from the current examination. This double segmentation is an indirect method in that volume change is not measured directly and will depend on the accuracy or consistency of the segmentation and the change assessment paradigm (34;42).

Tran et al. (92) analyzed 32 lesions from 15 patients at baseline and two follow up visits, yielding 30 response classifications for each measurement technique. Measurements in 1D, 2D and 3D were concordant in 21 of 30 classifications. The 1D and 3D

measurements were concordant in 29 of 30 classifications and 2D and 3D measurements were concordant in 23 of 30 classifications.

Three-dimensional quantification may be considered as more representative than 1D measurements for the actual size changes of lesions (23). Segmentation is the most crucial and also most challenging step in the analysis of pulmonary nodules from CT image data (34). Lesions are frequently attached to other structures including the local pulmonary vasculature and the pleural surface adjoining the thoracic wall (34).

1.5 Image Analysis Software: 3D Quantify

The volumetric measurements were performed using in-house customized software created in the Imaging Research Laboratories, at Robarts Research Institute (London, ON), called 3D Quantify. This software has been used to perform 3D measurements of plaque in the carotid artery (93-96), as well as 3D measurements of the prostate (97-99). The algorithm within 3D Quantify uses triangular 3D meshes (100;101) to calculate the volume of the contoured image.

1.6 Research Hypotheses

The hypotheses tested in this thesis are as follows:

I. 3D measurements are accurate and precise, when measuring lung tumour phantoms of known dimensions

II. RECIST, WHO, and 3D measurements have high inter-observer and intra-observer reproducibility.

III. RECIST, WHO, and 3D measurements of expert medical professionals, and trained research students and technicians are not statistically different.

IV. 3D measurements may be used to stratify tumour response longitudinally.

V. RECIST, WHO, and 3D measurements are highly correlated with one another.

In Chapter 2, the objectives were to determine the accuracy and precision of the 1D, 2D, and 3D measurements with the use of a chest phantom, and lung tumour phantoms of known dimensions used as ground truth measurements. Four observers performed repeated measurements of tumours of various shapes, sizes, and CT numbers.

In Chapter 3, the objectives were to determine the reproducibility of the 1D, 2D and 3D measurements of pulmonary metastases in seven subjects by trained and expert observers. In a cross-sectional analysis, twenty-nine pulmonary metastases were quantified by seven observers using 1D and 2D, and four observers quantified tumour size in 3D. Five subjects were available for a longitudinal analysis and four observers quantified tumour size at each time point using 1D, 2D, and 3D measurements. Intra- and inter-observer reproducibility was evaluated, as well as the use of 3D measurements to classify the change in tumour size.

Chapter 4 provides a summary of the work described in Chapters 2 and 3, as well as provides opportunities for future studies.

1.7 References

1. Statistics Canada. Leading Causes of Death in Canada. 84-215-X. 2005.
   Ref Type: Report

2. Canadian Cancer Society's Steering Committee. Canadian Cancer Statistics 2009. 4-2-0009. Toronto, Canadian Cancer Society.
   Ref Type: Report

3. Boyd DR, Genuis SJ. The environmental burden of disease in Canada: respiratory disease, cardiovascular disease, cancer, and congenital affliction. *Environ.Res.* 2008;106:240-9.

4. Health Canada. Economic Burden of Illness in Canada, 1998. H21-136/1998E. 2002.
   Ref Type: Report

5. World Health Organization. The World Health Organization's Fight Against Cancer: Strategies That Prevent, Cure and Care. 2007. Switzerland, World Health Organization.
   Ref Type: Report

6. World Health Organization and International Union Against Cancer. Global Action Against Cancer. 2005. Switzerland.
   Ref Type: Report

7. World Health Organization. Fact sheet No. 297: Cancer. 2009.
   Ref Type: Report

8. World Health Organization. WHO Handbook for Reporting Results for Cancer Treatment. 1979. Geneva, World Health Organization.
   Ref Type: Report

9. Chojniak R, Yu LS, Younes RN. Response to chemotherapy in patients with lung metastases: how many nodules should be measured? *Cancer Imaging* 2006;6:107-12.

10. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207-14.

11. Forner A, Ayuso C, Varela M, Rimola J, Hessheimer AJ, de Lope CR et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? *Cancer* 2009;115:616-23.

12. Park JO, Lee SI, Song SY, Kim K, Kim WS, Jung CW et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn.J.Clin.Oncol.* 2003;33:533-7.

13. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J.Natl.Cancer Inst.* 2000;92:205-16.

14. Watanabe H, Kunitoh H, Yamamoto S, Kawasaki S, Inoue A, Hotta K et al. Effect of the introduction of minimum lesion size on interobserver reproducibility using RECIST guidelines in non-small cell lung cancer patients. *Cancer Sci.* 2006;97:214-8.

15. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur.J.Cancer* 2009;45:228-47.

16. Darkeh MH, Suzuki C, Torkzad MR. The minimum number of target lesions that need to be measured to be representative of the total number of target lesions (according to RECIST). *Br.J.Radiol.* 2009;82:681-6.

17. Kalender WA. CT: the unexpected evolution of an imaging modality. *Eur.Radiol.* 2005;15 Suppl 4:D21-D24.

18. Kalra MK, Maher MM, D'Souza R, Saini S. Multidetector computed tomography technology: current status and emerging developments. *J.Comput.Assist.Tomogr.* 2004;28 Suppl 1:S2-S6.

19. Kohl G. The evolution and state-of-the-art principles of multislice computed tomography. *Proc.Am.Thorac.Soc.* 2005;2:470-500.

20. Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. *AJR Am.J.Roentgenol.* 2006;186:989-94.

21. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology* 2000;217:251-6.

22. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RT, Reznek RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br.J.Radiol.* 2000;73:1178-84.

23. Mantatzis M, Kakolyris S, Amarantidis K, Karayiannakis A, Prassopoulos P. Treatment response classification of liver metastatic disease evaluated on imaging. Are RECIST unidimensional measurements accurate? *Eur.Radiol.* 2009;19:1809-16.

24. Pauls S, Kurschner C, Dharaiya E, Muche R, Schmidt SA, Kruger S et al. Comparison of manual and automated size measurements of lung metastases on

MDCT images: potential influence on therapeutic decisions. *Eur.J.Radiol.* 2008;66:19-26.

25. Bubendorf L, Schopfer A, Wagner U, Sauter G, Moch H, Willi N et al. Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Hum.Pathol.* 2000;31:578-83.

26. Fabozzi SJ, Schellhammer PF, el-Mahdi AM. Pulmonary metastases from prostate cancer. *Cancer* 1995;75:2706-9.

27. Cahan WG, Shah JP, Castro EB. Benign solitary lung lesions in patients with cancer. *Ann.Surg.* 1978;187:241-4.

28. Billingsley KG, Lewis JJ, Leung DH, Casper ES, Woodruff JM, Brennan MF. Multifactorial analysis of the survival of patients with distant metastasis arising from primary extremity sarcoma. *Cancer* 1999;85:389-95.

29. Sivaramakrishna B, Gupta NP, Wadhwa P, Hemal AK, Dogra PN, Seth A et al. Pattern of metastases in renal cell carcinoma: a single institution study. *Indian J.Cancer* 2005;42:173-7.

30. Gross BH, Glazer GM, Bookstein FL. Multiple pulmonary nodules detected by computed tomography: diagnostic implications. *J.Comput.Assist.Tomogr.* 1985;9:880-5.

31. Dougherty DW, Gonsorcik VK, Harpster LE, Trussell JC, Drabick JJ. Superficial bladder cancer metastatic to the lungs: two case reports and review of the literature. *Urology* 2009;73:210-5.

32. Miyamoto H, Jones CE, Raymond DP, Wandtke JC, Strang JG, Bourne PA et al. Pulmonary metastases from uterine neoplasms after long tumour-free interval: four cases and review of the literature. *Pathology* 2009;41:234-41.

33. Shamim T, Varghese VI, Shameena PM, Sudha S. Primary intraosseous adenoid cystic carcinoma of the mandible with lung metastasis: a case report. *J.Oral Sci.* 2008;50:95-8.

34. Reeves AP, Chan AB, Yankelevitz DF, Henschke CI, Kressler B, Kostis WJ. On measuring the change in size of pulmonary nodules. *IEEE Trans.Med.Imaging* 2006;25:435-50.

35. Danila E, Zurauskas E, Loskutoviene G, Zablockis R, Nargela R, Birzietyte V et al. Significance of bronchoscopic lung biopsy in clinical practice. *Adv.Med.Sci.* 2008;53:11-6.

36. Eberhardt R, Ernst A, Herth FJ. Ultrasound-guided transbronchial biopsy of solitary pulmonary nodules less than 20 mm. *Eur.Respir.J.* 2009;34:1284-7.

37. Nakajima T, Yasufuku K, Fujiwara T, Chiyo M, Sekine Y, Shibuya K et al. Endobronchial ultrasound-guided transbronchial needle aspiration for the diagnosis of intrapulmonary lesions. *J.Thorac.Oncol.* 2008;3:985-8.

38. Sakairi Y, Yasufuku K, Iyoda A, Suzuki M, Nakajima T, Sekine Y et al. A solitary metastatic lung tumor from thyroid papillary carcinoma diagnosed by endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA): report of a case. *Surg.Today* 2008;38:46-8.

39. Parada D, Ugas G, Pena K, Caricote L, Mujica N. Lung metastases of low grade phyllodes tumor of the prostate: histopathologic confirmation. *Arch.Esp.Urol.* 2008;61:658-62.

40. Bushberg JT, Seibert JA, Leidholdt EM, Boone JM. The Essential Physics of Medical Imaging, Second Edition ed. Philadelphia: Lippincott Williams & Wilkins, 2002.

41. Plathow C, Klopp M, Thieke C, Herth F, Thomas A, Schmaehl A et al. Therapy response in malignant pleural mesothelioma-role of MRI using RECIST, modified RECIST and volumetric approaches in comparison with CT. *Eur.Radiol.* 2008;18:1635-43.

42. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: computerized quantification of tumor response--initial results. *Radiology* 2006;241:892-8.

43. Miwa S, Kadono Y, Sugata T, Mizokami A, Namiki M. Successful treatment for metastases from renal cell carcinoma with alternation of interferon-alpha subtypes. *Int.J.Clin.Oncol.* 2010;15:97-100.

44. Chang AE, Schaner EG, Conkle DM, Flye MW, Doppman JL, Rosenberg SA. Evaluation of computed tomography in the detection of pulmonary metastases: a prospective study. *Cancer* 1979;43:913-6.

45. Schaner EG, Chang AE, Doppman JL, Conkle DM, Flye MW, Rosenberg SA. Comparison of computed and conventional whole lung tomography in detecting pulmonary nodules: a prospective radiologic-pathologic study. *AJR Am.J.Roentgenol.* 1978;131:51-4.

46. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur.Radiol.* 2006;16:781-90.

47. Chronik BA, Zinke-Allmang M. Concepts of Medical Imaging, First Edition ed. London: The University of Western Ontario, 2008.

48. Schroeder T, Ruehm SG, Debatin JF, Ladd ME, Barkhausen J, Goehde SC. Detection of pulmonary nodules using a 2D HASTE MR sequence: comparison with MDCT. *AJR Am.J.Roentgenol.* 2005;185:979-84.

49. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001;285:914-24.

50. Kersjes W, Mayer E, Buchenroth M, Schunk K, Fouda N, Cagil H. Diagnosis of pulmonary metastases with turbo-SE MR imaging. *Eur.Radiol.* 1997;7:1190-4.

51. Choi JI, Choi HJ, Jung DC, Kim MJ, Hong EK, Park JW et al. Diagnostic value of early-phase-enhanced computed tomography for the differentiation of pulmonary metastases from hepatocellular carcinoma and primary lung cancer. *Acta Radiol.* 2009;50:1005-10.

52. Jung DC, Choi HJ, Kim HY, Lee KH. Pulmonary metastasis from renal cell carcinoma: characterization using contrast-enhanced CT attenuation value measurements. *J.Comput.Assist.Tomogr.* 2009;33:54-7.

53. Schillaci O, Travascio L, Bolacchi F, Calabria F, Bruni C, Ciccio C et al. Accuracy of early and delayed FDG PET-CT and of contrast-enhanced CT in the evaluation of lung nodules: a preliminary study on 30 patients. *Radiol.Med.* 2009;114:890-906.

54. Brown MA, Semelka RC. MRI Basic Principles and Applications, Third Edition ed. Hoboken: John Wiley and Sons, Inc., 2010.

55. Vogt FM, Herborn CU, Hunold P, Lauenstein TC, Schroder T, Debatin JF et al. HASTE MRI versus chest radiography in the detection of pulmonary nodules: comparison with MDCT. *AJR Am.J.Roentgenol.* 2004;183:71-8.

56. Yi CA, Jeon TY, Lee KS, Lee JH, Seo JB, Kim YK et al. 3-T MRI: usefulness for evaluating primary lung cancer and small nodules in lobes not containing primary tumors. *AJR Am.J.Roentgenol.* 2007;189:386-92.

57. Fink C, Puderbach M, Biederer J, Fabel M, Dietrich O, Kauczor HU et al. Lung MRI at 1.5 and 3 Tesla: observer preference study and lesion contrast using five different pulse sequences. *Invest Radiol.* 2007;42:377-83.

58. Bruegel M, Gaa J, Woertler K, Ganter C, Waldt S, Hillerer C et al. MRI of the lung: value of different turbo spin-echo, single-shot turbo spin-echo, and 3D gradient-echo pulse sequences for the detection of pulmonary metastases. *J.Magn Reson.Imaging* 2007;25:73-81.

59. Biederer J, Graessner J, Heller M. Magnetic resonance imaging of the lung with a volumetric interpolated 3D-gradient echo sequence. *Rofo* 2001;173:883-7.

60. Rofsky NM, Lee VS, Laub G, Pollack MA, Krinsky GA, Thomasson D et al. Abdominal MR imaging with a volumetric interpolated breath-hold examination. *Radiology* 1999;212:876-84.

61. Thomson V, Pialat JB, Gay F, Coulon A, Voloch A, Granier A et al. Whole-body MRI for metastases screening: a preliminary study using 3D VIBE sequences with automatic subtraction between noncontrast and contrast enhanced images. *Am.J.Clin.Oncol.* 2008;31:285-92.

62. Pauls S, Mottaghy FM, Schmidt SA, Kruger S, Moller P, Brambs HJ et al. Evaluation of lung tumor perfusion by dynamic contrast-enhanced MRI. *Magn Reson.Imaging* 2008;26:1334-41.

63. Duhaylongsod FG, Lowe VJ, Patz EF, Jr., Vaughn AL, Coleman RE, Wolfe WG. Lung tumor growth correlates with glucose metabolism measured by fluoride-18 fluorodeoxyglucose positron emission tomography. *Ann.Thorac.Surg.* 1995;60:1348-52.

64. Heron DE, Andrade RS, Beriwal S, Smith RP. PET-CT in radiation oncology: the impact on diagnosis, treatment planning, and assessment of treatment response. *Am.J.Clin.Oncol.* 2008;31:352-62.

65. Lowe VJ, Naunheim KS. Current role of positron emission tomography in thoracic oncology. *Thorax* 1998;53:703-12.

66. Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J.Nucl.Med.* 2006;47:1059-66.

67. Fletcher JW, Kymes SM, Gould M, Alazraki N, Coleman RE, Lowe VJ et al. A comparison of the diagnostic accuracy of 18F-FDG PET and CT in the characterization of solitary pulmonary nodules. *J.Nucl.Med.* 2008;49:179-85.

68. Khalaf M, bdel-Nabi H, Baker J, Shao Y, Lamonica D, Gona J. Relation between nodule size and 18F-FDG-PET SUV for malignant and benign pulmonary nodules. *J.Hematol.Oncol.* 2008;1:13.

69. Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB et al. Staging non-small cell lung cancer by whole-body positron emission tomographic imaging. *Ann.Thorac.Surg.* 1995;60:1573-81.

70. Stahl A, Ott K, Schwaiger M, Weber WA. Comparison of different SUV-based methods for monitoring cytotoxic therapy with FDG PET. *Eur.J.Nucl.Med.Mol.Imaging* 2004;31:1471-8.

71. Bernier DR, Christian PE, Langan JK. Nuclear Medicine: Technology and Techniques, Fourth Edition ed. Toronto: Mosby-Year Book, Inc., 1997.

72. Cherry SR, Sorenson JA, Phelps MA. Physics in Nuclear Medicine, 3rd Edition ed. Philadelphia: Elselvier Science, 2003.

73. Buscombe JR, Bombardieri E. Imaging cancer using single photon techniques. *Q.J.Nucl.Med.Mol.Imaging* 2005;49:121-31.

74. Sergiacomi G, Schillaci O, Leporace M, Laviani F, Carlani M, Manni C et al. Integrated multislice CT and Tc-99m Sestamibi SPECT-CT evaluation of solitary pulmonary nodules. *Radiol.Med.* 2006;111:213-24.

75. Spanu A, Schillaci O, Pirina P, Arru A, Madeddu G, Chessa F et al. 99mTc-tetrofosmin SPECT in solitary pulmonary nodule evaluation. *Oncol.Rep.* 2006;16:763-9.

76. Boundas D, Karatzas N, Moralidis E, Arsos G, Drevelengas A, Pistevou-Gompaki K et al. Comparative evaluation of 99mTc-depreotide and 201Tl chloride single photon emission tomography in the characterization of pulmonary lesions. *Nucl.Med.Commun.* 2007;28:533-40.

77. Halley A, Hugentobler A, Icard P, Porret E, Sobrio F, Lerochais JP et al. Efficiency of 18F-FDG and 99mTc-depreotide SPECT in the diagnosis of malignancy of solitary pulmonary nodules. *Eur.J.Nucl.Med.Mol.Imaging* 2005;32:1026-32.

78. Komori T, Narabayashi I, Matsui R, Tatsu Y, Sueyoshi K, Adachi I et al. Evaluation of uptake and release of technetium-99m MIBI SPECT of pulmonary and mediastinal lesions. *Ann.Nucl.Med.* 1997;11:227-32.

79. Suga K, Kume N, Orihashi N, Nishigauchi K, Uchisako H, Matsumoto T et al. Difference in 201Tl accumulation on single photon emission computed tomography in benign and malignant thoracic lesions. *Nucl.Med.Commun.* 1993;14:1071-8.

80. Higashi K, Ueda Y, Sakuma T, Seki H, Oguchi M, Taniguchi M et al. Comparison of [(18)F]FDG PET and (201)Tl SPECT in evaluation of pulmonary nodules. *J.Nucl.Med.* 2001;42:1489-96.

81. Birchard KR, Hoang JK, Herndon JE, Jr., Patz EF, Jr. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. *Cancer* 2009;115:581-6.

82. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J.Natl.Cancer Inst.* 1999;91:523-8.

83. Therasse P. Measuring the clinical response. What does it mean? *Eur.J.Cancer* 2002;38:1817-23.

84. Gurland J, Johnson RO. Case for using only maximum diameter in measuring tumors. *Cancer Chemother.Rep.* 1966;50:119-24.

85. Spears CP. Volume doubling measurement of spherical and ellipsoidal tumors. *Med.Pediatr.Oncol.* 1984;12:212-7.

86. Scher HI, Morris MJ, Kelly WK, Schwartz LH, Heller G. Prostate cancer clinical trial end points: "RECIST"ing a step backwards. *Clin.Cancer Res.* 2005;11:5223-32.

87. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann.Oncol.* 2004;15:257-60.

88. van Klaveren RJ, Aerts JG, de BH, Giaccone G, Manegold C, van Meerbeeck JP. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. *Lung Cancer* 2004;43:63-9.

89. Watanabe H, Yamamoto S, Kunitoh H, Sekine I, Yamamoto N, Ohe Y et al. Tumor response to chemotherapy: the validity and reproducibility of RECIST guidelines in NSCLC patients. *Cancer Sci.* 2003;94:1015-20.

90. Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin.Cancer Res.* 2003;9:4318-23.

91. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J.Clin.Oncol.* 2003;21:2574-82.

92. Tran LN, Brown MS, Goldin JG, Yan X, Pais RC, Nitt-Gray MF et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. *Acad.Radiol.* 2004;11:1355-60.

93. Egger M, Krasinski A, Rutt BK, Fenster A, Parraga G. Comparison of B-mode ultrasound, 3-dimensional ultrasound, and magnetic resonance imaging measurements of carotid atherosclerosis. *J.Ultrasound Med.* 2008;27:1321-34.

94. Krasinski A, Chiu B, Spence JD, Fenster A, Parraga G. Three-dimensional ultrasound quantification of intensive statin treatment of carotid atherosclerosis. *Ultrasound Med.Biol.* 2009;35:1763-72.

95. Landry A, Spence JD, Fenster A. Quantification of carotid plaque volume measurements using 3D ultrasound imaging. *Ultrasound Med.Biol.* 2005;31:751-62.

96. Landry A, Ainsworth C, Blake C, Spence JD, Fenster A. Manual planimetric measurement of carotid plaque volume using three-dimensional ultrasound imaging. *Med.Phys.* 2007;34:1496-505.

97. Ding M, Gyacskov I, Yuan X, Drangova M, Fenster A, Downey D. Slice-Based Prostate Segmentation in 3D US Images Using Continuity Constraint. *Conf.Proc.IEEE Eng Med.Biol.Soc.* 2005;1:662-5.

98. Karnik VV, Fenster A, Bax J, Cool DW, Gardi L, Gyacskov I et al. Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy. *Med.Phys.* 2010;37:802-13.

99. Smith WL, Lewis C, Bauman G, Rodrigues G, D'Souza D, Ash R et al. Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR. *Int.J.Radiat.Oncol.Biol.Phys.* 2007;67:1238-47.

100. Lewis RT, Zheng Y, Gethin DT. Three-dimensional unstructured mesh generation: Part 3. Volume meshes. *Comput.Methods Appl.Mech.Engrg.* 1996;134:285-310.

101. Zhang, C. and Chen, T. Efficient Feature Extraction for 2D/3D Objects in Mesh Representation. ICIP 2001 . 2001. Thessaloniki, Greece.
Ref Type: Conference Proceeding

Chapter 2: Accuracy and Reproducibility of Multi-Dimensional X-ray Computed Tomography Measurements of Pulmonary Tumour Phantoms

2.1 INTRODUCTION

Cancer is the leading cause of death in Canada, accounting for 29% of all deaths (1;2). As cancer metastasizes throughout the body, diagnostic imaging is often used to detect the presence of metastatic tumours (3), and monitor their changes longitudinally (4). Upon diagnosis, accurate measurements are necessary as the change in tumour size may be indicative of therapeutic efficacy (5;6). Measurement accuracy is also critical in clinical trials, as the efficacy of cytotoxic drugs is assessed by the change in tumour size (7;8).

To facilitate measurement reproducibility among clinicians and scientists, tumour measurements should be standardized and accurate, with high intra- and inter-observer reproducibility (9). Universal guidelines for the reproducible evaluation of the change in tumour size are essential to evaluate therapeutics and overall disease management (5;10). Two major criteria for the measurement and longitudinal assessment of tumour burden currently exist: the World Health Organization (WHO), which introduced a bidimensional (2D) measurement in 1979 (11), and the Response Evaluation Criteria in Solid Tumours (RECIST) which implemented a unidimensional (1D) measurement in 2000 (12). The 2D measurement described by the WHO criteria was defined as the product of the longest axis of the tumour and its longest perpendicular bisector. The sum of this measurement for each tumour was used to assess the change in tumour size. Despite the use of the WHO criteria in 20 years worth of clinical trials, there were response classification and

measurements limitations which eventually precluded reproducibility among clinicians and scientists (10;12-14). In particular, a limitation in the WHO criteria accuracy was the failure to establish a minimum lesion size allowable; the longest dimensions of a small tumour may be lost in the out-of-plane resolution in cross-sectional imaging. The Response Evaluation Criteria in Solid Tumours (RECIST) harmonised a method of tumour quantification (14), and provided solutions to some of the limitations of the WHO criteria, including the definition of the minimum lesion size allowable as no less than double the slice thickness of x-ray CT or MRI (12). RECIST defined a 1D measurement of the longest diameter of each tumour, based on x-ray CT and MRI (12). The sum of the longest axis of each tumour for all tumours measured was used to assess response.

Although the clinical and scientific communities adopted RECIST to evaluate response (15), investigators questioned some of the foundations of the criteria used to stratify tumour response (4;15;16). In 2009, the RECIST working group published revised RECIST 1.1 (15) to address to address these concerns; however, the improvements to the RECIST criteria were solely related to the concerns of response classification and the 1D measurement of used to quantify tumour size remained unchanged.

In daily practice, linear measurements in the axial plane are commonly used to evaluate tumour size changes (17) and has shown high inter-observer reproducibility (9). However, 1D measurements were developed under the assumption of tumour sphericity. Tumours may be lobular in shape (17), and grow asymmetrically at different growth rates (18). Although WHO has shown good inter-observer reproducibility (9), there may be an increased risk of error when performing a bi-dimensional measurement, and subsequently

calculating their products and sums of all tumour deposits (13). Direct volume (3D) measurements may potentially be more representative of tumour size (5;17;19). Improvement in the accuracy of tumour size may suggest consequent implications for response assessment (5). Technological advances of imaging modalities, such as multi-detector computed tomography (MDCT), provide an improvement in in-plane and out-of-plane resolution, suggesting increased accuracy for 3D measurements (17). In this chapter, we determine measurement accuracy using 1D, 2D, and 3D measurement tools, as well intra- and inter-observer reproducibility. The accuracy of our 3D technique is validated for accuracy and precision using a phantom with pulmonary tumour phantoms.

2.2 METHODS

2.2.1 Overview

A detailed overview of the data collected per analysis is provided in Table 2-1. The total number of measurements per analysis is calculated in Figure 2-1.

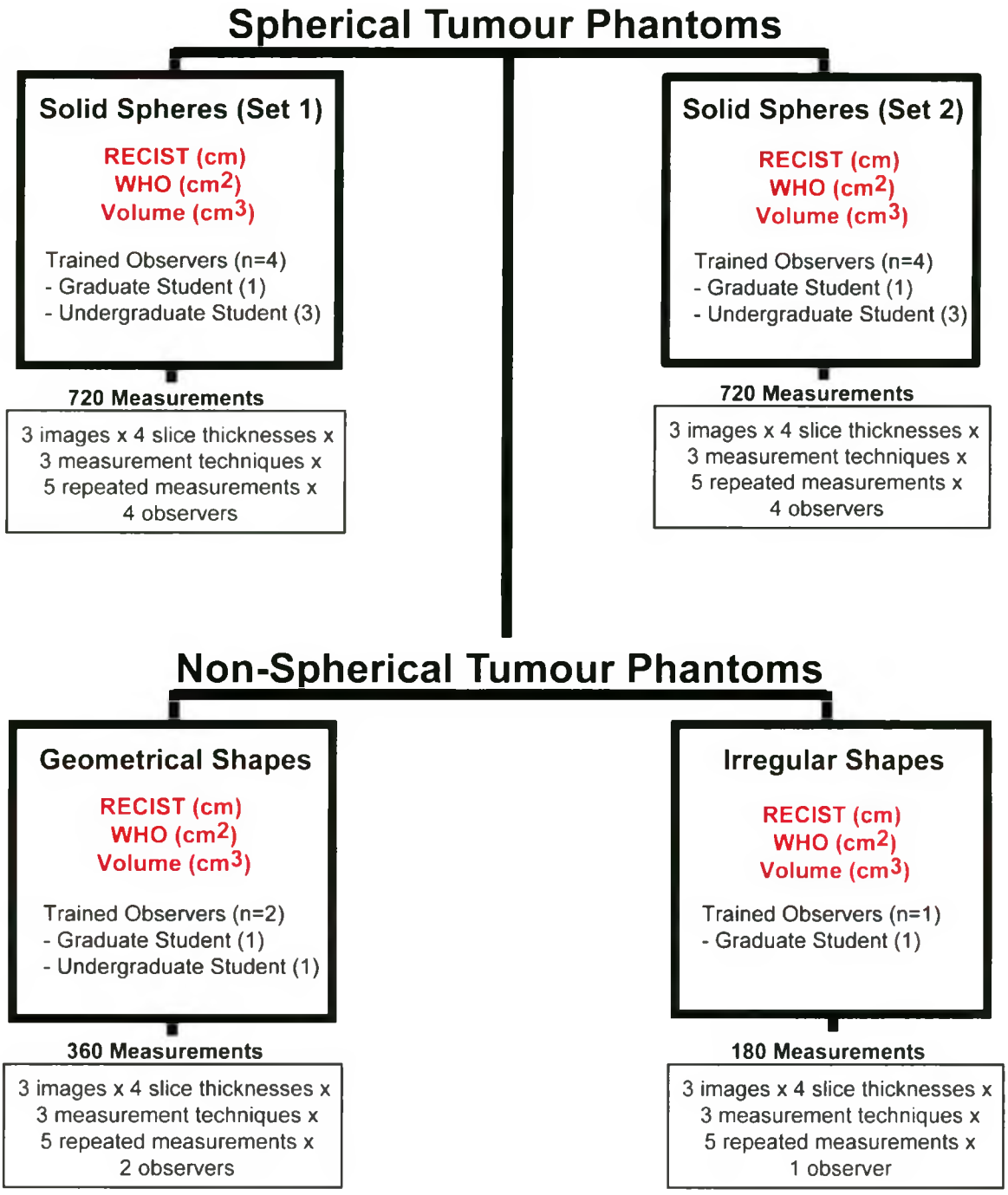
Lung Tumour Phantoms	Slice thicknesses (n)	Tumours (n)	Observers (n)			Measurements per Observer (n)		
			1D	2D	3D	1D	2D	3D
Solid Spheres (Set 1)	4	3	4	4	4	60	60	60
Solid Spheres (Set 2)	4	3	4	4	4	60	60	60
Geometrical Shapes	4	3	2	2	2	60	60	60
Irregular Shapes	4	3	1	1	1	60	60	60

1D = RECIST (cm); 2D = WHO (cm^2); 3D = Volume (cm^3)

Table 2-1: Overview of Data Collected per Analysis

The number of slice thicknesses, tumour phantoms, observers, and total measurements per observer are provided for each analysis. The number of observers and total measurements per observer are further categorized by measurement technique.

Figure 2-1: Flow Chart of Observer Classifications and Responsibilities
All observers are listed under the specific areas of the study in which they were involved. The number of subjects, tumours, time points and total tumour measurements per analysis are also listed.

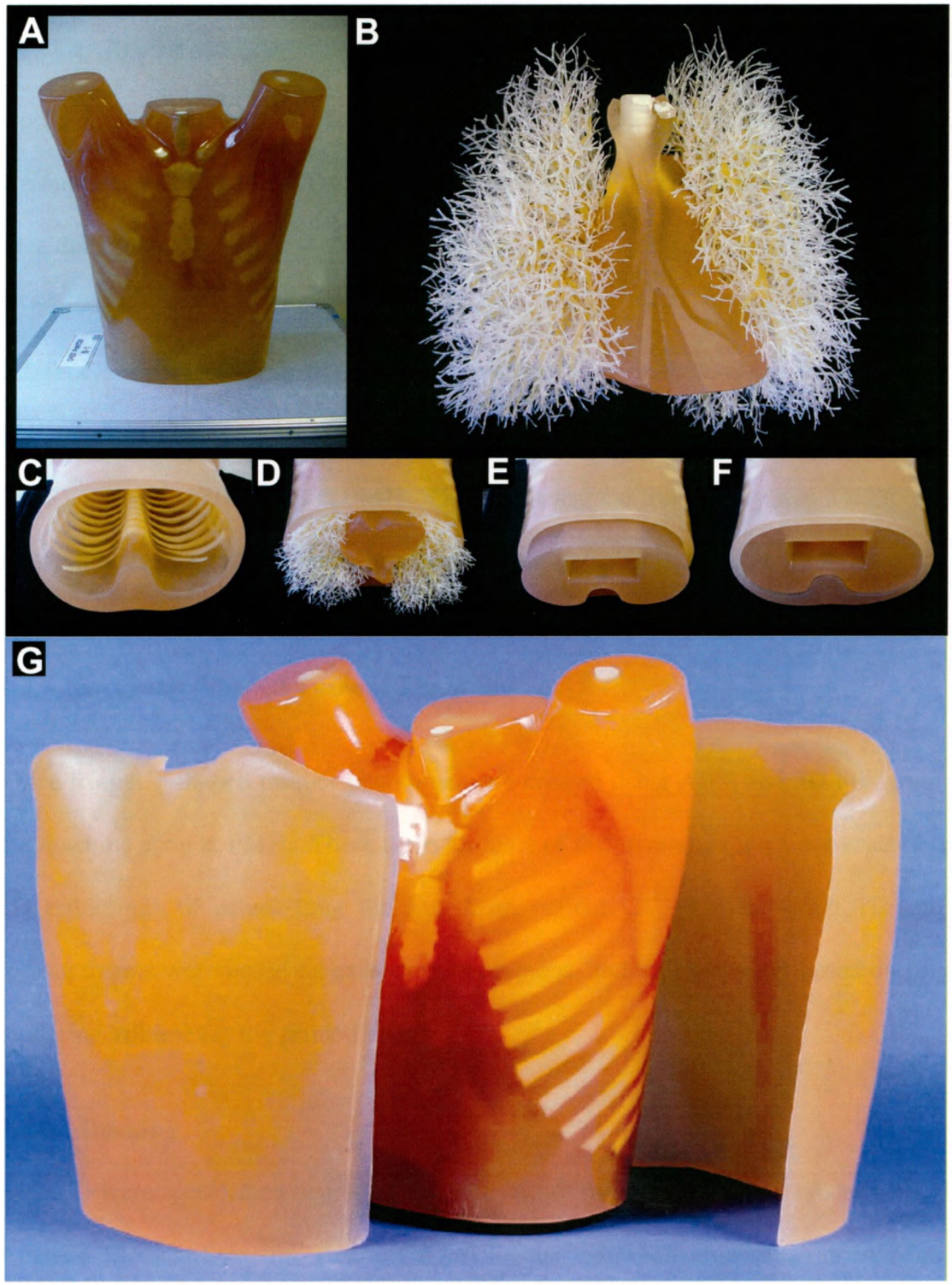


2.2.2 Chest Phantom

A commercially-available chest phantom was made available to us by Dr. H. Keller at Princess Margaret Hospital (Toronto, Ontario; chest phantom: Kyoto Kagaku Co. Ltd., Japan). The phantom was constructed with the relative proportions of a Japanese human adult thorax simulating soft tissue, bone, mediastinum (including the heart and trachea), and the pulmonary arterial tree. The soft tissue was composed of polyurethane, and the bones of epoxy resin. Figure 2.2 is a photograph of the chest phantom used in this study. Figure 2.2.G displays the additional chest plates used to replicate the relative proportions of a North American human adult thorax. The anterior and posterior plates were 60 mm and 30 mm respectively. Four phantom studies were conducted to quantify the size and shape of realistic lung tumours; various objects of known dimensions were glued inside the pulmonary vessel tree of the chest phantom. Each set of objects was imaged using x-ray CT and analyzed using RECIST and WHO criteria, as well as the volumetric techniques employed by our lab.

Figure 2-2: Chest phantom

A. Anterior view of chest phantom; B. Mediastinum and pulmonary vessel tree; C-F. Insertion of the pulmonary vessel into the phantom torso; G. Additional chest plates used to replicate the relative proportions of a North American human adult thorax.



2.2.3 Spherical Tumour Phantoms

Six spherical tumour phantoms were purchased with the chest phantom (Kyoto Kagaku Co. Ltd., Japan), and glued inside the pulmonary vessel tree of the chest phantom. These tumours were of various sizes and Hounsfield units (HU). Each observer performed five repeated measurements of each tumour using 1D, 2D, and 3D measurements. Each tumour was evaluated in CT scans of 0.5mm, 1.0mm, 2.0mm, and 5.0mm reconstructed slice thicknesses.

2.2.4 Geometrical-Shaped Tumour Phantoms

Three tumour phantoms were constructed using a plastic, acrylonitrile butadiene styrene (ABS), in the machine shop at PMH. Three-dimensional lithography was used to created each tumour as a combination of geometrical shapes (spheres, cylinders, cubes).

2.2.5 Irregular-Shaped Tumour Phantoms

Three irregular-shaped tumour phantoms were created using ABS, and 3D lithography in the machine shop at PMH. These three tumours of different sizes were constructed using 3D renderings of actual lung tumours from patient CT. Ground truth 1D and 2D measurements were based on calliper measuremens, and volumes were based on original rendering volume for the printed tumors.

2.2.6 Imaging

All CT were performed using a Toshiba Aquilion One 320-slice system (Toshiba America Medical Systems, Tustin, CA). All scans were performed at Princess Margaret Hospital (PMH; Toronto, Canada), and the images were sent to the Imaging Research

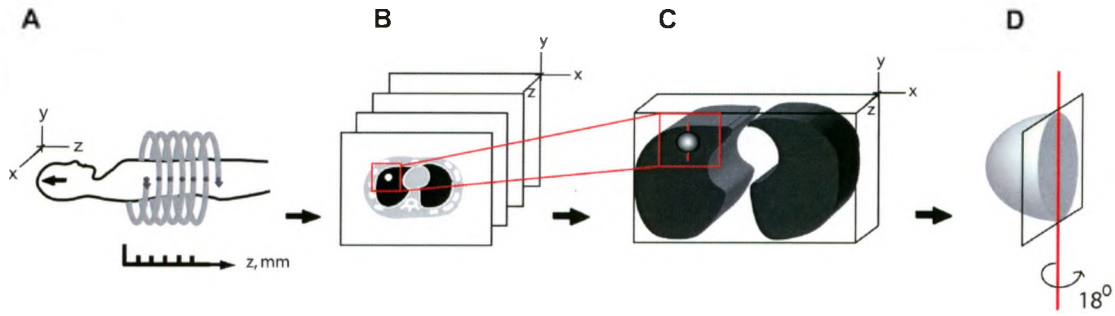
Laboratories at Robarts Research Institute (Robarts; London, Canada) for analysis. Scans were acquired in helical mode at x-ray tube voltage, 120kV, x-ray tube current, 200mA, and were reconstructed with 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses. The pixel matrix was 512 x 512. Detailed scanning parameters are described in Appendix B.

2.2.7 Image Analysis

Image analysis was performed blinded to ground truth measurements. CT images were displayed on LCD screens using conventional parameters in lung (window width: 1600HU, window centre: -550) (9;20). Each observer could magnify and manipulate window setting to optimize the display of each tumour deposit (7). Image analysis using the RECIST and WHO criteria was performed using electronic calipers in an open-source picture archiving and communication system (PACS), ClearCanvas (ClearCanvas, Inc., Toronto, Canada).

Volumetric analysis was performed using manual segmentation of tumour boundaries, after establishing an axis of rotation in the centre of the tumour, and rotating this tumour about an angle of 18°.

Figure 2-3: Image Reconstruction for Three-Dimensional Measurements
A. The phantom is scanned in the cranial-caudal direction; B. Image slices are displayed on computer monitors using ClearCanvas; C. All slices are reconstructed into a volume in 3DQuantify; D. The user-defined rotational axis rotates the tumour volume by 18o and the tumour boundary is delineated.



A schematic of the process for performing the 3D measurements is shown in Figure 2-3. Ten tumour boundaries were contoured for each tumour, and a smoothing algorithm within the software platform calculated the volume. All manual segmentation was performed using an in-house image processing software platform as previously described (21). RECIST, WHO, and volumetric image segmentation were repeated five times and the mean values, standard deviation, and coefficient of variation were recorded. For the RECIST, WHO, and volumetric measurements, the longest axis, the cross-product of longest axis and perpendicular bisector, and volumes were respectively summed to obtain the total tumour burden per subject. All measurements were computed for each tumour independently as well as combined.

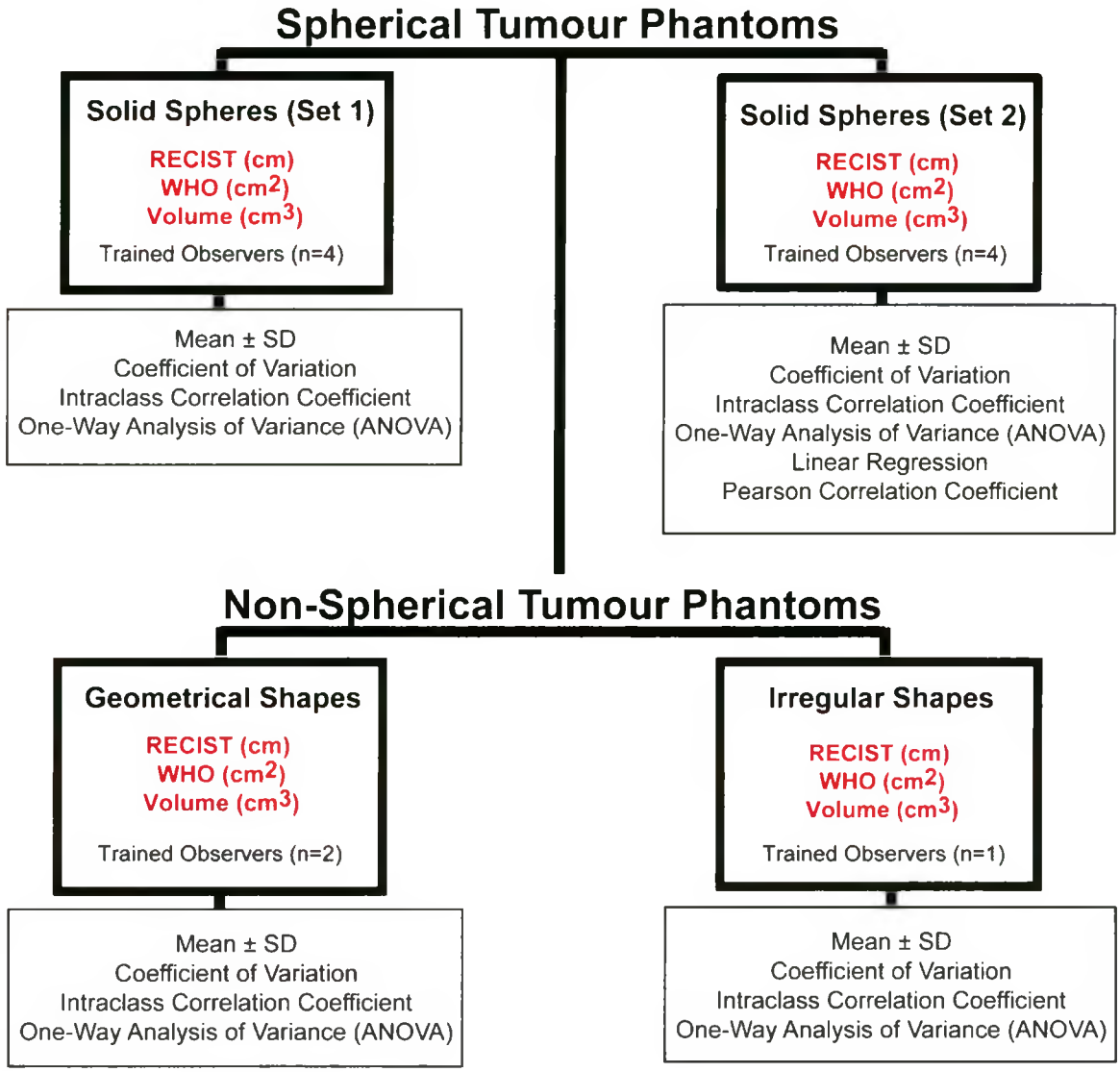
Four observers at Robarts participated in this study. Table 2-1 outlines the data collected for each phantom scan, and Figure 2-1 lists the total number of measurements performed for each phantom scan. Each observer was blinded to the ground truth dimensions of the

tumour phantoms sent from PMH until after all measurements were completed. Measurement trials were performed with at least one day between measurements.

2.2.8 Statistical Analysis

An overview of the statistical analyses performed in this chapter is shown in Figure 2-4. Mean RECIST, WHO, and volumes, and standard deviations were calculated from five repeated measures for all lung tumour phantoms (all data presented as mean ± one SD). Differences between reconstructed slice thicknesses and observers were evaluated using a one-way analysis of variance (ANOVA). Inter- and intra-observer reproducibility was evaluated using the Interclass Correlation Coefficient (consistency: ICC(C); absolute agreement: ICC(A)) and the coefficient of variation (CV), which is the standard deviation divided by the mean of the measurements. All statistical analysis was performed using SPSS 17.0 (SPSS Inc, Chicago, IL) and results were considered significant when the probability of making a Type I error was less than 5% ($p < 0.05$).

Figure 2-4: Flow Chart of Statistical Analyses per Study Component
All study components and respective statistical analyses are listed.



2.3 RESULTS

2.3.1 RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms

The descriptive statistics for the solid spherical tumour phantoms are presented, and the mean ± SD of each tumour measured in 1D, 2D, and 3D, are reported for each observer. The CVs are reported for each observer, and ICC (A) and ICC(C) are reported to show

intra- and inter-observer reproducibility. ANOVAs are presented to show significant differences between the data.

The 1D, 2D, and 3D mean ± SD measurements of Tumours 1 is reported in Table 2-2. Observer measurements for Tumours 2 and 3 are in Appendix C. Each tumour was measured by four observers, and each observer performed five repeated measurements. The three tumours were quantified at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses.

2.3.2 Reproducibility of RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms (Set 1)

To assess measurement reproducibility, each tumour was measured five times. The coefficients of variation (CV) for Tumour 1 are reported in Table 2-3. This table shows the CV of each observer at each slice thickness. The CVs reported in this table were calculated from the means and standard deviations of the individual tumours measured by each observer in Table 2-2.

Table 2-4 reports the intra- and inter-observer reproducibility of 1D, 2D, and 3D measurements. The ICC value for each individual observer is the intra-observer reproducibility, and the inter-observer reproducibility is reported as the ICC values for all observers. The ICC(A) measures the absolute agreement of the repeated measures, and the ICC(C) measures the consistency of repeated measures. An ICC value of 1.000 is the highest value achievable (22;23); the ICC values in Table 2-4 show high intra- and inter-observer reproducibility for each observer 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses and for 1D, 2D, and 3D measurements.

Mean ± SD observer measurements			
	Tumour 1		
	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm			
Observer 1 (JM)	1.0 ± 0.0	1.00 ± 0.00	0.53 ± 0.05
Observer 2 (FS)	1.0 ± 0.0	1.00 ± 0.00	0.54 ± 0.03
Observer 3 (LW)	1.0 ± 0.0	1.00 ± 0.00	0.52 ± 0.02
Observer 4 (SS)	1.0 ± 0.1	0.92 ± 0.10	0.45 ± 0.01
1.0mm			
Observer 1 (JM)	1.0 ± 0.0	1.00 ± 0.00	0.49 ± 0.06
Observer 2 (FS)	1.0 ± 0.0	1.00 ± 0.00	0.57 ± 0.08
Observer 3 (LW)	1.0 ± 0.0	1.00 ± 0.00	0.59 ± 0.05
Observer 4 (SS)	0.9 ± 0.1	0.89 ± 0.10	0.51 ± 0.04
2.0mm			
Observer 1 (JM)	1.0 ± 0.0	1.00 ± 0.00	0.50 ± 0.05
Observer 2 (FS)	1.0 ± 0.0	1.00 ± 0.00	0.58 ± 0.08
Observer 3 (LW)	1.0 ± 0.0	1.00 ± 0.00	0.53 ± 0.03
Observer 4 (SS)	1.0 ± 0.0	1.00 ± 0.00	0.64 ± 0.07
5.0mm			
Observer 1 (JM)	1.0 ± 0.0	1.00 ± 0.00	0.51 ± 0.06
Observer 2 (FS)	1.0 ± 0.0	1.00 ± 0.00	0.62 ± 0.09
Observer 3 (LW)	1.0 ± 0.0	1.00 ± 0.00	0.57 ± 0.03
Observer 4 (SS)	1.0 ± 0.0	1.00 ± 0.00	0.64 ± 0.11

Table 2-2: Mean 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 1)
The mean 1D, 2D, and 3D measurements for Tumour 1 are shown in this table.

Intra-observer coefficient of variation (%)			
	Tumour 1		
	RECIST (%)	WHO (%)	VOLUME (%)
0.5mm			
Observer 1 (JM)	0.0	0.00	9.44
Observer 2 (FS)	0.0	0.00	6.04
Observer 3 (LW)	0.0	0.00	3.71
Observer 4 (SS)	5.7	11.26	3.28
1.0mm			
Observer 1 (JM)	0.0	0.00	11.68
Observer 2 (FS)	0.0	0.00	13.42
Observer 3 (LW)	0.0	0.00	8.24
Observer 4 (SS)	5.8	11.75	7.23
2.0mm			
Observer 1 (JM)	0.0	0.00	9.67
Observer 2 (FS)	0.0	0.00	13.61
Observer 3 (LW)	0.0	0.00	5.50
Observer 4 (SS)	0.0	0.00	11.26
5.0mm			
Observer 1 (JM)	0.0	0.00	11.91
Observer 2 (FS)	0.0	0.00	14.44
Observer 3 (LW)	0.0	0.00	4.80
Observer 4 (SS)	0.0	0.00	17.64

Table 2-3: Mean CVs of 1D, 2D, and 3D Measurements of Solid Spherical Tumour 1 (Set 1)

Mean coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumour 1 is reported. This tumour was measured at all slice thickness by each observer. Values are expressed as percents.

	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC(A)	ICC(C)	ICC(A)	ICC(C)	ICC(A)	ICC(C)
0.5mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.975	0.981
Observer 2 (FS)	0.985	0.985	0.978	0.978	0.987	0.995
Observer 3 (LW)	1.000	1.000	0.997	0.997	0.998	0.999
Observer 4 (SS)	0.948	0.963	0.942	0.957	0.998	0.997
All Observers	**0.969**	**0.985**	**0.970**	**0.984**	**0.975**	**0.989**
1.0mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.981	0.990
Observer 2 (FS)	0.985	0.985	0.978	0.978	0.972	0.979
Observer 3 (LW)	0.985	0.985	0.981	0.981	0.988	0.990
Observer 4 (SS)	0.964	0.974	0.953	0.967	0.984	0.986
All Observers	**0.953**	**0.985**	**0.952**	**0.977**	**0.968**	**0.982**
2.0mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.967	0.963
Observer 2 (FS)	0.985	0.985	0.991	0.991	0.937	0.986
Observer 3 (LW)	0.985	0.985	0.993	0.996	0.986	0.990
Observer 4 (SS)	0.986	0.986	1.000	1.000	0.966	0.968
All Observers	**0.970**	**0.980**	**0.970**	**0.980**	**0.923**	**0.966**
5.0mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.891	0.902
Observer 2 (FS)	1.000	1.000	1.000	1.000	0.926	0.963
Observer 3 (LW)	0.985	0.985	0.986	0.985	0.980	0.985
Observer 4 (SS)	0.984	0.984	1.000	1.000	0.960	0.971
All Observers	**0.957**	**0.974**	**0.968**	**0.980**	**0.938**	**0.957**

Table 2-4: Correlation Coefficients for Each Observer at Each Slice Thickness (Solid Spheres Set 1)

Absolute agreement and consistency intra-class correlation coefficients (ICC) for each observer at each slice thickness are reported. Inter-observer and intra-observer reliability are assessed for RECIST, WHO, and volumetric measurements.

2.3.3 Accuracy of RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms (Set 1)

The ground truth measurements for the three solid spherical tumour phantoms are shown in Table 2-5. The 1D measurement for the three tumours are 1.0cm, 0.8cm, and 1.2cm, respectively. The 2D measurements for the three tumours are 1.00cm^2, 0.64cm^2, and 1.44cm^2, respectively. The 3D measurements for the three tumours are 0.52cm^3, 0.27cm^3, and 0.90cm^3, respectively. The mean measurements of all observers were compared to the phantom measurements in this table to determine observer accuracy.

	RECIST (cm)	WHO (cm^2)	Volume (cm^3)
Tumour 1	1.0	1.00	0.52
Tumour 2	0.8	0.64	0.27
Tumour 3	1.2	1.44	0.90

Table 2-5: Phantom Ground Truth Measurements for Solid Spherical Tumours (Set 1)
RECIST, WHO, and 3D ground truth measurements are reported. These measurements were used as ground truth when compared to the multiple observers for accuracy testing.

RECIST and WHO measurements for solid spherical Tumour 3 are shown in Figure 2-5. The RECIST measurement is shown the longest axis of the tumour (red line). The length of the red line is 1.2cm; therefore, the RECIST measurement for this tumour is 1.2 cm. The World Health Organization (WHO) measurement is calculated as the product the longest axis of the tumour (red line) and its longest perpendicular bisector (blue line). The lengths of the red and blue lines are both 1.2cm; the WHO measurement for this tumour is 1.44cm^2.

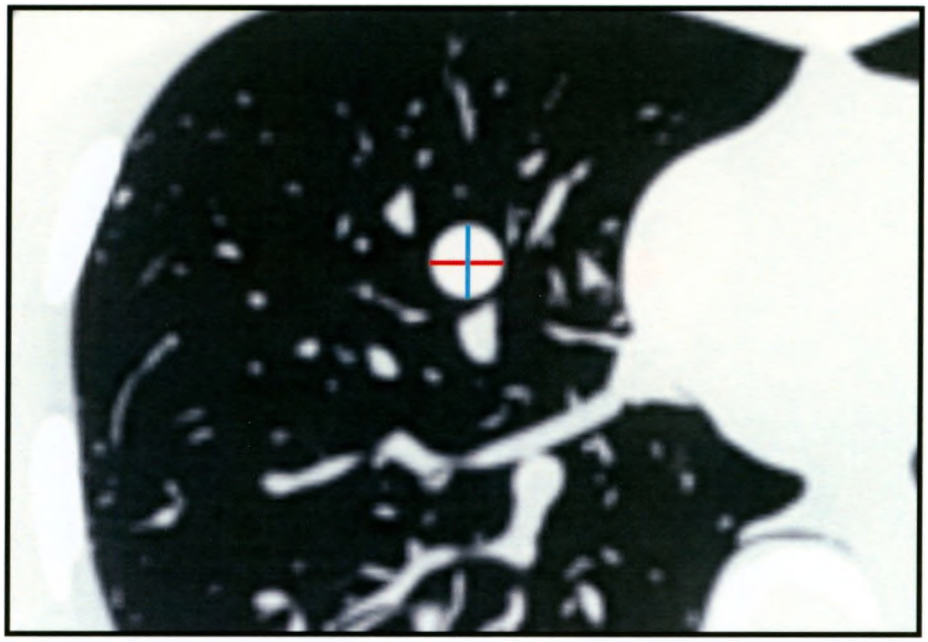


Figure 2-5: RECIST and WHO Measurements for Solid Spherical Tumour 3
The longest axis measurement (RECIST) is shown in red, and the longest perpendicular bisector is shown in blue. The WHO measurement is calculated as the product of the red and blue lines. This image was reconstructed with 0.5mm slice thickness.

Volume rendering of the three solid spherical tumour phantoms are shown in Figure 2-6. Tumour 1 was located in the parenchyma of the left lung, and Tumour 2 and 3 were located in the right lung parenchyma. This image is a 2D representation of three 3D objects hanging in space, and as such, it appears that Tumour 1 is located in the heart, and instead of the left lung parenchyma.

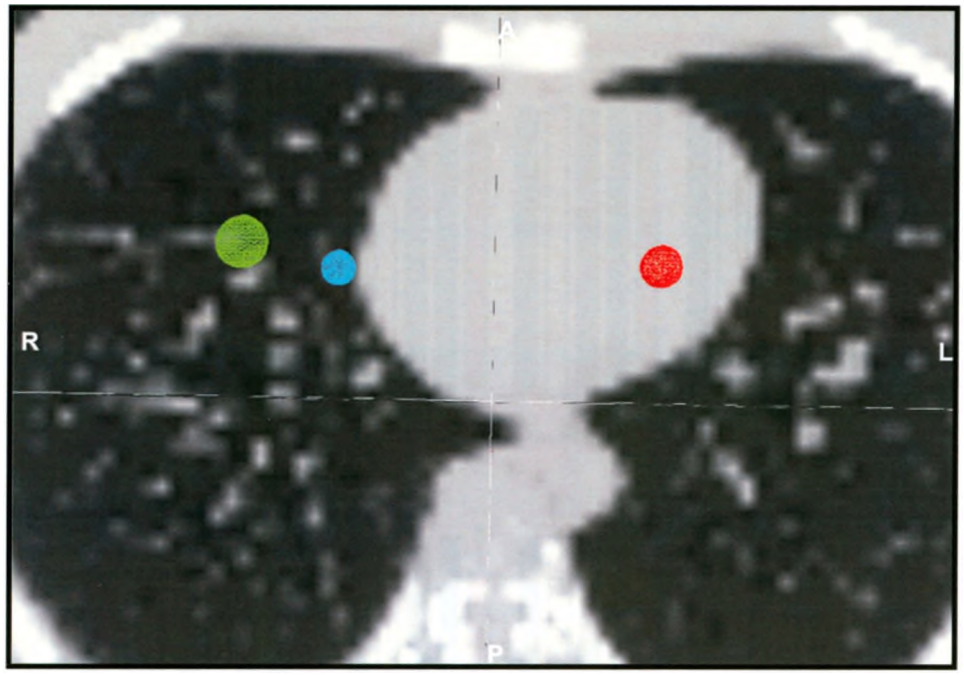


Figure 2-6: Volume Rendering of Solid Spherical Tumour Phantoms (Set 1)
The volumetric images of the three solid spherical tumour phantoms are shown. The ground truth volumes for Tumours 1, 2, and 3, are 0.52cm3 (red), 0.27cm3 (blue), and 0.90cm3 (green), respectively. This image was reconstructed with 0.5mm slice thickness.

To determine the accuracy of observer measurements, an ANOVA was used to test for significant differences between observer measurements and the known phantom dimensions. The mean ± SD values for 1D, 2D, and 3D measurements, and the results of the ANOVA for Tumour 1 at each slice thickness are reported in Figure 2-7. The results for Tumours 2 and 3 are in Appendix C.

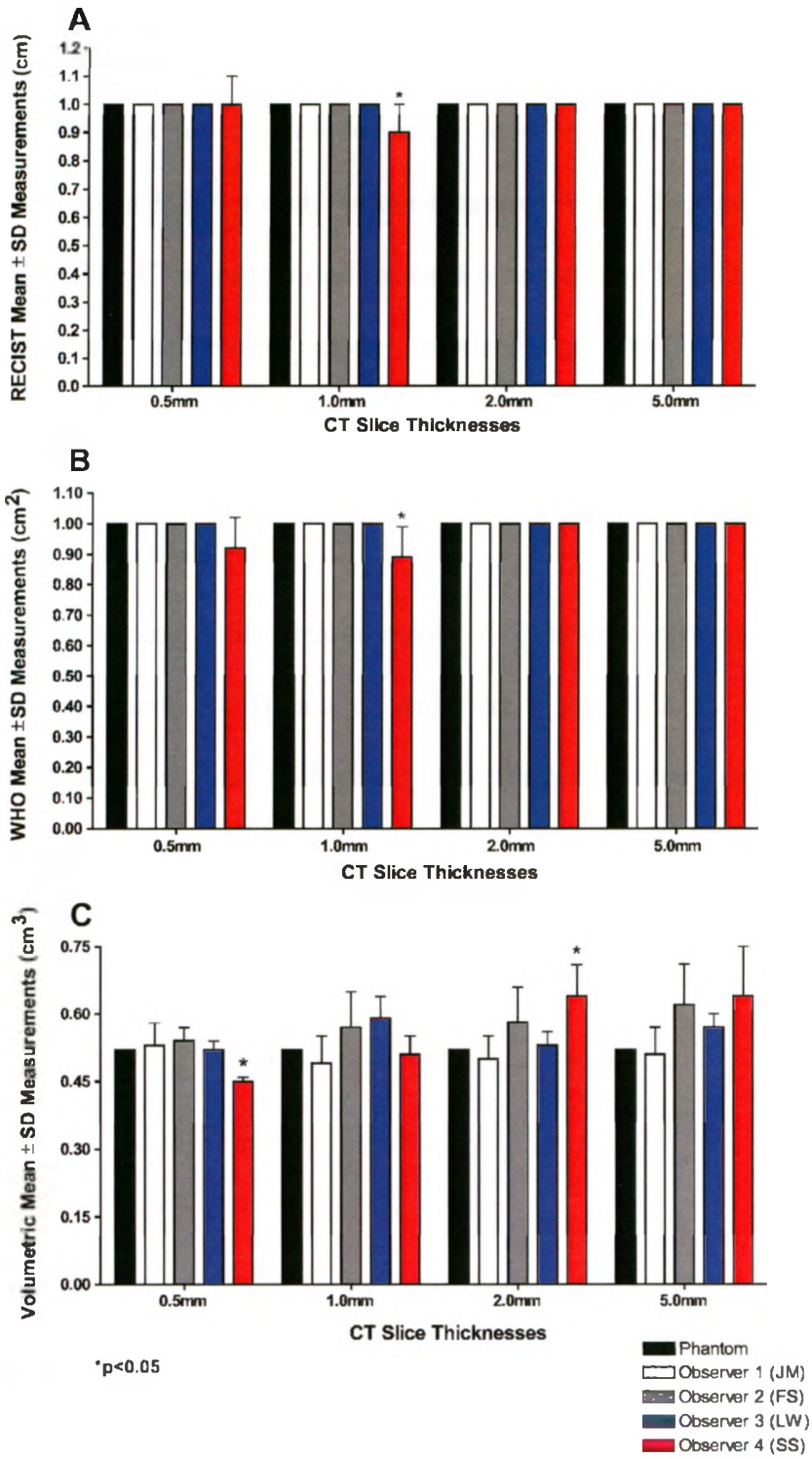


Figure 2-7: Statistical Differences Between Phantom and Observer Volume Measurements of Solid Spherical Tumour 1 (Set 1)

(A) Observer 4 at 1.0mm slice thickness (0.9 ± 0.1cm, $p<0.001$)

(B) Observer 4 at 1.0mm slice thickness (0.89 ± 0.10cm^2, $p=0.007$)

(C) Observer 4 at 0.5mm (0.45 ± 0.01cm^3, $p=0.011$); and 2.0mm slice thicknesses (0.64 ± 0.07cm^3, $p=0.021$)

2.3.4 RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms (Set 2)

The descriptive statistics for the second set of solid spherical tumour phantoms are presented, and the mean ± SD of each tumour measured in 1D, 2D, and 3D, are reported for each observer. The CVs are reported for each observer, and ICC (A) and ICC(C) are reported to show intra- and inter-observer reproducibility. ANOVAs are presented to show significant differences between the data, and linear regressions and Pearson correlations report relationships between observer measurements.

The 1D, 2D, and 3D mean ± SD measurements of Tumour 1 is reported in Table 2-6. The observer measurements for Tumours 2 and 3 are in Appendix C. Each tumour was measured by four observers, and each observer performed five repeated measurements. The three tumours were quantified at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses.

	Tumour 1		
	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm			
Observer 1 (JM)	0.6 ± 0.0	0.36 ± 0.00	0.08 ± 0.02
Observer 2 (FS)	0.6 ± 0.0	0.36 ± 0.00	0.10 ± 0.01
Observer 3 (LW)	0.5 ± 0.0	0.25 ± 0.00	0.07 ± 0.01
Observer 4 (SS)	0.5 ± 0.0	0.25 ± 0.00	0.08 ± 0.01
1.0mm			
Observer 1 (JM)	0.6 ± 0.0	0.36 ± 0.00	0.07 ± 0.02
Observer 2 (FS)	0.6 ± 0.0	0.34 ± 0.05	0.08 ± 0.02
Observer 3 (LW)	0.5 ± 0.0	0.25 ± 0.00	0.06 ± 0.01
Observer 4 (SS)	0.5 ± 0.0	0.25 ± 0.00	0.08 ± 0.01
2.0mm			
Observer 1 (JM)	0.6 ± 0.0	0.36 ± 0.00	0.11 ± 0.03
Observer 2 (FS)	0.6 ± 0.0	0.36 ± 0.00	0.09 ± 0.02
Observer 3 (LW)	0.5 ± 0.0	0.25 ± 0.00	0.08 ± 0.01
Observer 4 (SS)	0.5 ± 0.0	0.25 ± 0.00	0.09 ± 0.01
5.0mm			
Observer 1 (JM)	0.6 ± 0.0	0.36 ± 0.00	0.14 ± 0.02
Observer 2 (FS)	0.5 ± 0.1	0.29 ± 0.06	0.11 ± 0.02
Observer 3 (LW)	0.5 ± 0.0	0.25 ± 0.00	0.10 ± 0.02
Observer 4 (SS)	0.5 ± 0.0	0.25 ± 0.00	0.09 ± 0.01

Table 2-6: Mean 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 2) Mean RECIST, WHO, and volumetric measurements for Tumours 1, 2, and 3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thickness by each observer. Values are expressed as mean ± SD of five repeated measurements

2.3.5 Reproducibility of RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms (Set 2)

To assess measurement reproducibility, each tumour was measured five times. The coefficients of variation (CV) for Tumour 1 are reported in Table 2-7. The CVs for Tumours 2 and 3 are in Appendix C. This table shows the CV of each observer at each

slice thickness. The CVs reported in this table were calculated from the means and standard deviations of the individual tumours measured by each observer in Table 2-6.

Table 2-8 compares the intra- and inter-observer reproducibility of 1D, 2D, and 3D measurements. The ICC value for each individual observer is the intra-observer reproducibility, and the inter-observer reproducibility is reported as the ICC values for all observers. The ICC(A) measures the absolute agreement of the repeated measures, and the ICC(C) measures the consistency of repeated measures. An ICC value of 1.000 is the highest value achievable (22;23); the ICC values in Table 2-8 show high intra- and inter-observer reproducibility for each observer 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses and for 1D, 2D, and 3D measurements.

Intra-Observer Coefficient of Variation (%)			
	Tumour 1		
	RECIST (%)	WHO (%)	VOLUME (%)
0.5mm			
Observer 1 (JM)	0.0	0.00	23.46
Observer 2 (FS)	0.0	0.00	5.71
Observer 3 (LW)	0.0	0.00	18.13
Observer 4 (SS)	0.0	0.00	0.10
1.0mm			
Observer 1 (JM)	0.0	0.00	28.57
Observer 2 (FS)	7.7	14.55	27.95
Observer 3 (LW)	0.0	0.00	8.56
Observer 4 (SS)	0.0	0.00	0.05
2.0mm			
Observer 1 (JM)	0.0	0.00	31.00
Observer 2 (FS)	0.0	0.00	17.86
Observer 3 (LW)	0.0	0.00	6.52
Observer 4 (SS)	0.0	0.00	0.08
5.0mm			
Observer 1 (JM)	0.0	0.00	12.62
Observer 2 (FS)	10.1	20.49	20.33
Observer 3 (LW)	0.0	0.00	20.00
Observer 4 (SS)	0.0	0.00	0.10

Table 2-7: Mean CVs of 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 2)

Mean coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumour 1 are reported. This tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by each observer. Values are expressed as

	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC(A)	ICC(C)	ICC(A)	ICC(C)	ICC(A)	ICC(C)
0.5mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.938	0.943
Observer 2 (FS)	0.957	0.957	0.938	0.940	0.984	0.979
Observer 3 (LW)	0.980	0.980	0.992	0.992	0.941	0.961
Observer 4 (SS)	1.000	1.000	1.000	1.000	0.990	0.988
All Observers	**0.918**	**0.967**	**0.907**	**0.968**	**0.938**	**0.958**
1.0mm						
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.942	0.977
Observer 2 (FS)	0.928	0.949	0.922	0.940	0.959	0.962
Observer 3 (LW)	1.000	1.000	1.000	1.000	0.966	0.966
Observer 4 (SS)	1.000	1.000	1.000	1.000	0.970	0.977
All Observers	**0.921**	**0.979**	**0.907**	**0.972**	**0.918**	**0.953**
2.0mm						
Observer 1 (JM)	1.000	1.000	0.992	0.992	0.843	0.866
Observer 2 (FS)	0.948	0.948	0.930	0.930	0.938	0.958
Observer 3 (LW)	0.972	0.972	0.980	0.975	0.976	0.981
Observer 4 (SS)	1.000	1.000	1.000	1.000	0.951	0.970
All Observers	**0.920**	**0.965**	**0.908**	**0.963**	**0.862**	**0.875**
5.0mm						
Observer 1 (JM)	0.977	0.977	0.969	0.969	0.914	0.905
Observer 2 (FS)	0.935	0.954	0.933	0.952	0.872	0.898
Observer 3 (LW)	1.000	1.000	1.000	1.000	0.878	0.957
Observer 4 (SS)	1.000	1.000	1.000	1.000	0.972	0.987
All Observers	**0.893**	**0.966**	**0.876**	**0.954**	**0.808**	**0.866**

Table 2-8: Correlation Coefficients for Each Observer at Each Slice Thickness (Solid Spheres Set 2)
Absolute agreement and consistency intra-class correlation coefficients (ICC) for each observer at each slice thickness are reported. Inter-observer and intra-observer reliability are assessed for RECIST, WHO, and volumetric measurements.

Inter-observer reproducibility for 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses are reported in **Table 2-**. The repeated measures of all four observers were analyzed together to compare measurement reproducibility between slice thicknesses. This table shows there is high inter-observer reproducibility between slice thicknesses. **Table 2-** also shows high inter-observer reproducibility between 1D, 2D, and 3D measurements.

This table shows that 1D, 2D, and 3D measurements are highly reproducible at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses.

2.3.6 Accuracy of RECIST, WHO, and Volumetric Measurements of Solid Spherical Tumour Phantoms (Set 2)

The ground truth measurements for the three solid spherical tumour phantoms are shown in Table 2-9. The 1D measurement for the three tumours are 0.5cm, 0.8cm, and 0.8cm, respectively. The 2D measurements for the three tumours are $0.25cm^2$, $0.64cm^2$, and $0.64cm^2$, respectively. The 3D measurements for the three tumours are $0.07cm^3$, $0.27cm^3$, and $0.27cm^3$, respectively. The mean measurements of all observers were compared to the phantom measurements in this table to determine observer accuracy.

	RECIST (cm)	WHO (cm^2)	Volume (cm^3)
Tumour 1	0.5	0.25	0.07
Tumour 2	0.8	0.64	0.27
Tumour 3	0.8	0.64	0.27

Table 2-9: Phantom Ground Truth Measurements for Solid Spherical Tumours (Set 2)
RECIST, WHO, and 3D ground truth measurements are reported. These measurements were used as ground truth when compared to the multiple observers for accuracy testing.

Tumour 1 had a CT number of +100HU, Tumour 2 had a CT number of +100HU, and Tumour 3 had a CT number of -600HU. Volume rendering of the three solid spherical tumour phantoms are shown in Figure 2-8. All three tumour phantoms were located in the right lung parenchyma.

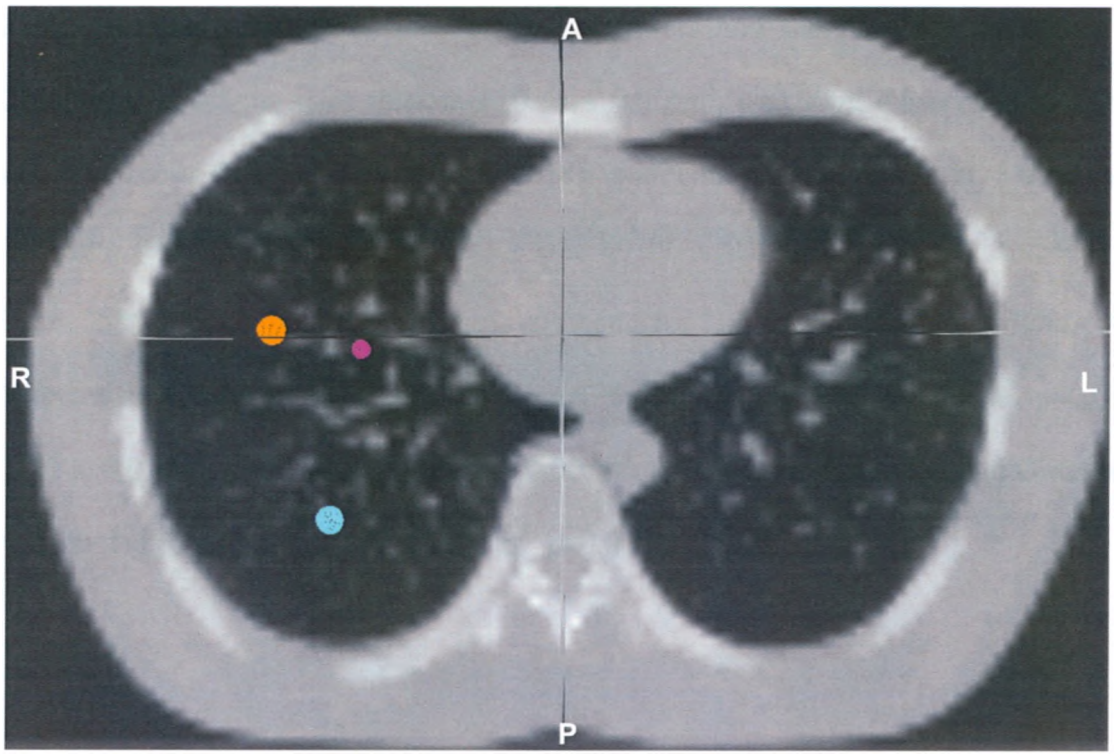


Figure 2-8: Volume Rendering of Solid Spherical Tumour Phantoms (Set 2)
The volumetric images of the three spherical tumour phantoms are shown. The ground truth volumes for Tumours 1, 2, and 3, are 0.07cm^3 (purple), 0.27cm^3 (orange), and 0.27cm^3 (blue), respectively. This image was reconstructed with 0.5mm slice thickness.

Volumetric measurements were performed for Tumours 1-3 by each observer. All four observers repeated five 3D measurements of the tumour phantoms at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses. Figure 2-9 presents mean ± SD 3D measurements, as well as the results of an ANOVA used to determine significant differences between observer measurements and ground truth phantom dimensions. In Figure 2-9A, Observer 2 shows a significant difference from the phantom measurement at 0.5mm slice thickness (0.10 ± 0.01cm3, p=0.023); Observer 1 at 2.0mm (0.11 ± 0.03cm3, p=0.025); Observers 1 and 2 at 5.0mm slice thickness (0.14 ± 0.02cm3, p<0.0001), and (0.11 ± 0.02cm3, p=0.008), respectively. Mean ± SD 3D measurements for Tumour 2 are shown in Figure

2-9B. A significant difference between observers and the phantom are shown by Observer 2 at 0.5mm slice thickness (0.33 ± 0.02cm^3, p=0.006); Observer 1 at 2.0mm slice thickness (0.37 ± 0.09cm3, p=0.038); and Observers 1 and 2 at 5.0mm slice thickness (0.54 ± 0.09cm^3, p<0.0001), and (0.39 ± 0.08cm^3, p=0.046), respectively. Figure 2-9C reports the mean ± SD measurements for Tumour 3, which were significantly different for Observers 2, 3, and 4 at 1.0mm slice thickness (0.32 ± 0.03cm^3, p=0.011), (0.21 ± 0.01cm^3, p=0.002), and (0.31 ± 0.02cm^3, p=0.044), respectively; and Observer 2 at 2.0mm slice thickness (0.37 ± 0.05cm^3, p<0.0001).

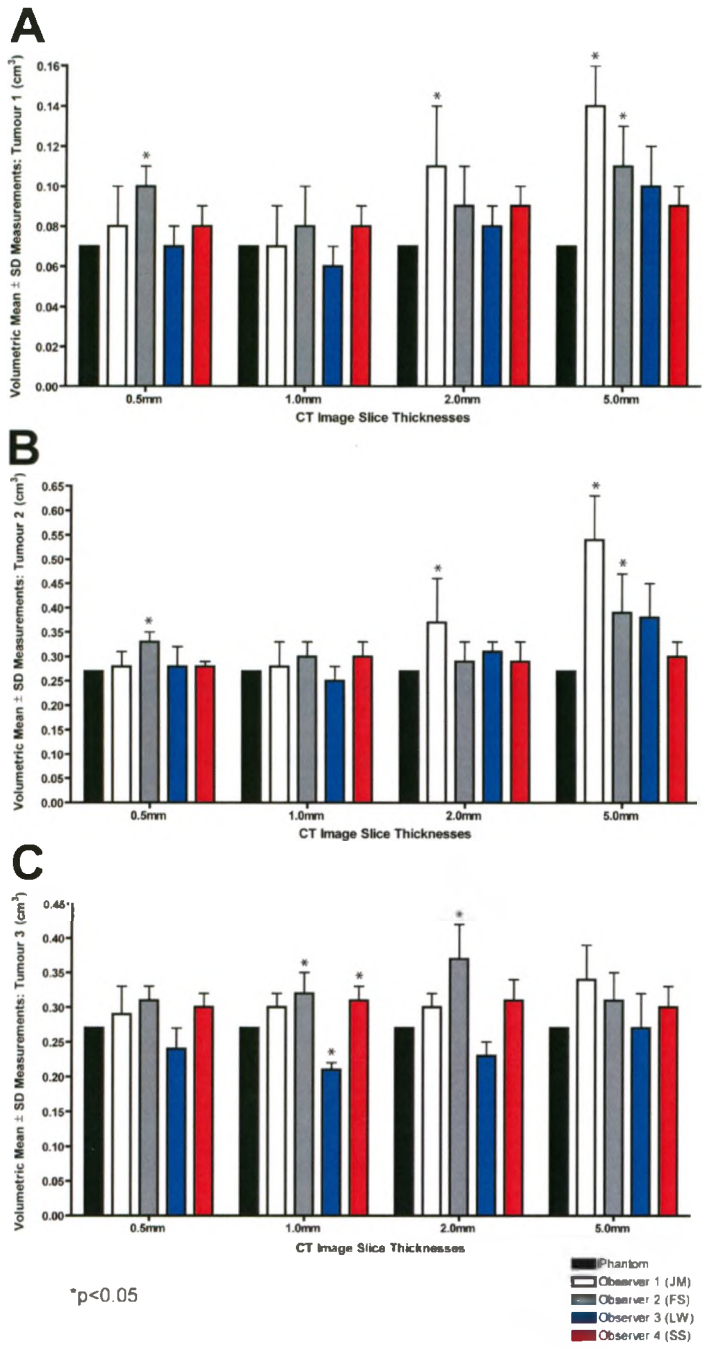


Figure 2-9: Statistical Differences Between Phantom and Observer Volume Measurements of Solid Spherical Tumours (Set 2)

(A) Observer 2 at 0.5mm slice thickness (0.10 ± 0.01cm^3, p=0.023); Observer 1 at 2.0mm (0.11 ± 0.03cm^3, p=0.025); Observers 1 and 2 at 5.0mm slice thickness (0.14 ± 0.02cm^3, p<0.0001), and (0.11 ± 0.02cm^3, p=0.008), respectively.

(B) Observer 2 at 0.5mm slice thickness (0.33 ± 0.02cm^3, p=0.006); Observer 1 at 2.0mm slice thickness (0.37 ± 0.09cm^3, p=0.038); Observers 1 and 2 at 5.0mm slice thickness (0.54 ± 0.09cm^3, p<0.0001), and (0.39 ± 0.08cm^3, p=0.046), respectively.

(C) Observers 2, 3, and 4 at 1.0mm slice thickness (0.32 ± 0.03cm^3, p=0.011), (0.21 ± 0.01cm^3, p=0.002), and (0.31 ± 0.02cm^3, p=0.044), respectively; Observer 2 at 2.0mm slice thickness (0.37 ± 0.05cm^3, p<0.0001).

2.3.7 RECIST, WHO, and Volumetric Measurements of Geometrically-Shaped Tumour Phantoms

The descriptive statistics for the geometrical-shaped tumour phantoms are presented, and the mean ± SD of each tumour measured in 1D, 2D, and 3D, are reported for each observer. The CVs are reported for each observer, and ICC (A) and ICC(C) are reported to show intra- and inter-observer reproducibility. ANOVAs are presented to show significant differences between the data.

The 1D, 2D, and 3D mean ± SD measurements of Tumours 1 and 2 are reported in Table 2-10. Scaled images of each tumour are shown in Appendix C. Each tumour was measured by two observers, and both observers performed five repeated measurements. The three tumours were quantified at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses.

	Tumour 1			Tumour 2		
	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm						
Observer 1 (JM)	3.8 ± 0.0	12.46 ± 0.20	8.94 ± 0.38	3.0 ± 0.0	7.68 ± 0.81	4.46 ± 0.22
Observer 2 (LW)	3.5 ± 0.1	11.62 ± 0.29	7.41 ± 0.23	2.9 ± 0.0	6.23 ± 0.45	3.66 ± 0.19
1.0mm						
Observer 1 (JM)	3.6 ± 0.0	12.39 ± 0.20	9.63 ± 0.89	3.0 ± 0.1	8.04 ± 0.65	4.55 ± 0.30
Observer 2 (LW)	3.5 ± 0.0	11.62 ± 0.29	7.70 ± 0.41	2.9 ± 0.0	6.23 ± 0.45	3.86 ± 0.20
2.0mm						
Observer 1 (JM)	3.6 ± 0.0	12.39 ± 0.31	9.53 ± 0.25	3.0 ± 0.1	8.39 ± 0.99	4.80 ± 0.25
Observer 2 (LW)	3.4 ± 0.0	11.63 ± 0.30	7.35 ± 0.57	2.9 ± 0.0	6.05 ± 0.23	3.67 ± 0.23
5.0mm						
Observer 1 (JM)	3.6 ± 0.0	12.39 ± 0.31	11.58 ± 0.46	3.0 ± 0.1	7.57 ± 0.45	6.68 ± 0.23
Observer 2 (LW)	3.4 ± 0.1	11.29 ± 0.28	7.36 ± 0.20	2.9 ± 0.0	6.21 ± 0.26	3.40 ± 0.25

Table 2-10: Mean 1D, 2D, and 3D Measurements of Geometrical-Shaped Tumours
Mean RECIST, WHO, and volumetric measurements for Tumours 1 and 2 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by each observer. Values are expressed as mean ± SD of five repeated measurements.

2.3.8 Reproducibility of RECIST, WHO, and Volumetric Measurements of Geometrically-Shaped Tumour Phantoms

Five repeated measurements performed by the observers allowed for the evaluation of measurement reproducibility. The coefficients of variation (CV) for each tumour measurement are reported in Table 2-11. This table shows the CV of each observer at each slice thickness.

	Tumour 1			Tumour 2		
	RECIST (%)	WHO (%)	VOLUME (%)	RECIST (%)	WHO (%)	VOLUME (%)
0.5mm						
Observer 1 (JM)	0.5	1.58	4.21	0.0	10.55	4.98
Observer 2 (LW)	1.6	2.46	3.13	0.0	7.16	5.11
1.0mm						
Observer 1 (JM)	0.0	1.58	9.22	0.0	8.09	6.59
Observer 2 (LW)	0.0	2.46	5.33	0.0	7.16	5.29
2.0mm						
Observer 1 (JM)	1.2	2.54	2.66	1.9	11.84	5.16
Observer 2 (LW)	0.0	2.68	7.72	0.0	3.73	6.25
5.0mm						
Observer 1 (JM)	1.2	2.54	3.99	3.0	5.96	3.39
Observer 2 (LW)	0.0	2.47	2.76	0.0	4.18	7.37

Table 2-11: Mean CVs of 1D, 2D, and 3D Measurements of Geometrical-Shaped Tumours
Mean coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumours 1, 2, and 3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by each observer. Values are expressed as percents.

Intra- and inter-observer reproducibility for two observers are reported in Table 2-12. ICC values for both observers show high intra- and inter-observer reproducibility at each slice thickness.

	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC(A)	ICC(C)	ICC(A)	ICC(C)	ICC(A)	ICC(C)
0.5mm						
Observer 1 (JM)	1.000	1.000	0.995	0.995	0.996	0.996
Observer 2 (LW)	0.999	0.999	0.999	0.999	0.997	0.996
All Observers	**0.996**	**0.999**	**0.987**	**0.992**	**0.976**	**0.986**
1.0mm						
Observer 1 (JM)	0.998	0.998	0.996	0.997	0.984	0.988
Observer 2 (LW)	0.998	0.999	0.997	0.998	0.994	0.997
All Observers	**0.995**	**0.998**	**0.988**	**0.993**	**0.963**	**0.978**
2.0mm						
Observer 1 (JM)	0.999	0.999	0.992	0.992	0.996	0.997
Observer 2 (LW)	0.998	0.999	0.998	0.998	0.988	0.991
All Observers	**0.996**	**0.999**	**0.980**	**0.986**	**0.953**	**0.979**
5.0mm						
Observer 1 (JM)	0.998	0.998	0.998	0.998	0.995	0.996
Observer 2 (LW)	0.998	0.999	0.998	0.999	0.997	0.997
All Observers	**0.977**	**0.984**	**0.989**	**0.995**	**0.849**	**0.955**

Table 2-12: Correlation Coefficients for Geometrical-Shaped Tumour Phantoms
The ICC(A) and ICC(C) values are reported in this table. The ICC values corresponding to individual observers reports the intra-observer reproducibility, and the inter-observer reproducibility is also reported.

2.3.9 Accuracy of RECIST, WHO, and Volumetric Measurements of Geometrically-Shaped Tumour Phantoms

Ground truth measurements used for significance testing to observer measurements are shown in Table 2-13. The mean measurements of all observers were compared to the phantom measurements in this table to determine observer accuracy.

	RECIST (cm)	WHO (cm^2)	Volume (cm^3)
Tumour 1	4.3	18.06	12.18
Tumour 2	3.0	9.00	5.76
Tumour 3	2.3	1.50	1.70

Table 2-13: Phantom Ground Truth Measurements for Geometrical-Shaped Tumours
RECIST and 3D ground truth measurements are reported. These measurements were used as ground truth when compared to the multiple observers for accuracy testing.

The mean ± SD of five repeated 3D measurements are shown in Figure 2-10. A one-way ANOVA was performed to determine any significant differences between observer measurements and ground truth. For each geometrically-shaped tumour phantom, the volumetric measurements were significantly different from the known phantom dimensions. The measurements for Observer 1 were significantly less than ground truth for Tumours 1-3 at 0.5mm, 1.0mm, and 2.0mm slice thicknesses. At 5.0mm slice thickness, the measurements for Tumour 1 were significantly less than ground truth, and for Tumours 2 and 3, 3D measurements were significantly greater. The 3D measurements of Tumours 1-3 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm were consistently significantly less than ground truth for Observer 2. Neither observer showed accurate 3D measurements of the geometrically-shaped tumour phantoms.

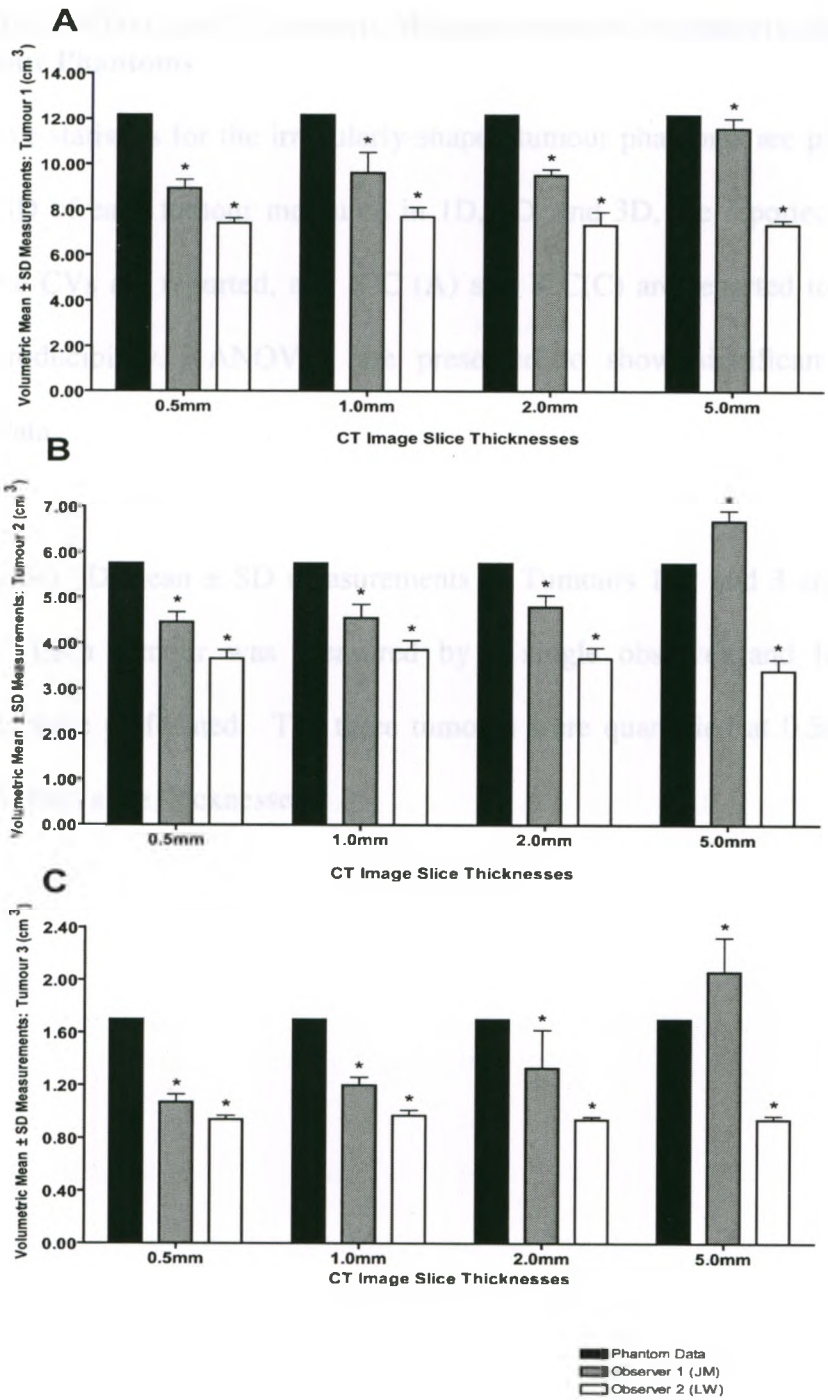


Figure 2-10: Statistical Differences Between Phantom and Observer Volumetric Measurements of Geometrical-Shaped Tumours

(A) Observer 1 at 0.5mm, 1.0mm, and 2.0mm slice thicknesses, $p<0.0001$, and 5.0mm slice thickness, $p=0.018$; Observer 2 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses, $p<0.0001$.

(B) Observer 1 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses, $p<0.0001$; Observer 2 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses, $p<0.0001$.

(C) Observer 1 at 0.5mm and 1.0mm slice thicknesses, $p<0.0001$, 2.0mm slice thickness, $p=0.012$, and 5.0mm slice thickness, $p=0.006$; Observer 2 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses, $p<0.0001$.

2.3.10 RECIST, WHO, and Volumetric Measurements of Irregularly-Shaped Tumour Phantoms

The descriptive statistics for the irregularly-shaped tumour phantoms are presented, and the mean ± SD of each tumour measured in 1D, 2D, and 3D, are reported for a single observer. The CVs are reported, and ICC (A) and ICC(C) are reported to show intra-observer reproducibility. ANOVAs are presented to show significant differences between the data.

The 1D, 2D, and 3D mean ± SD measurements of Tumours 1, 2 and 3 are reported in Table 2-14. Each tumour was measured by a single observer and five repeated measurements were performed. The three tumours were quantified at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses.

Tumour #	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm			
1	2.6 ± 0.1	4.10 ± 0.09	2.68 ± 0.12
2	4.1 ± 0.0	11.23 ± 0.47	14.44 ± 0.65
3	1.3 ± 0.0	1.66 ± 0.06	1.09 ± 0.06
1.0mm			
1	2.5 ± 0.0	4.00 ± 0.00	2.61 ± 0.06
2	4.1 ± 0.0	11.54 ± 0.32	14.69 ± 0.54
3	1.3 ± 0.0	1.66 ± 0.06	1.08 ± 0.05
2.0mm			
1	2.6 ± 0.0	3.97 ± 0.18	2.48 ± 0.16
2	4.2 ± 0.1	11.57 ± 0.42	13.71 ± 0.41
3	1.3 ± 0.0	1.66 ± 0.06	1.06 ± 0.02
5.0mm			
1	2.5 ± 0.0	3.95 ± 0.11	2.99 ± 0.19
2	3.6 ± 0.2	8.64 ± 0.77	15.58 ± 0.86
3	1.3 ± 0.0	1.54 ± 0.05	1.27 ± 0.11

Table 2-14: Mean 1D, 2D, and 3D Measurements of Irregular-Shaped Tumours
Mean RECIST, WHO, and volumetric measurements for Tumours 1–3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by a single observer. Values are expressed as mean ± SD of five repeated measurements.

2.3.11 Reproducibility of RECIST, WHO, and Volumetric Measurements of Irregularly- Shaped Tumour Phantoms

Coefficients of variation are shown in Table 2-15. The CVs were calculated as the quotient of the standard deviation and mean of the five repeated measurements. Based on the values in Table 2-15, 3D measurements had higher CVs than 1D and 2D measurements.

Tumour #	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm			
1	2.1	2.14	4.33
2	0.0	4.16	4.48
3	0.0	3.49	5.81
1.0mm			
1	0.0	0.00	2.14
2	1.1	2.74	3.70
3	0.0	3.49	4.34
2.0mm			
1	1.7	4.54	6.31
2	1.3	3.67	2.98
3	0.0	3.49	1.81
5.0mm			
1	0.0	2.83	6.29
2	6.2	8.92	5.52
3	3.5	3.49	8.91

Table 2-15: Mean CVs of 1D, 2D, and 3D Measurements of Irregular-Shaped Tumours
Mean coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumours 1–3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by each observer. Values are expressed as percents.

2.3.12 Accuracy of RECIST, WHO, and Volumetric Measurements of Irregularly-Shaped Tumour Phantoms

Table 2-16 shows the 1D, 2D, and 3D ground truth measurements for the irregularly-shaped tumours. An ANOVA was used to determine the significant differences of observer measurements. The volume rendering of the tumour phantoms after manual segmentation is shown in Figure 2-11

	RECIST (cm)	WHO (cm^2)	Volume (cm^3)
Tumour 1	3.2 ± 0.0	7.36 ± 0.0	2.93 ± 0.00
Tumour 2	5.2 ± 0.0	22.15 ± 0.0	16.96 ± 0.00
Tumour 3	2.1 ± 0.0	3.15 ± 0.0	1.39 ± 0.00

Table 2-16: Phantom Ground Truth Measurements for Irregular-Shaped Tumours
RECIST, WHO, and 3D ground truth measurements are reported. These measurements were used as ground truth when compared to the multiple observers for accuracy testing.

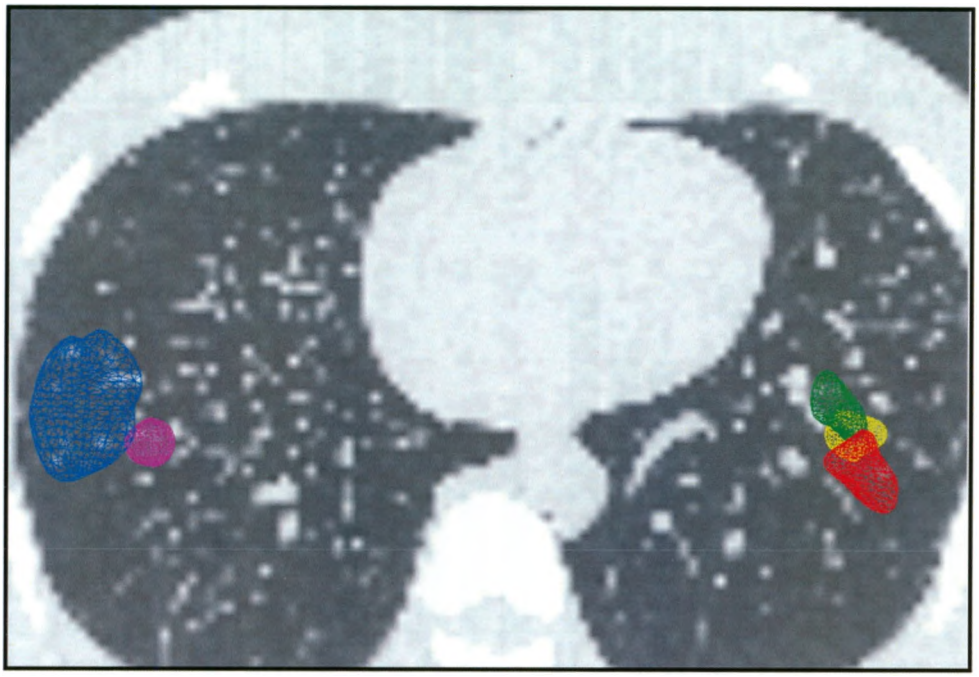


Figure 2-11: Volume Rendering of Irregularly-Shaped Tumour Phantoms
The volume rendering of the irregularly-shaped tumour phantoms after manual segmentation are shown. The large phantom is shown in blue ($V=16.96cm^3$), the medium phantom is shown in red ($V=2.93\ cm^3$), and the small tumour is shown in green ($V=1.39\ cm^3$). The purple and yellow volumes are tumour phantoms made of water and paste, which did not have ground truth values.

The results of an ANOVA to test observer measurement accuracy are shown in Figure 2-12. RECIST, WHO, and 3D measurements are shown for Tumours 1 and 3 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses. The results for Tumour 2 are in Appendix C. There was no significant difference between the 3D measurements at 5.0mm slice thickness for the large tumour (Tumour 1), and the small tumour (Tumour 3).

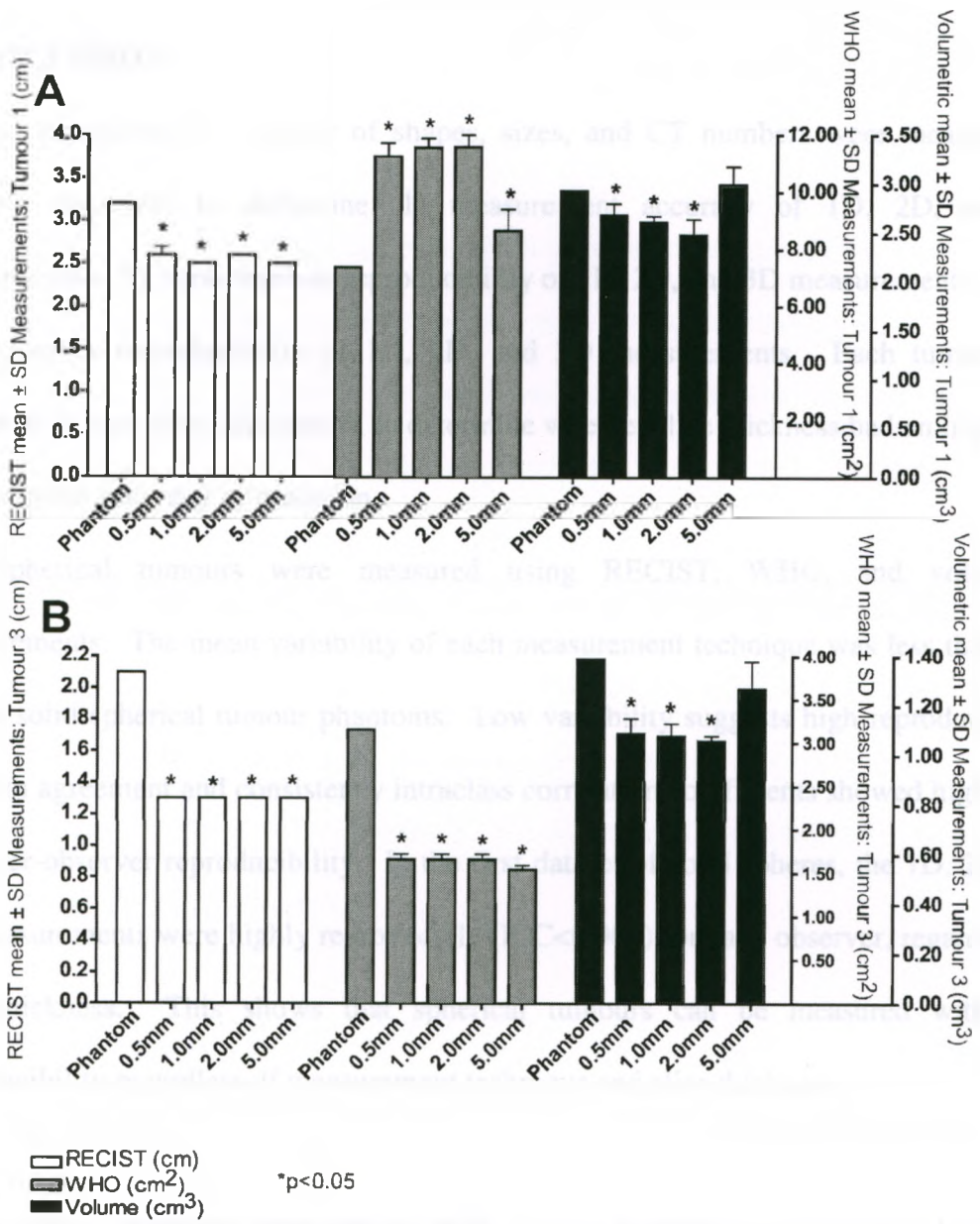


Figure 2-12: Statistical Differences between a Single Observer and Ground Truth for Tumours 1 and 3 (Irregularly-Shaped Tumour Phantoms)

(A) RECIST at 0.5mm (2.6 ± 0.1cm, p<0.0001), at 1.0mm (2.5 ± 0.0cm, p<0.0001), at 2.0mm (2.6 ± 0.0cm, p<0.0001), and at 5.0mm slice thicknesses (2.5 ± 0.0, p<0.001); WHO at 0.5mm (4.10 ± 0.09cm^2, p<0.0001), at 1.0mm (4.00 ± 0.00cm^2, p<0.0001), at 2.0mm (3.97 ± 0.18cm^2, p<0.0001), and at 5.0mm slice thicknesses (3.95 ± 0.11cm^2, p<0.0001); 3D at 0.5mm (2.68 ± 0.12cm^3, p=0.031), at 1.0mm (2.61 ± 0.06cm^3, p=0.011), at and 2.0mm slice thicknesses (2.48 ± 0.16cm^3, p<0.0001).

(B) RECIST at 0.5mm (1.3 ± 0.0cm, p<0.0001), at1.0mm (1.3 ± 0.0cm, p<0.0001), at 2.0mm (1.3 ± 0.0cm, p<0.0001), and at 5.0mm slice thicknesses (1.3 ± 0.0cm, p<0.0001); WHO at 0.5mm (1.66 ± 0.06cm^2, p<0.0001), at 1.0mm (1.66 ± 0.06cm^2, p<0.0001), at 2.0mm (1.66 ± 0.06cm^2, p<0.0001), and at 5.0mm slice thicknesses (1.54 ± 0.05cm^2, p<0.0001); 3D at 0.5mm (1.09 ± 0.06cm^3, p<0.0001), at 1.0mm (1.08 ± 0.05cm^3, p<0.0001), at and 2.0mm slice thicknesses (1.06 ± 0.02cm^3, p<0.0001).

2.4 DISCUSSION

Tumour phantoms of a variety of shapes, sizes, and CT numbers were measured by multiple observers to determine: 1) measurement accuracy of 1D, 2D, and 3D measurements; 2) intra-observer reproducibility of 1D, 2D, and 3D measurements; and 3) inter-observer reproducibility of 1D, 2D, and 3D measurements. Each tumour was evaluated at four slice thicknesses, to determine whether slice thickness had an impact on measurement accuracy or precision.

Six spherical tumours were measured using RECIST, WHO, and volumetric measurements. The mean variability of each measurement technique was less than 10% for the solid spherical tumour phantoms. Low variability suggests high reproducibility, and both agreement and consistency intraclass correlation coefficients showed high intra- and inter-observer reproducibility. In the first dataset of solid spheres, the 1D, 2D, and 3D measurements were highly reproducible (ICC<0.900) for each observer, regardless of slice thickness. This shows that spherical tumours can be measured with high reproducibility regardless of measurement technique and slice thickness.

Three tumour phantoms were created with various geometries, consisting of spheres, cylinders, and concave cubes. The measurement variability was greater than 10% for the 2D measurements of Tumour 3C performed by one observer at 2.0mm and 5.0mm slice thicknesses. The volumetric measurements for Tumour 3 at 2.0mm and 5.0mm slice thicknesses were greater than 10% for observer 1. The inter-observer ICC(A) value for 3D measurements was 0.849, and the remaining ICC(A) and ICC(C) values were greater

than 0.900. This shows that 1D, 2D, and 3D measurements were highly reproducible for tumour phantoms of shapes that are unrealistically found in any patient population.

The irregularly-shaped tumour phantoms which were constructed with the dimensions of three primary tumours in another database were evaluated using 1D, 2D, and 3D measurements. All measurements were performed by a single observer, and the CVs for each tumour were less than 10%. When testing for measurement accuracy, RECIST consistently underestimated tumour size, regardless of tumour and slice thickness. The WHO measurements were consistently overestimating tumour size for Tumour 1 at all slice thicknesses, and were consistently underestimating tumour size in Tumours 2 and 3. Volumetric measurements showed the ability to accurately measure tumour size for Tumours 1 and 3 at 5.0mm slice thickness. For all other 3D measurements, tumour size was significantly underestimated.

A limitation of this study was, for the irregular tumour phantoms, the ground truth longest and perpendicular axes were orientation-specific: the significant underestimation for RECIST measurements may be related to the positioning of tumours within the chest phantom. The longest axis displayed on the CT slice may not have been the same longest axis measured on the tumour.

2.5 CONCLUSION

RECIST, WHO, and 3D measurements were reproducible when evaluating tumours of various shapes and sizes. Volumetric measurements can provide measurement accuracy in tumours when RECIST and WHO show consistent measurement underestimation.

2.6 REFERENCES

1. Canadian Cancer Society's Steering Committee. Canadian Cancer Statistics 2009. 4-2-0009. Toronto, Canadian Cancer Society.
   Ref Type: Report

2. Statistics Canada. Leading Causes of Death in Canada. 84-215-X. 2005.
   Ref Type: Report

3. Petrou M, Quint LE, Nan B, Baker LH. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am.J.Roentgenol.* 2007;188:306-12.

4. Chojniak R, Yu LS, Younes RN. Response to chemotherapy in patients with lung metastases: how many nodules should be measured? *Cancer Imaging* 2006;6:107-12.

5. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RT, Reznek RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br.J.Radiol.* 2000;73:1178-84.

6. Wormanns D, Diederich S, Lentschig MG, Winter F, Heindel W. Spiral CT of pulmonary nodules: interobserver variation in assessment of lesion size. *Eur.Radiol.* 2000;10:710-3.

7. Schwartz LH, Colville JA, Ginsberg MS, Wang L, Mazumdar M, Kalaigian J et al. Measuring tumor response and shape change on CT: esophageal cancer as a paradigm. *Ann.Oncol.* 2006;17:1018-23.

8. Sun JM, Ahn MJ, Park MJ, Yi JH, Kim TS, Chung MJ et al. Accuracy of RECIST 1.1 for non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors. *Lung Cancer* 2009.

9. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J.Clin.Oncol.* 2003;21:2574-82.

10. Forner A, Ayuso C, Varela M, Rimola J, Hessheimer AJ, de Lope CR et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? *Cancer* 2009;115:616-23.

11. World Health Organization. WHO Handbook for Reporting Results for Cancer Treatment. 1979. Geneva, World Health Organization.
    Ref Type: Report

12. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J.Natl.Cancer Inst.* 2000;92:205-16.

13. Park JO, Lee SI, Song SY, Kim K, Kim WS, Jung CW et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn.J.Clin.Oncol.* 2003;33:533-7.

14. Watanabe H, Kunitoh H, Yamamoto S, Kawasaki S, Inoue A, Hotta K et al. Effect of the introduction of minimum lesion size on interobserver reproducibility using RECIST guidelines in non-small cell lung cancer patients. *Cancer Sci.* 2006;97:214-8.

15. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur.J.Cancer* 2009;45:228-47.

16. Darkeh MH, Suzuki C, Torkzad MR. The minimum number of target lesions that need to be measured to be representative of the total number of target lesions (according to RECIST). *Br.J.Radiol.* 2009;82:681-6.

17. Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. *AJR Am.J.Roentgenol.* 2006;186:989-94.

18. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology* 2000;217:251-6.

19. Mantatzis M, Kakolyris S, Amarantidis K, Karayiannakis A, Prassopoulos P. Treatment response classification of liver metastatic disease evaluated on imaging. Are RECIST unidimensional measurements accurate? *Eur.Radiol.* 2009;19:1809-16.

20. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR Am.J.Roentgenol.* 1996;167:851-4.

21. Landry A, Spence JD, Fenster A. Quantification of carotid plaque volume measurements using 3D ultrasound imaging. *Ultrasound Med.Biol.* 2005;31:751-62.

22. Armstrong GD. The intraclass correlation as a measure of interrater reliability of subjective judgments. *Nurs.Res.* 1981;30:314-5, 320A.

23. Laschinger HK. Intraclass correlations as estimates of interrater reliability in nursing research. *West J.Nurs.Res.* 1992;14:246-51.

Chapter 3: Multi-Dimensional X-Ray Computed Tomography Measurements of Pulmonary Metastases in Patients

3.1 Introduction

Objective tumour response is a common endpoint used in the development of potential anticancer agents, and has played a role in the development of drugs approved for use in cancer treatment (1-5). Tumour response is also essential for the evaluation of the results of chemotherapy or radiotherapy in individual patients (6-10). Conventional response criteria are based on x-ray computed tomography (CT) images (11-15), and the change in tumour size is determined on sequential radiological evaluations (15-17). Because the results can affect the treatment plan of a patient, or influence the outcome of an investigational drug evaluation, accurate and reproducible measurements of tumours are of paramount importance (16).

In 1979, the World Health Organization (WHO) recommended a standardized approach of quantifying tumour size, and classifying the clinical implications of the changes in tumour size (18). The WHO measurement was defined as the product of the longest axis of the tumour and its longest perpendicular bisector, measured from a CT or x-ray image. Since its application 20 years ago, many limitations of the WHO measurements were discovered; neither the minimum lesion size (12;19), nor the number of target lesions per subject to be counted (12;19;20) were stated in the original document, and this resulted in many variations in the measurement among investigators. More limitations included inconsistency in the definition of progressive disease (1;19), and the risk of error inherent in measuring in two dimensions and calculating their sums (1;20). Recent advances in

image acquisition strategies, including those related to positron emission tomography (PET), single photon emission computed tomography (SPECT) and CT, called for an improvement in the measurement paradigm (19;20).

To provide a solution to the many limitations of the WHO measurement, the Response Evaluation Criteria in Solid Tumours (RECIST) was first implemented in 2000. RECIST measurements were defined as 1) the unidimensional (1D) measurement of the tumour's longest axis in the plane of image acquisition, 2) the sum of 1D measurements for 10 target lesions, and 3) the minimum size to be measured was twice the slice thickness of image acquisition. RECIST guideline specified the minimum tumour size and number of target lesions (12;21), and accounted for subject positioning (21), differences in scan thickness (21), and difficulty in estimating size of irregular or confluent lesions (21). In 2009, RECIST 1.1 (16) was introduced with the modifications of a 1) reduced number of target lesions from 10 to five, and, 2) revised definition of progressive disease (which includes an absolute 5mm increase in the longest diameter in addition to the original 20% increase in longest diameter); however, the foundation of this paradigm remained quantifying tumour size with a 1D measurement.

As imaging technologies continue to evolve and improve, it is of great importance that image analysis tools must improve in their sophistication, yet remain reproducible, such that the information provided by the medical images may be exploited. To that end, a number of investigators have initiated the development of three-dimensional (3D) tumour measurements. Sohaib et al. (7) confirmed that volume measurements determined from

helical CT data with appropriate image processing software are accurate and reproducible. Three-dimensional measurements may provide a more accurate representation of tumour size (1), with consequent implications for assessment of response (16). The established 1D and 2D measurements evaluate tumour size under the assumptions of tumour sphericity (12;18), and that the longest axis of the tumour is located in the axial plane (15).

In an extension of these previous attempts, we aim to: 1) determine the inter-observer and intra-observer reproducibility of pulmonary metastases in 1D, 2D, and 3D; 2) determine the intra- and inter-observer reproducibility of trained and expert observers; 3) determine the interscan precision of 3D measurements for multiple timepoints.

3.2 METHODS

3.2.1 Overview

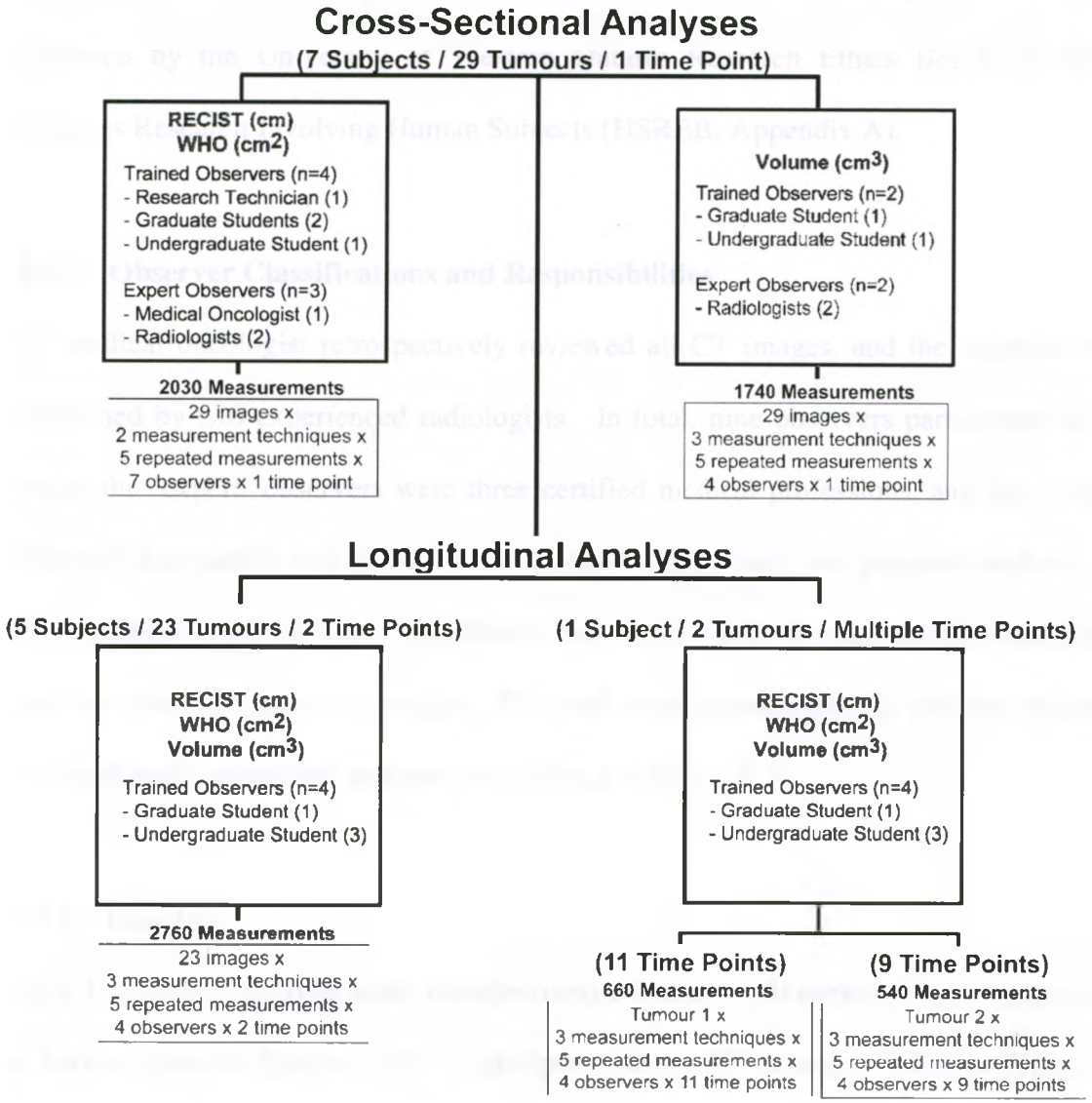
A detailed overview of the data collected per analysis is provided in Table 3-1. The total number of measurements per analysis is provided in Figure 3-1.

	Subjects (n)	Tumours (n)	Observers (n)			Observer Measurements (n)		
			1D	2D	3D	1D	2D	3D
Cross-Sectional Analysis	7	29	7	7	4	145	145	145
Longitudinal Analysis: 2 Time Points	5	23	4	4	4	230	230	230
Longitudinal Analysis: Multiple Time Points	1	2	4	4	4	100	100	100

1D = RECIST (cm); 2D = WHO (cm^2); 3D = Volume (cm^3)

Table 3-1: Overview of Data Collected per Analysis
The number of subjects, tumours, observers, and total measurements per observer are provided for each analysis and categorized by measurement dimension.

Figure 3-1: Flow Chart of Observer Classifications and Responsibilities
All observers are listed under the specific areas of the study in which they were involved. The number of subjects, tumours, time points and total tumour measurements per analysis are also listed.



3.2.2 Study Subjects

Written informed consent was not required from the study subjects, because this was a retrospective study of pre-existing, anonymized images. The study protocol was approved by the University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB; Appendix A).

3.2.3 Observer Classifications and Responsibilities

On medical oncologist retrospectively reviewed all CT images, and the locations were confirmed by two experienced radiologists. In total, nine observers participated in this study: the 'expert' observers were three certified medical professions, and the 'trained' observer designation was given to one research technician, two graduate students, and three undergraduate students. Specifically, one expert observer was a medical oncologist, and two observers were radiologists. The total measurements used to perform the cross-sectional and longitudinal analyses are outlined in Figure 3-1.

3.2.4 Imaging

All CT protocols and data were retrospectively collected. All subjects were imaged using a helical General Electric (GE) LightSpeed Series CT scanner (VCT or Ultra; GE HealthCare, Waukesha, WI). Three CT image data were reconstructed with 2.5mm slice thickness, and the remaining scans were reconstructed with 5.0mm slice thickness. In three scans, contrast agents were administered to subjects either orally, intravenously, or both. Images were displayed on LCD monitors using a Picture Archiving and Communications System (PACS; ClearCanvas, Inc., Toronto, ON).

3.2.5 Image Analysis

Image analysis was performed blinded to subject identity, clinical status and timepoint. Lung windows were used to measure lung lesions (22). Image analysis using the RECIST and WHO criteria was performed using electronic calipers in an open-source PACS system, ClearCanvas (ClearCanvas, Inc., Toronto, Canada).

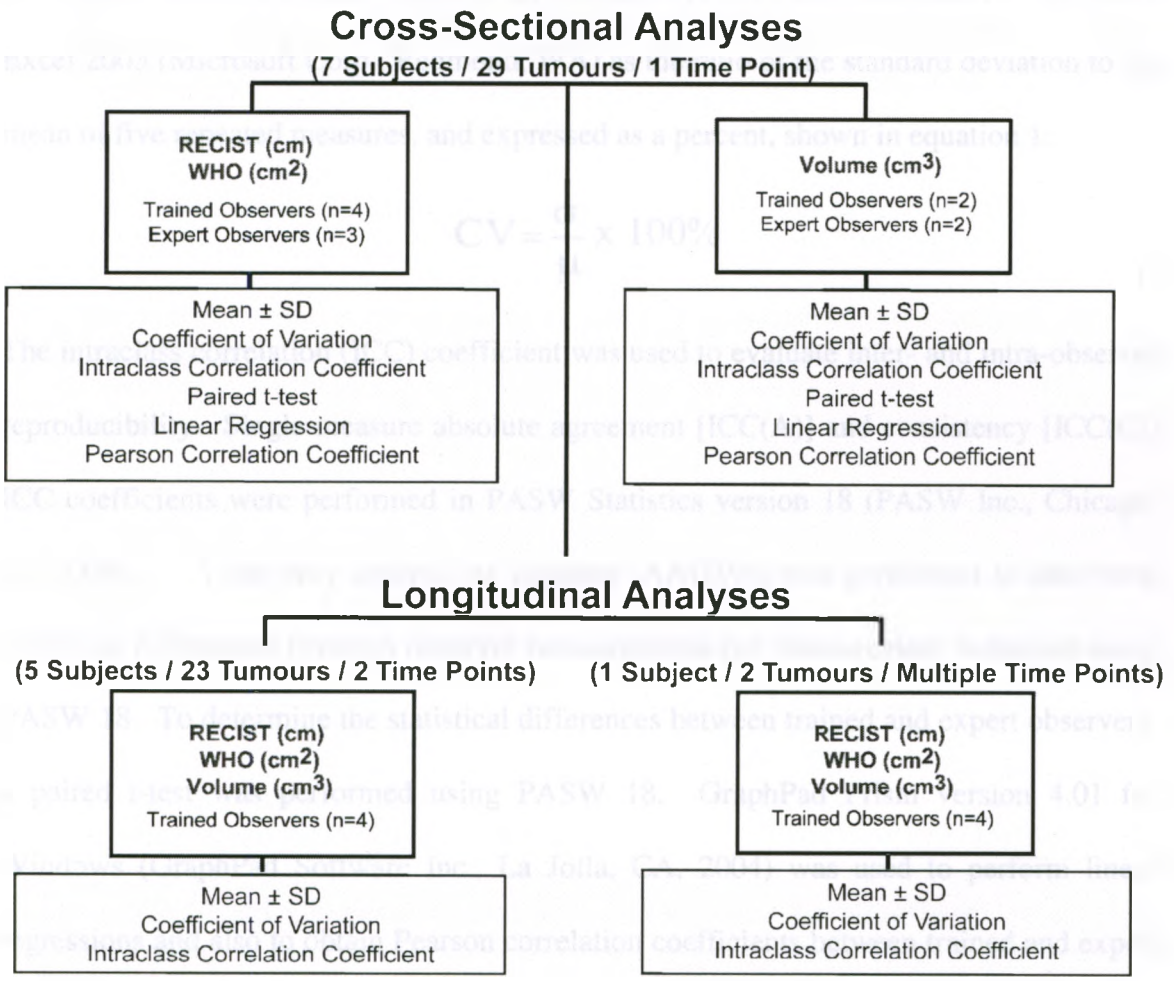
Volumetric analysis was performed using a customized visualization and segmentation software developed in-house, 3D Quantify (Robarts Research Institute, London, ON), as previously described (23). Lung tumour volumes were calculated from the manual segmentation of tumour boundaries using VTK (Visualization Toolkit; Kitware, Inc, Clifton Park, NY). This technique was previously developed for the quantification of 3DUS prostate volumes (23-25). After establishing user-defined axis of rotation in the centre of the tumour, the tumour was rotated about this axis at an angle of 18o. Ten tumour boundaries were contoured for each tumour, and each contour set was converted to a 3D mesh, where the radial distance from the center of mass was calculated at each angle (24). RECIST, WHO, and volumetric image segmentation were repeated five times and the mean values (x), standard deviation (SD), and coefficient of variation (CV) were recorded. To calculate the tumour burden for each subject, the RECIST, WHO and tumour volumes were summed per subject.

For the RECIST, WHO, and volumetric measurements, the longest axis, the cross-product of longest axis and perpendicular bisector, and volumes were respectively summed to obtain the total tumour burden per subject.

3.2.6 Statistical Analysis

An overview of the statistical analysis in this chapter is provided in Figure 3-2. The statistical tests used for each section are listed in the figure, and are further explained in this section.

Figure 3-2: Flow Chart of Statistical Analyses per Study Component
All study components and respective statistical analyses are listed.



3.2.6.1 Cross-Sectional Analyses

Mean RECIST, WHO, and volumes, and standard deviations were calculated from five repeated measures for each individual lung tumour. Total tumour burden was also calculated as the sum of RECIST, WHO, and volume measurements. Total tumour burden for volumetric measurements was determined as the sum of measured tumour volumes per subject. Coefficients of variation (CV) were used to assess the variability among five repeated measurements of all tumours. All CVs were calculated in Microsoft Excel 2003 (Microsoft Corp., Redmond, WA) as the ratio of the standard deviation to the mean of five repeated measures, and expressed as a percent, shown in equation 1:

$$CV = \frac{\sigma}{\mu} \text{ x } 100\% \tag{1}$$

The intraclass correlation (ICC) coefficient was used to evaluate inter- and intra-observer reproducibility. Single measure absolute agreement [ICC(A)] and consistency [ICC(C)] ICC coefficients were performed in PASW Statistics version 18 (PASW Inc., Chicago, IL, 2009). A one-way analysis of variance (ANOVA) was performed to determine statistical differences between observer measurements per measurement technique using PASW 18. To determine the statistical differences between trained and expert observers, a paired t-test was performed using PASW 18. GraphPad Prism version 4.01 for Windows (GraphPad Software Inc., La Jolla, CA, 2004) was used to perform linear regressions and also to obtain Pearson correlation coefficients between trained and expert observers for each measurement technique. The results of all statistical analyses were considered significant when the probability of making a Type I error was less than 5% ($p < 0.05$).

3.2.6.2 Longitudinal Analyses

Mean tumour size and total tumour burden using RECIST, WHO, and volumetric measurements were calculated and represented as mean ± standard deviation of five repeated measurements at each time point.

Subjects were sorted into the tumour response classifications outlined in RECIST 1.1, and we focused on three of the four classifications: progressive disease (PD), partial response (PR), and stable disease (SD). Absolute change (%) in tumour size was calculated as $[(V_{fu}- V_{bl})/V_{bl}]*100\%$ (7;13). The subjects were then sorted into tumour response classifications according to the volume categorization derived from RECIST.

3.3 RESULTS

3.3.1 Study Subjects

Subject demographics are provided in Table 3-2. The database is a compilation of CT images acquired within a three-year period, from January 2005 to December 2007.

	Scan (n=7)	Rescan (n=5)
Mean Age, yrs (±SD)[Range]	58 (8)[50-74]	61 (10)[50-75]
Male sex	6	4
Total number of lung tumours	29	23
Time between scans, weeks (±SD)[Range]	--	23 (16)[7-43]

Table 3-2: Subject Demographics and Scan and Rescan Characteristics
The scan demographics for seven subjects are summarized. The subjects' age at scan and rescan, as well the time between scans are expressed as a mean ± standard deviation, and includes the range.

Seven study subjects were retrospectively selected: six subjects with a clinical diagnosis of renal cell carcinoma and lung metastases, and one subject had an unknown primary

cancer diagnosis with lung metastases. Of the seven subjects scanned at scan, one subject could not be contacted, one subject was deceased, and therefore, five subjects returned for a rescan visit.

Subject 4 was a 54 year old male, who received a total of 13 thoracic CT scans between January 2005 and December 2007. This subject had two posterior subpleural metastases, both located in the lower lobe of the right lung. Tumour 1 was evaluated at 11 of the 13 timepoints, and tumour 2 was evaluated at 9 timepoints. At the timepoints when the tumours were measurable, four observers measured both tumours using RECIST, WHO, and 3D measurements to retrospectively observe the response and progression of the tumour burden longitudinally, over two years.

3.3.2 Cross-Sectional Analysis

3.3.2.1 RECIST and WHO Measurements by Trained and Expert Observers

Total tumour burden was calculated for each of the seven subjects. To calculate the RECIST tumour burden, the longest dimension of each tumour deposit for each subject was summed. To calculate the WHO tumour burden, the product of the longest dimension and longest perpendicular bisector of each tumour deposit were summed. The multiplicity of lung metastases per subject ranges from 1-11, and the distribution of tumour deposits is reported in Table 3-3.

Subject #	Pulmonary Metastases (n)	Tumour Numbers
1	1	1
2	2	2-3
3	5	4-8
4	2	9-10
5	3	11-13
6	11	14-24
7	5	25-29
TOTAL	**29**	

Table 3-3: Distribution of Tumours among Subjects
The number of tumour deposits per subject is reported. Tumour burden is calculated by the sum of RECIST and WHO measurements of all tumour deposits per patient. The number of tumours used to calculated tumour burden per subject is reported.

The tumour burden per subject as calculated by trained and expert observers is reported in Table 3-4. The mean values of RECIST and WHO measurements for both trained and expert observers suggests that the measurements of the trained observers are consistently less than those of the expert observers.

Subject	RECIST (cm)			WHO (cm^2)		
	Trained Observers	Expert Observers	All Observers	Trained Observers	Expert Observers	All Observers
1	1.6 ± 0.1	3.9 ± 3.9	2.6 ± 2.8	2.20 ± 0.27	13.39 ± 18.96	6.99 ± 13.40
2	6.3 ± 0.3	6.6 ± 0.3	6.4 ± 0.3	18.64 ± 1.58	19.65 ± 0.88	19.07 ± 1.40
3	14.2 ± 0.5	14.0 ± 0.7	14.1 ± 0.6	40.14 ± 2.49	39.47 ± 2.68	39.85 ± 2.56
4	3.6 ± 0.2	3.9 ± 0.7	3.7 ± 0.5	5.80 ± 0.69	6.51 ± 1.15	6.10 ± 0.97
5	4.1 ± 0.6	4.8 ± 0.4	4.4 ± 0.6	4.50 ± 1.07	5.79 ± 0.89	5.05 ± 1.17
6	19.7 ± 1.9	21.1 ± 1.9	20.3 ± 2.0	30.17 ± 4.83	34.06 ± 4.27	31.8 4± 4.93
7	7.3 ± 0.7	8.0 ± 0.8	7.6 ± 0.8	8.84 ± 1.45	11.09 ± 1.68	9.80 ± 1.90
Mean	**8.1 ± 0.3**	**8.9 ± 0.6**	**8.4 ± 0.5**	**15.75 ± 0.86**	**18.56 ± 2.82**	**16.96 ± 2.11**

Table 3-4: Trained and Expert Observer Means of RECIST and WHO Measurements
Total tumour burden measurements are expressed as the mean ± SD using RECIST and WHO measurements. The means of trained (n=4) and expert (n=3) observer measurements are provided for each subject.

3.3.2.2 Reproducibility of RECIST and WHO Measurements by Trained and Expert Observers

Table 3-5 shows the coefficients of variation total tumour burden per subject as calculated by trained and expert observers. The CVs reported in this table were calculated from the mean ± SD values of RECIST and WHO measurements reported in Table 3-4. The data in this table suggests that expert observers have higher variability for both RECIST and WHO measurements.

	RECIST (cm)			WHO (cm^2)		
Subject	Trained Observers	Expert Observers	All Observers	Trained Observers	Expert Observers	All Observers
1	5.4	100.3	107.7	141.6	12.46	191.67
2	4.5	4.5	5.0	4.5	8.45	7.35
3	3.4	4.9	4.1	6.8	6.21	6.42
4	5.0	19.0	14.1	17.7	11.98	15.90
5	13.7	8.7	14.0	15.3	23.66	23.19
6	9.9	8.8	9.9	12.5	16.01	15.50
7	9.4	9.4	10.5	15.2	16.44	19.43
Mean	**3.9**	**7.3**	**6.2**	**5.43**	**15.20**	**12.45**

Table 3-5: Trained and Expert Observer CVs of RECIST and WHO Measurements Coefficients of variation (CV) of total tumour burden measurements are expressed for RECIST and WHO measurements. Each CV was calculated using the mean ± SD of trained (n=4) and expert (n=3) observer measurements for each subject.

To assess inter-observer and intra-observer reliability, the Intra-Class Correlation (ICC) coefficients for each observer are shown in Table 3-6. This table also shows the ICC values for trained and expert observers, as well as the ICC values for all observers combined. Trained observer ICC(A) values for RECIST range from 0.949 (Observer 3) to 0.992 (Observers 1 and 4), and ICC(C) values range from 0.964 (Observer 3) to 0.994 (Observer 4). Expert observer ICC(A) values for RECIST measurements range from

0.758 (Observer 5) to 0.991 (Observer 7), and ICC(C) values range from 0.765 (Observer 5) to 0.991 (Observer 7). Trained observer ICC(A) values for WHO measurements range from 0.982 (Observer 3) to 0.995 (Observer 1), and ICC(C) values range from 0.989 (Observer 3) to 0.995 (Observer 1). Expert observer ICC(A) values for WHO measurements range from 0.781 (Observer 5) to 0.994 (Observer 7), and ICC(C) values range from 0.785 (Observer 3) to 0.994 (Observer 7).

	RECIST (cm)		WHO (cm2)	
	ICC (A)	ICC (C)	ICC (A)	ICC (C)
Observer 1 (AW)	0.992	0.992	0.995	0.995
Observer 2 (JM)	0.989	0.990	0.992	0.992
Observer 3 (LM)	0.949	0.964	0.982	0.989
Observer 4 (LW)	0.992	0.994	0.994	0.994
Observer 5 (MM)	0.758	0.765	0.781	0.785
Observer 6 (RER)	0.978	0.982	0.989	0.991
Observer 7 (EO)	0.991	0.991	0.994	0.994
Trained Observers (n=4)	0.940	0.960	0.979	0.989
Expert Observers (n=3)	0.509	0.506	0.542	0.540
Inter-observer (n=7)	**0.645**	**0.653**	**0.675**	**0.679**

Table 3-6: Correlation Coefficients for RECIST and WHO Measurements
Intra-class correlation coefficients (ICC) for repeated measurements of RECIST and WHO measurements are reported. According to these reliability statistics, trained observers demonstrate higher reproducibility than expert observers for both RECIST and WHO measurements.

Linear regressions were performed on the mean RECIST measurements for trained and expert observers, and WHO measurements for trained and expert observers. Linear regressions were used to quantify the ability to predict trained observer measurements with the knowledge of expert observer measurements. Figure 3-3 shows the linear regression performed on the RECIST measurements by expert and trained observers.

There is a significant correlation ($r = 0.8539$; $p < 0.0001$) between trained and expert observer RECIST measurements. Figure 3-4 shows WHO measurements also performed by both categories of observers, and there is a significant correlation ($r = 0.8589$; $p < 0.001$) between trained and expert observer WHO measurements.

Figure 3-3: Linear regression of RECIST measurements
Correlation of trained and expert observer RECIST measurements for seven subjects.
$R^2 = 0.7291$; $r = 0.8539$; $y = 0.8280 + 0.1751$; $p < 0.0001$

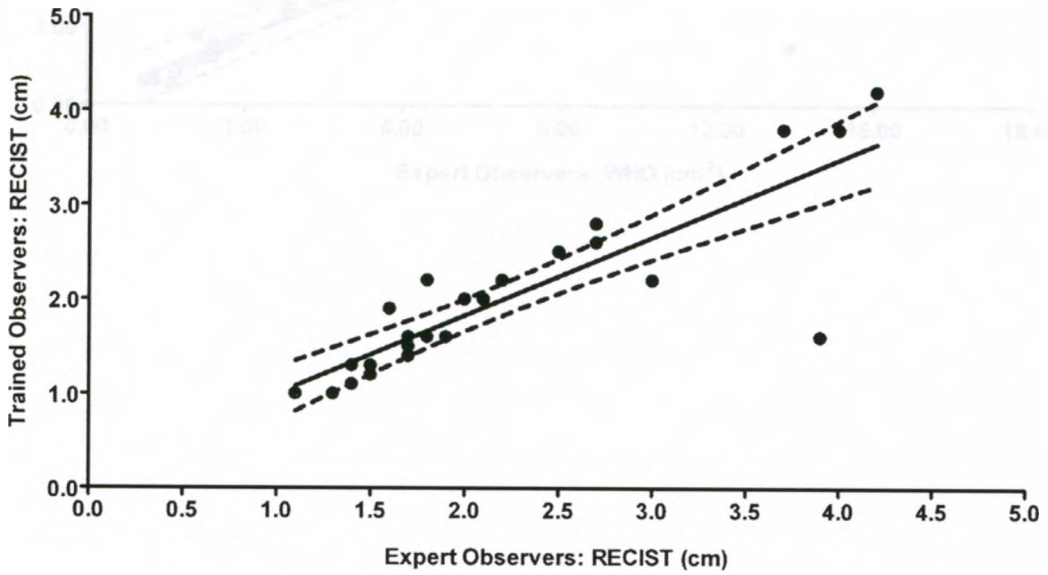
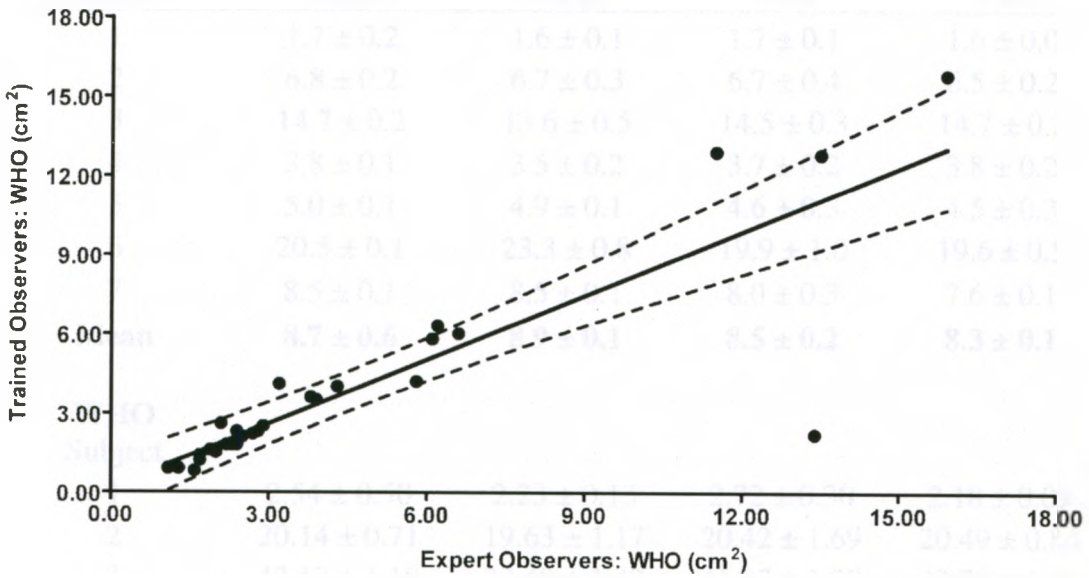


Figure 3-4: Linear regression of WHO measurements

Correlation of trained and expert observer WHO measurements for seven subjects.
$R^2 = 0.7377$; $r = 0.8589$; $y = 0.8088 + 0.1901$; $p < 0.0001$



3.3.2.3 Volumetric Measurements by Trained and Expert Observers

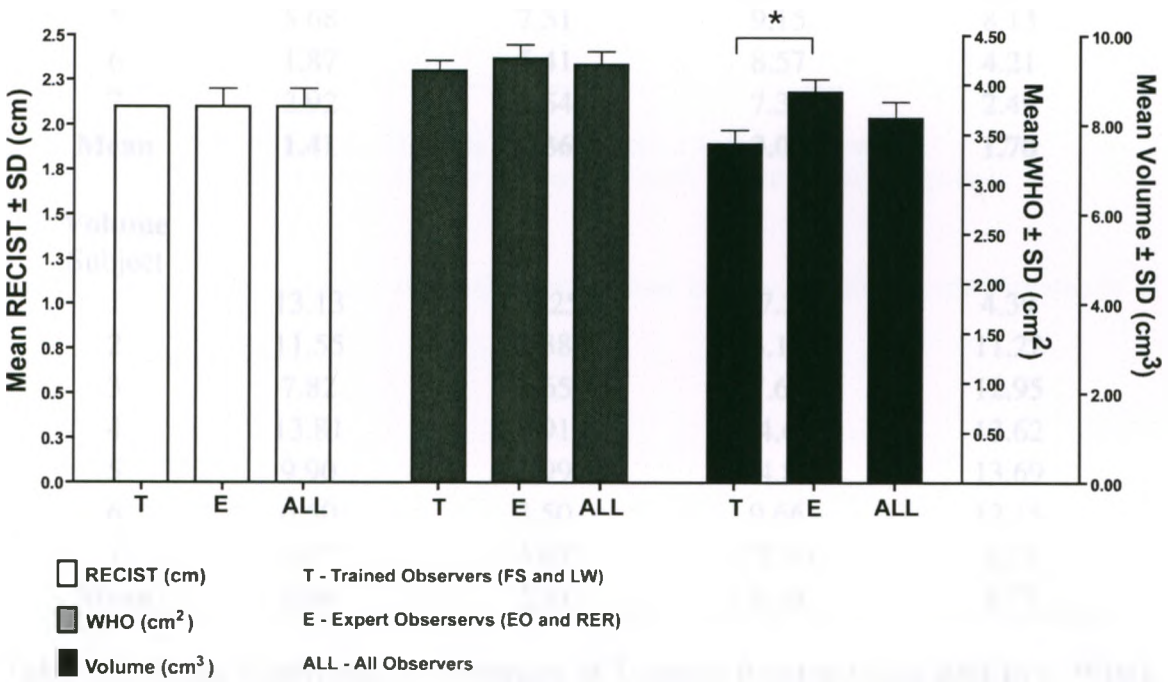
RECIST Subject	Observer 1 (EO)	Observer 2 (RER)	Observer 3 (FS)	Observer 4 (LW)
1	1.7 ± 0.2	1.6 ± 0.1	1.7 ± 0.1	1.6 ± 0.0
2	6.8 ± 0.2	6.7 ± 0.3	6.7 ± 0.4	6.5 ± 0.2
3	14.7 ± 0.2	13.6 ± 0.5	14.5 ± 0.3	14.7 ± 0.3
4	3.8 ± 0.1	3.5 ± 0.2	3.7 ± 0.2	3.8 ± 0.2
5	5.0 ± 0.1	4.9 ± 0.1	4.6 ± 0.3	4.5 ± 0.3
6	20.5 ± 0.1	23.3 ± 0.8	19.9 ± 1.0	19.6 ± 0.5
7	8.5 ± 0.1	8.5 ± 0.1	8.0 ± 0.3	7.6 ± 0.1
Mean	**8.7 ± 0.6**	**8.9 ± 0.1**	**8.5 ± 0.2**	**8.3 ± 0.1**
WHO Subject				
1	2.54 ± 0.50	2.23 ± 0.13	2.72 ± 0.30	2.18 ± 0.09
2	20.14 ± 0.71	19.63 ± 1.17	20.42 ± 1.69	20.49 ± 0.84
3	42.12 ± 1.18	37.49 ± 1.42	41.87 ± 1.28	42.70 ± 1.12
4	6.30 ± 0.40	6.05 ± 0.50	6.05 ± 0.63	6.39 ± 0.35
5	6.14 ± 0.35	5.85 ± 0.44	5.40 ± 0.49	5.54 ± 0.45
6	32.74 ± 0.61	39.21 ± 2.12	30.66 ± 2.63	30.11 ± 1.27
7	11.96 ± 0.35	12.38 ± 0.32	10.60 ± 0.77	9.81 ± 0.24
Mean	**17.42 ± 0.24**	**17.55 ± 0.41**	**16.82 ± 0.51**	**16.74 ± 0.28**
Volume Subject				
1	3.49 ± 0.46	4.40 ± 1.16	3.84 ± 0.66	3.27 ± 0.14
2	38.69 ± 4.47	22.24 ± 1.39	38.27 ± 2.35	37.60 ± 4.22
3	96.70 ± 7.56	106.90 ± 4.97	85.02 ± 6.49	79.17 ± 10.25
4	9.19 ± 1.27	9.43 ± 0.65	8.26 ± 1.21	7.03 ± 0.96
5	9.71 ± 0.96	9.19 ± 0.83	6.94 ± 1.03	5.96 ± 0.82
6	37.42 ± 2.43	46.61 ± 1.63	45.21 ± 4.37	37.91 ± 4.61
7	40.40 ± 10.82	49.27 ± 2.30	38.93 ± 8.76	39.58 ± 1.30
Mean	**33.66 ± 2.03**	**38.61 ± 0.97**	**32.35 ± 1.73**	**30.07 ± 1.73**

Table 3-7: Mean Total Tumour Burden Using RECIST, WHO, and Volumetric Measurements

Mean RECIST, WHO, and volumetric measurements for tumour measurements using RECIST and WHO criteria, and volumetric measurements are expressed as mean ± SD of five repeated measurements by all observers. Total tumour burden was calculated as the sum of all tumours per subject.

Table 3-7 shows the mean ± SD tumour burden for each subject as measured by each observer. Tumour burden for 3D measurements is calculated as the sum of volumes for all tumours per subject. Figure 3-5 shows the total mean measurement of all tumours evaluated with RECIST, WHO, and 3D measurements for trained and expert observers. A paired t-test shows that there is a significant difference (p=0.005) between the mean 3D measurements performed by trained and expert observers.

Figure 3-5: Mean Tumour Size by Trained and Expert Observers.
The mean ± SD of all tumours (n=29) were evaluated using RECIST, WHO, and 3D measurements, and the results were categorized by trained and expert observers for comparisons (p=0.005).



3.3.2.4 Reproducibility of Volumetric Measurements by Trained and Expert Observers

Table 3-8 shows coefficients of variation of tumour burden per subject as measured by each observer.

RECIST Subject	Observer 1 (EO)	Observer 2 (RER)	Observer 3 (FS)	Observer 4 (LW)
1	10.9	8.0	7.7	0.0
2	2.8	4.2	5.4	2.6
3	1.4	3.9	1.9	2.2
4	2.2	4.7	4.1	5.9
5	3.0	2.0	5.8	6.1
6	0.3	3.4	5.2	2.7
7	1.0	1.3	3.9	0.7
Mean	**0.6**	**1.7**	**2.1**	**1.3**
WHO Subject				
1	19.65	5.86	11.00	4.03
2	3.55	5.94	8.29	4.08
3	2.81	3.79	3.05	2.62
4	6.40	8.29	10.46	5.42
5	5.68	7.51	9.15	8.13
6	1.87	5.41	8.57	4.21
7	2.92	2.54	7.31	2.49
Mean	**1.41**	**2.36**	**3.03**	**1.70**
Volume Subject				
1	13.13	26.25	17.22	4.39
2	11.55	7.38	6.14	11.22
3	7.82	4.65	7.64	12.95
4	13.81	6.91	14.60	13.62
5	9.90	8.99	14.82	13.69
6	6.50	3.50	9.66	12.15
7	26.77	4.67	22.50	3.29
Mean	**6.04**	**2.51**	**5.34**	**5.77**

Table 3-8: Mean Coefficient of Variation of Tumour Burden Using RECIST, WHO, and Volumetric Measurements

Mean CV of RECIST, WHO, and volumetric measurements for tumour measurements using RECIST and WHO criteria, and volumetric measurements are expressed as mean ± SD of five repeated measurements by four observers.

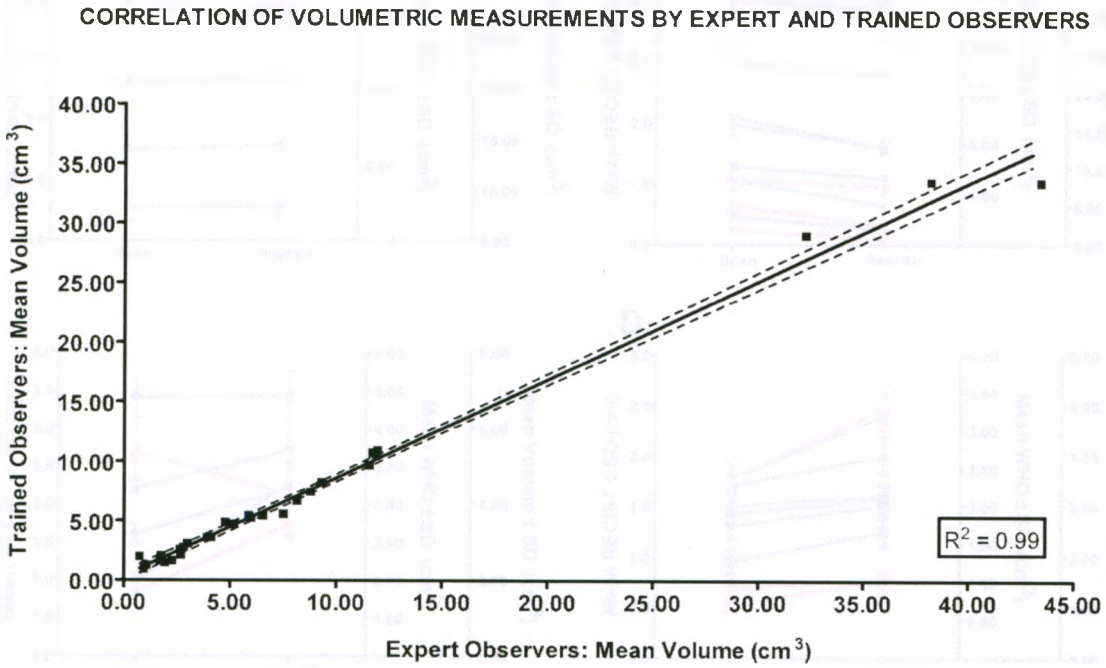
To further examine inter-observer and intra-observer reproducibility, ICC values were determined for all volumetric measurements. Table 3-9 shows the ICC values for each observer, and for trained and expert observers.

	Volume (cm^3)	
	ICC (A)	ICC (C)
Observer 1 (EO)	0.980	0.987
Observer 2 (RER)	0.992	0.993
Observer 3 (FS)	0.988	0.970
Observer 4 (LW)	0.966	0.969
Trained Observers (n=2)	0.966	0.970
Expert Observers (n=2)	0.981	0.988
All Observers (n=2)	**0.962**	**0.971**

Table 3-9: Correlation Coefficients of Volumetric Measurements
Intra-class correlation coefficients (ICC) for repeated volumetric measurements are reported.

Figure 3-6 presents a linear regression performed on the volumetric measurements of expert and trained observers.

Figure 3-6: Linear Regression of volumetric measurements
Correlation of volume measurements by expert and trained observers.
$R^2 = 0.9914$; $r = 0.9957$; $y = 0.8269 + 0.3439$; $p < 0.0001$

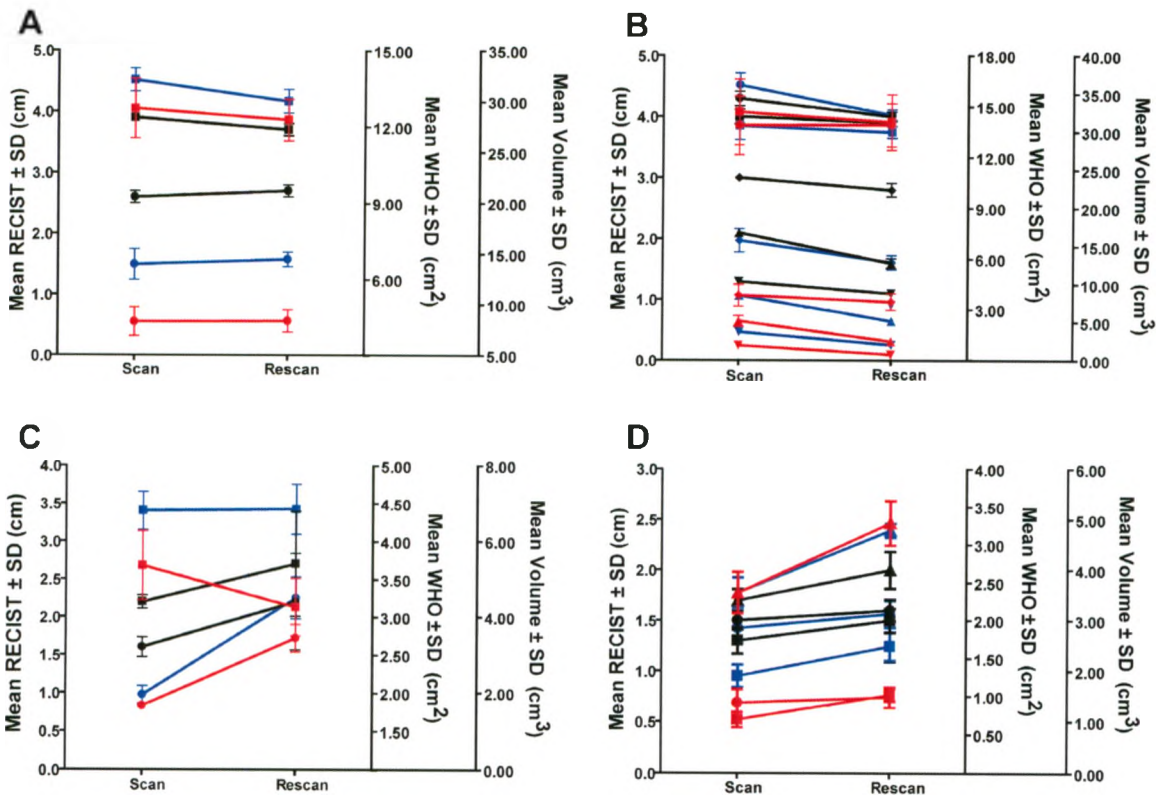


3.3.3 Longitudinal Analysis

3.3.3.1 RECIST, WHO, and Volumetric Measurements at Scan and Rescan

The changes in tumour size between two time points are shown in Figure 3-7. The measurements of a single observer show the mean ± SD of five repeated measurements of all each tumour using RECIST, WHO, and volumetric measurements. The multiplicity of tumours for each subject is represented in this figure, as well as the measurements at scan and rescan

Figure 3-7: Scan and rescan measurements by a single observer



3.3.3.2 Reproducibility of RECIST, WHO, and Volumetric Measurements at Scan and Rescan

To evaluate observer reproducibility, ICC values were determined for each observer. The ICC values for scan and rescan and 1D, 2D, and 3D measurements are shown in Table 3-10. ICC(A) and ICC(C) values for each observer were greater than 0.900. Each observer showed high intra-observer reproducibility at scan and rescan. High intra-observer reproducibility was shown for RECIST, WHO and 3D measurements. The observers also reported high inter-observer reproducibility, and the ICC(A) and ICC(C) values for RECIST, WHO, and 3D measurements are all greater than 0.900.

	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC (A)	ICC (C)	ICC (A)	ICC (C)	ICC (A)	ICC (C)
SCAN						
Observer 1 (JM)	0.994	0.994	0.996	0.996	0.988	0.990
Observer 2 (FS)	0.970	0.974	0.988	0.991	0.981	0.983
Observer 3 (LW)	0.992	0.994	0.994	0.995	0.977	0.983
Observer 4 (SS)	0.996	0.997	0.996	0.996	0.996	0.996
All observers	**0.984**	**0.989**	**0.991**	**0.994**	**0.981**	**0.985**
RESCAN						
Observer 1 (JM)	0.983	0.985	0.991	0.992	0.990	0.991
Observer 2 (FS)	0.973	0.976	0.987	0.988	0.982	0.982
Observer 3 (LW)	0.949	0.953	0.996	0.997	0.954	0.957
Observer 4 (SS)	0.993	0.993	0.990	0.990	0.995	0.995
All Observers	**0.966**	**0.971**	**0.985**	**0.988**	**0.975**	**0.978**

Table 3-10: Correlation Coefficients by Each Observer at Scan and Rescan
The ICC values for all tumour measurements using both Absolute Agreement (A) and Consistency (C) Intraclass Correlation Coefficient values

3.3.3.3 Response Classifications of Pulmonary Metastases using RECIST, WHO, and Volumetric Measurements at Scan and Rescan

Response classifications fall into four categories: stable disease, progressive disease, partial response and complete response. Each of these stratifications are described with their respective criteria using RECIST, WHO, and RECIST 1.1 guidelines.

The measurements of a single observer are shown in Table 3-11. This table shows the mean tumour burden at scan and rescan for the five observers. The absolute change (%) is reported, and these values were used to stratify tumour response. Positive values of absolute change represent tumour growth, and negative values represent tumour shrinkage. Mean tumour burden is reported for RECIST, WHO, and 3D measurements.

These results show that the change of direct 3D measurements can classify the change of tumour size.

Subject #	# of tumours	Scan Mean Tumour Burden (±SD)	Rescan Mean Tumour Burden (±SD)	Absolute Change (%)	Response Classifications
RECIST (cm)					
2	2	6.84 ± 0.06	6.52 ± 0.04	-4.68	SD
3	5	14.8 ± 0.03	13.66 ± 0.04	-7.70	SD
4	2	3.86 ± 0.06	4.5 ± 0.05	16.58	SD
5	3	4.52 ± 0.04	5.34 ± 0.04	18.14	SD
6	11	20.34 ± 0.02	19.98 ± 0.05	-1.77	SD
WHO (cm^2)					
2	2	20.33 ± 0.46	19.16 ± 0.41	-5.76	SD
3	5	43.32 ± 0.15	36.68 ± 0.23	-15.33	SD
4	2	6.56 ± 0.16	8.17 ± 0.19	24.54	SD
5	3	5.53 ± 0.06	7.01 ± 0.08	26.76	PD
6	11	31.88 ± 0.05	26.18 ± 0.33	-17.88	SD
Volume (cm^3)					
2	2	38.26 ± 0.80	39.99 ± 1.11	4.52	SD
3	5	84.98 ± 1.20	77.91 ± 0.73	-8.32	SD
4	2	7.67 ± 0.87	8.72 ± 0.40	13.69	SD
5	3	6.30 ± 0.51	8.50 ± 0.29	34.92	SD
6	11	40.4 ± 0.23	40.79 ± 0.22	0.97	SD

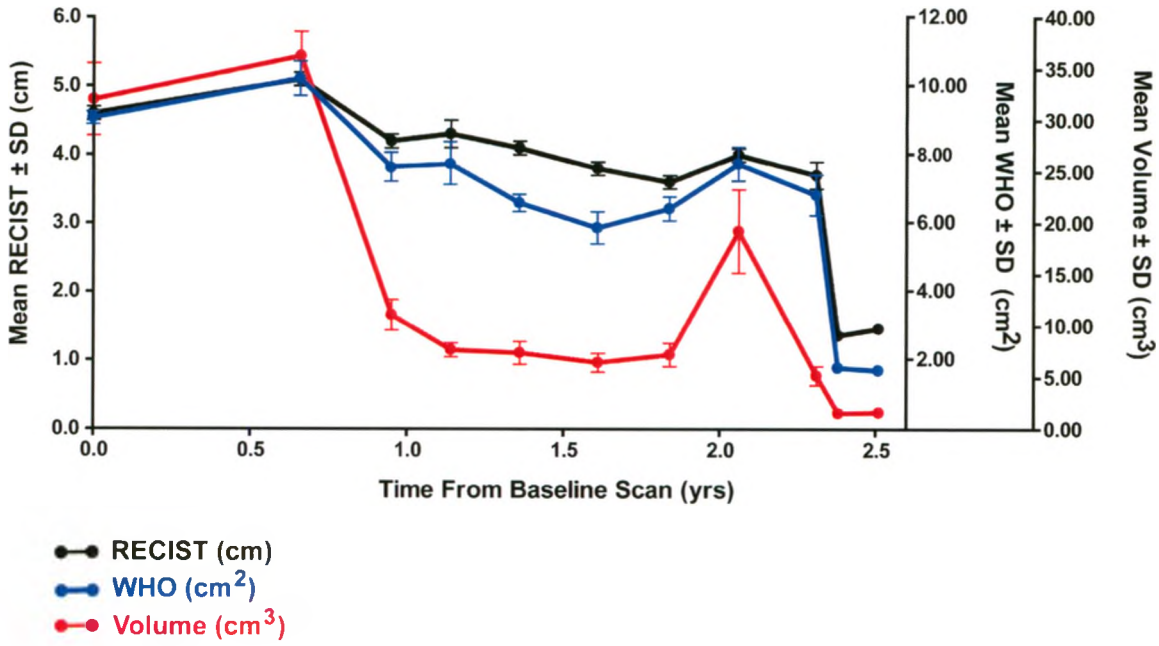
Table 3-11: Response Classifications of Tumour Burden in 1D, 2D, and 3D
Tumour burden measurements at scan and rescan and the absolute change in measurements at both timepoints are presented.

3.3.3.4 Longitudinal Analysis of RECIST, WHO, and Volumetric Measurements

Subject 4 had a tumour load of two tumours. This subject received 13 CT scans; the images for Tumour 1 were measurable at 11 time points, and Tumour 2 was measurable at nine time points.

The sum of 1D, 2D, and 3D measurements for Tumours 1 and 2 are shown in Figure 3-8. Tumour 1 is evaluated from time points 1-11, and Tumour 2 was evaluated from time point 2-10. The changes in tumour size were classified into established response classifications by calculating the absolute change (%) between scans. The absolute change is tumour size at time points 1-11, and the corresponding response classifications are reported in Table 3-12. Tumour measurements in 1D, 2D, and 3D are classified into response categories.

Figure 3-8: Subject 4 – Total tumour burden
The longitudinal changes of the total tumour burden (sum of tumour 1 and tumour 2) are displayed over nine timepoints using RECIST, WHO, and 3D measurements. Mean ± SD are shown by a single observer.



Time Point	Mean Tumour Burden (±SD)	Absolute Change (%)	Response Classification
RECIST (cm)			
1	1.9 ± 0.2		
2	4.6 ± 0.1	142.1	PD
3	5.1 ± 0.1	10.9	SD
4	4.2 ± 0.1	-17.6	SD
5	4.3 ± 0.2	2.4	SD
6	4.1 ± 0.1	-4.7	SD
7	3.8 ± 0.1	-7.3	SD
8	3.6 ± 0.1	-5.3	SD
9	4.0 ± 0.1	11.1	SD
10	3.7 ± 0.2	-7.5	SD
11	1.5 ± 0.1	-60.5	PR
WHO (cm^2)			
1	1.98 ± 0.31		
2	9.03 ± 0.19	355.14	PD
3	10.18 ± 0.50	12.74	SD
4	7.61 ± 0.43	-25.25	SD
5	7.69 ± 0.60	1.05	SD
6	6.55 ± 0.25	-14.82	SD
7	5.83 ± 0.47	-10.99	SD
8	6.39 ± 0.35	9.61	SD
9	7.7 ± 0.50	20.50	SD
10	6.79 ± 0.59	-11.82	SD
11	5.05 ± 0.69	-25.68	SD
Volume (cm^3)			
1	1.78 ± 0.32		
2	32.01 ± 3.51	1698.31	PD
3	36.24 ± 2.36	13.21	SD
4	11.06 ± 1.48	-69.48	PR
5	7.64 ± 0.68	-30.92	SD
6	7.36 ± 1.12	-3.66	SD
7	6.42 ± 0.90	-12.77	SD
8	7.19 ± 1.14	11.99	SD
9	19.22 ± 4.05	167.32	PD
10	1.67 ± 0.13	-91.33	PR
11	1.63 ± 0.21	-2.28	SD

Table 3-12: Subject 4: Longitudinal Response Classifications of Tumour Burden
The longitudinal changes in tumour size were retrospectively classified according to RECIST and WHO criteria. The corresponding volumetric percent changes of each response classification described in RECIST and WHO criteria were applied to determine the response classification of the volumetric measurements.

3.3.3.3 Reproducibility of RECIST, WHO, and Volumetric Measurements at Multiple Time Points

ICC values were used to assess intra- and inter-observer reproducibility. RECIST, WHO, and volumetric measurements of Tumour 1 showed high intra- and inter-observer reproducibility, as all ICC(A) and ICC(C) values were greater than 0.900. RECIST, WHO, and volumetric measurements of Tumour 2 showed low intra- and inter-observer reproducibility. Four ICC(A) values were less than 0.500, and four ICC(C) values were less than 0.700 for RECIST measurements. Intra- and inter-observer reproducibility was also low for WHO and 3D measurements of Tumour 2.

	Tumour 1		Tumour 2	
	ICC(A)	ICC(C)	ICC (A)	ICC(C)
RECIST (cm)				
Observer 1 (JM)	0.958	0.962	0.241	0.498
Observer 2 (FS)	0.953	0.949	0.331	0.551
Observer 3 (LW)	0.964	0.963	0.793	0.782
Observer 4 (SS)	0.976	0.981	0.495	0.635
Inter-observer	**0.944**	**0.961**	**0.405**	**0.577**
WHO (cm^2)				
Observer 1 (JM)	0.967	0.980	0.284	0.608
Observer 2 (FS)	0.970	0.980	0.294	0.585
Observer 3 (LW)	0.982	0.984	0.816	0.811
Observer 4 (SS)	0.988	0.992	0.542	0.655
Inter-observer	**0.971**	**0.979**	**0.343**	**0.485**
Volume (cm^3)				
Observer 1 (JM)	0.979	0.983	0.865	0.865
Observer 2 (FS)	0.962	0.974	0.823	0.879
Observer 3 (LW)	0.990	0.993	0.900	0.923
Observer 4 (SS)	0.992	0.992	0.961	0.965
Inter-observer	**0.929**	**0.949**	**0.685**	**0.792**

Table 3-13: Correlation Coefficients for Each Observer
Absolute agreement and consistency ICC values are reported for each observer. These reliability statistics demonstrate high inter-observer and intra-observer reproducibility.

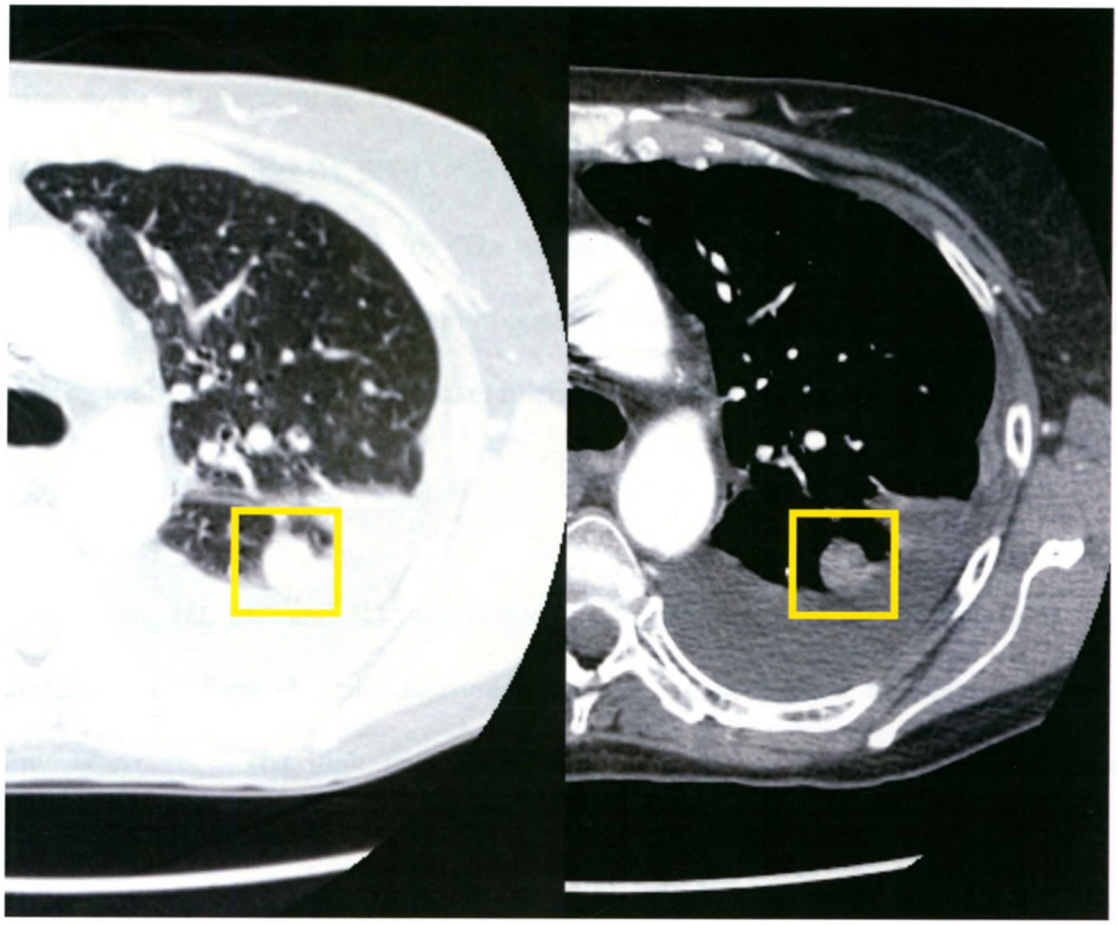
3.4 DISCUSSION

As a first step towards evaluating the potential of 3D measurements to quantify lung tumour size, we compared imaging measurements of pulmonary metastases in 1D, 2D, and 3D. Because there is variability among observers regarding the site selection of target lesions (22), all tumours were selected by one medical oncologist, and subsequently confirmed by two radiologists. There were four main results of this chapter: 1) intra- and inter-observer measurements in 1D, 2D, and 3D were highly reproducible; 2) intra- and inter-observer measurements by trained and expert observers in 1D, 2D, and 3D were highly reproducible, with the exception of one expert observer; 3) interscan precision of 3D measurements were reproducible; and 4) 3D measurements were used to stratify tumour response according to RECIST and WHO criteria.

Twenty-nine pulmonary metastases were evaluated using 1D, 2D, and 3D measurements. Seven observers evaluated tumour size using 1D and 2D measurements; the mean CV for each observer was less than 10%. The lowest RECIST CV was 0.7% (Observer 1) and the highest was 46.1% (Observer 5). The lowest WHO CV was 1.73% (Observer 1) and the highest was 52.86% (Observer 5). We observed high intra- and inter-observer reproducibility in for the 1D, 2D, and 3D measurements, shown by the ICC values. Six observers reported ICC(A) and ICC(C) values greater than 0.900. Previous studies that have evaluated the reproducibility of 1D, 2D, and 3D measurements also observed high inter-observer (13;22) and intra-observer reproducibility (22). Our findings in this study are consistent with the previous results.

In the cross-sectional analysis, seven observers performed RECIST and WHO measurements; four were trained observers (Observers 1-4) and three were expert observers (Observers 5-7). The trained observers had lower RECIST CVs than the expert observers (trained: 3.9%; expert: 7.3%), and also had lower WHO CVs than the expert observers (trained: 5.43%; expert: 15.20%). The high variability of the expert observers was also reflected in the low inter-observer reproducibility: low inter-observer reproducibility for expert observers with ICC(A) and ICC(C) values ranging from 0.509 – 0.542. High inter-observer reproducibility for trained observers was shown in ICC(A) and ICC(C) values greater than 0.900. The measurements of a single tumour (Subject 1) performed by one expert observer (Observer 5), ranged from 1.5 – 10.6cm for RECIST measurements, and 2.1 – 47.0cm2 for WHO measurements. These large ranges resulted in high variability (RECIST: 46.1%; WHO: 52.86%), and low reproducibility (ICC values ranging from 0.758 – 0.785). Figure 3-9 is an image of the lesion, in the lung and soft tissue windows.

Figure 3-9: Subject 1 Solitary Tumour
RECIST and WHO measurements of the solitary tumour of Subject 1 are highly varied by one observer. This tumour is shown using lung and soft tissue windows. It is possible that pleural effusion at the base of the tumour may impede precise measurements. The tumour is enclosed in the yellow box.



Five subjects were evaluated longitudinally, and four observers evaluated tumour sizes in 1D, 2D, and 3D at scan and rescan. Observers had low variability for RECIST, WHO, and 3D measurements; the mean CV of total tumour burden for 1D, 2D, and 3D measurements at scan and rescan were less than 2%.

As Subject 4 was being evaluated longitudinally, pleural effusion in the right lung observed in the CT scans of the last two time points precluded the differentiation of the tumour boundary from the surrounding fluid. When reviewing the subject's medical

history, we found no clinical notes from time points 1 to 9 (January 11, 2005 to May 04, 2007). At time point 10 (May 30, 2007), the subject was prescribed Sunitinib, and classified with PD. At time point 11 (July 17, 2007), the subject was enrolled in a chemotherapy clinical trial at Victoria Hospital, and was on the active drug, Everolimus. At time point 12 (September 11, 2007), the subject was stopped using Sunitinib, and at time point 13 (December 04, 2007), the subject was at home, with an oxygen respiratory aid, and was prescribed Sorafenib. The subject's records did not reveal any treatment (prescription drugs, radiation or chemotherapy) between time points 8-10, as a result, the reasoning for the spike in the 3D measurements of Tumour 2 is inconclusive.

3.5 Conclusions

In conclusion, 1D, 2D, and 3D measurements can quantify lung tumour size with high reproducibility. Reproducible measurements may be performed by both expert and trained observers. The high reproducibility and ability to classify tumour response suggest this technique may be used in clinical trials.

3.6 References

1. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J.Natl.Cancer Inst.* 1999;91:523-8.

2. Sun JM, Ahn MJ, Park MJ, Yi JH, Kim TS, Chung MJ et al. Accuracy of RECIST 1.1 for non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors. *Lung Cancer* 2009.

3. van Klaveren RJ, Aerts JG, de BH, Giaccone G, Manegold C, van Meerbeeck JP. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. *Lung Cancer* 2004;43:63-9.

4. Thiam R, Fournier LS, Trinquart L, Medioni J, Chatellier G, Balvay D et al. Optimizing the size variation threshold for the CT evaluation of response in metastatic renal cell carcinoma treated with sunitinib. *Ann.Oncol.* 2009.

5. Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. *J.Clin.Epidemiol.* 2004;57:358-65.

6. Yamamoto Y, Kameyama R, Murota M, Bandoh S, Ishii T, Nishiyama Y. Early assessment of therapeutic response using FDG PET in small cell lung cancer. *Mol.Imaging Biol.* 2009;11:467-72.

7. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RT, Reznek RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br.J.Radiol.* 2000;73:1178-84.

8. Pauls S, Kurschner C, Dharaiya E, Muche R, Schmidt SA, Kruger S et al. Comparison of manual and automated size measurements of lung metastases on MDCT images: potential influence on therapeutic decisions. *Eur.J.Radiol.* 2008;66:19-26.

9. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur.Radiol.* 2006;16:781-90.

10. Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin.Cancer Res.* 2003;9:4318-23.

11. Plathow C, Klopp M, Thieke C, Herth F, Thomas A, Schmaehl A et al. Therapy response in malignant pleural mesothelioma-role of MRI using RECIST, modified RECIST and volumetric approaches in comparison with CT. *Eur.Radiol.* 2008;18:1635-43.

12. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J.Natl.Cancer Inst.* 2000;92:205-16.

13. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* 2009;252:263-72.

14. Fraioli F, Bertoletti L, Napoli A, Calabrese FA, Masciangelo R, Cortesi E et al. Volumetric evaluation of therapy response in patients with lung metastases. Preliminary results with a computer system (CAD) and comparison with unidimensional measurements. *Radiol.Med.* 2006;111:365-75.

15. Iwano S, Okada T, Koike W, Matsuo K, Toya R, Yamazaki M et al. Semi-automatic volumetric measurement of lung cancer using multi-detector CT effects of nodule characteristics. *Acad.Radiol.* 2009;16:1179-86.

16. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur.J.Cancer* 2009;45:228-47.

17. Birchard KR, Hoang JK, Herndon JE, Jr., Patz EF, Jr. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. *Cancer* 2009;115:581-6.

18. World Health Organization. WHO Handbook for Reporting Results for Cancer Treatment. 1979. Geneva, World Health Organization.
Ref Type: Report

19. Galanis E, Buckner JC, Maurer MJ, Sykora R, Castillo R, Ballman KV et al. Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods. *Neuro.Oncol.* 2006;8:156-65.

20. Park JO, Lee SI, Song SY, Kim K, Kim WS, Jung CW et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn.J.Clin.Oncol.* 2003;33:533-7.

21. Tran LN, Brown MS, Goldin JG, Yan X, Pais RC, Nitt-Gray MF et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. *Acad.Radiol.* 2004;11:1355-60.

22. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR Am.J.Roentgenol.* 1996;167:851-4.

23. Landry A, Spence JD, Fenster A. Quantification of carotid plaque volume measurements using 3D ultrasound imaging. *Ultrasound Med.Biol.* 2005;31:751-62.

24. Smith WL, Lewis C, Bauman G, Rodrigues G, D'Souza D, Ash R et al. Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR. *Int.J.Radiat.Oncol.Biol.Phys.* 2007;67:1238-47.

25. Landry A, Ainsworth C, Blake C, Spence JD, Fenster A. Manual planimetric measurement of carotid plaque volume using three-dimensional ultrasound imaging. *Med.Phys.* 2007;34:1496-505.

Chapter 4: Conclusions and Future Work

4.1 Summary

The development of accurate and precise measurements of pulmonary metastatic lesions is critical for monitoring tumour changes over time. Although established methods of quantifying metastatic tumour size *in vivo*, Response Evaluation Criteria in Solid Tumours (RECIST) (1;2), and World Health Organization (WHO) (3), have reported high intra- and inter-observer reproducibility (4), it is possible that these 1D and 2D measurements are not encompassing the full geometry of the tumour. The limitations of existing measurement techniques have served as motivation for the development of a three-dimensional technique to evaluate tumour size.

Listed below are the hypotheses tested in this thesis, as well as the important results:

I. RECIST, WHO, and 3D measurements are accurate and precise, when measuring lung tumour phantoms of known dimensions.

In Chapter 2, four observers evaluated two sets of solid spherical tumour phantoms, two observers evaluated three geometrically-shaped tumour phantoms, and 3 irregularly-shaped tumour phantoms were evaluated by a single observer. For the first set of solid spherical tumours, there were eight instances of a significant difference from ground truth measurements 1D measurements, by a single observer. There was a significant difference at 1.0mm slice thickness (0.9 ± 0.1cm, $p<0.001$) for Tumour 1; at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses (0.7 ± 0.0cm, $p<0.001$) for Tumour 2; and at 1.0mm, 2.0mm, and 5.0mm slice thicknesses (1.1 ± 0.0cm, $p<0.001$) for Tumour 3.

For WHO measurements, there were 12 instances of a significant difference from ground truth, by two observers. For Tumour 1 there was a significant difference in the measurements by Observer 4 at 1.0mm slice thickness ($0.89 \pm 0.10cm^2$, $p=0.007$); Tumour 2, Observer 4 had significantly different measurements at 0.5mm ($0.52 \pm 0.07cm^2$, $p<0.001$), 1.0mm ($0.49 \pm 0.00cm^2$, $p<0.001$), 2.0mm ($0.49 \pm 0.00cm^2$, $p<0.001$), and 5.0mm slice thicknesses ($0.49 \pm 0.00cm^2$, $p<0.001$). Also for Tumour 2, Observer 3 had significantly different measurements at 1.0mm ($0.56 \pm 0.00cm^2$, $p<0.000$), 2.0mm ($0.55 \pm 0.03cm^2$, $p=0.002$), and 5.0mm slice thicknesses ($0.56 \pm 0.05cm^2$, $p<0.001$). For Tumour 3, Observer 4 had significantly different at 1.0mm ($1.26 \pm 0.10cm^2$, $p=0.019$); and 2.0mm and 5.0mm slice thicknesses ($1.21 \pm 0.00cm^2$, $p<0.001$). Observer 3 had significantly different measurements at 5.0mm slice thickness ($1.37 \pm 0.07cm^2$, $p=0.007$). In the volumetric analysis, there were seven instances of a significant difference from ground truth measurements, by three observers. Observer 4 had significantly different 3D measurements at 0.5mm ($0.45 \pm 0.01cm^3$, $p=0.011$); and 2.0mm slice thicknesses ($0.64 \pm 0.07cm^3$, $p=0.021$) for Tumour 1. For Tumour 2, Observers 1 and 4 showed significant differences at 0.5mm slice thickness ($0.24 \pm 0.02cm^3$, $p=0.038$), and ($0.22 \pm 0.01cm^3$, $p<0.001$), respectively. For Tumour 3, Observer 4 had significantly different measurements at 0.5mm ($0.89 \pm 0.10cm^3$, $p=0.015$); 2.0mm ($0.89 \pm 0.10cm^3$, $p=0.006$); and 5.0mm slice thicknesses ($0.89 \pm 0.10cm^3$, $p=0.022$). For each measurement technique, there were 48 mean ± SD observer measurements which were compared to ground truth measurements using a one-way ANOVA. For RECIST measurements, 83% of observer measurements were accurate; for WHO measurements, 75% of observer

measurements were accurate; and for 3D measurements, 85% of observer measurements were accurate.

For the analysis of the irregularly-shaped tumours performed by a single observer, there was a significant difference between observer measurements and ground truth for 1D, 2D, and 3D measurements at all slice thicknesses for Tumour 2. The measurements for Tumours 1 and 3 were also significantly different for 1D and 2D measurements at all slice thicknesses; however the 3D measurements at 5.0mm slice thickness were not significantly different. Although these measurements showed poor accuracy, 3D measurements showed the ability to accurately quantify irregular tumour size, when the 1D and 2D measurements showed significant differences.

II. RECIST, WHO, and 3D measurements have high inter-observer and intra-observer reproducibility.

In Chapter 2, four observers evaluated both sets of spherical tumour phantoms using five repeated 1D, 2D, and 3D measurements. For the first set of spherical phantoms, the inter-observer reproducibility was ICC(A) = 0.962 and ICC(C) = 0.980 for RECIST measurements; ICC(A) = 0.964 and ICC(C) = 0.979 for WHO measurements; and ICC(A) = 0.926 and ICC(C) = 0.959 for 3D measurements. All ICC(A) and ICC(C) values of individual observers were greater than 0.900 for 1D, 2D, and 3D measurements, except for the 3D measurements performed by Observer 4 [ICC(A) = 0.898], and the ranges were: ICC(A) = 0.962-1.000 and ICC(C) = 0.966-1.000 for RECIST; ICC(A) = 0.962-1.000 and ICC(C) = 0.966-1.000 for WHO; and ICC(A) = 0.898-0.928 and ICC(C)

= 0.937-0.981 for 3D measurements. The inter-observer CVs were less than 10% for 1D, 2D, and 3D measurements, and intra-observer CVs ranged from 0.0 - 4.0% for RECIST, 0.00 - 7.86% for WHO, and 0.77 – 13.55% for 3D measurements. The 1D, 2D, and 3D measurements for the second set of spherical tumour phantoms also showed high inter-observer reproducibility; ICC(A) and ICC(C) values for RECIST and WHO measurements were greater than 0.900, and ICC(A) and ICC(C) values for volumetric measurements were greater than 0.800. Intra-observer ICC(A) and ICC(C) values for 1D and 2D measurements were greater than 0.900, and the ICC(A) and ICC(C) values for 3D measurements ranged from 0.751 – 0.972 and 0.820 – 0.982, respectively. The geometrically-shaped tumour phantoms also showed high inter- and intra-observer reproducibility with all ICC values greater than 0.900. One observer performed repeated measurements of the irregularly-shaped tumour phantoms, and the intra-observer CVs for 1D, 2D, and 3D measurements were less than 10%.

In Chapter 3, seven observers performed a cross-sectional analysis of seven subjects with pulmonary metastases using 1D and 2D measurements. Six observers had ICC values greater than 0.900, and one observer had ICC values that ranged from 0.758 – 0.785. The inter-observer ICC values for all observers ranged from 0.645 – 0.679. Observers were classified into two categories: trained and expert observers. The inter-observer reproducibility for trained observers had ICC values that ranged from 0.940 – 0.989. The expert observer ICC values ranged from 0.506 – 0.542. The RECIST CVs for all observers was 6.2% and the WHO CVs were 12.45%. Two trained and expert observers also evaluated the same tumours using 3D measurements. Each category of observers

(trained, expert, all observers) showed high reproducibility, with ICC values greater than 0.900. Inter-observer CVs for these four observers were less than 10%.

III. RECIST, WHO, and 3D measurements of expert medical professionals, and trained research students and technicians are not statistically different.

A paired t-test was used to determine any significant differences between trained and expert observer measurements. There were no differences between trained and expert observers for RECIST (trained: 2.1 ± 0.0cm; expert: 2.1 ± 0.1cm), and WHO (trained: 4.14 ± 0.10cm^2; expert: 4.26 ± 0.14cm^2) measurements, but there was a significant difference for the 3D measurements (trained: 7.57 ± 0.31cm^3; expert: 8.73 ± 0.28cm^3, p=0.005). It can be concluded that the trained observer measurements were significantly less than the expert observers.

IV. 3D measurements may be used to stratify tumour response at longitudinally.

Longitudinal analyses of five subjects were performed by four trained observers. The established RECIST and WHO criteria categorize the changes in tumour size as a function of the sum of all longest axes, and product of longest axes and longest perpendicular bisectors, respectively; these response classifications also provide corresponding volumetric changes. The absolute percent change of tumour size was calculated from the direct 3D measurements performed at scan and rescan. Using the volumetric percent changes corresponding with the 1D and 2D measurements, the tumour burden of each subject was classified into a response category. All five response

classifications determined by direct 3D measurements agreed with the five RECIST classifications. Four 3D classifications agreed with WHO classifications.

Two tumours in Subject 4 were evaluated longitudinally over 11 time points. RECIST and WHO criteria and direct 3D measurements were used to classify the change in tumour size at each time point. Six of 10 (60%) classifications based on 3D measurements were in agreement with RECIST measurements, and seven of 10 (70%) 3D measurements were in agreement with WHO measurements. Nine of 10 (90%) of RECIST measurements were in agreement with WHO measurements.

V. RECIST, WHO, and 3D measurements are highly correlated with one another.

Trained and expert observer measurements for the seven observers who performed 1D and 2D measurements were significantly correlated:1D ($R^2 = 0.7291$; $r = 0.8539$; $y = 0.8280 + 0.1751$; $p < 0.0001$), and 2D ($R^2 = 0.7377$; $r = 0.8589$; $y = 0.8088 + 0.1901$; $p < 0.0001$). Four observers, two trained and two experts, performed 3D measurements and they were highly correlated ($R^2 = 0.9914$; $r = 0.9957$; $y = 0.8269 + 0.3439$; $p < 0.0001$).

The overarching objective of this thesis was to quantify pulmonary metastases using 1D, 2D, and 3D measurements. A secondary goal was to determine the intra- and inter-observer reproducibility and accuracy of each measurement. Ultimately, the goal of using 3D measurements is to encompass the entire tumour volume when quantifying tumour size, unveiling the possibility of implementing this technique into the clinical

setting, such as radiology viewing and clinical trials, to assess longitudinal tumour changes. Future work to be undertaken utilizing the data and approaches in Chapters 2 and 3 are described in this chapter, which provides a roadmap for the laboratory to semi-automate the 3D measurements, making it a step closer to clinical integration.

4.2 Limitations of the Current Study and Analysis

As a result of the research completed in this thesis, we established that our 3D measurement provides accurate and precise measurements of pulmonary metastases. The limitations of the study presented in this thesis, as well as suggestions for improvements to the imaging and analysis methods outlined in Chapters 2 and 3 are described in the section. In particular, there are limitations with our sample size, as well as our 3D measurement technique.

We found that the manual 3D segmentations were very time consuming, which is a limitation that has been acknowledged; Mantatzis et al. (5) noted the time consuming nature of 3D measurements, and the obstacle that it presents from being translated into clinical practice. In addition, the axis about which the tumour rotated in 18^{o} increments, as described in Chapters 2 and 3, was observer dependent. The image processing platform did not provide the feature to record the coordinates the rotational axis placement in each measurement trial. This is another possible source of measurement variability within the five repeated measures of each tumour.

As a review, the axis of rotation was established for each tumour, and the tumour was rotated about 18°; ten contours were delineated which were used to calculate the tumour volume. Once the ten-contour series began, 3DQuantify did not provide the option of modifying the window widths; all of the juxtapleural nodules (n=13) had image slices in which the tumour boundary and pleural were not clearly resolved. This may also have been the cause of some measurement variability.

A solution to the limitations of 3D measurement variability and extensive analysis duration may be addressed with the development of automated segmentation tools. With the implementation of automated or semi-automated measurement tools, we expect a reduction in intra- and inter-observer variability. The reduced observer variability and analysis time anticipated with automation or semi-automation could enable the application of 3D measurements as a clinical analysis tool.

4.3 Future Studies

4.3.1 Semi-Automated Segmentation

Currently, our lab is working on the development and validation of a semi-automated algorithm to segment pulmonary metastases in x-ray CT images. The algorithm is a combination of localizing region-based active contours, and active contours without edges (6;7). Figure 4-1 shows the different steps in the semi-automated algorithm, as well as the visual

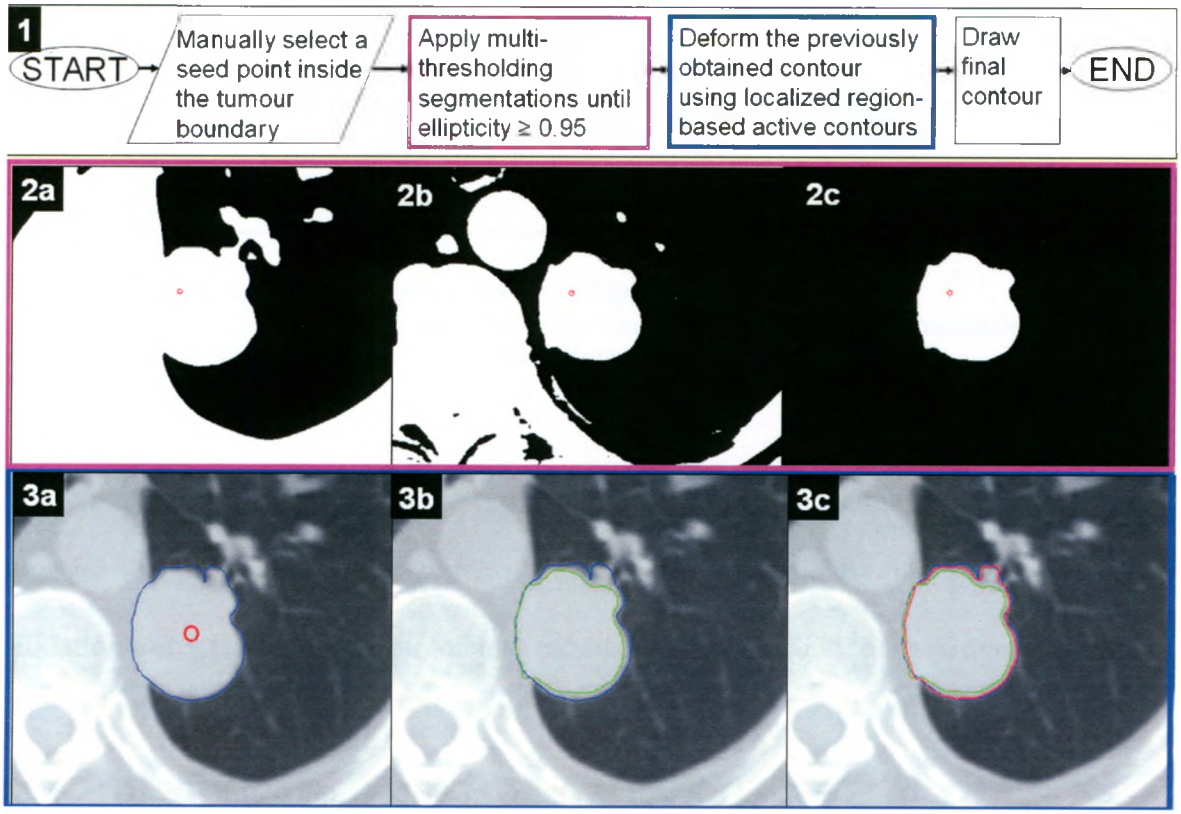


Figure 4-1: Semi-automated segmentation algorithm

1. Flow chart of the semi-automated algorithm, identifying the two techniques used to perform the segmentation: multi-thresholding (enclosed in purple box); and localized region-based active contours (blue box). **2.** When the seed point is chosen inside the tumour boundary, the threshold value is increased until the ratio between the area of the region containing the seed point (tumour) to the area of the ellipse that has the same normalized second central moments as the region is greater than 0.95. **3.** User-determined seeding point (red circle), and manually segmented contour (blue) are shown in **3a**. The final contour achieved using the multi-threshold technique is used as the initial contour (green) of the localized region-based active contours method, shown in **3b**. The final contour (red) is based on a smooth version of uniform modeling energy with multiple localized radii.

Preliminary results for the application of the application are based on 3D measurements of total tumour burden for the solid spherical tumour phantoms evaluated in Chapter 2. Three observers performed five repeated 3D measurements of each tumour, and evaluated the total tumour burden at each measurement trial as the volumetric sum of each tumour. Each observer also performed five repeated volumetric measurements of the three solid spherical tumour phantoms using the semi-automated segmentation algorithm. The results are found in Table 4-1.

	Tumour Burden (Manual)		Tumour Burden (Semi-automated)	
	Mean(±SD) (cm^3)	CV %	Mean(±SD) (cm^3)	CV %
Observer 1	1.62(0.10)	6.4	1.55(0.00)	0.0
Observer 2	1.78(0.10)	5.7	1.56(0.01)	0.4
Observer 3	1.68(0.04)	2.3	1.53(0.00)	0.4

Table 4-1: Tumour burden volumes by multiple observers
Three observers used manual segmentation techniques and the semi-automated segmentation software to determine total tumour burden volumes in the chest phantom.

Preliminary results show that, as expected, the semi-automated technique has lower intra-observer variability. For future studies, this algorithm could be modified for further validation using phantom measurements, and may also be applied to the patient cohort.

4.3.2 Three-Dimensional Measurements of Brain Tumours

Similar techniques as the ones performed in this thesis are currently being used in a retrospective analysis of subjects with cerebral lesions. This study is a longitudinal analysis of sixteen subjects, who were imaged with contrast-enhanced MRI or CT. The observer is currently blinded between time points to prevent any measurement bias. An example of a brain lesion is shown in Figure 4-2.

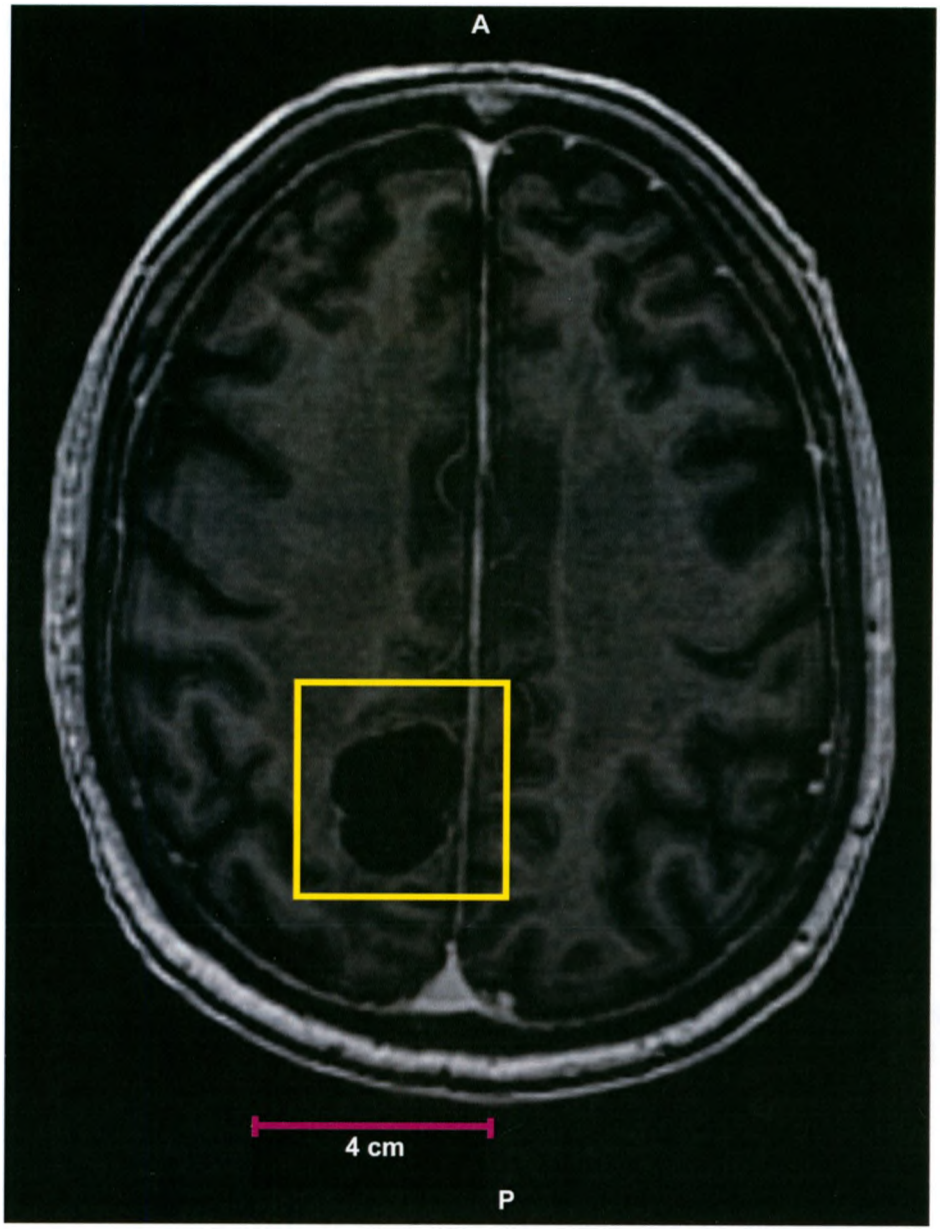


Figure 4-2: Brain Lesion
A brain lesion is shown, enclosed in the yellow box. This is an MR image, post-contrast agent.

4.4 Conclusions

The work in this thesis has shown the potential of using volumetric assessment to accurately evaluate pulmonary metastatic lesions, with high inter- and intra-observer reproducibility. While there are some limitations to the current method, semi-automated

segmentation techniques have the potential to provide volumetric measurements with low variability, and may be applied to other areas of interest, such as the brain. With the continued development of this 3D measurement technique, it has the potential of clinical translation.

4.5 References

1. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur.J.Cancer* 2009;45:228-47.

2. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J.Natl.Cancer Inst.* 2000;92:205-16.

3. World Health Organization. The World Health Organization's Fight Against Cancer: Strategies That Prevent, Cure and Care. 2007. Switzerland, World Health Organization.
   Ref Type: Report

4. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J.Clin.Oncol.* 2003;21:2574-82.

5. Mantatzis M, Kakolyris S, Amarantidis K, Karayiannakis A, Prassopoulos P. Treatment response classification of liver metastatic disease evaluated on imaging. Are RECIST unidimensional measurements accurate? *Eur.Radiol.* 2009;19:1809-16.

6. Chan TF, Vese LA. Active contours without edges. *IEEE Trans.Image Process* 2001;10:266-77.

7. Lankton S, Tannenbaum A. Localizing region-based active contours. *IEEE Trans.Image Process* 2008;17:2029-39.

Appendix A: Ethics Approvals for Study

Office of Research Ethics
The University of Western Ontario
Room 00045 Dental Sciences Building, London, ON, Canada N6A 5C1
Telephone: (519) 661-3036 Fax: (519) 850-2466 Email: ethics@uwo.ca
Website: www.uwo.ca/research/ethics

Western

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. G. Parraga
Review Number: 12851E **Review Date:** November 15, 2006 **Revision Number:**
Protocol Title: Variability of RECIST Measurements
Department and Institution: Diagnostic Imaging Nuclear Medicine, Robarts Research Institute
Sponsor:
Ethics Approval Date: December 13, 2006 **Expiry Date:** November 30, 2011
Documents Reviewed and Approved: UWO protocol
Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted expedited approval to the above named research study on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

This approval shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. John W. McDonald
Deputy Chair: Susan Hoddinott

Ethics Officer to Contact for Further Information		
☑ Denise Grafton	☐ Janice Sutherland	☐ Jennifer McEwen

This is an official document. Please retain the original in your files.

cc: ORE File
LHRI

Appendix B: CT Parameters for Chest Phantom Imaging

The imaging parameters used for imaging the chest phantom are listed in Table B-1. All images were acquired at Princess Margaret Hospital (Toronto, Ontario), using a 320-slice Toshiba Aquilion One CT scanner. Twelve scans were performed using helical CT, and four scans were acquired using volume CT. All 16 scans were performed with 120 kVp and 300 mA. All images had 512x512 pixel matrix.

Scan Name	Scanning Option	# of images	Slice Thickness (mm)	x-pixel length (mm)	y-pixel length (mm)	FOV(x) (mm)	FOV(y) (mm)	FOV(z) (mm)
Spherical Solid (Set 1)	Volume	736	0.5	0.976	0.976	500	500	368
Spherical Solid (Set 1)	Volume	368	1	0.976	0.976	500	500	368
Spherical Solid (Set 1)	Volume	45	2	0.976	0.976	500	500	90
Spherical Solid (Set 1)	Volume	18	5	0.976	0.976	500	500	90
Spherical Solid (Set 2)	Volume	368	0.5	0.976	0.976	500	500	184
Spherical Solid (Set 2)	Volume	184	1	0.976	0.976	500	500	184
Spherical Solid (Set 2)	Volume	92	2	0.976	0.976	500	500	184
Spherical Solid (Set 2)	Volume	37	5	0.976	0.976	500	500	185
Geometrical Shapes	Volume	363	0.5	0.976	0.976	500	500	181.5
Geometrical Shapes	Volume	184	1	0.976	0.976	500	500	184
Geometrical Shapes	Volume	91	2	0.976	0.976	500	500	182
Geometrical Shapes	Volume	37	5	0.976	0.976	500	500	185
Irregular	Helical	629	0.5	0.671	0.671	344	344	314.5
Irregular	Helical	315	1	0.671	0.671	344	344	315
Irregular	Helical	158	2	0.671	0.671	344	344	316
Irregular	Helical	63	5	0.671	0.671	344	344	315

Table B-1: CT Scanning Parameters of Chest Phantom
All CT imaging was performed on a 320-slice Toshiba Aquilion One scanner, located at Princess Margaret Hospital (Toronto, ON).

Appendix C: Supplemental Data from Phantom Analyses (Chapter 2)

C.1. Solid Spherical Tumour Phantoms (Set 1)

	Tumour 2			Tumour 3		
	RECIST (cm)	WHO (cm2)	VOLUME (cm3)	RECIST (cm)	WHO (cm2)	VOLUME (cm3)
0.5mm						
Observer 1 (JM)	0.8 ± 0.0	0.64 ± 0.00	0.24 ± 0.02	1.2 ± 0.0	1.44 ± 0.00	0.85 ± 0.06
Observer 2 (FS)	0.8 ± 0.0	0.64 ± 0.00	0.28 ± 0.02	1.2 ± 0.0	1.49 ± 0.11	0.95 ± 0.06
Observer 3 (LW)	0.8 ± 0.0	0.62 ± 0.04	0.26 ± 0.01	1.2 ± 0.0	1.44 ± 0.00	0.90 ± 0.02
Observer 4 (SS)	0.7 ± 0.0	0.52 ± 0.07	0.22 ± 0.01	1.2 ± 0.1	1.35 ± 0.13	0.81 ± 0.02
1.0mm						
Observer 1 (JM)	0.8 ± 0.0	0.64 ± 0.00	0.24 ± 0.03	1.2 ± 0.0	1.44 ± 0.00	0.85 ± 0.04
Observer 2 (FS)	0.8 ± 0.0	0.64 ± 0.00	0.29 ± 0.02	1.2 ± 0.0	1.49 ± 0.01	0.97 ± 0.06
Observer 3 (LW)	0.8 ± 0.0	0.56 ± 0.00	0.31 ± 0.02	1.2 ± 0.0	1.49 ± 0.11	0.88 ± 0.01
Observer 4 (SS)	0.7 ± 0.0	0.49 ± 0.00	0.24 ± 0.01	1.1 ± 0.0	1.26 ± 0.10	0.87 ± 0.06
2.0mm						
Observer 1 (JM)	0.8 ± 0.0	0.64 ± 0.00	0.23 ± 0.02	1.2 ± 0.0	1.44 ± 0.00	0.82 ± 0.08
Observer 2 (FS)	0.8 ± 0.0	0.61 ± 0.07	0.30 ± 0.05	1.2 ± 0.0	1.44 ± 0.00	1.00 ± 0.13
Observer 3 (LW)	0.8 ± 0.0	0.55 ± 0.03	0.27 ± 0.02	1.2 ± 0.0	1.42 ± 0.05	0.89 ± 0.05
Observer 4 (SS)	0.7 ± 0.0	0.49 ± 0.00	0.33 ± 0.04	1.1 ± 0.0	1.21 ± 0.00	1.11 ± 0.10
5.0mm						
Observer 1 (JM)	0.8 ± 0.0	0.64 ± 0.00	0.27 ± 0.05	1.2 ± 0.0	1.44 ± 0.00	1.03 ± 0.22
Observer 2 (FS)	0.8 ± 0.0	0.64 ± 0.00	0.33 ± 0.05	1.2 ± 0.0	1.44 ± 0.00	1.07 ± 0.15
Observer 3 (LW)	0.8 ± 0.0	0.56 ± 0.05	0.30 ± 0.03	1.2 ± 0.0	1.37 ± 0.07	1.12 ± 0.09
Observer 4 (SS)	0.7 ± 0.0	0.49 ± 0.00	0.30 ± 0.03	1.1 ± 0.0	1.21 ± 0.00	1.19 ± 0.11

Table C-1: Mean 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 1)
Mean RECIST, WHO, and volumetric measurements for Tumours 2 and 3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thickness by each observer. Values are expressed as mean ± SD of five repeated measurements.

Intra-observer Coefficients of Variation (%)						
	Tumour 2			Tumour 3		
	RECIST (%)	WHO (%)	VOLUME (%)	RECIST (%)	WHO (%)	VOLUME (%)
0.5mm						
Observer 1 (JM)	0.0	0.00	9.87	0.0	0.00	7.52
Observer 2 (FS)	0.0	0.00	6.34	3.7	7.50	5.91
Observer 3 (LW)	0.0	5.73	2.14	0.0	0.00	2.13
Observer 4 (SS)	6.2	12.90	3.99	4.7	9.35	2.31
1.0mm						
Observer 1 (JM)	0.0	0.00	13.60	0.0	0.00	4.16
Observer 2 (FS)	0.0	0.00	8.05	3.7	7.50	6.52
Observer 3 (LW)	0.0	0.00	6.37	3.7	7.50	1.69
Observer 4 (SS)	0.0	0.00	5.48	0.2	8.19	6.60
2.0mm						
Observer 1 (JM)	0.0	0.00	10.92	0.0	0.00	9.41
Observer 2 (FS)	5.7	11.00	16.48	0.0	0.00	12.83
Observer 3 (LW)	5.7	5.73	6.13	0.0	3.79	6.05
Observer 4 (SS)	0.0	0.00	12.38	4.0	0.00	8.84
5.0mm						
Observer 1 (JM)	0.0	0.00	18.91	0.0	0.00	21.49
Observer 2 (FS)	0.0	0.00	16.17	0.0	0.00	14.08
Observer 3 (LW)	5.7	9.45	8.26	0.0	4.80	8.42
Observer 4 (SS)	0.0	0.00	10.58	0.2	0.00	8.91

Table C-2: Mean CVs of 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 1)

Mean intra-observer coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumours 2 and 3 are reported. Each tumour was measured at all slice thickness by each observer. Values are expressed as percents.

Mean ± SD Observer Measurements				
RECIST Observer	0.5mm	1.0mm	2.0mm	5.0mm
Observer 1 (JM)	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
Observer 2 (FS)	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
Observer 3 (LW)	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
Observer 4 (SS)	2.8 ± 0.1	2.8 ± 0.1	2.8 ± 0.0	2.8 ± 0.0
Mean	**3.0 ± 0.0**	**3.0 ± 0.0**	**2.9 ± 0.0**	**2.9 ± 0.0**
WHO Observer				
Observer 1 (JM)	3.08 ± 0.00	3.08 ± 0.00	3.08 ± 0.00	3.08 ± 0.00
Observer 2 (FS)	3.13 ± 0.11	3.13 ± 0.11	3.05 ± 0.07	3.08 ± 0.00
Observer 3 (LW)	3.06 ± 0.04	3.05 ± 0.11	2.96 ± 0.08	2.93 ± 0.08
Observer 4 (SS)	2.79 ± 0.22	2.63 ± 0.19	2.70 ± 0.00	2.70 ± 0.00
Mean	**3.02 ± 0.06**	**2.97 ± 0.06**	**2.95 ± 0.03**	**2.95 ± 0.02**
Volume Observer				
Observer 1 (JM)	1.62 ± 0.10	1.58 ± 0.10	1.55 ± 0.08	1.81 ± 0.26
Observer 2 (FS)	1.78 ± 0.10	1.83 ± 0.13	1.88 ± 0.25	2.02 ± 0.26
Observer 3 (LW)	1.68 ± 0.04	1.77 ± 0.06	1.69 ± 0.08	1.99 ± 0.12
Observer 4 (SS)	1.49 ± 0.01	1.63 ± 0.08	2.07 ± 0.14	2.13 ± 0.20
Mean	**1.64 ± 0.04**	**1.70 ± 0.05**	**1.80 ± 0.08**	**1.99 ± 0.11**

Table C-3: Mean Total Tumour Burden of Solid Spherical Tumours (Set 1) per Slice Thickness Using RECIST, WHO, and Volumetric Measurements

Mean total tumour burden for each observer at each slice thickness are reported. Total tumour burden is calculated as the sum of all tumour measurements per observer. Values are expressed as mean ± SD of five repeated measurements of all tumour sums.

Mean ± SD Observer Measurements				
RECIST Observer	0.5mm	1.0mm	2.0mm	5.0mm
Observer 1 (JM)	0.0	0.0	0.0	0.0
Observer 2 (FS)	1.5	1.5	1.5	0.0
Observer 3 (LW)	0.0	1.5	1.5	1.5
Observer 4 (SS)	4.0	3.2	1.6	1.6
Mean	**1.0**	**0.9**	**0.6**	**0.5**
WHO Observer				
Observer 1 (JM)	0.00	0.00	0.00	0.00
Observer 2 (FS)	3.57	3.57	2.20	0.00
Observer 3 (LW)	1.17	3.67	2.87	2.84
Observer 4 (SS)	7.86	7.06	0.00	0.00
Mean	**2.06**	**2.05**	**0.92**	**0.70**
Volume Observer				
Observer 1 (JM)	6.44	6.58	5.17	14.43
Observer 2 (FS)	5.73	6.84	13.55	13.00
Observer 3 (LW)	2.30	3.41	4.54	6.22
Observer 4 (SS)	0.77	4.87	6.57	9.24
Mean	**2.31**	**2.80**	**4.30**	**5.50**

Table C-4: Mean CV of Total Tumour Burden of Solid Spherical Tumours (Set 1) per Slice Thickness Using RECIST, WHO, and Volumetric Measurements
Mean coefficients of variation of mean total tumour burden for each observer at each slice thickness are reported. Values are expressed as a percent.

Intra-class Correlation Coefficients for Intra- and Inter-Observer Reproducibility						
	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC (A)	ICC (C)	ICC (A)	ICC (C)	ICC (A)	ICC (C)
Observer 1 (JM)	1.000	1.000	1.000	1.000	0.928	0.937
Observer 2 (FS)	0.989	0.989	0.987	0.987	0.952	0.981
Observer 3 (LW)	0.989	0.989	0.985	0.986	0.956	0.965
Observer 4 (SS)	0.962	0.966	0.962	0.966	0.898	0.952
Inter-Observer	**0.962**	**0.980**	**0.964**	**0.979**	**0.926**	**0.959**

Table C-5: Correlation Coefficients for Each Observer (Solid Spheres Set 1)
Absolute agreement and consistency intra-class correlation coefficients (ICC) for each observer at all slice thickness are reported. Inter-observer and intra-observer reliability is assessed for RECIST, WHO, and volumetric measurements.

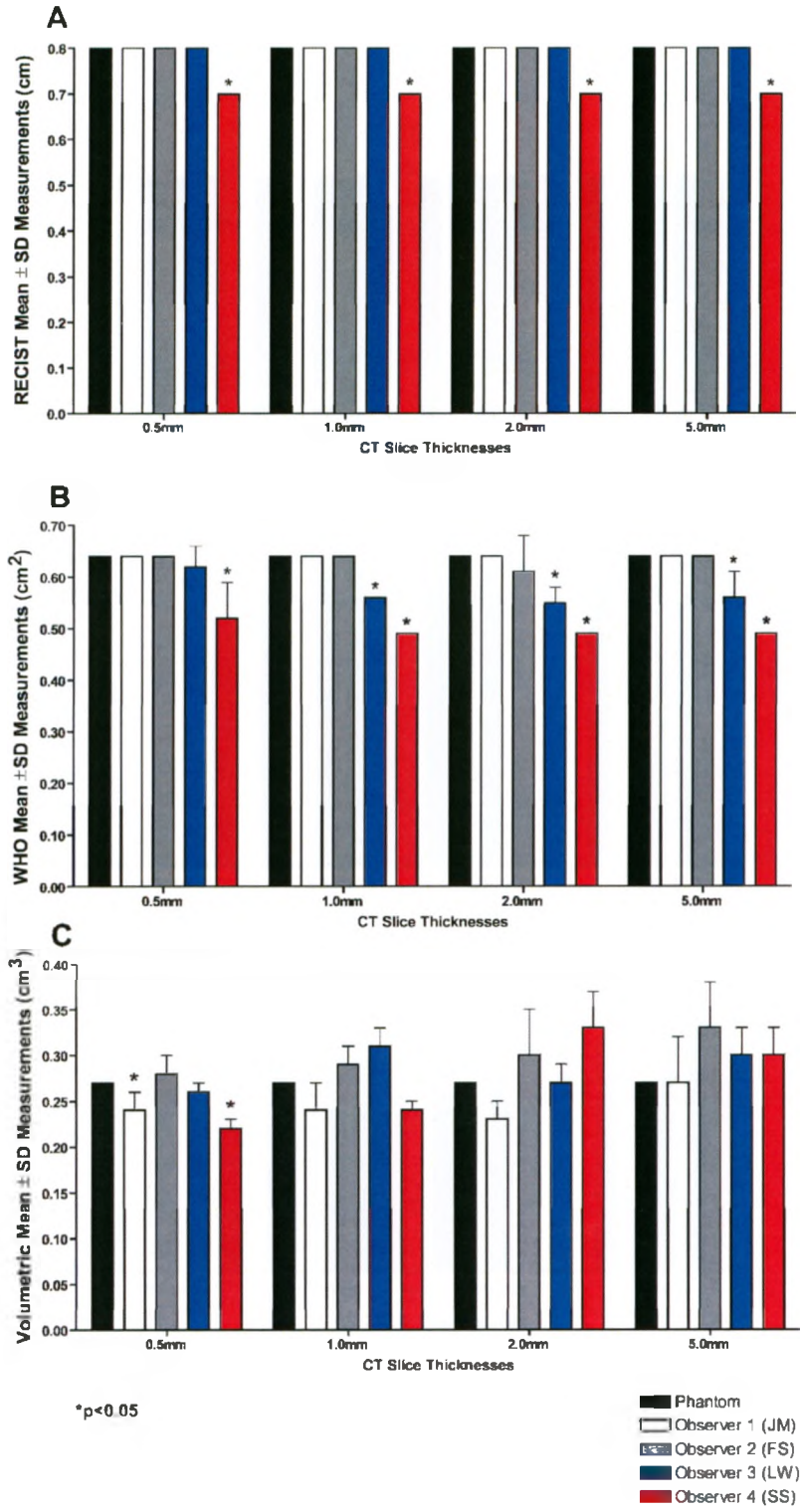


Figure C-1: Statistical Differences Between Phantom and Observer RECIST Measurements of Solid Spherical Tumours (Set 1)

(A) Observer 4 at 1.0mm slice thickness (0.9 ± 0.1cm, $p<0.001$)

(B) Observer 4 at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses (0.7 ± 0.0cm, $p<0.001$)

(C) Observer 4 at 1.0mm, 2.0mm, and 5.0mm slice thicknesses (1.1 ± 0.0cm, $p<0.001$)

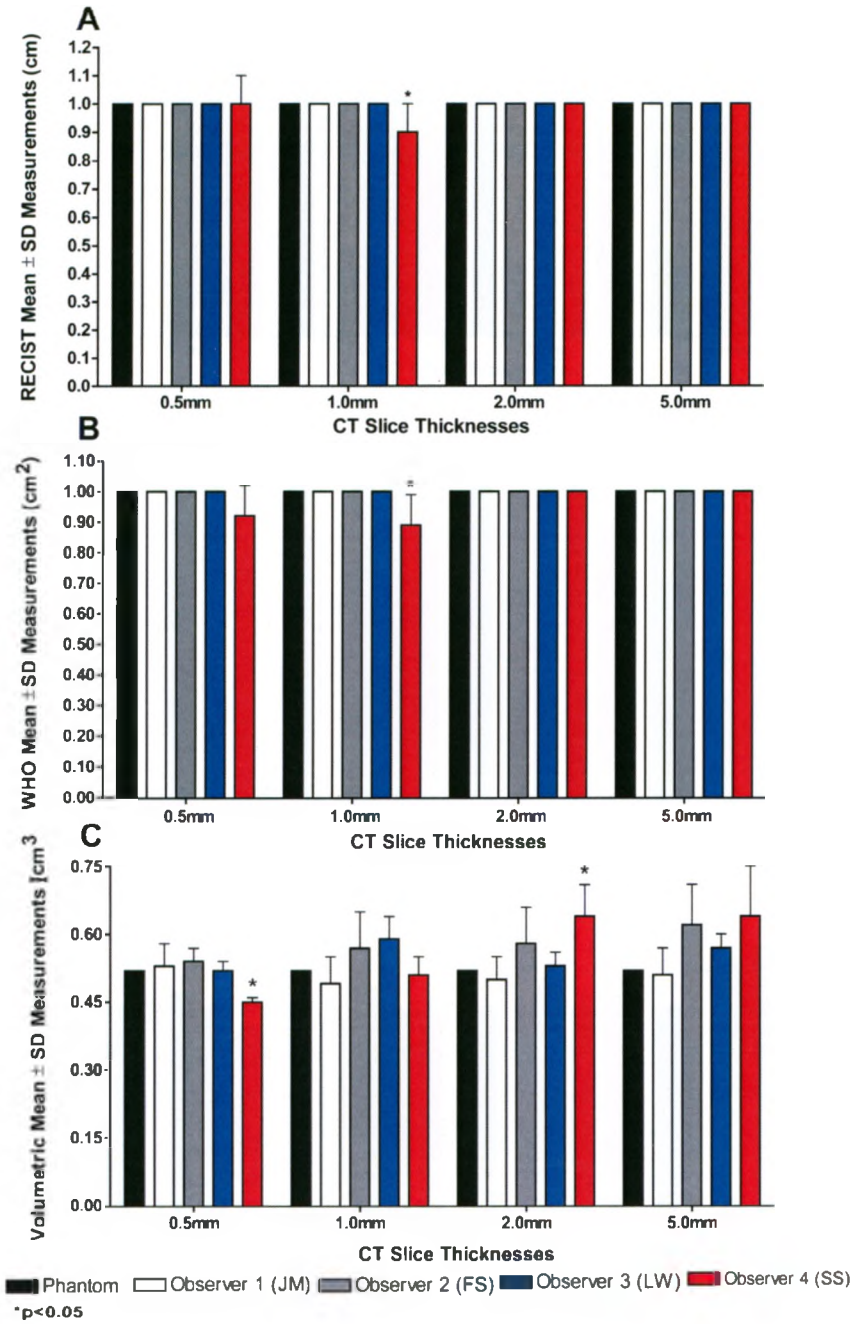


Figure C-2: Statistical Differences Between Phantom and Observer WHO Measurements of Solid Spherical Tumours (Set 1)

(A) Observer 4 at 1.0mm slice thickness ($0.89 \pm 0.10 cm^2$, p=0.007)

(B) Observer 4 at 0.5mm ($0.52 \pm 0.07 cm^2$, p<0.001); 1.0mm ($0.49 \pm 0.00 cm^2$, p<0.001); 2.0mm ($0.49 \pm 0.00 cm^2$, p<0.001); and 5.0mm slice thicknesses ($0.49 \pm 0.00 cm^2$, p<0.001). Observer 3 at 1.0mm ($0.56 \pm 0.00 cm^2$, p<0.000); 2.0mm ($0.55 \pm 0.03 cm^2$, p=0.002); and 5.0mm slice thicknesses ($0.56 \pm 0.05 cm^2$, p<0.001)

(C) Observer 4 at 1.0mm ($1.26 \pm 0.10 cm^2$, p=0.019); and 2.0mm and 5.0mm slice thicknesses ($1.21 \pm 0.00 cm^2$, p<0.001). Observer 3 at 5.0mm slice thickness ($1.37 \pm 0.07 cm^2$, p=0.007)

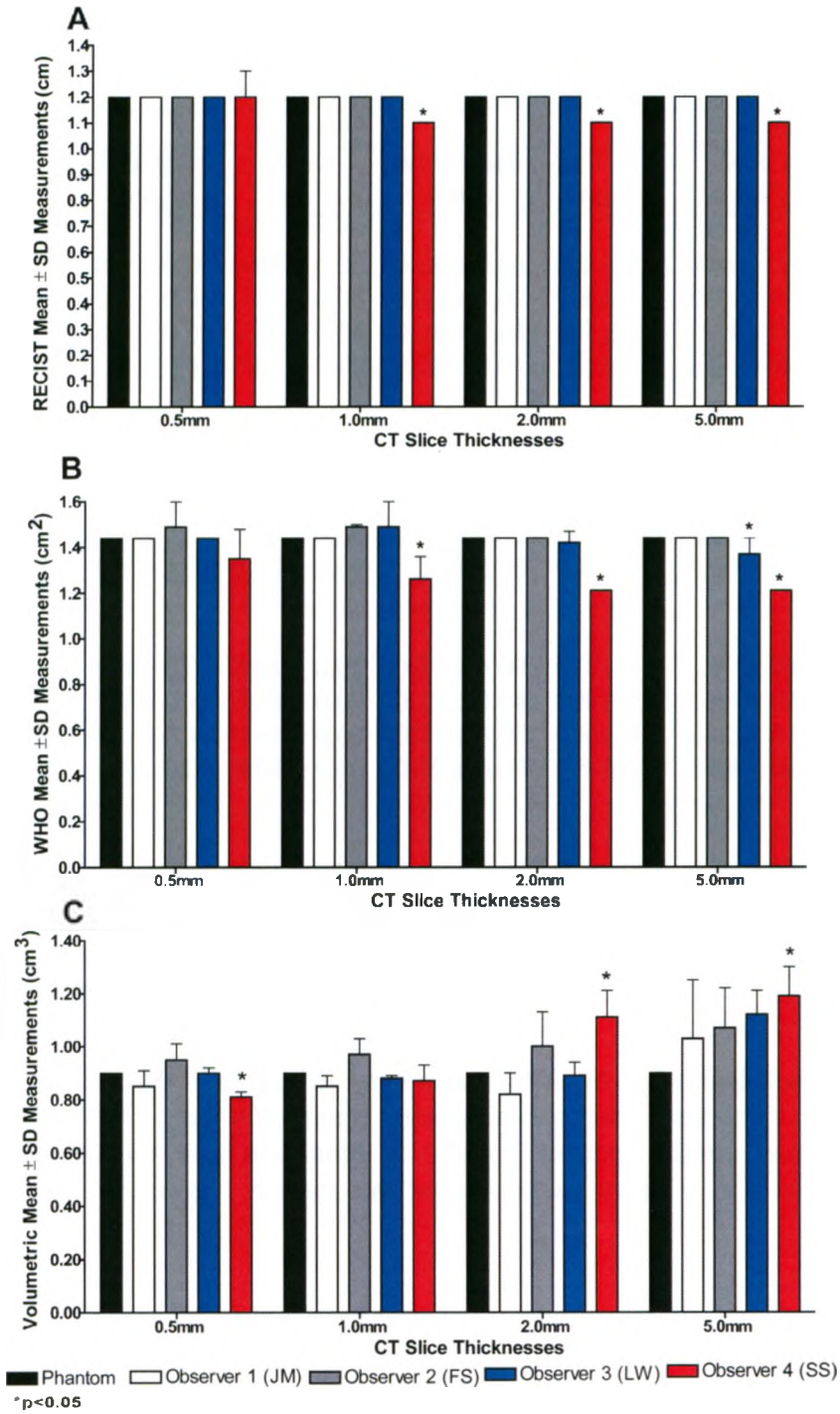


Figure C-3: Statistical Differences Between Phantom and Observer Volume Measurements of Solid Spherical Tumours (Set 1)

(A) Observer 4 at 0.5mm (0.45 ± 0.01cm^3, p=0.011); and 2.0mm slice thicknesses (0.64 ± 0.07cm^3, p=0.021)

(B) Observers 1 and 4 at 0.5mm slice thickness (0.24 ± 0.02cm^3, p=0.038), and (0.22 ± 0.01cm^3, p<0.001), respectively.

(C) Observer 4 at 0.5mm (0.89 ± 0.10cm^3, p=0.015); 2.0mm (0.89 ± 0.10cm^3, p=0.006); and 5.0mm slice thicknesses (0.89 ± 0.10cm^3, p=0.022)

C.2 Solid Spherical Tumour Phantoms (Set 2)

	Mean ± SD Observer Measurements					
	Tumour 2			Tumour 3		
	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)	RECIST (cm)	WHO (cm^2)	VOLUME (cm^3)
0.5mm						
Observer 1 (JM)	0.9 ± 0.0	0.81 ± 0.00	0.28 ± 0.03	0.9 ± 0.0	0.81 ± 0.00	0.29 ± 0.04
Observer 2 (FS)	0.8 ± 0.0	0.67 ± 0.08	0.33 ± 0.02	0.8 ± 0.0	0.64 ± 0.00	0.31 ± 0.02
Observer 3 (LW)	0.8 ± 0.0	0.66 ± 0.04	0.28 ± 0.04	0.8 ± 0.0	0.64 ± 0.00	0.24 ± 0.03
Observer 4 (SS)	0.8 ± 0.0	0.64 ± 0.00	0.28 ± 0.01	0.8 ± 0.0	0.64 ± 0.00	0.30 ± 0.02
1.0mm						
Observer 1 (JM)	0.9 ± 0.0	0.81 ± 0.00	0.28 ± 0.05	0.9 ± 0.0	0.81 ± 0.00	0.30 ± 0.02
Observer 2 (FS)	0.9 ± 0.1	0.74 ± 0.09	0.30 ± 0.03	0.8 ± 0.0	0.64 ± 0.00	0.32 ± 0.03
Observer 3 (LW)	0.8 ± 0.0	0.64 ± 0.00	0.25 ± 0.03	0.8 ± 0.0	0.64 ± 0.00	0.21 ± 0.01
Observer 4 (SS)	0.8 ± 0.0	0.64 ± 0.00	0.30 ± 0.03	0.8 ± 0.0	0.64 ± 0.00	0.31 ± 0.02
2.0mm						
Observer 1 (JM)	0.9 ± 0.0	0.81 ± 0.00	0.37 ± 0.09	0.9 ± 0.0	0.79 ± 0.04	0.30 ± 0.02
Observer 2 (FS)	0.9 ± 0.1	0.74 ± 0.09	0.29 ± 0.04	0.8 ± 0.0	0.64 ± 0.00	0.37 ± 0.05
Observer 3 (LW)	0.8 ± 0.1	0.67 ± 0.04	0.31 ± 0.02	0.8 ± 0.0	0.62 ± 0.04	0.23 ± 0.02
Observer 4 (SS)	0.8 ± 0.0	0.64 ± 0.00	0.29 ± 0.04	0.8 ± 0.0	0.64 ± 0.00	0.31 ± 0.03
5.0mm						
Observer 1 (JM)	0.9 ± 0.0	0.81 ± 0.00	0.54 ± 0.09	0.9 ± 0.0	0.76 ± 0.08	0.34 ± 0.05
Observer 2 (FS)	0.8 ± 0.0	0.67 ± 0.08	0.39 ± 0.08	0.8 ± 0.0	0.64 ± 0.00	0.31 ± 0.04
Observer 3 (LW)	0.8 ± 0.0	0.64 ± 0.00	0.38 ± 0.07	0.8 ± 0.0	0.64 ± 0.00	0.27 ± 0.05
Observer 4 (SS)	0.8 ± 0.0	0.64 ± 0.00	0.30 ± 0.03	0.7 ± 0.0	0.49 ± 0.00	0.30 ± 0.03

Table C-6: Mean 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 2)
Mean RECIST, WHO, and volumetric measurements for Tumours 1, 2, and 3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thickness by each observer. Values are expressed as mean ± SD of five repeated measurements

	Mean ± SD Observer Measurements			
RECIST Observer	0.5mm	1.0mm	2.0mm	5.0mm
Observer 1 (JM)	2.4 ± 0.0	2.4 ± 0.0	2.4 ± 0.0	2.4 ± 0.0
Observer 2 (FS)	2.2 ± 0.0	2.2 ± 0.1	2.3 ± 0.1	2.2 ± 0.1
Observer 3 (LW)	2.1 ± 0.0	2.1 ± 0.0	2.1 ± 0.1	2.1 ± 0.0
Observer 4 (SS)	2.1 ± 0.0	2.1 ± 0.0	2.1 ± 0.0	2.0 ± 0.0
Mean	**2.2 ± 0.0**	**2.2 ± 0.0**	**2.2 ± 0.0**	**2.2 ± 0.0**
WHO Observer				
Observer 1 (JM)	1.98 ± 0.00	1.98 ± 0.00	1.96 ± 0.04	1.93 ± 0.08
Observer 2 (FS)	1.67 ± 0.08	1.72 ± 0.13	1.74 ± 0.09	1.61 ± 0.12
Observer 3 (LW)	1.55 ± 0.04	1.53 ± 0.00	1.55 ± 0.04	1.53 ± 0.00
Observer 4 (SS)	1.53 ± 0.00	1.53 ± 0.00	1.53 ± 0.00	1.38 ± 0.00
Mean	**1.68 ± 0.02**	**1.69 ± 0.03**	**1.70 ± 0.03**	**1.61 ± 0.04**
Volume Observer				
Observer 1 (JM)	0.65 ± 0.06	0.65 ± 0.08	0.77 ± 0.11	1.03 ± 0.09
Observer 2 (FS)	0.74 ± 0.02	0.70 ± 0.05	0.76 ± 0.08	0.82 ± 0.12
Observer 3 (LW)	0.59 ± 0.06	0.53 ± 0.03	0.62 ± 0.04	0.75 ± 0.14
Observer 4 (SS)	0.65 ± 0.02	0.69 ± 0.05	0.69 ± 0.06	0.69 ± 0.05
Mean	**0.66 ± 0.02**	**0.64 ± 0.03**	**0.71 ± 0.04**	**0.82 ± 0.05**

Table C-7: Mean Total Tumour Burden of Solid Spherical Tumours (Set 2) per Slice Thickness Using RECIST, WHO, and Volumetric Measurements

Mean total tumour burden per slice thickness is reported. Values are expressed as mean ± SD of five repeated measurements of each measurement technique.

Intra-observer Coefficients of Variation (%)						
	Tumour 2			Tumour 3		
	RECIST (%)	WHO (%)	VOLUME (%)	RECIST (%)	WHO (%)	VOLUME (%)
0.5mm						
Observer 1 (JM)	0.0	0.00	10.71	0.0	0.00	13.08
Observer 2 (FS)	5.5	11.28	7.11	0.0	0.00	5.10
Observer 3 (LW)	5.5	5.45	13.21	0.0	0.00	10.64
Observer 4 (SS)	0.0	0.00	0.15	0.0	0.00	0.45
1.0mm						
Observer 1 (JM)	0.0	0.00	16.52	0.0	0.00	7.07
Observer 2 (FS)	6.4	12.55	10.27	0.0	0.00	8.84
Observer 3 (LW)	0.0	0.00	12.04	0.0	0.00	4.18
Observer 4 (SS)	0.0	0.00	0.85	0.0	0.00	0.77
2.0mm						
Observer 1 (JM)	0.0	0.00	25.09	0.0	5.08	6.00
Observer 2 (FS)	6.4	12.55	13.09	0.0	0.00	13.09
Observer 3 (LW)	6.5	6.52	7.90	0.0	5.73	8.13
Observer 4 (SS)	0.0	0.00	1.03	0.0	0.00	0.85
5.0mm						
Observer 1 (JM)	0.0	0.00	16.57	5.1	10.11	14.82
Observer 2 (FS)	5.5	11.28	20.83	0.0	0.00	14.17
Observer 3 (LW)	0.0	0.00	19.00	0.0	0.00	19.07
Observer 4 (SS)	0.0	0.00	0.50	0.0	0.00	0.85

Table C-8: Mean CVs of 1D, 2D, and 3D Measurements of Solid Spherical Tumours (Set 2)

Mean coefficients of variation for all RECIST, WHO, and volumetric measurements for Tumours 1, 2, and 3 are reported. Each tumour was measured at 0.5mm, 1.0mm, 2.0mm, and 5.0mm slice thicknesses by each observer. Values are expressed as

Intra- and Inter-observer Coefficients of Variation (%)				
	0.5mm	1.0mm	2.0mm	5.0mm
RECIST				
Observer 1 (JM)	0.0	0.0	0.0	1.9
Observer 2 (FS)	2.0	4.0	2.4	4.1
Observer 3 (LW)	2.1	0.0	2.6	0.0
Observer 4 (SS)	0.0	0.0	0.0	0.0
Inter-Observer	**0.7**	**1.0**	**0.9**	**1.2**
WHO				
Observer 1 (JM)	0.00	0.00	2.05	3.97
Observer 2 (FS)	4.54	7.51	5.35	7.62
Observer 3 (LW)	2.31	0.00	2.31	0.00
Observer 4 (SS)	0.00	0.00	0.00	0.00
Inter-Observer	**1.25**	**1.91**	**1.59**	**2.24**
Volume				
Observer 1 (JM)	8.63	12.51	14.89	8.90
Observer 2 (FS)	2.47	7.37	10.97	14.15
Observer 3 (LW)	10.35	5.97	6.06	18.63
Observer 4 (SS)	0.02	6.80	9.04	7.40
Inter-Observer	**3.28**	**4.34**	**5.61**	**6.38**

Table C-9: Mean CV of Total Tumour Burden of Solid Spherical Tumours (Set 2) per Slice Thickness Using RECIST, WHO, and Volumetric Measurements
Mean coefficients of variation of mean tumour size for each observer at each slice thickness are reported. Values are expressed as a percent.

Intra-Class Correlation Coefficients						
	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC (A)	ICC (C)	ICC (A)	ICC (C)	ICC (A)	ICC (C)
Observer 1 (JM)	0.994	0.994	0.990	0.990	0.751	0.820
Observer 2 (FS)	0.933	0.946	0.925	0.937	0.903	0.911
Observer 3 (LW)	0.986	0.986	0.992	0.991	0.873	0.942
Observer 4 (SS)	0.977	0.977	0.969	0.969	0.972	0.982
Inter-Observer	**0.913**	**0.969**	**0.900**	**0.964**	**0.831**	**0.882**

Table C-10: Correlation Coefficients for Each Observer (Solid Spheres Set 2)
ICC(A) and ICC(C) values are reported for each observer. Intra- and inter-observer reproducibility is assessed for RECIST, WHO, and 3D measurements.

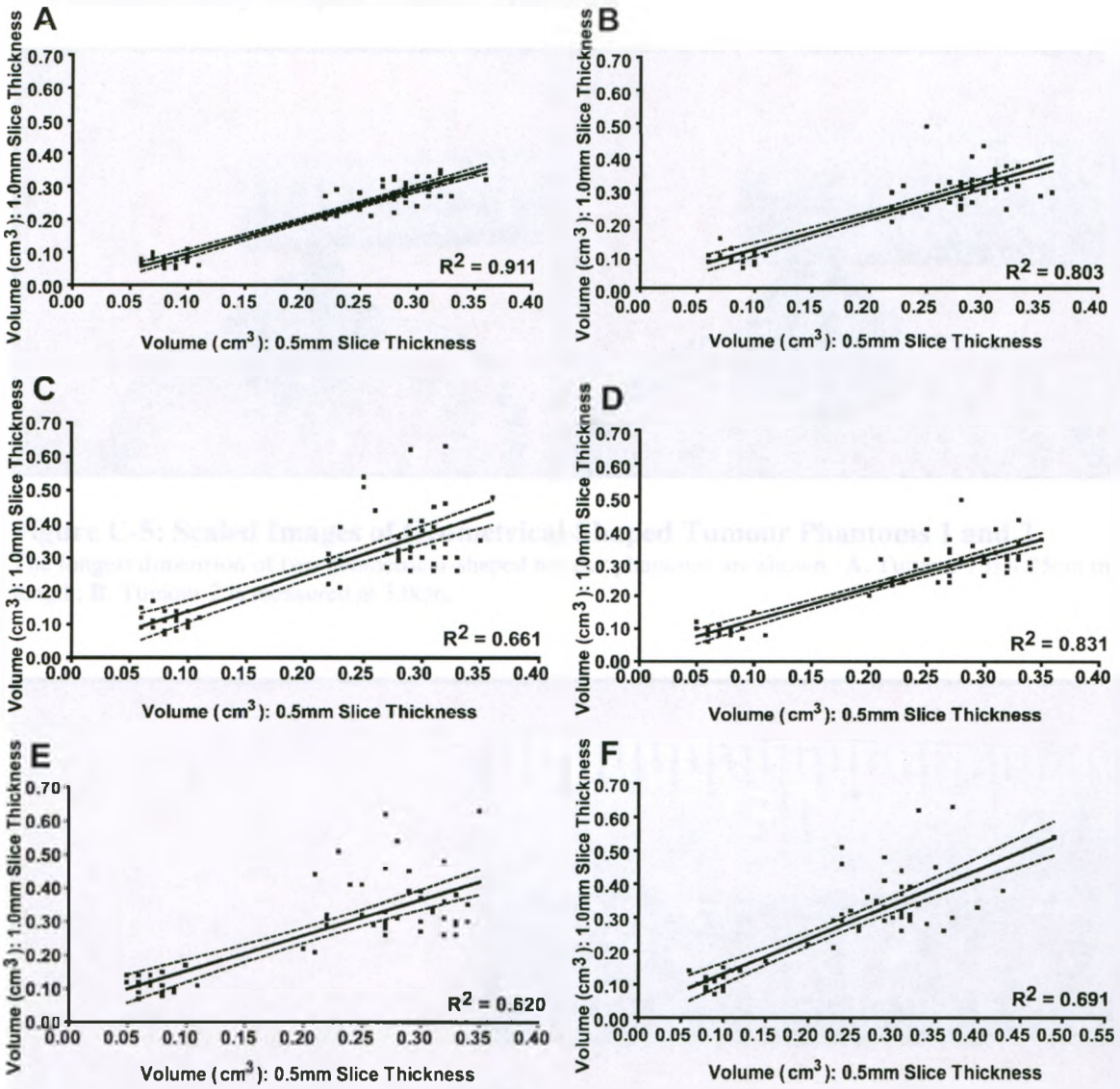


Figure C-4: Linear Regressions of Volumetric Measurements of Solid Spherical Tumours (Set 2) at Each Slice Thickness

(A) $R^2 = 0.911$; $r = 0.956$; $y = 0.990 - 0.003$; $p < 0.0001$
(B) $R^2 = 0.803$; $r = 0.896$; $y = 1.132 + 0.026$; $p < 0.0001$
(C) $R^2 = 0.661$; $r = 0.813$; $y = 0.996 + 0.019$; $p < 0.0001$
(D) $R^2 = 0.831$; $r = 0.911$; $y = 0.978 + 0.028$; $p < 0.0001$
(E) $R^2 = 0.620$; $r = 0.787$; $y = 1.057 + 0.048$; $p < 0.0001$
(F) $R^2 = 0.691$; $r = 0.831$; $y = 1.041 + 0.027$; $p < 0.0001$

C.3 Geometrically-Shaped Tumour Phantoms

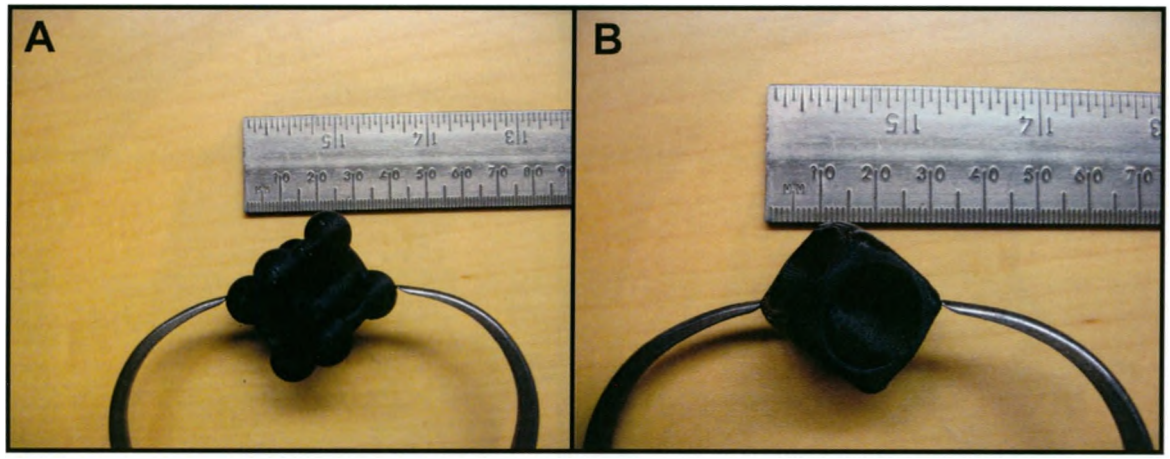


Figure C-5: Scaled Images of Geometrical-Shaped Tumour Phantoms 1 and 2
The longest dimension of two geometrical-shaped tumour phantoms are shown. **A.** Tumour 1 is 4.25cm in length; **B.** Tumour 2 is measured at 3.0cm.

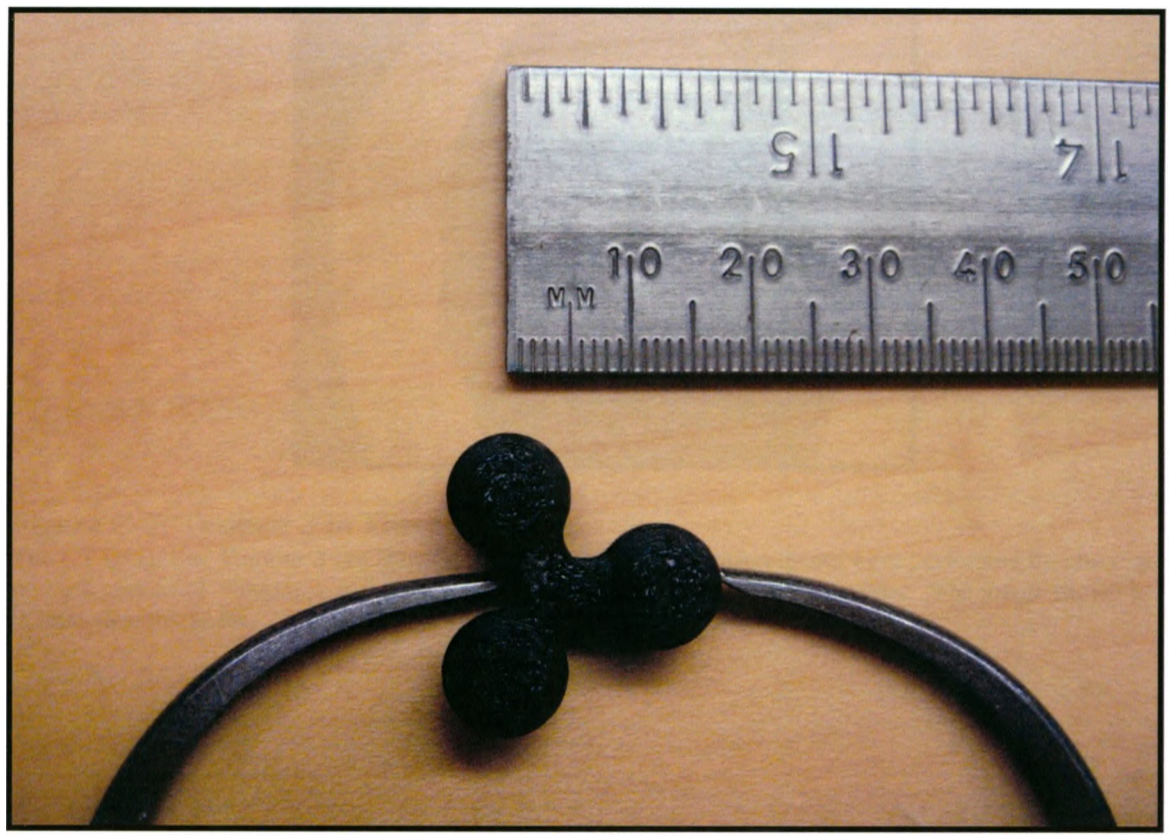


Figure C-6: Scaled Image of Geometrical-Shaped Tumour Phantom 3
The longest dimension of Tumour 3 is shown. The longest axis is measured at 1.5cm.

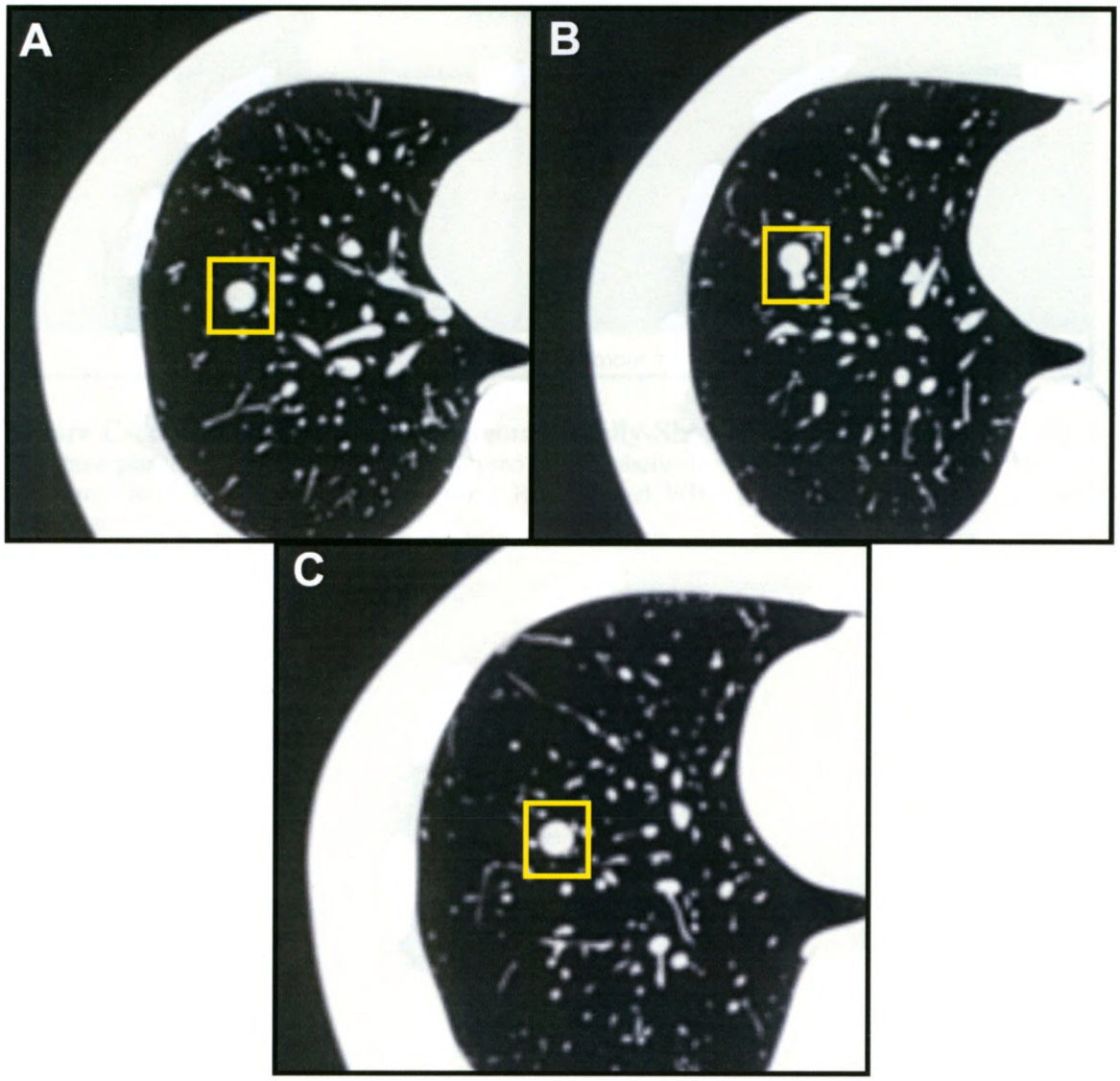


Figure C-7: Tumour 3 in Three CT Slices
The CT images of Tumour 3 are shown. RECIST and WHO measurements were performed by observers on each section of Tumour 3, and each section was treated as an individual tumour. The image shown has 0.5mm slice thickness.

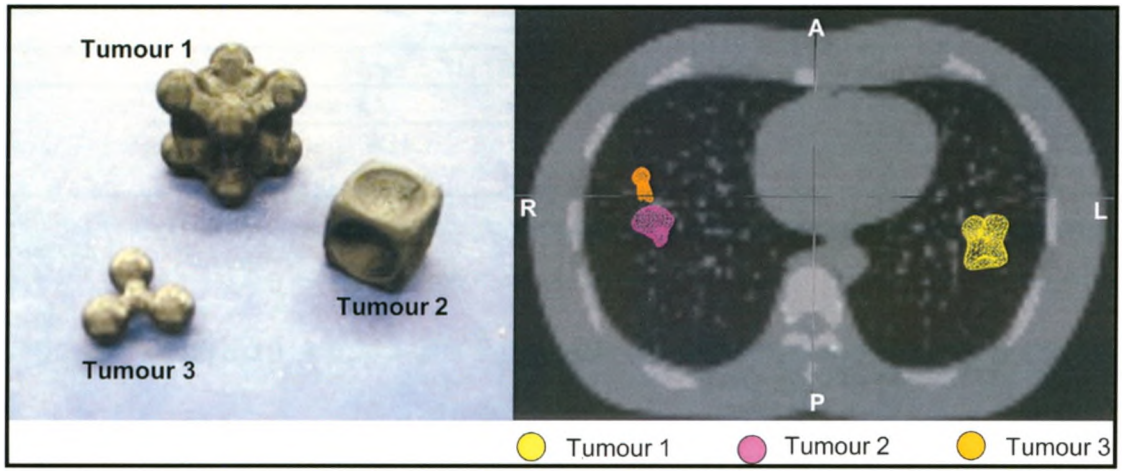


Figure C-8: Volume Rendering of Geometrically-Shaped Tumour Phantoms
The three plastic objects in this figure were used as irregularly-shaped tumour phantoms inside the chest phantom. Each tumour was measured using RECIST and WHO criteria, as well as our volumetric technique.

	Mean ± SD Observer Measurements					
	Tumour 3A		Tumour 3B		Tumour 3C	
	RECIST (cm)	WHO (cm^2)	RECIST (cm)	WHO (cm^2)	RECIST (cm)	WHO (cm^2)
0.5mm						
Observer 1 (JM)	1.1 ± 0.0	1.21 ± 0.00	1.6 ± 0.1	1.80 ± 0.06	1.1 ± 0.0	1.21 ± 0.00
Observer 2 (LW)	1.0 ± 0.0	0.96 ± 0.08	1.5 ± 0.0	1.47 ± 0.13	0.9 ± 0.1	0.81 ± 0.00
1.0mm						
Observer 1 (JM)	1.1 ± 0.0	1.21 ± 0.00	1.6 ± 0.0	1.78 ± 0.05	1.1 ± 0.0	1.19 ± 0.05
Observer 2 (LW)	1.0 ± 0.0	0.96 ± 0.08	1.5 ± 0.1	1.47 ± 0.13	0.9 ± 0.0	0.81 ± 0.00
2.0mm						
Observer 1 (JM)	1.1 ± 0.0	1.21 ± 0.00	1.6 ± 0.0	1.76 ± 0.00	1.0 ± 0.1	1.04 ± 0.05
Observer 2 (LW)	1.0 ± 0.0	0.96 ± 0.08	1.5 ± 0.1	1.41 ± 0.10	0.9 ± 0.0	0.85 ± 0.08
5.0mm						
Observer 1 (JM)	1.1 ± 0.0	1.19 ± 0.05	1.6 ± 0.1	1.74 ± 0.10	1.0 ± 0.0	1.00 ± 0.0
Observer 2 (LW)	1.0 ± 0.0	0.94 ± 0.09	0.9 ± 0.1	0.87 ± 0.08	1.0 ± 0.1	0.89 ± 0.12

Table C-11: RECIST and WHO Measurements of Geometrical-Shaped Tumour 3
Three segments of Tumour 3 appeared in three separate CT image slices. Each of these segments was treated as individual tumours for RECIST and WHO measurements. The values shown in the table are the mean ± SD of five repeated measurements.

	Mean ± SD Observer measurements			
Volume (cm^3)	0.5mm	1.0mm	2.0mm	5.0mm
Observer 1 (JM)	1.07 ± 0.06	1.20 ± 0.06	1.33 ± 0.29	2.06 ± 0.26
Observer 2 (LW)	0.94 ± 0.03	0.97 ± 0.04	0.94 ± 0.02	0.94 ± 0.03

Table C-12: Volumetric Measurements of Geometrical-Shaped Tumour 3
Tumour 3 was evaluated as a single tumour during the volumetric analyses. The values shown in the table are the mean ± SD of five repeated measurements.

Intra-Observer Coefficients of Variation (%)						
	Tumour 3A		Tumour 3B		Tumour 3C	
	RECIST (%)	WHO (%)	RECIST (%)	WHO (%)	RECIST (%)	WHO (%)
0.5mm						
Observer 1 (JM)	0.0	0.00	3.3	3.34	0.0	0.00
Observer 2 (LW)	4.6	8.83	2.9	8.57	5.8	0.00
1.0mm						
Observer 1 (JM)	0.0	0.00	2.8	2.76	0.0	4.14
Observer 2 (LW)	4.6	8.83	4.7	8.57	0.0	0.00
2.0mm						
Observer 1 (JM)	0.0	0.00	0.0	0.00	5.3	5.27
Observer 2 (LW)	4.6	8.83	4.7	7.37	4.9	10.02
5.0mm						
Observer 1 (JM)	0.0	4.14	3.3	5.67	0.0	0.00
Observer 2 (LW)	4.6	9.08	5.8	9.75	5.7	13.77

Table C-13: RECIST and WHO CV of Tumour 3 Segments
The CV of five repeated RECIST and WHO measurements of each Tumour 3 segment are reported.

Intra-Observer Coefficients of Variation (%)				
Volume (%)	0.5mm	1.0mm	2.0mm	5.0mm
Observer 1 (JM)	5.17	4.92	22.2	12.52
Observer 2 (LW)	3.23	4.28	2.34	2.79

Table C-14: CV of Volumetric Measurements of Tumour 3
The CV of five repeated 3D measurements of Tumour 3 are reported.

Intra-Class Correlation Coefficients						
	RECIST (cm)		WHO (cm^2)		Volume (cm^3)	
	ICC (A)	ICC (C)	ICC (A)	ICC (C)	ICC (A)	ICC (C)
Observer 1 (JM)	0.989	0.999	0.995	0.998	0.958	0.995
Observer 2 (LW)	0.989	0.990	0.998	0.995	0.993	0.989
All Observers	**0.989**	**0.993**	**0.986**	**0.992**	**0.925**	**0.974**

Table C-15: Correlation Coefficients for Each Observer (Geometrical-Shaped Tumour Phantoms)
ICC(A) and ICC(C) values are reported for each observer. Intra- and inter-observer reproducibility is assessed for RECIST, WHO, and 3D measurements.

C.4 Irregularly-Shaped Tumour Phantoms

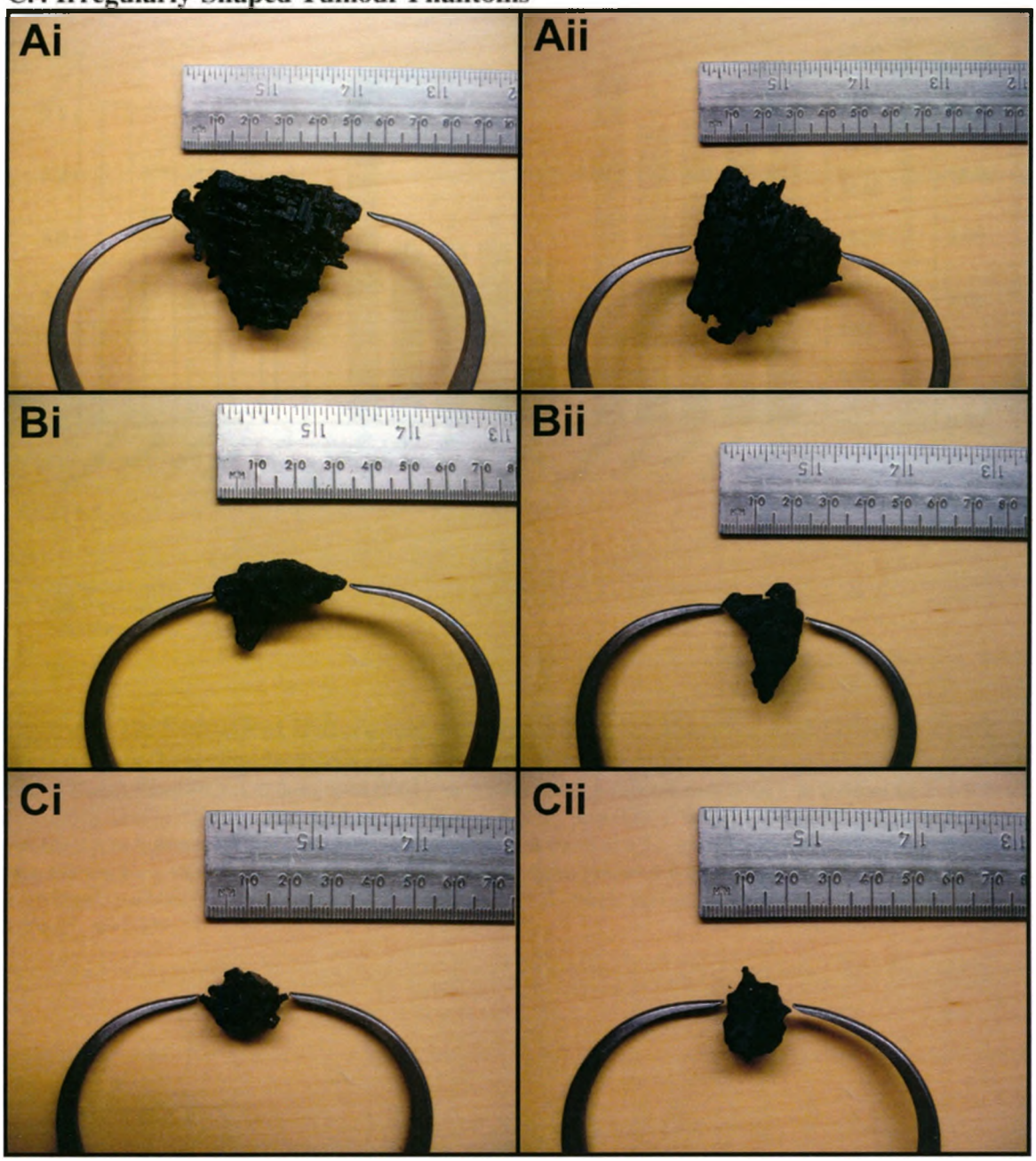


Figure C-9: Scaled Images of Irregular-Shaped Tumour Phantoms
The three irregular tumour phantoms are shown. The longest axis of the large, medium, and small sized phantoms are shown in Ai, Bi, and Ci, respectively. The longest dimension perpendicular to the longest axis of the large, medium, and small phantoms are shown in Aii, Bii, and Cii, respectively.

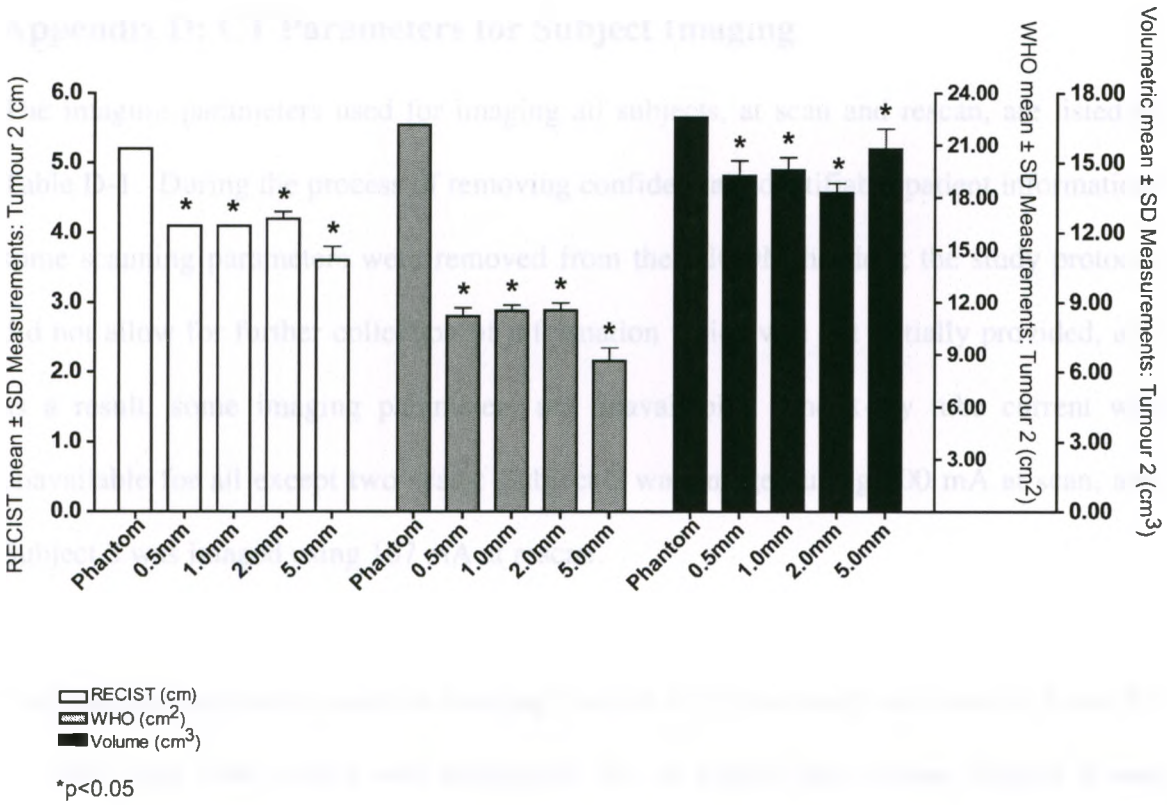


Figure C-10: Statistical Differences between a Single Observer and Ground Truth for Tumour 2 (Irregularly-Shaped Tumour Phantoms)

RECIST at 0.5mm (4.1 ± 0.0cm, $p<0.0001$), at 1.0mm (4.1 ± 0.0cm, $p<0.0001$), at 2.0mm (4.1 ± 0.0cm, $p<0.0001$), and at 5.0mm slice thicknesses (3.6 ± 0.2cm, $p<0.0001$); WHO at 0.5mm (11.23 ± 0.47cm^2, $p<0.0001$), at 1.0mm (11.54 ± 0.32cm^2, $p<0.0001$), at 2.0mm (11.57 ± 0.42cm^2, $p<0.0001$), and at 5.0mm slice thicknesses (8.64 ± 0.77cm^2, $p<0.0001$); 3D at 0.5mm (14.44 ± 0.65cm^3, $p<0.0001$), at 1.0mm (14.69 ± 0.54cm^3, $p<0.0001$), at 2.0mm (13.71 ± 0.41cm^3, $p<0.0001$), and at 5.0mm slice thicknesses (15.58 ± 0.86cm^3, $p=0.009$).

Appendix D: CT Parameters for Subject Imaging

The imaging parameters used for imaging all subjects, at scan and rescan, are listed in Table D-1. During the process of removing confidential, identifiable patient information, some scanning parameters were removed from the DICOM headers; the study protocol did not allow for further collection of information which was not initially provided, and as a result, some imaging parameters are unavailable. The x-ray tube current was unavailable for all except two scans: Subject 5 was imaged using 300 mA at scan, and Subject 3 was imaged using 167 mA at rescan.

The imaging parameters used for imaging Subject 4 longitudinally are listed in Table D-2. The x-ray tube current was unavailable for all except three scans: Subject 4 was imaged using 280 mA, 250 mA, and 350 mA on 11-Jan-2005, 2-Mar-2006, and 4-Dec-2007, respectively.

Subject #	Scan Date	CT Scanner	# of images	Slice Thickness (mm)	x-pixel length (mm)	y-pixel length (mm)	FOV(x) (mm)	FOV(y) (mm)	FOV(z) (mm)	Contrast-Bolus Agent	Contrast-Bolus Route
1	28-Oct-2006	GE LightSpeed VCT	104	2.5	0.587891	0.587891	301	301	260	YES	Oral & IV
2	6-Jan-2007	GE LightSpeed VCT	136	5	0.859375	0.859375	440	440	680	YES	Oral
3	19-Dec-2006	GE LightSpeed VCT	83	5	0.647662	0.647662	332	332	415		
4	4-May-2007	GE LightSpeed VCT	68	5	0.706368	0.706368	362	362	340	YES	IV
5	19-Jan-2007	GE LightSpeed Ultra	90	5	0.722656	0.722656	370	370	450		
6	15-Feb-2007	GE LightSpeed VCT	73	5	0.637269	0.637269	326	326	365		
7	24-Mar-2006	GE LightSpeed VCT	132	5	0.673828	0.673828	345	345	660	YES	Oral & IV
	Rescan Date										
2	3-Nov-2007	GE LightSpeed VCT	122	5	0.820312	0.820312	420	420	610		
3	7-Sep-2007	GE LightSpeed VCT	75	5	0.726916	0.726916	372	372	375		
4	2-Feb-2007	GE LightSpeed VCT	58	5	0.703125	0.703125	360	360	290	YES	Oral & IV
5	9-Mar-2007	GE LightSpeed VCT	87	5	0.746408	0.746408	382	382	435	YES	Oral & IV
6	1-Nov-2006	GE LightSpeed VCT	72	5	0.703125	0.703125	360	360	360	YES	Oral & IV

Table D-1: CT Scanning Parameters of Subjects at Scan-Rescan
All CT imaging was performed at Victoria Hospital (London, Ontario).

Scan Date	# of images	Slice Thickness (mm)	x-pixel length (mm)	y-pixel length (mm)	FOV(x) (mm)	FOV(y) (mm)	FOV(z) (mm)	Contrast-Bolus Agent	Contrast-Bolus Route
11-Jan-2005	138	5	0.722656	0.722656	370	370	690	YES	Oral & IV
9-Sep-2005	127	5	0.703125	0.703125	360	360	635	YES	IV
22-Dec-2005	128	5	0.703125	0.703125	360	360	640	YES	IV
2-Mar-2006	164	5	0.703125	0.703125	360	360	820	YES	Oral & IV
23-May-2006	58	5	0.703125	0.703125	360	360	290	YES	Oral & IV
21-Aug-2006	72	5	0.688009	0.688009	352	352	360		
14-Nov-2006	71	5	0.643030	0.643030	329	329	355		
2-Feb-2007	68	5	0.706368	0.706368	362	362	340	YES	IV
4-May-2007	67	5	0.764040	0.764040	391	391	335		Oral
30-May-2007	58	5	0.707102	0.707102	362	362	290	YES	Oral & IV
17-Jul-2007	77	5	0.697046	0.697046	357	357	385	YES	Oral & IV
11-Sep-2007	64	5	0.691653	0.691653	354	354	320		
4-Dec-2007	75	5	0.664062	0.664062	340	340	375	YES	IV

Table D-2: CT Scanning Parameters of Subject 4
All CT imaging was performed at Victoria Hospital (London, Ontario).

Appendix E: Supplemental Data from Patient Analyses (Chapter 3)

E.1 Cross-Sectional Analyses: RECIST and WHO

		Mean ± SD Observer Measurements	
Tumour #	**Radiological locations**	**RECIST (cm)**	**WHO (cm^2)**
1	L lower lobe superior segment	2.6 ± 2.8	7.0 ± 13.4
2	R upper lobe basal segment	2.5 ± 0.2	6.0 ± 0.9
3	L lower lobe superior segment subpleural	3.9 ± 0.2	13.1 ± 0.9
4	R lower paratracheal (mediastinum)	4.2 ± 0.1	15.8 ± 1.1
5	R middle lobe lateral segment subpleural	3.8 ± 0.2	12.3 ± 1.2
6	R middle lobe medial segment pericardial	2.0 ± 0.1	3.7 ± 0.4
7	R lower lobe medial basal segment pericardial	1.4 ± 0.3	1.8 ± 0.8
8	L lower lobe subpleural/perifissural	2.8 ± 0.2	6.3 ± 0.6
9	R lower lobe posterior subpleural	1.5 ± 0.5	1.9 ± 0.8
10	R lower lobe posterior subpleural	2.2 ± 0.1	4.2 ± 0.5
11	L upper lobe subpleural	1.4 ± 0.2	1.5 ± 0.4
12	L upper lobe subpleural	1.2 ± 0.3	1.1 ± 0.4
13	R lower lobe superior basal segment subpleural	1.7 ± 0.2	2.4 ± 0.5
14	R lower lobe posterior basal segment near diaphragm	2.6 ± 0.2	6.3 ± 0.7
15	R lower lobe medial basal segment near diaphragm	2.2 ± 0.7	2.7 ± 1.0
16	L lower lobe posteromedial	2.1 ± 0.2	3.7 ± 0.7
17	R lower lobe lateral basal segment subpleural	1.7 ± 0.2	2.3 ± 0.4
18	Lingular (L upper lobe)	2.5 ± 0.8	4.9 ± 1.6
19	R lower lobe lateral basal segment subpleural	2.0 ± 0.4	3.7 ± 1.1
20	R upper lobe perifissural	1.8 ± 0.4	2.4 ± 0.7
21	R lower lobe posterior basal segment subpleural	1.3 ± 0.2	1.4 ± 0.4
22	R lower lobe posterior basal segment subpleural	1.1 ± 0.3	1.1 ± 0.6
23	R lower lobe medial basal segment near diaphragm	1.4 ± 0.2	1.7 ± 0.6
24	R lower lobe medial basal segment near diaphragm	1.5 ± 0.4	1.7 ± 0.6
25	R lower lobe superior basal segment	1.6 ± 0.3	2.0 ± 0.5
26	Lingular (L upper lobe)	1.7 ± 0.2	2.2 ± 0.4
27	L lower lobe posterior basal segment	1.6 ± 0.2	2.0 ± 0.5
28	R lower lobe lateral basal segment	1.1 ± 0.2	1.0 ± 0.3
29	R lower lobe lateral basal segment subpleural	1.7 ± 0.1	2.5 ± 0.4
	Mean	**2.0 ± 0.1**	**4.1 ± 0.5**

Table E-1: RECIST and WHO Means for All Observers
The mean ± SD for the RECIST and WHO measurements of seven observers are displayed. Each observer performed five repeated measurement of each tumour.

Figure E-1: Images of RECIST and WHO Measurements

A. Tumour 2 is shown with RECIST and WHO measurements; and B. Tumour 13 is shown with RECIST and WHO measurements.

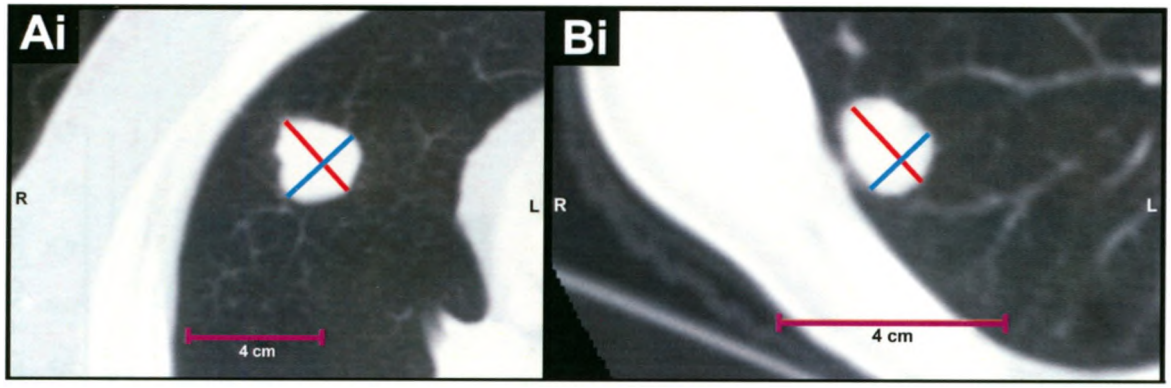
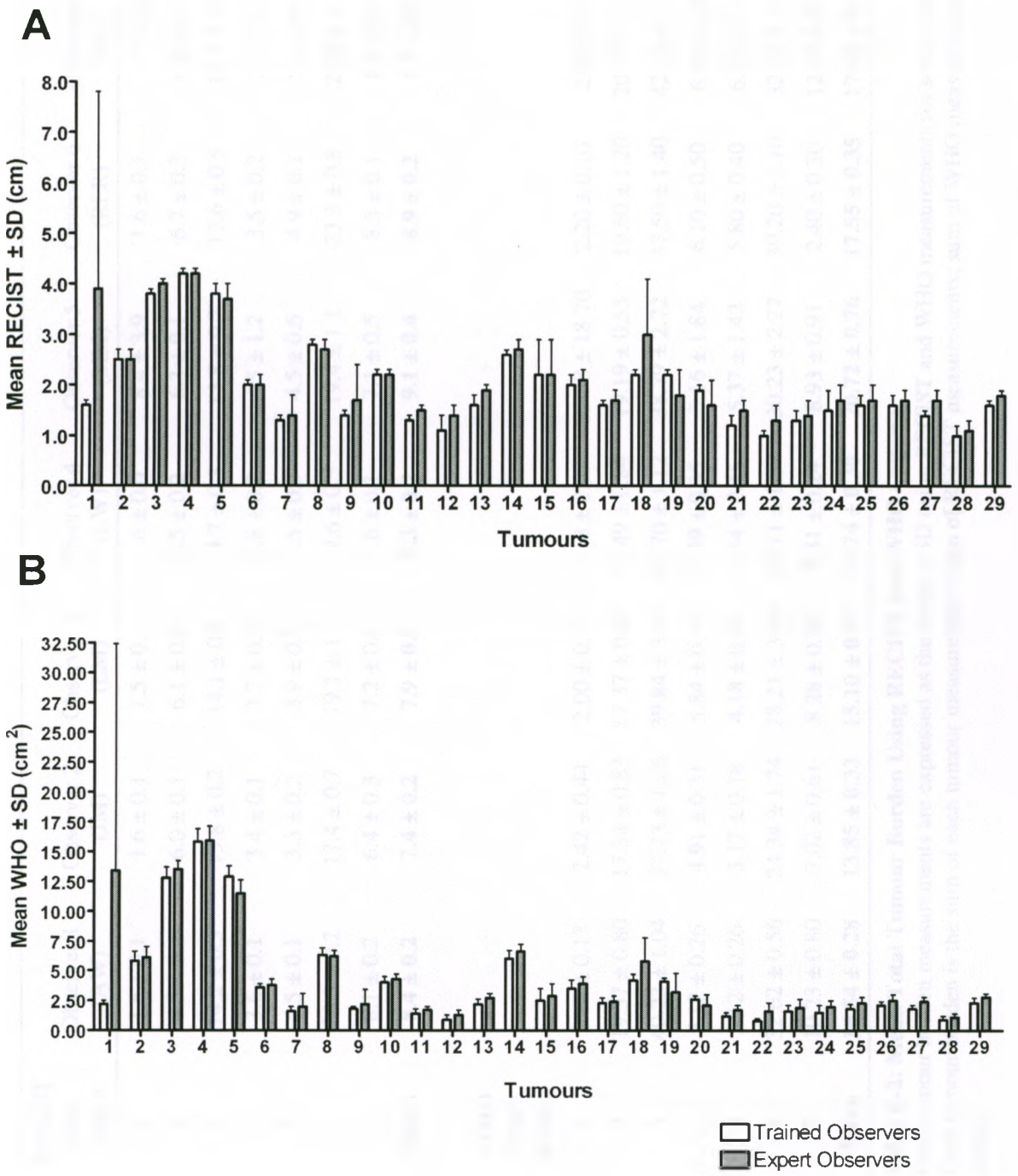


Figure E-2: Tumour measurements

RECIST (cm) and WHO (cm^2) measurements for all 29 tumours are displayed for A) trained observers (n=4), and B) expert observers (n=3). Measurements are expressed as the mean of five repeated measurements ± standard deviation



RECIST (cm) Subject	Observer 1 (AW)	Observer 2 (JM)	Observer 3 (LM)	Observer 4 (LW)	Observer 5 (MM)	Observer 6 (RER)	Observer 7 (EO)
1	1.6 ± 0.1	1.6 ± 0.1	1.5 ± 0.1	1.6 ± 0.0	8.4 ± 3.9	1.6 ± 0.1	1.7 ± 0.2
2	6.5 ± 0.2	6.0 ± 0.1	6.1 ± 0.2	6.5 ± 0.2	6.3 ± 0.1	6.7 ± 0.3	6.8 ± 0.2
3	14.2 ± 0.3	13.8 ± 0.2	14.1 ± 0.6	14.7 ± 0.3	13.8 ± 0.7	13.6 ± 0.5	14.7 ± 0.2
4	3.6 ± 0.1	3.4 ± 0.1	3.7 ± 0.2	3.8 ± 0.2	4.5 ± 1.2	3.5 ± 0.2	3.8 ± 0.1
5	4.5 ± 0.1	3.3 ± 0.2	3.9 ± 0.5	4.5 ± 0.3	4.5 ± 0.6	4.9 ± 0.1	5.0 ± 0.1
6	20.3 ± 0.2	17.4 ± 0.7	19.1 ± 1.7	19.6 ± 0.5	19.4 ± 1.1	23.3 ± 0.8	20.5 ± 0.1
7	8.1 ± 0.2	6.4 ± 0.3	7.2 ± 0.4	7.6 ± 0.1	7.1 ± 0.5	8.5 ± 0.1	8.5 ± 0.1
Mean	**8.4 ± 0.2**	**7.4 ± 0.2**	**7.9 ± 0.3**	**8.3 ± 0.2**	**9.1 ± 0.4**	**8.9 ± 0.2**	**8.7 ± 0.1**
WHO (cm^2) Subject							
1	2.18 ± 0.13	2.42 ± 0.44	2.00 ± 0.17	2.18 ± 0.09	35.38 ± 18.70	2.20 ± 0.10	2.50 ± 0.50
2	19.37 ± 0.80	17.34 ± 0.83	17.37 ± 0.84	20.49 ± 0.84	19.19 ± 0.55	19.60 ± 1.20	20.10 ± 0.70
3	40.23 ± 1.04	37.73 ± 1.26	39.84 ± 3.08	42.70 ± 1.12	38.79 ± 2.72	37.50 ± 1.40	42.10 ± 1.20
4	5.12 ± 0.26	4.91 ± 0.31	5.84 ± 0.70	6.39 ± 0.35	7.16 ± 1.84	6.10 ± 0.50	6.30 ± 0.40
5	5.12 ± 0.26	3.17 ± 0.18	4.18 ± 0.96	5.54 ± 0.45	5.37 ± 1.43	5.80 ± 0.40	6.10 ± 0.30
6	32.62 ± 0.56	24.34 ± 1.74	28.21 ± 3.99	30.11 ± 1.27	30.23 ± 2.27	39.20 ± 2.10	32.70 ± 0.60
7	10.23 ± 0.80	7.02 ± 0.61	8.28 ± 0.92	9.81 ± 0.24	8.93 ± 0.91	2.40 ± 0.30	12.00 ± 0.30
Mean	**16.54 ± 0.28**	**13.85 ± 0.33**	**15.10 ± 0.47**	**16.74 ± 0.30**	**20.72 ± 0.76**	**17.55 ± 0.35**	**17.42 ± 0.29**

Table E-2: Mean Total Tumour Burden Using RECIST and WHO Criteria
Total tumour burden measurements are expressed as the mean ± SD using RECIST and WHO measurements for seven subjects. Total tumour burden is the sum of each tumour measurement (sum of RECIST measurements; sum of WHO measurements) per subject.

Coefficients of Variation (%)				
Tumour #	Tumour Location	Trained Observers	Expert Observers	All Observers
1	L lower lobe superior segment	12.46	141.60	191.67
2	R upper lobe basal segment	14.55	14.98	14.78
3	L lower lobe superior segment subpleural	6.86	4.85	6.51
4	R lower paratracheal (mediastinum)	7.02	7.39	7.09
5	R middle lobe lateral segment subpleural	7.36	9.94	10.00
6	R middle lobe medial segment pericardial	8.12	12.83	10.58
7	R lower lobe medial basal segment pericardial	22.70	55.91	45.33
8	L lower lobe subpleural/perifissural	10.34	8.62	9.52
9	R lower lobe posterior subpleural	13.06	55.50	42.41
10	R lower lobe posterior subpleural	13.02	9.69	11.96
11	L upper lobe subpleural	28.02	15.81	24.76
12	L upper lobe subpleural	38.72	33.59	39.77
13	R lower lobe superior basal segment subpleural	18.86	13.46	19.81
14	R lower lobe posterior basal segment near diaphragm	11.82	8.73	11.44
15	R lower lobe medial basal segment near diaphragm	40.24	35.39	38.07
16	L lower lobe posteromedial	19.57	14.88	17.94
17	R lower lobe lateral basal segment subpleural	16.23	22.37	19.21
18	Lingular (L upper lobe)	12.30	34.95	32.75
19	R lower lobe lateral basal segment subpleural	7.34	49.11	30.32
20	R upper lobe perifissural	10.00	45.24	28.82
21	R lower lobe posterior basal segment subpleural	22.99	26.07	30.71
22	R lower lobe posterior basal segment subpleural	28.59	44.88	54.95
23	R lower lobe medial basal segment near diaphragm	34.31	38.72	37.22
24	R lower lobe medial basal segment near diaphragm	42.33	25.21	36.99
25	R lower lobe superior basal segment	22.79	22.23	25.32
26	Lingular (L upper lobe)	15.68	18.29	19.09
27	L lower lobe posterior basal segment	16.14	17.53	23.50
28	R lower lobe lateral basal segment	32.38	26.75	32.08
29	R lower lobe lateral basal segment subpleural	15.76	11.51	16.28
	Inter-Observer	**2.73**	**14.98**	**11.77**

Table E-3: CV for RECIST Measurements

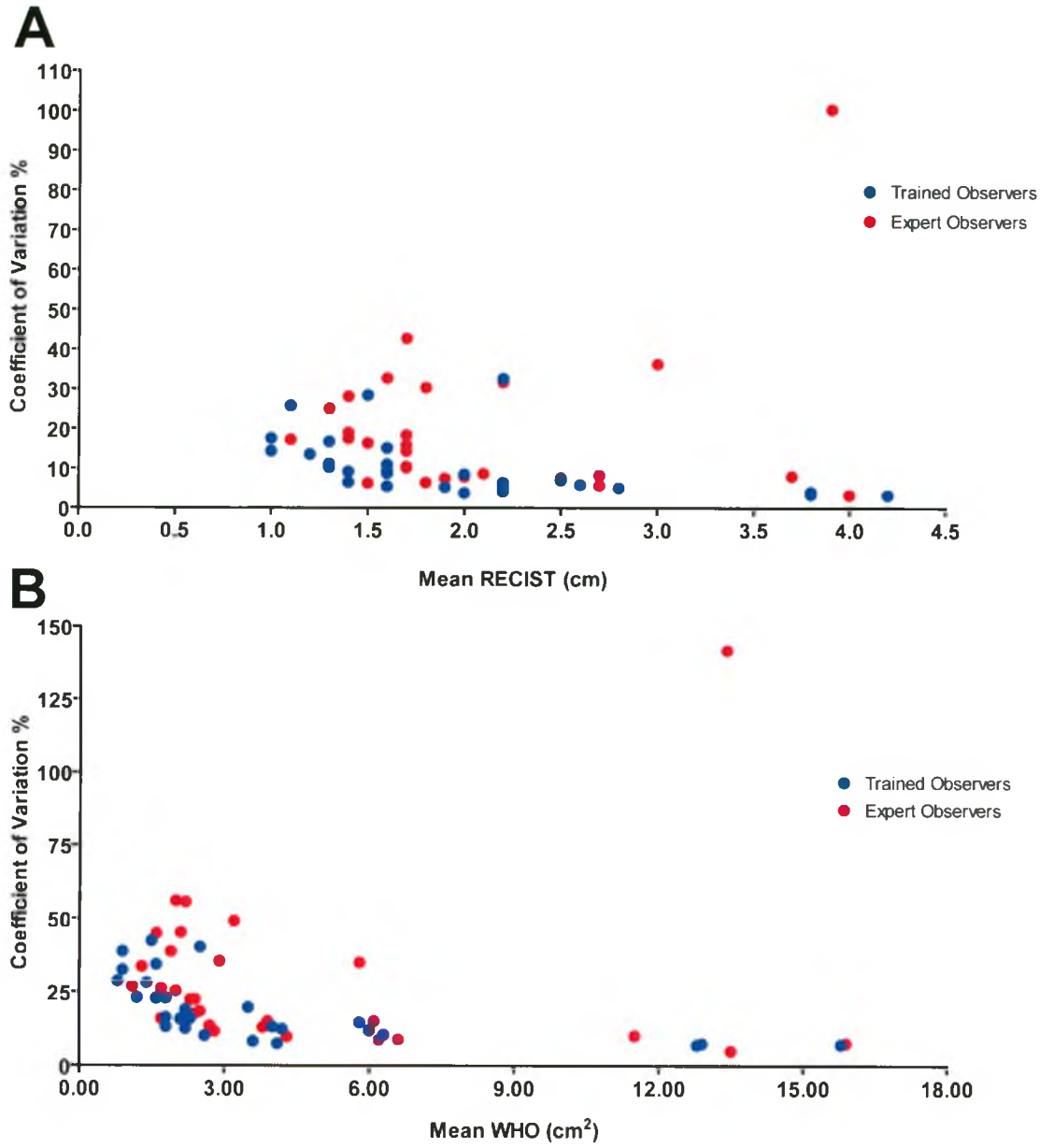
Mean coefficient of variation of RECIST measurements for each tumour is expressed for trained (n=4), expert (n=3), and all observers.

Subject #	Tumour Location	Coefficients of Variation (%) Trained Observers	Expert Observers	All Observers
1	L lower lobe superior segment	12.46	141.60	191.67
2	R upper lobe basal segment	14.55	14.98	14.78
3	L lower lobe superior segment subpleural	6.86	4.85	6.51
4	R lower paratracheal (mediastinum)	7.02	7.39	7.09
5	R middle lobe lateral segment subpleural	7.36	9.94	10.00
6	R middle lobe medial segment pericardial	8.12	12.83	10.58
7	R lower lobe medial basal segment pericardial	22.70	55.91	45.33
8	L lower lobe subpleural/perifissural	10.34	8.62	9.52
9	R lower lobe posterior subpleural	13.06	55.50	42.41
10	R lower lobe posterior subpleural	13.02	9.69	11.96
11	L upper lobe subpleural	28.02	15.81	24.76
12	L upper lobe subpleural	38.72	33.59	39.77
13	R lower lobe superior basal segment subpleural	18.86	13.46	19.81
14	R lower lobe posterior basal segment near diaphragm	11.82	8.73	11.44
15	R lower lobe medial basal segment near diaphragm	40.24	35.39	38.07
16	L lower lobe posteromedial	19.57	14.88	17.94
17	R lower lobe lateral basal segment subpleural	16.23	22.37	19.21
18	Lingular (L upper lobe)	12.30	34.95	32.75
19	R lower lobe lateral basal segment subpleural	7.34	49.11	30.32
20	R upper lobe perifissural	10.00	45.24	28.82
21	R lower lobe posterior basal segment subpleural	22.99	26.07	30.71
22	R lower lobe posterior basal segment subpleural	28.59	44.88	54.95
23	R lower lobe medial basal segment near diaphragm	34.31	38.72	37.22
24	R lower lobe medial basal segment near diaphragm	42.33	25.21	36.99
25	R lower lobe superior basal segment	22.79	22.23	25.32
26	Lingular (L upper lobe)	15.68	18.29	19.09
27	L lower lobe posterior basal segment	16.14	17.53	23.50
28	R lower lobe lateral basal segment	32.38	26.75	32.08
29	R lower lobe lateral basal segment subpleural	15.76	11.51	16.28
	Inter-observer	**2.73**	**14.98**	**11.77**

Table E-4: CV for WHO Measurements

Mean coefficient of variation of WHO measurements for each tumour is expressed for trained (n=4), expert (n=3), and all observers.

Figure E-3: Coefficient of Variation Maps for RECIST and WHO Measurements
The coefficients of variation for trained and expert observers are shown. A) CV vs. mean RECIST measurements show how CV changes as the longest axis of the tumour increases. B) CV vs. mean WHO measurements show how CV changes as the bi-dimensional measurement increases.



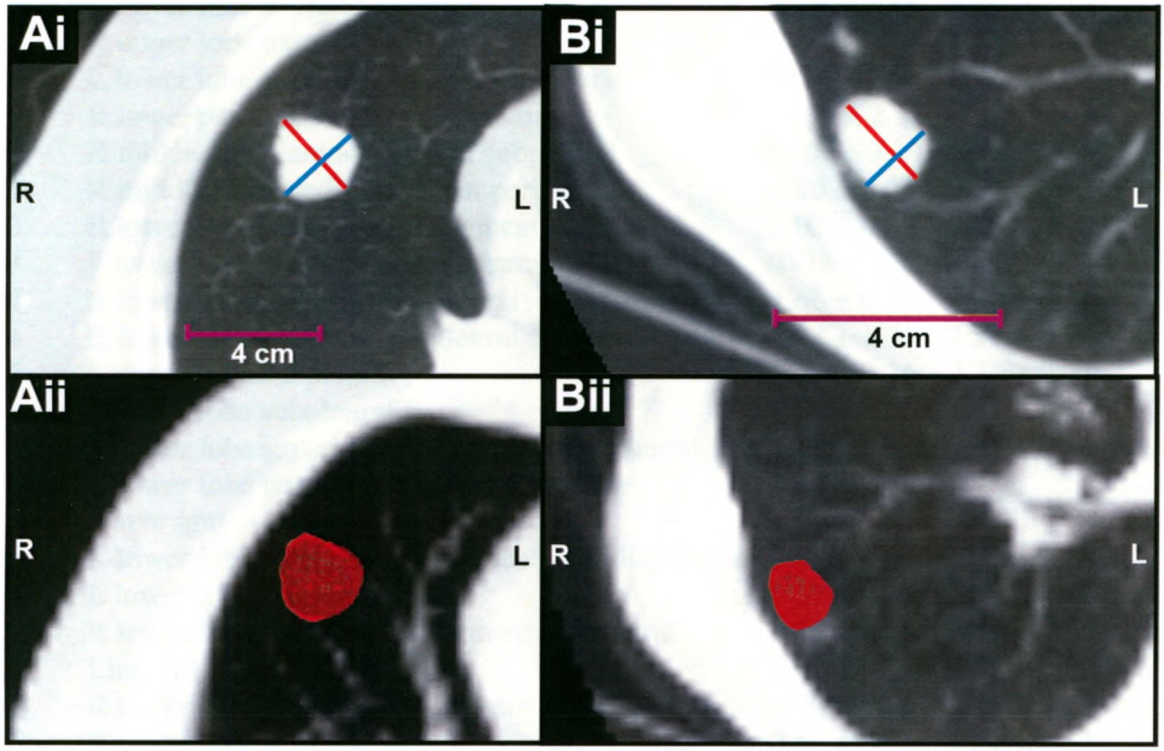
	Intra-Observer Coefficients of Variation (%)						
RECIST Subject	Observer 1 (AW)	Observer 2 (JM)	Observer 3 (LM)	Observer 4 (LW)	Observer 5 (MM)	Observer 6 (RER)	Observer 7 (EO)
1	5.7	8.8	3.6	0.0	46.1	8.0	10.9
2	2.9	1.5	3.4	2.6	0.9	4.2	2.8
3	2.2	1.3	4.0	2.2	5.1	3.9	1.4
4	1.5	2.5	6.6	5.9	25.8	4.7	2.2
5	3.3	4.5	12.3	6.1	13.7	2.0	3.0
6	0.7	3.8	8.8	2.7	5.5	3.4	0.3
7	2.4	5.1	6.0	0.7	6.8	1.3	1.0
Inter-Observer	**2.4**	**2.7**	**3.5**	**2.4**	**4.4**	**2.2**	**1.1**
WHO Subject							
1	5.73	18.40	8.23	4.03	52.86	5.90	19.70
2	4.13	4.80	4.82	4.08	2.86	5.90	3.50
3	2.58	3.35	7.73	2.62	7.01	3.80	2.80
4	5.06	6.41	12.04	5.42	25.66	8.30	6.40
5	5.06	5.57	23.09	8.13	26.67	7.50	5.70
6	1.73	7.14	14.14	4.21	7.52	5.40	1.90
7	7.80	8.70	11.06	2.49	10.16	2.50	2.90
Inter-Observer	**1.69**	**2.38**	**3.11**	**1.79**	**3.67**	**1.99**	**1.66**

Table E-5: CV for Reproducibility of RECIST and WHO Measurements
Coefficients of variation (CV) for mean RECIST and WHO measurements of the total tumour burden per subject are reported. CV was calculated by using the mean ± SD of the total tumour burden per subject.

E.2 Cross-Sectional Analyses: RECIST, WHO, Volumetry

Figure E-4: RECIST, WHO and Volumetric Measurements

A. Tumour 2; and B. Tumour 13 are shown. Ai and Bi show RECIST and WHO measurements of each lung tumour, as in Figure E-3. Aii and Bii show volume rendering of the same respective tumours.



Mean ± SD Observer Measurements			
		Expert Observers	
Subject #	Radiological locations	Observer 1 (EO)	Observer 2 (RER)
1	L lower lobe superior segment	3.49 ± 0.46	4.40 ± 1.16
2	R upper lobe basal segment	8.30 ± 1.61	10.36 ± 0.64
2	L lower lobe superior segment subpleural	30.38 ± 2.92	34.12 ± 2.70
3	R lower paratracheal (mediastinum)	40.84 ± 2.91	46.01 ± 3.57
3	R middle lobe lateral segment subpleural	37.09 ± 3.17	39.28 ± 0.95
3	R middle lobe medial segment pericardial	5.96 ± 0.44	5.84 ± 0.13
3	R lower lobe medial basal segment pericardial	2.03 ± 0.17	3.40 ± 0.90
3	L lower lobe subpleural/perifissural	10.78 ± 1.28	12.37 ± 0.63
4	R lower lobe posterior subpleural	1.66 ± 0.28	1.92 ± 0.31
4	R lower lobe posterior subpleural	7.52 ± 1.05	7.51 ± 0.78
5	L upper lobe subpleural	2.28 ± 0.21	2.31 ± 0.31
5	L upper lobe subpleural	2.06 ± 0.24	1.87 ± 0.29
5	R lower lobe superior basal segment subpleural	5.37 ± 0.53	5.01 ± 0.54
6	R lower lobe posterior basal segment near diaphragm	11.21 ± 0.90	12.59 ± 0.68
6	R lower lobe medial basal segment near diaphragm	3.88 ± 0.14	5.70 ± 0.13
6	L lower lobe posteromedial	5.37 ± 0.98	6.62 ± 0.44
6	R lower lobe lateral basal segment subpleural	2.87 ± 0.19	3.15 ± 0.08
6	Lingular (L upper lobe)	6.01 ± 0.24	7.13 ± 0.59
6	R lower lobe lateral basal segment subpleural	0.70 ± 0.08	1.24 ± 0.64
6	R upper lobe perifissural	0.49 ± 0.04	1.02 ± 0.52
6	R lower lobe posterior basal segment subpleural	1.53 ± 0.07	1.93 ± 0.14
6	R lower lobe posterior basal segment subpleural	0.94 ± 0.11	1.24 ± 0.12
6	R lower lobe medial basal segment near diaphragm	2.34 ± 0.25	3.02 ± 0.48
6	R lower lobe medial basal segment near diaphragm	2.07 ± 0.20	3.60 ± 0.61
7	R lower lobe superior basal segment	7.07 ± 2.73	9.26 ± 0.44
7	Lingular (L upper lobe)	7.84 ± 2.09	9.76 ± 0.94
7	L lower lobe posterior basal segment	10.82 ± 3.65	12.69 ± 1.15
7	R lower lobe lateral basal segment	3.58 ± 1.01	4.67 ± 0.66
7	R lower lobe lateral basal segment subpleural	11.07 ± 1.71	12.88 ± 0.72

Table E-6: Cross-Sectional Mean Volume Measurements (Expert Observers)
This table reports the mean ± SD measurements of each tumour, performed by expert observers.

Mean ± SD Observer Measurements			
		Trained Observers	
Subject #	Radiological locations	Observer 3 (FS)	Observer 2 (LW)
1	L lower lobe superior segment	3.84 ± 0.66	3.27 ± 0.14
2	R upper lobe basal segment	8.05 ± 1.53	8.41 ± 1.04
2	L lower lobe superior segment subpleural	30.50 ± 3.25	27.46 ± 1.45
3	R lower paratracheal (mediastinum)	35.21 ± 2.95	31.55 ± 4.02
3	R middle lobe lateral segment subpleural	33.51 ± 4.45	33.39 ± 4.09
3	R middle lobe medial segment pericardial	5.57 ± 0.81	5.26 ± 0.69
3	R lower lobe medial basal segment pericardial	2.14 ± 0.40	2.07 ± 0.38
3	L lower lobe subpleural/perifissural	10.13 ± 1.34	9.15 ± 1.81
4	R lower lobe posterior subpleural	2.32 ± 0.30	1.73 ± 0.22
4	R lower lobe posterior subpleural	5.62 ± 0.80	5.46 ± 1.00
5	L upper lobe subpleural	1.76 ± 0.45	1.45 ± 0.43
5	L upper lobe subpleural	1.26 ± 0.39	1.60 ± 0.99
5	R lower lobe superior basal segment subpleural	4.10 ± 0.60	5.28 ± 3.42
6	R lower lobe posterior basal segment near diaphragm	11.79 ± 0.50	9.33 ± 3.13
6	R lower lobe medial basal segment near diaphragm	5.13 ± 0.68	4.60 ± 0.86
6	L lower lobe posteromedial	5.41 ± 0.69	5.03 ± 1.54
6	R lower lobe lateral basal segment subpleural	3.04 ± 0.25	3.07 ± 1.05
6	Lingular (L upper lobe)	6.38 ± 0.26	4.42 ± 2.39
6	R lower lobe lateral basal segment subpleural	0.96 ± 0.10	0.86 ± 0.20
6	R upper lobe perifissural	3.08 ± 3.30	0.84 ± 0.36
6	R lower lobe posterior basal segment subpleural	1.87 ± 0.15	1.35 ± 0.41
6	R lower lobe posterior basal segment subpleural	1.09 ± 0.14	1.33 ± 0.55
6	R lower lobe medial basal segment near diaphragm	2.66 ± 0.19	2.45 ± 0.29
6	R lower lobe medial basal segment near diaphragm	2.80 ± 0.46	2.67 ± 0.82
7	R lower lobe superior basal segment	6.13 ± 2.33	7.21 ± 0.14
7	Lingular (L upper lobe)	7.30 ± 2.26	7.51 ± 0.73
7	L lower lobe posterior basal segment	10.86 ± 2.09	10.57 ± 0.73
7	R lower lobe lateral basal segment	3.40 ± 1.18	3.75 ± 0.26
7	R lower lobe lateral basal segment subpleural	11.24 ± 1.83	10.54 ± 0.57

Table E-7: Cross-Sectional Mean Volume Measurements (Trained Observers)
This table reports the mean ± SD measurements of each tumour, performed by trained observers.

Figure E-5: Mean ± SD of RECIST, WHO, and volumetric measurements by four observers.

Observers 1 (EO) and 2 (RER) are experts observers, and observers 3 (FS) and 4 (LW) are trained observers.

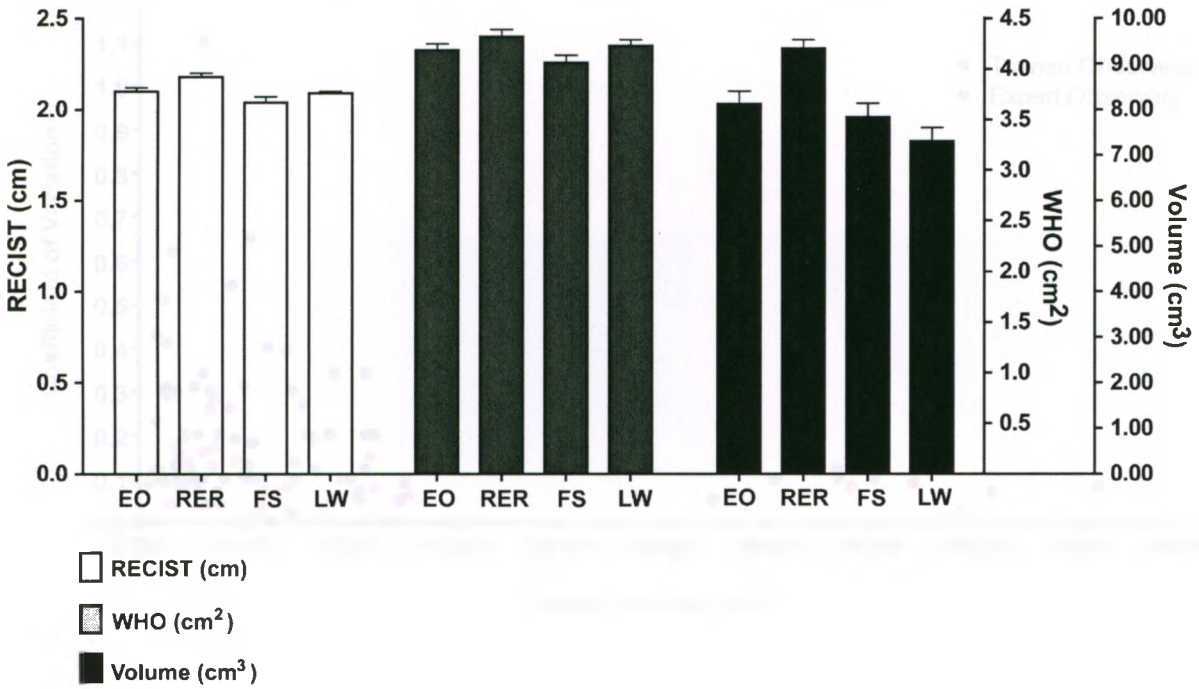
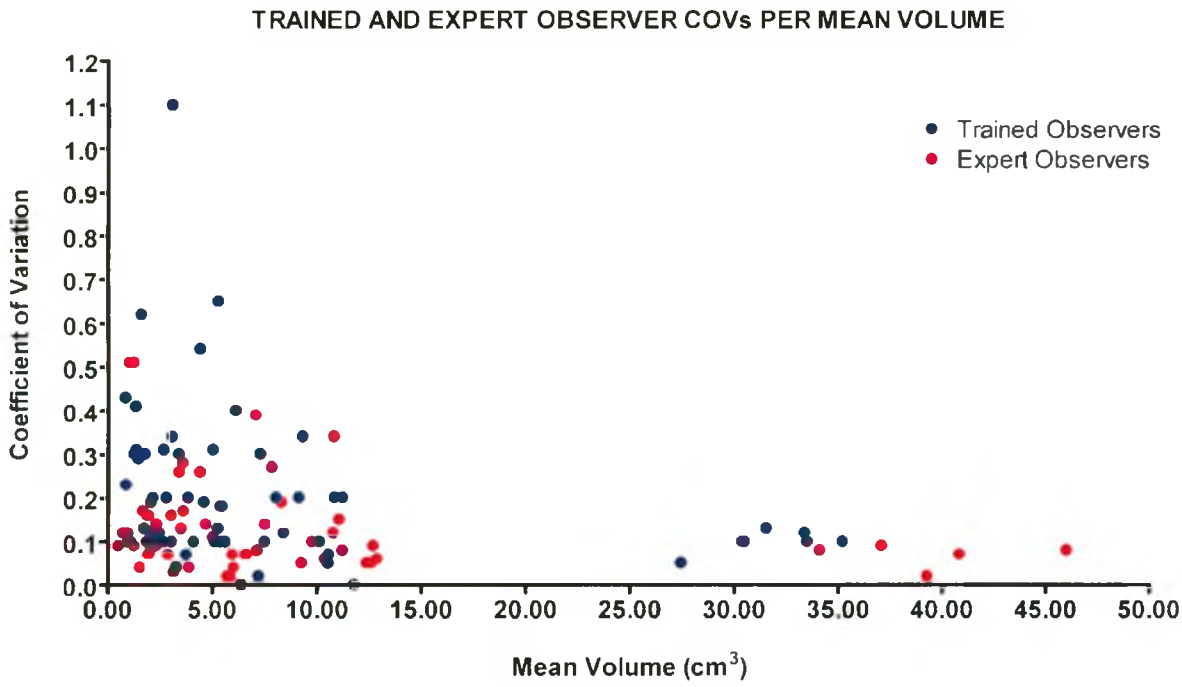


Figure E-6: COV Map

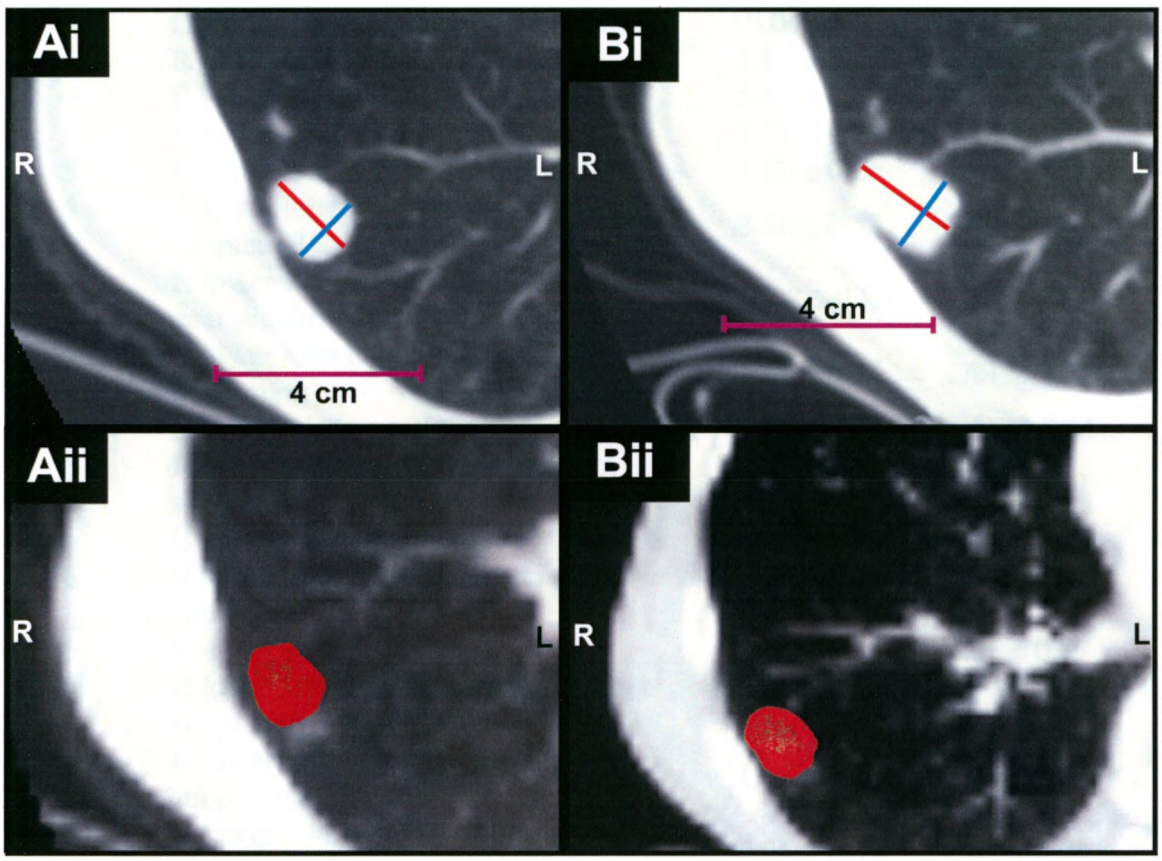
The mean volume of each tumour measurement was plotted against its corresponding coefficient of variation (COV). The results are categorized by trained and expert observers.



E.3 Longitudinal Analyses: Scan and Rescan

Figure E-7: Subject 6 Tumour Images at Scan and Rescan

Representative images of RECIST, WHO, and volumetric measurements for Subject 6 are presented. Ai shows RECIST and WHO measurements at scan, and Bi shows these measurements at rescan. Aii and Bii show the volume rendering of the same tumour at scan and rescan, respectively. Subject 6 has 11 pulmonary metastases.



Subject #	Radiological locations	Observer 1 (JM)	
		Scan	Rescan
2	R upper lobe basal segment	9.08 ± 0.53	10.65 ± 1.23
2	L lower lobe superior segment subpleural	29.18 ± 2.35	29.34 ± 1.84
3	R lower paratracheal (mediastinum)	34.67 ± 2.70	33.96 ± 2.08
3	R middle lobe lateral segment subpleural	33.14 ± 3.06	33.74 ± 2.27
3	R middle lobe medial segment pericardial	5.37 ± 0.54	2.65 ± 0.42
3	R lower lobe medial basal segment pericardial	1.76 ± 0.22	0.61 ± 0.16
3	L lower lobe subpleural/perifissural	10.04 ± 1.38	6.94 ± 1.87
4	R lower lobe posterior subpleural	1.82 ± 0.42	3.68 ± 0.07
4	R lower lobe posterior subpleural	5.85 ± 0.92	5.03 ± 0.80
5	L upper lobe subpleural	1.41 ± 0.29	1.73 ± 0.34
5	L upper lobe subpleural	1.40 ± 0.26	1.77 ± 0.26
5	R lower lobe superior basal segment subpleural	3.48 ± 0.57	5.00 ± 0.77
6	R lower lobe posterior basal segment near diaphragm	11.98 ± 0.88	7.33 ± 0.51
6	R lower lobe medial basal segment near diaphragm	3.97 ± 0.83	1.21 ± 0.23
6	L lower lobe posteromedial	5.95 ± 0.50	20.06 ± 2.11
6	R lower lobe lateral basal segment subpleural	3.01 ± 0.43	1.03 ± 0.21
6	Lingular (L upper lobe)	5.89 ± 0.51	2.77 ± 0.13
6	R lower lobe lateral basal segment subpleural	0.90 ± 0.22	0.50 ± 0.06
6	R upper lobe perifissural	0.80 ± 0.14	1.41 ± 0.37
6	R lower lobe posterior basal segment subpleural	1.61 ± 0.27	1.11 ± 0.19
6	R lower lobe posterior basal segment subpleural	1.05 ± 0.22	0.95 ± 0.22
6	R lower lobe medial basal segment near diaphragm	2.36 ± 0.33	2.42 ± 0.94
6	R lower lobe medial basal segment near diaphragm	2.87 ± 0.54	2.01 ± 0.23

Table E-8: Observer 1 Volumetric Measurements at Scan-Rescan
This table reports the mean ± SD of volumetric measurements performed by Observer 1 for each tumour.

Subject #	Radiological locations	Observer 2 (FS)	
		Scan	Rescan
2	R upper lobe basal segment	8.43 ± 1.82	9.65 ± 0.58
2	L lower lobe superior segment subpleural	29.84 ± 2.37	28.98 ± 0.98
3	R lower paratracheal (mediastinum)	34.62 ± 2.17	32.23 ± 3.80
3	R middle lobe lateral segment subpleural	32.74 ± 3.77	31.70 ± 3.43
3	R middle lobe medial segment pericardial	5.61 ± 0.84	2.53 ± 0.37
3	R lower lobe medial basal segment pericardial	2.19 ± 0.39	0.81 ± 0.11
3	L lower lobe subpleural/perifissural	9.86 ± 1.41	8.37 ± 0.78
4	R lower lobe posterior subpleural	2.36 ± 0.28	4.61 ± 0.38
4	R lower lobe posterior subpleural	5.90 ± 1.02	4.87 ± 0.89
5	L upper lobe subpleural	1.71 ± 0.40	1.60 ± 0.32
5	L upper lobe subpleural	1.21 ± 0.31	1.38 ± 0.29
5	R lower lobe superior basal segment subpleural	4.03 ± 0.50	5.01 ± 0.66
6	R lower lobe posterior basal segment near diaphragm	11.79 ± 0.50	7.89 ± 0.21
6	R lower lobe medial basal segment near diaphragm	5.09 ± 0.69	1.89 ± 0.11
6	L lower lobe posteromedial	5.25 ± 0.71	17.72 ± 1.43
6	R lower lobe lateral basal segment subpleural	3.03 ± 0.25	1.28 ± 0.29
6	Lingular (L upper lobe)	6.44 ± 0.16	3.09 ± 0.54
6	R lower lobe lateral basal segment subpleural	0.94 ± 0.10	0.69 ±0.09
6	R upper lobe perifissural	4.24 ± 3.24	4.51 ± 3.11
6	R lower lobe posterior basal segment subpleural	1.86 ± 0.15	1.28 ± 0.26
6	R lower lobe posterior basal segment subpleural	1.05 ± 0.14	0.63 ± 0.20
6	R lower lobe medial basal segment near diaphragm	2.52 ± 0.22	2.33 ± 0.94
6	R lower lobe medial basal segment near diaphragm	3.00 ± 0.38	2.19 ± 0.62

Table E-9: Observer 2 Volumetric Measurements at Scan-Rescan
This table reports the mean ± SD of volumetric measurements performed by Observer 2 for each tumour.

Subject #	Radiological locations	Observer 3 (LW)	
		Scan	Rescan
2	R upper lobe basal segment	8.29 ± 1.40	8.36 ± 1.08
2	L lower lobe superior segment subpleural	29.31 ± 2.93	28.17 ± 2.05
3	R lower paratracheal (mediastinum)	30.85 ± 3.83	30.93 ± 2.85
3	R middle lobe lateral segment subpleural	32.61 ± 4.31	31.27 ± 3.66
3	R middle lobe medial segment pericardial	5.15 ± 0.71	2.50 ± 0.24
3	R lower lobe medial basal segment pericardial	1.98 ± 0.38	0.78 ± 0.27
3	L lower lobe subpleural/perifissural	8.57 ± 1.46	7.69 ± 1.10
4	R lower lobe posterior subpleural	1.66 ± 0.10	3.44 ± 0.36
4	R lower lobe posterior subpleural	5.36 ± 0.91	4.26 ± 0.76
5	L upper lobe subpleural	1.37 ± 0.27	1.48 ± 0.20
5	L upper lobe subpleural	1.04 ± 0.15	1.53 ± 0.13
5	R lower lobe superior basal segment subpleural	3.55 ± 0.41	4.93 ± 0.44
6	R lower lobe posterior basal segment near diaphragm	11.05 ± 1.10	7.31 ± 0.52
6	R lower lobe medial basal segment near diaphragm	4.30 ± 0.60	1.37 ± 0.43
6	L lower lobe posteromedial	5.75 ± 0.81	12.67 ± 6.36
6	R lower lobe lateral basal segment subpleural	2.70 ± 0.31	3.78 ± 5.50
6	Lingular (L upper lobe)	4.89 ± 1.14	2.30 ± 0.72
6	R lower lobe lateral basal segment subpleural	0.88 ± 0.15	1.00 ± 0.75
6	R upper lobe perifissural	0.65 ± 0.15	0.89 ± 0.33
6	R lower lobe posterior basal segment subpleural	1.45 ± 0.24	1.04 ± 0.29
6	R lower lobe posterior basal segment subpleural	0.94 ± 0.25	0.77 ± 0.23
6	R lower lobe medial basal segment near diaphragm	2.33 ± 0.12	2.13 ± 0.92
6	R lower lobe medial basal segment near diaphragm	2.96 ± 0.33	1.94 ± 0.48

Table E-10: Observer 3 Volumetric Measurements at Scan-Rescan
This table reports the mean ± SD of volumetric measurements performed by Observer 3 for each tumour.

Subject #	Radiological locations	Observer 4 (SS)	
		Scan	Rescan
2	R upper lobe basal segment	10.03 ± 0.78	11.22 ± 0.47
2	L lower lobe superior segment subpleural	32.64 ± 1.60	31.59 ± 1.24
3	R lower paratracheal (mediastinum)	38.01 ± 1.80	33.67 ± 1.44
3	R middle lobe lateral segment subpleural	36.78 ± 1.48	35.33 ± 1.79
3	R middle lobe medial segment pericardial	6.20 ± 0.32	2.65 ± 0.13
3	R lower lobe medial basal segment pericardial	2.37 ± 0.15	0.87 ± 0.03
3	L lower lobe subpleural/perifissural	10.57 ± 0.44	8.40 ± 0.84
4	R lower lobe posterior subpleural	2.25 ± 0.24	4.66 ± 0.68
4	R lower lobe posterior subpleural	7.88 ± 0.42	6.87 ± 0.91
5	L upper lobe subpleural	1.82 ± 0.22	1.74 ± 0.54
5	L upper lobe subpleural	1.43 ± 0.37	2.01 ± 0.32
5	R lower lobe superior basal segment subpleural	3.74 ± 0.35	5.96 ± 0.43
6	R lower lobe posterior basal segment near diaphragm	13.09 ± 0.79	8.80 ± 0.80
6	R lower lobe medial basal segment near diaphragm	5.59 ± 0.28	2.24 ± 0.18
6	L lower lobe posteromedial	6.37 ± 0.54	20.17 ± 1.45
6	R lower lobe lateral basal segment subpleural	3.41 ± 0.18	1.29 ± 0.20
6	Lingular (L upper lobe)	6.89 ± 0.49	3.28 ± 0.21
6	R lower lobe lateral basal segment subpleural	1.10 ± 0.13	0.79 ± 0.04
6	R upper lobe perifissural	0.76 ± 0.14	1.34 ± 0.19
6	R lower lobe posterior basal segment subpleural	1.83 ± 0.61	1.37 ± 0.10
6	R lower lobe posterior basal segment subpleural	1.35 ± 0.28	1.14 ± 0.26
6	R lower lobe medial basal segment near diaphragm	3.31 ± 0.63	3.59 ± 0.38
6	R lower lobe medial basal segment near diaphragm	2.87 ± 0.27	2.65 ± 0.38

Table E-11: Observer 4 Volumetric Measurements at Scan-Rescan
This table reports the mean ± SD of volumetric measurements performed by Observer 4 for each tumour.

	Observer 1 (JM)		Observer 2 (FS)	
RECIST				
Subject	**SCAN**	**RESCAN**	**SCAN**	**RESCAN**
2	6.8 ± 0.2	6.5 ± 0.0	6.7 ± 0.4	6.6 ± 0.2
3	14.8 ± 0.2	13.7 ± 0.2	14.5 ± 0.3	13.3 ± 0.5
4	3.9 ± 0.1	4.5 ± 0.1	3.7 ± 0.2	4.5 ± 0.2
5	4.5 ± 0.1	5.3 ± 0.1	4.6 ± 0.3	5.4 ± 0.2
6	20.3 ± 0.4	20.0 ± 0.9	19.9 ± 1.0	18.9 ± 0.9
Mean	**10.1 ± 0.1**	**10.0 ± 0.2**	**9.9 ± 0.2**	**9.7 ± 0.3**
WHO				
Subject				
2	20.33 ± 0.62	19.16 ± 0.69	20.42 ± 1.69	20.34 ± 0.99
3	43.32 ± 0.55	36.68 ± 1.23	41.87 ± 1.28	36.58 ± 1.13
4	6.56 ± 0.44	8.17 ± 0.48	6.05 ± 0.63	7.81 ± 0.69
5	5.53 ± 0.13	7.01 ± 0.23	5.40 ± 0.49	7.01 ± 0.33
6	31.88 ± 0.82	26.18 ± 1.86	30.66 ± 2.63	24.23 ± 1.92
Mean	**21.52 ± 0.32**	**19.44 ± 0.42**	**20.88 ± 0.52**	**19.19 ± 0.45**
Volume				
Subject				
2	38.26 ± 2.78	39.99 ± 2.93	38.27 ± 2.35	38.63 ± 1.20
3	84.98 ± 7.47	77.91 ± 5.20	85.02 ± 6.49	75.63 ± 6.51
4	7.67 ± 1.29	8.72 ± 0.73	8.26 ± 1.21	9.48 ± 1.15
5	6.30 ± 0.84	8.50 ± 0.93	6.94 ± 1.03	7.99 ± 1.11
6	40.40 ± 2.32	40.79 ± 2.63	45.21 ± 4.37	43.51 ± 3.77
Mean	**35.52 ± 0.77**	**35.18 ± 0.70**	**36.74 ± 0.79**	**35.05 ± 0.74**

Table E-12: Observers 1 and 2 Tumour Burden Measurements at Scan-Rescan
This table reports the mean ± SD of 1D, 2D, and 3D tumour burden measurements performed by Observers 1 and 2 for each subject.

	Mean ± SD Observer Measurements			
	Observer 3 (LW)		Observer 4 (SS)	
RECIST				
Subject	**SCAN**	**RESCAN**	**SCAN**	**RESCAN**
2	6.5 ± 0.2	6.4 ± 0.1	7.1 ± 0.1	6.8 ± 0.1
3	14.7 ± 0.3	13.4 ± 0.1	15.3 ± 0.3	14.0 ± 0.1
4	3.8 ± 0.2	4.9 ± 1.3	4.1 ± 0.1	4.8 ± 0.1
5	4.5 ± 0.3	5.1 ± 0.4	5.0 ± 0.1	5.6 ± 0.1
6	19.6 ± 0.5	18.8 ± 0.5	21.1 ± 0.3	20.8 ± 0.5
Mean	**9.8 ± 0.1**	**9.7 ± 0.3**	**10.5 ± 0.1**	**10.4 ± 0.2**
WHO Subject				
2	20.49 ± 0.84	19.84 ± 0.66	22.08 ± 0.43	21.00 ± 1.50
3	42.70 ± 1.12	36.97 ± 0.51	45.57 ± 1.49	39.84 ± 1.16
4	6.39 ± 0.35	7.69 ± 0.60	7.23 ± 0.58	9.18 ± 0.40
5	5.54 ± 0.45	6.95 ± 0.23	6.45 ± 0.28	7.48 ± 0.17
6	30.11 ± 1.27	24.46 ± 1.24	34.22 ± 0.81	30.34 ± 0.84
Mean	**21.05 ± 0.40**	**19.18 ± 0.36**	**23.11 ± 0.38**	**21.57 ± 0.40**
Volume				
Subject				
2	37.60 ± 4.22	36.53 ± 2.52	42.68 ± 2.21	42.81 ± 1.43
3	79.17 ± 10.25	73.18 ± 7.85	93.94 ± 3.52	80.92 ± 2.31
4	7.03 ± 0.96	7.70 ± 0.98	10.13 ± 0.65	11.53 ± 1.55
5	5.96 ± 0.82	7.93 ± 0.66	6.99 ± 0.83	9.70 ± 1.19
6	37.91 ± 4.61	35.20 ± 4.97	46.58 ± 2.01	46.66 ± 2.38
Mean	**33.53 ± 0.91**	**32.11 ± 0.82**	**40.06 ± 0.61**	**38.32 ± 0.60**

Table E-13: Observers 3 and 4 Tumour Burden Measurements at Scan-Rescan
This table reports the mean ± SD of 1D, 2D, and 3D tumour burden measurements performed by Observers 3 and 4 for each subject.

Intra- and Inter-Observer Coefficients of Variation (%)								
	Observer 1 (JM)		**Observer 2 (FS)**		**Observer 3 (LW)**		**Observer 4 (SS)**	
RECIST								
Subject	**SCAN**	**RESCAN**	**SCAN**	**RESCAN**	**SCAN**	**RESCAN**	**SCAN**	**RESCAN**
2	2.2	0.7	5.4	2.9	2.6	1.8	1.3	0.8
3	1.1	1.7	1.9	3.6	2.2	1.1	1.9	0.6
4	2.3	2.7	4.1	4.1	5.9	26.9	3.2	3.1
5	2.9	2.5	5.8	3.1	6.1	7.4	1.7	2.1
6	2.1	4.4	5.2	4.5	2.7	2.9	1.6	2.3
Inter-Observer	**3.9**	**2.7**	**5.5**	**5.4**	**4.8**	**6.6**	**0.0**	**0.0**
WHO								
Subject								
2	3.05	3.60	8.29	4.85	4.08	3.33	1.94	7.12
3	1.28	3.35	3.05	3.09	2.62	1.37	3.28	2.90
4	6.69	5.89	10.46	8.82	5.42	7.74	8.02	4.37
5	2.41	3.25	9.15	4.64	8.13	3.30	4.33	2.30
6	2.58	7.11	8.57	7.94	4.21	5.05	2.36	2.78
Inter-Observer	**3.50**	**3.91**	**4.42**	**4.63**	**3.65**	**3.46**	**3.14**	**4.13**
Volume								
Subject								
2	7.26	7.34	6.14	3.10	11.22	6.89	5.17	3.33
3	8.79	6.68	7.64	8.60	12.95	10.72	3.75	2.85
4	16.78	8.42	14.60	12.14	13.62	12.73	6.44	13.45
5	13.28	10.90	14.82	13.87	13.69	8.37	11.82	12.22
6	5.74	6.46	9.66	8.66	12.15	14.12	4.33	5.11
Inter-Observer	**3.76**	**3.57**	**4.01**	**3.57**	**4.38**	**4.13**	**2.68**	**2.87**

Table E-14: CVs of Tumour Burden Measurements at Scan-Rescan
This table reports the intra- and inter-observer CVs of tumour burden measurements performed by all observers.

E.4: Longitudinal Analyses: Multiple Timepoints

	Tumour 1	Tumour 2
RECIST (cm)		
Observer 1 (JM)	1.9 ± 0.0	2.3 ± 0.1
Observer 2 (FS)	2.0 ± 0.0	2.4 ± 0.1
Observer 3 (LW)	1.8 ± 0.0	2.3 ± 0.0
Observer 4 (SS)	1.9 ± 0.0	2.4 ± 0.0
All Observers	**1.9 ± 0.0**	**2.3 ± 0.1**
WHO (cm^2)		
Observer 1 (JM)	2.99 ± 0.08	4.17 ± 0.22
Observer 2 (FS)	3.09 ± 0.09	4.58 ± 0.25
Observer 3 (LW)	2.91 ± 0.06	4.38 ± 0.12
Observer 4 (SS)	2.91 ± 0.05	4.47 ± 0.11
All Observers	**2.97 ± 0.08**	**4.40 ± 0.21**
Volume (cm^3)		
Observer 1 (JM)	6.59 ± 0.33	9.66 ± 0.69
Observer 2 (FS)	7.90 ± 0.51	9.87 ± 0.86
Observer 3 (LW)	5.76 ± 0.21	8.04 ± 0.55
Observer 4 (SS)	8.73 ± 0.23	14.88 ± 0.60
All Observers	**7.25 ± 0.63**	**10.61 ± 0.50**

Table E-15: Mean RECIST, WHO, and Volumetric Measurements of Two Tumours

Measurements are expressed as mean ± SD of five repeated measurements over all time points. Tumour 1 was evaluated by each observer over 11 time points, and tumour was evaluated over nine time points.

	Tumour 1	Tumour 2
RECIST (cm)		
Observer 1 (JM)	1.5	3.3
Observer 2 (FS)	1.2	2.7
Observer 3 (LW)	1.4	1.5
Observer 4 (SS)	0.9	1.7
Inter-Observer	**1.6**	**2.5**
WHO (cm^2)		
Observer 1 (JM)	2.76	5.2
Observer 2 (FS)	2.83	5.54
Observer 3 (LW)	2.01	2.85
Observer 4 (SS)	1.8	2.45
Inter-Observer	**2.61**	**4.70**
Volume (cm^3)		
Observer 1 (JM)	5.02	7.14
Observer 2 (FS)	6.39	8.73
Observer 3 (LW)	3.65	6.89
Observer 4 (SS)	2.59	4.02
Inter-Observer	**8.73**	**12.61**

Table E-25: Mean CV RECIST, WHO, and Volumetric Measurements of Two Tumours

Coefficients of variation for mean RECIST, WHO, and volumetric measurements are reported as a percent.

Figure E-8: Subject 4 - Tumour 1

The longitudinal changes in one tumour are displayed over eleven time points using RECIST, WHO, and volumetric measurements.

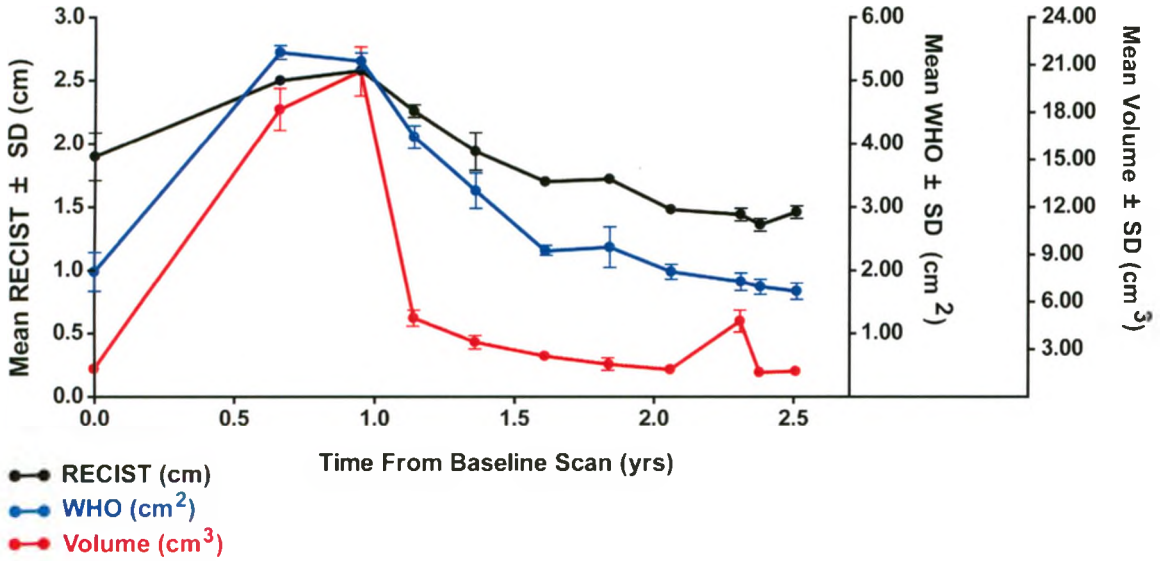


Figure E-9: Subject 4 – Tumour 2

The longitudinal changes in one tumour are displayed over nine time points using RECIST, WHO, and volumetric measurements.