Validation and Application of a Virtual Imaging Trial Platform for Accurate and Precise CT Quantifications in Lung Imaging

by

Sachin Shankar

Department of Electrical and Computer Engineering
Duke University

Date:_____

Approved:

_____
Ehsan Abadi, Supervisor

_____
Michael Gehm, Co-Supervisor

_____
Loren W Nolte

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Department of
Electrical and Computer Engineering in the Graduate School
of Duke University

2021

ABSTRACT

Validation and Application of a Virtual Imaging Trial Platform for Accurate and Precise CT Quantifications in Lung Imaging

by

Sachin Shankar

Department of Electrical and Computer Engineering
Duke University

Date:_____

Approved:

_____
Ehsan Abadi, Supervisor

_____
Michael Gehm, Co-Supervisor

_____
Loren W Nolte

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Department of
Electrical and Computer Engineering in the Graduate School of
Duke University

2021

Abstract

Computed Tomography (CT) is a prevalent imaging technique in modern medicine that provides physicians a non-invasive method to evaluate and diagnose various clinical conditions. To aid in diagnosis, it is important to have a high accuracy and reliability in these images. In the first phase of this study, the variability of clinically-relevant imaging biomarkers was analyzed across different scanners and imaging parameters through usage of a customized anthropomorphic chest phantom with several experimental sample inserts. This phantom was scanned across 10 different scanners. Imaging biomarkers were computed for each scan. Intra and inter-scan variability was assessed by computing coefficients of variation and standard deviations of the measurements. It was found that LAA -950 and LAA -856 were the biomarkers with the highest levels of variability, while the majority of other biomarkers had variability less than 10 HU or 10% CV in both inter and intra-scan measurements. No clear trend was found between the variability of the biomarkers and radiation dose (i.e., CTDI).

Traditional assessments of CT technologies are limited in the sense that they work with real patient data and are not efficient. Alternatively, Virtual Imaging Trials (VITs), which use virtual scanners and patients, are more efficient and avoid unnecessary radiation exposure. DukeSim is a CT simulator that has been validated with

simple cylindrical phantoms in the past, but not with more clinically-relevant phantoms and conditions. Biomarkers computed from real CT image data were compared to those from simulated CT scans of a computational version of an anthropomorphic chest phantom. Overall, relative percent errors ranged from 0.187% to 18.269%.

Having validated DukeSim in a clinically relevant context, the utility of DukeSim as a VIT tool was shown by investigating the effects of imaging and reconstruction parameters on the clinically relevant biomarkers. It was found that sharper reconstruction kernels and lower tube currents tended to reduce the accuracy of measured biomarkers. These findings will help to spark further studies in virtual imaging, which can help to yield further clinical insights to improve patient health outcomes.

Dedication

To my parents, brother, and grandparents, for their constant love and support throughout this journey, without which none of this would have been possible.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my family, to whom I owe everything, for the invaluable love and support that they have given me throughout this journey. I would also like to thank all the friends that I have made during my graduate studies at Duke University for their support as well.

I want to express my heartfelt gratitude to Dr. Ehsan Abadi for his continual mentorship, guidance, support, and advising that he has given me throughout my time at Duke University. I could not have asked for a better mentor for my graduate work.

I would also like to thank all of the faculty, staff, and students whom I had the pleasure of working with at the Carl E. Ravin Advanced Imaging Laboratories (RAI Labs) and the Center for Virtual Imaging Trials (CVIT). In particular, I would like to thank Dr. Ehsan Samei for his guidance and Brian Harrawood for his technical help. I would also like to thank Giavanna Jadick, Shobhit Sharma, Nicholas Felice, and Jayasai Rajagopal for their help. I would also like to thank Dr. Eric Hoffman, a collaborator at the University of Iowa.

I sincerely appreciated the input and guidance of Dr. Michael Gehm and Dr. Loren Nolte, who also served on my thesis committee.

1. Introduction

1.1 Background

Computed Tomography (CT) imaging, alternatively known as Computed Axial Tomography (CAT), provides physicians a non-invasive approach to diagnose and evaluate clinical diseases seen in patients. In a typical CT scan, a patient lies flat on a moving table while a rotating X-ray source scans the patient. The X-rays that pass through the patient are collected by a detector and a computer is used to reconstruct a 3D rendering of the patient's body part of interest. The resultant image can be represented in terms of Hounsfield Units (HU), with certain ranges of HU values corresponding to particular materials or organs (for standardization purposes, air is defined as -1000 HU and water is defined as 0 HU).

Given the fact that CT has only been in usage for about half a century or so, further studies on CT can help to further optimize and improve this technology in hopes of expanding its clinical application to help improve patient health outcomes in many fields of medicine [1]. Perhaps the greatest benefit of CT imaging is that it is a non-invasive and relatively painless procedure from the patient perspective.

As evidence of this advancement of CT technology, approximately 80 million CT scans are conducted every year in the United States, compared to just the three million scans back in 1980 [1, 2]. Currently, CT imaging is utilized in a variety of fields, including radiology, emergency care, surgery, pediatrics, pulmonology, and neurology

for a variety of disease s including stroke, colorectal cancer, and lung cancers [3-6]. Of particular interest for this study are lungs and respiratory diseases. These respiratory diseases represent the third leading cause of death globally, as of 2019 [7]. Examples of such diseases include Chronic Obstructive Pulmonary Disease (COPD) and chronic bronchitis.

Given that CT has become a widespread medical technique, there will be some amount of variation across images and imaging parameters, both in a qualitative and quantitative regard. For lung diseases in particular, analysis and evaluation of this variability using clinically-relevant biomarkers could aid in optimizing clinical diagnoses.

In addition to the advancements that have been made with real CT scanners, virtual trials of these imaging procedures have been developed. In place of using real patients and scanners in a traditional clinical setting, virtual trials use computerized virtual patient phantoms and acquisition techniques [8]. While these virtual trials avoid some of the challenges associated with working with real patient data (e.g., patient privacy concerns), it is still important to validate these virtual approaches and assess their variability in clinically-relevant contexts with respect to pulmonary disorders, in hopes of eventually improving patient health outcomes.

1.2 Purpose

The goal of this study is three-fold in regards to the analysis of CT image data for relevant lung diseases. The first phase of this study seeks to analyze variability of a set of clinically relevant biomarkers taken from real CT image data taken from multiple different scanners across a variety of imaging and reconstruction parameters. Variation of these biomarkers across multiple scanners and within one scanner will be analyzed. To facilitate this, a custom anthropomorphic chest phantom was utilized in place of real patients for efficiency purposes as well as to minimize radiation exposures.

The second phase of this study involved working with virtual CT simulations done through Virtual Imaging Trials (VITs) using DukeSim, a scanner-specific imaging simulator. A computational version of the anthropomorphic chest phantom was used to validate the simulation accuracy of DukeSim. Clinically-relevant biomarkers were derived from the virtual result and compared to biomarkers computed from the corresponding CT image produced by a real CT scanner obtained with the same imaging and reconstruction parameters.

The third phase of this study involved an application of these Virtual Imaging Trials. With a validated simulator from phase two of this study, the utility of DukeSim was shown through simulating more images of a computational chest phantom with varying tube current strengths and reconstruction kernels. The impact of these varied parameters on the clinically relevant biomarkers was investigated.

2. Demonstration of Variability of CT Imaging Biomarkers in Real Image Data

The first phase of this study was to analyze both the intra- and inter- scanner variability in biomarker measurements that were observed from CT Image data obtained from real scanners across a range of imaging and reconstruction parameters. The information presented in this chapter is planned to be submitted for SPIE Medical Imaging, so the contents of this chapter (text, tables, and figures) were largely adapted from that planned submission.

2.1Introduction

The role of CT imaging in clinical practice is to aid medical doctors in the diagnosis of various diseases. To reiterate, CT imaging utilizes an x-ray source that rotates around a patient. This process produces projection data, which can be reconstructed via various algorithms to provide physicians a computerized depiction of their patient.

With any clinical measurement, it is important to have qualitative accuracy as this will allow physicians to get a general idea of the underlying physiology of the relevant disease or condition. To provide an even more accurate and targeted diagnosis for patients, it is also better to have a high quantitative accuracy in the CT image, given the wide range of diseases that CT has the potential to diagnose.

Of particular interest are lower respiratory chronic diseases. According to the World Health Organization, this category of diseases was the third leading cause of death globally in 2019 [7]. As such, an examination of the quantitative accuracy of CT images of human lungs becomes of clinical relevance. Following this idea, an analysis of the variation of these quantifications across different scanner manufacturers and imaging parameters is also warranted.

Previous analysis of variability of medical images has been done before. Both qualitative and quantitative methods of analysis have been done in the past in the context of other imaging modalities such as Magnetic Resonance Imaging (MRI) [9-12]. However, analysis with MRI images has limitations, namely that MRI scans are significantly more expensive compared to CT scans [13].

In this regard, previous studies of CT lung image analysis have been performed. Previous work in variability has examined lung CT image feature analysis and nodule size measurements from real patient data [14-18]. Other studies in lung CT images have use simpler cylindrical phantoms and have performed variability analysis based on HU values [19]. Lung CT image variability analysis has also been performed on real patient data for relevant image segmentation and analysis tasks involving the usage of deep learning methods [20, 21]. Another study has used an anthropomorphic phantom to examine CT noise characteristics of the thorax [22]. To our knowledge, there have not been any studies on anthropomorphic phantoms that examine variability in lung CT

images using clinically relevant biomarker measurements on a range of different CT scanner models and imaging parameters.

In order to be of clinical relevance, the quantitative biomarkers computed from the lung CT images must be able to accurately and reliably reflect patient conditions [23]. However, there is the potential for the scanner model and imaging parameters to affect these CT quantifications [24]. This study comprehensively analyzes variability in biomarker measurements with respect to scanner and imaging parameters in the context of lung CT imaging. To this end, a clinically-relevant anthropomorphic chest phantom was used in order to reduce unnecessary radiation exposure and allow for images to be gathered relatively efficiently.

2.2 Methods

2.2.1Anthropomorphic Chest Phantom

For this study, an anthropomorphic chest phantom was used. More specifically, a customized version (designed by the University of Iowa) of the commercially available Multipurpose Chest Phantom N1 "LUNGMAN" (PH-1) manufactured by the Kyoto Kagaku company (based in Kyoto, Japan) was used [25-27]. This chest phantom contains models for the two lungs, ribs, spine, sternum, and scapulae. There is also a urethane resin that surrounds the models of the right and left lungs and thoracic bones, and the whole phantom is surrounded by detachable chest plates. An acrylic foam cover is used for the CT table that the phantom rests on while scanning.

For the customization, three vertically-spanning tubes containing sample material inserts were placed into the lung models: one 59 mm diameter tube passed through the right lung and two 28 mm diameter tubes passed through the left lung (one in the back and one in the front of the left lung). Three different sets of either 15 or 16 experimental sample inserts were placed into the vertical tubes. Each set is denoted either by Configuration 1, Configuration 2, or Configuration 3. The sample inserts in each Configuration are outlined in Table 1 below:

Table 1: Contents of Each Tube in Each Configuration of the Customized Kyoto Kagaku Phantom

Configuration Number	Right Tube	Left Posterior Tube	Left Anterior Tube
1	Water, Lung Samples (A, B, and C), Air	Bone 20%, Bone 50%, Air	NIST Foam (Density= 4, 8, 12, 14,20 lb/ft^3), Acrylic, Air
2	Water, Lung Samples (D, E, and F), Air	Bone 20%, Bone 50%, Air	NIST Foam (Density= 4, 8, 12, 14, 20 lb/ft3), Acrylic, Air
3	Low Contrast Agent, MTF Bead, MTF Cube, Air	Acrylic, Bone 20%, Bone 50%, Air	NIST Foam (Density= 4, 8, 12, 14, 20 lb/ft3), Polymethylpentene (PMP), Air

The physical setup of each configuration, along with a schematic of the chest phantom, can all be seen below in Figure 1. Regarding the sample inserts shown below

in Figure 1, the lung samples are fixed porcine lung cored samples from an inflated and an over-inflated pair of lungs. Additionally, the NIST foams sample inserts are standardized foams obtained from the National Institute of Standards and Technology (NIST) [28].

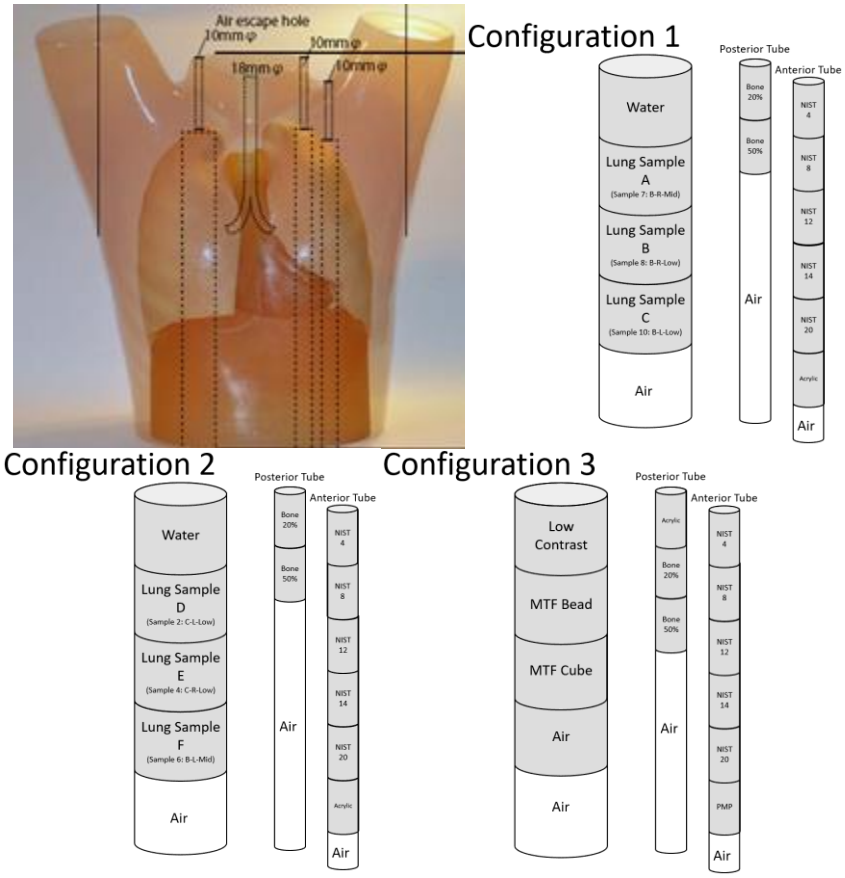

Figure 1: Different Schematics for the Kyoto-Kagaku chest phantom: Physical Phantom (top row, left) and Configurations 1, 2, and 3 for the three tube inserts (top row right and bottom row left and right, respectively). Note that the region marked as "MTF Bead" in Configuration 3 is just one tungsten carbide bead surrounded by low contrast. PMP is polymethylpentene polymer.

An anthropomorphic phantom was used in this study in place of traditional patient data in order to be able to efficiently obtain more images for analysis. With an artificial phantom, there is no issue of too much exposure to radiation to the test subject.

2.2.2 CT Imaging

The CT scanners used in this study came from multiple vendors, including General Electric (GE), Siemens, Canon, and Philips (courtesy of University of Iowa). Imaging and reconstruction parameters that varied across the datasets included reconstruction method and kernel type, tube current, and CT dose index (CTDIvol). A summary table of the scanners used in this study to image the Kyoto Kagaku chest phantom can be seen below in Table 2 (total 209 images).

Table 2: Summary of CT Scanners Used in this study

SCANNER NAME	Number of Images	kVp	Pixel Spacing (mm) x (mm)	Slice Thickness (mm)
GE Discovery CT750HD	18	120	0.5078 x 0.5078	0.625
GE Revolution CT	20	120	0.5078 x 0.5078	0.625
Siemens SOMATOM Definition Flash	15	120	0.5078 x 0.5078	0.75
Siemens SOMATOM Force	17	100 or 120	0.5078 x 0.5078 or 0.9766 x 0.9766	0.75
Siemens SOMATOM Definition AS	12	120	0.5078 x 0.5078	0.75

Siemens SOMATOM Definition AS+	15	120	0.5078 x 0.5078	0.75
Siemens SOMATOM Definition Edge	25	120	0.5078 x 0.5078	0.75
Siemens SOMATOM Drive	21	120	0.5078 x 0.5078	0.75
Philips Brilliance iCT	22	120	0.5078 x 0.5078	0.67
Canon Aquilion Precision	44	120	0.5074 x 0.5074	0.5

Sample scans from each vendor can be seen below in Figure 2:

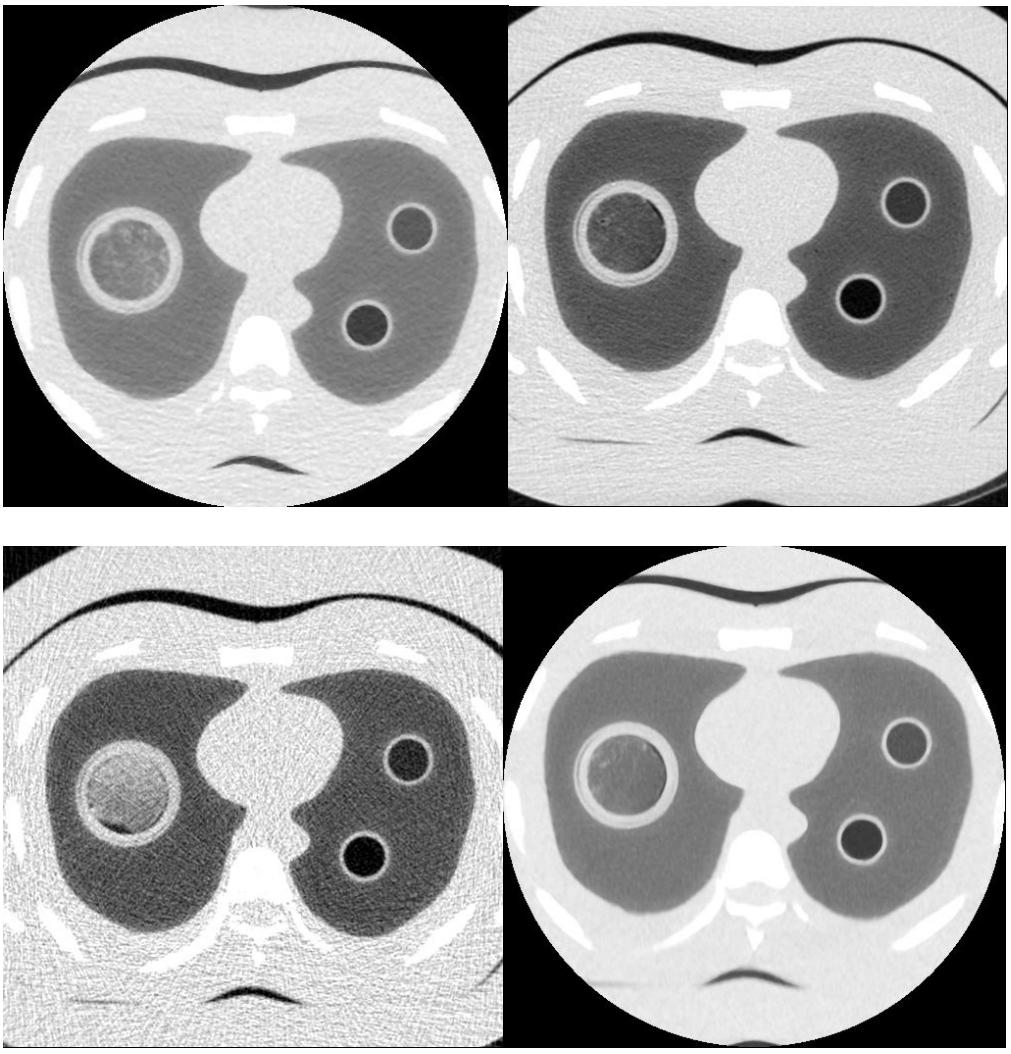

Figure 2: Sample mid-level axial views of the Kyoto Kagaku phantom from four different vendors (from scanner A, B,C, and D), shown using a window level of -600 HU and window of 1500 HU.

In regards to the window and window level in Figure 2 above, any voxel in the CT image with HU less than -1350 HU is displayed as black, and any voxel with HU greater than 150 is displayed as white. Anything in between -1350 HU and 150 HU is displayed as the appropriate gray scale.

2.2.3 Biomarker Analysis

From each image taken by each scanner in the above Table 2, several clinically relevant biomarkers were computed. In order to do this, a MATLAB code was written that processed and analyzed the reconstructed dicom images from the scanners. For each image, the lungs were segmented through a combination of Otsu's method and binary connected component labeling. Because the lungs have an irregularly shaped morphology, the fact that lungs typically have HU values ranging between -500 and -800 was utilized to aid in the segmentation process. In a similar regard, the vertical cylindrical tubes were segmented through a combination of using circular Hough transforms and canny edge detection.

From here, binary masks of both the lungs and vertical tubes were made via a simple in-built binarization function in MATLAB. These, in tandem with image-specific information such as voxel size, were used to compute the clinically relevant biomarkers.

Based on previous literature regarding pulmonary biomarkers [29], several biomarkers were computed, as summarized below in Table 3:

Table 3: A summary of pulmonary biomarkers used in this study, adapted from [29]. Note that N is the total number of relevant voxels and HUi represents the HU value for the voxel of interest.

Biomarker Name	Definition	Clinical Significance
LAA -950	Percentage of lung voxels with HU values ≤ -950	Is a marker of pulmonary airflow obstruction
LAA -856	Percentage of lung voxels with HU values ≤ -856	Measure of gas trapping
Perc 15	The lung HU value at the 15th percentile	Is a marker of emphysema progression
Lung Mass (g)	$\sum_{i=1}^{N}(\frac{HU_i + 1024}{1024})(\text{Voxel Size in cm}^3)$	Is a measure of progression of various pulmonary diseases
Lung Volume (cm^3)	$(\text{N})\ (\text{Voxel Size in cm}^3)$	

Even though these quantifications were performed on an anthropomorphic phantom, results of these analyses can still be applied towards gaining a better understanding of lower respiratory diseases commonly seen in clinical practice.

In addition, the distribution of the HU values of the lung and for each of the tube sample inserts was computed as a mean±std. In total, there were 7 lung biomarkers computed (Lung Mean HU, Lung Std HU, LAA -950, LAA -856, Perc 15, Lung Mass, and Lung Volume) and 21 mean and standard deviations computed for each sample in the tube inserts (Note that a mean and standard deviation of the MTF Bead wasn't computed since this was scanned using just 1 bead, as opposed to using an entire insert).

With these quantifications in hand, the effects of varying scanners and imaging parameters for the same phantom can be better understood.

2.3 Results

Both biomarker variability within (intra-) and across (inter-) scanners were analyzed for this study. All scanners (with the exception of 4 Siemens SOMATOM Force images) were scanned using 120 kVp. All scanner names are depicted anonymously.

2.3.1 Intra-Scan Variability

For the analysis of intra-scanner variability, either the coefficient of variation (CV) or a raw standard deviation was computed, depending upon the unit of the biomarker. In CT, the Hounsfield Unit (HU) for any given material is defined as follows:

$$HU = 1000\frac{\mu - \mu_{Water}}{\mu_{Water} - \mu_{Air}}$$

In which μ represents the linear attenuation coefficient for that material for any particular given energy. The coefficient of variation (CV) was expressed as a percentage and is defined as follows:

$$CV = 100\left|\frac{Standard\ Deviation\ of\ Biomarker}{Mean\ of\ Biomarker}\right|$$

The purpose of using CV as a measure of variation as opposed to a traditional measure of variability (ex. variance or standard deviation) is that it provides a standardized measure of the distribution of the dataset [30].

This standardization comes through dividing the standard deviation by the mean of any given measurement. Additionally, the CV is a useful metric in this particular scenario because there is no known ground truth with which to standardize measurements.

However, as given by the formula for HU above, there is already a normalization term in the denominator ($\mu_{Water} - \mu_{Air}$), so just presenting a standard deviation is sufficient for measures of HU variability. For each CT Dose Index Volume (CTDIvol) used in each scanner, the standard deviation of measurements involving HU was computed.

All other biomarker measurements (ex. lung volume or lung mass) were expressed as a normalized CV. For each CT Dose Index Volume (CTDIvol) used in each scanner, CV across multiple images was computed using the formula for CV above. Due to the nature of CV requiring a standard deviation of any given measurement across multiple images, the CV was computed for only images which had multiple measurements at the same CTDIvol value.

The key results of these intra-scanner variability measurements can be seen below in Figures 3 through 10 below, to show the wide range in image variability that was seen in this study. Each Figure represents a different reconstruction setting for that particular scanner. Each bar represents the aggregate measurement for only three images with all the same imaging parameters.

Each biomarker, conditioned to a particular CTDIvol (though the range of CTDIvol values in Scanner B was relatively small), represents the horizontal axis, and either CV or the standard deviation (for HU measurements) served as the vertical axis. Because of the nature of the biomarkers computed, the plots in each Figure have been bifurcated: one plot all measurements in HU (Expressed as a standard deviation) and one plot for all other data types (which can easily be converted to CV).

In general, while intra-scan variability was generally quite low, across both scanners LAA -950 and LAA -856 see to show the highest level of intra-scan variability, with CV values ranging up to 50% Note that some experimental inserts were omitted from these plots, as there weren't sufficient images with these inserts present to compute a variability.

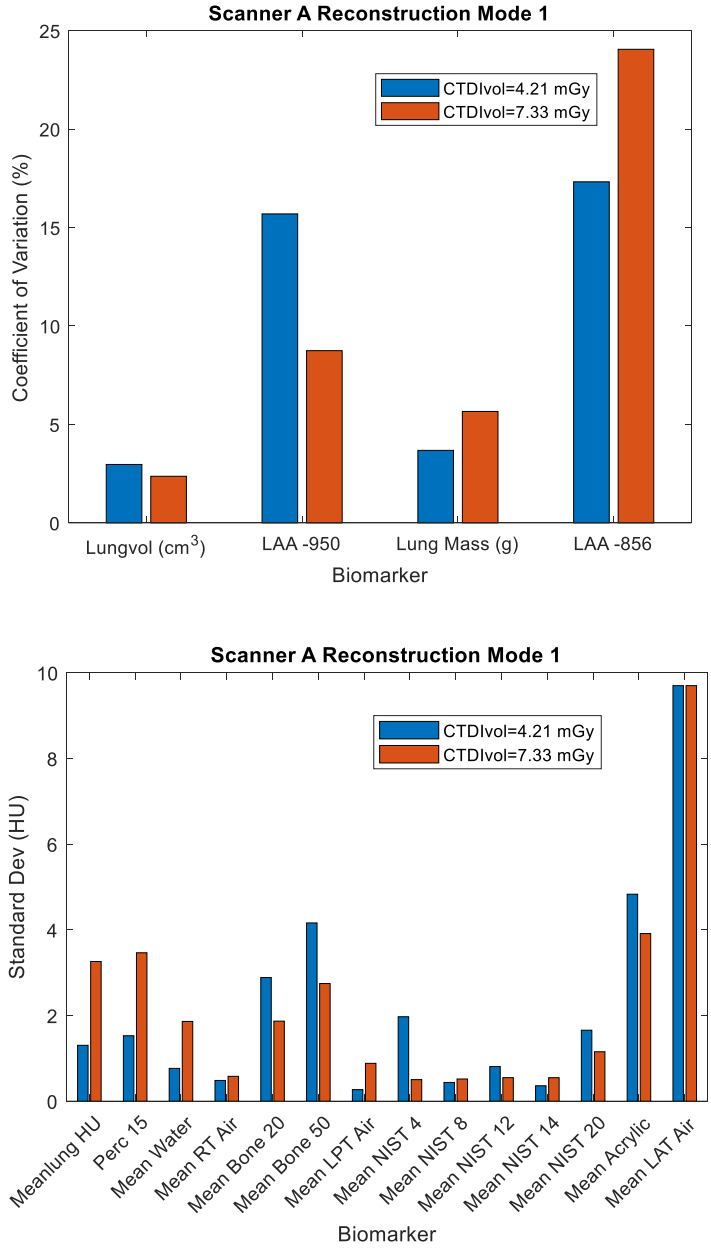

Figure 3: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 1 of Scanner A

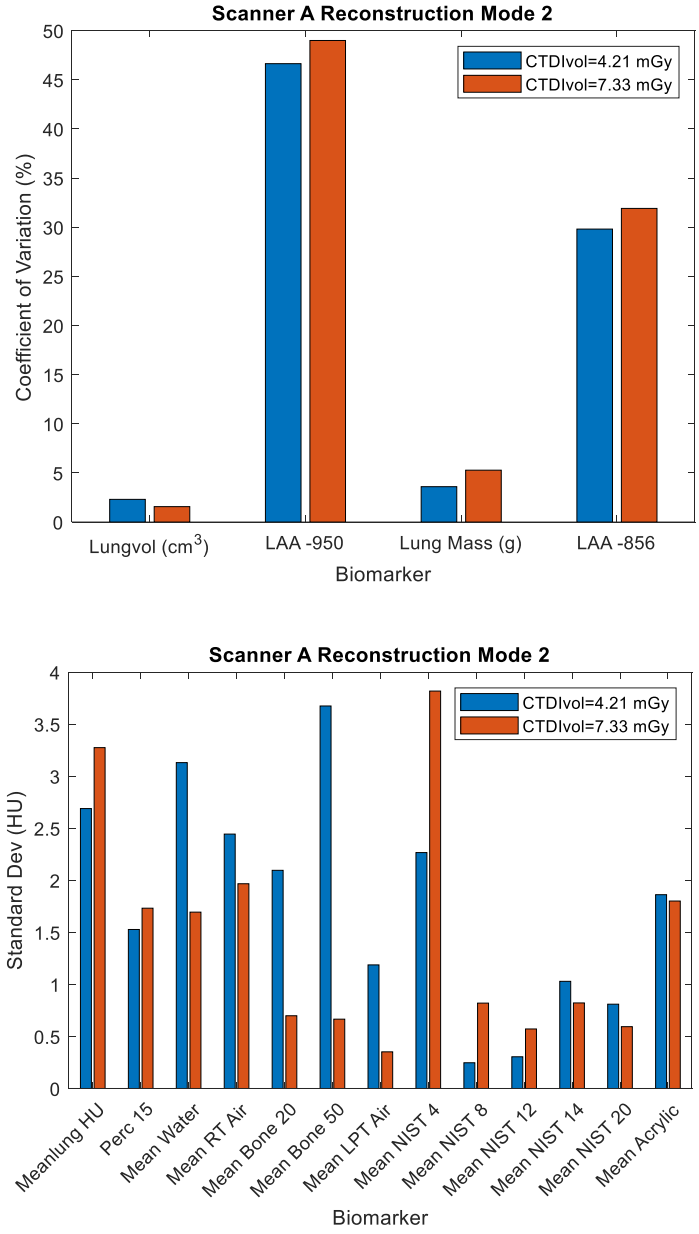

Figure 4: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 2 of Scanner A

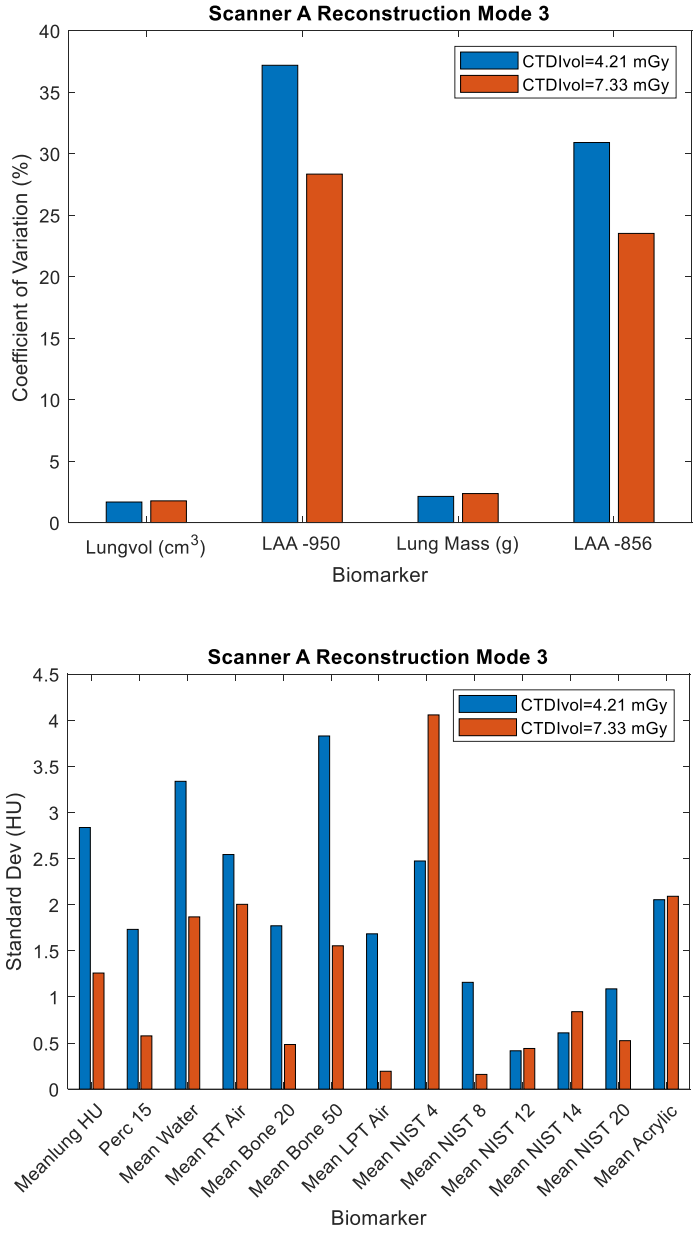

Figure 5: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 3 of Scanner A

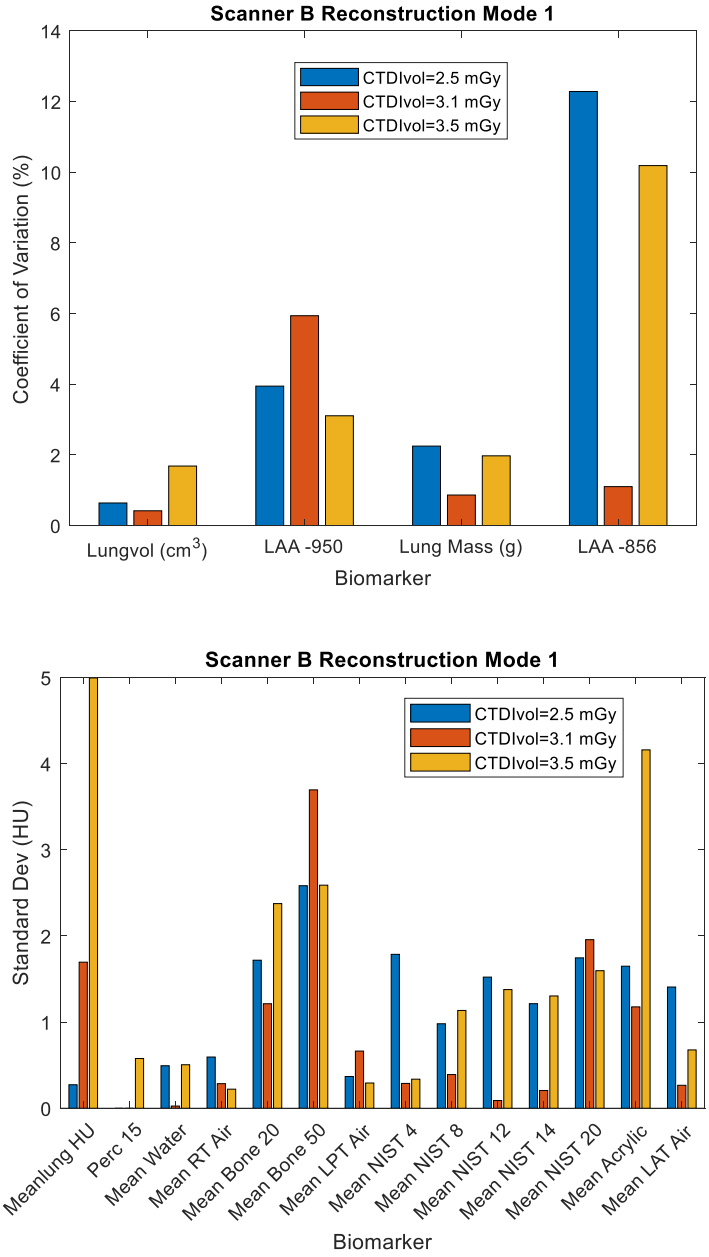

Figure 6: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 1 of Scanner B

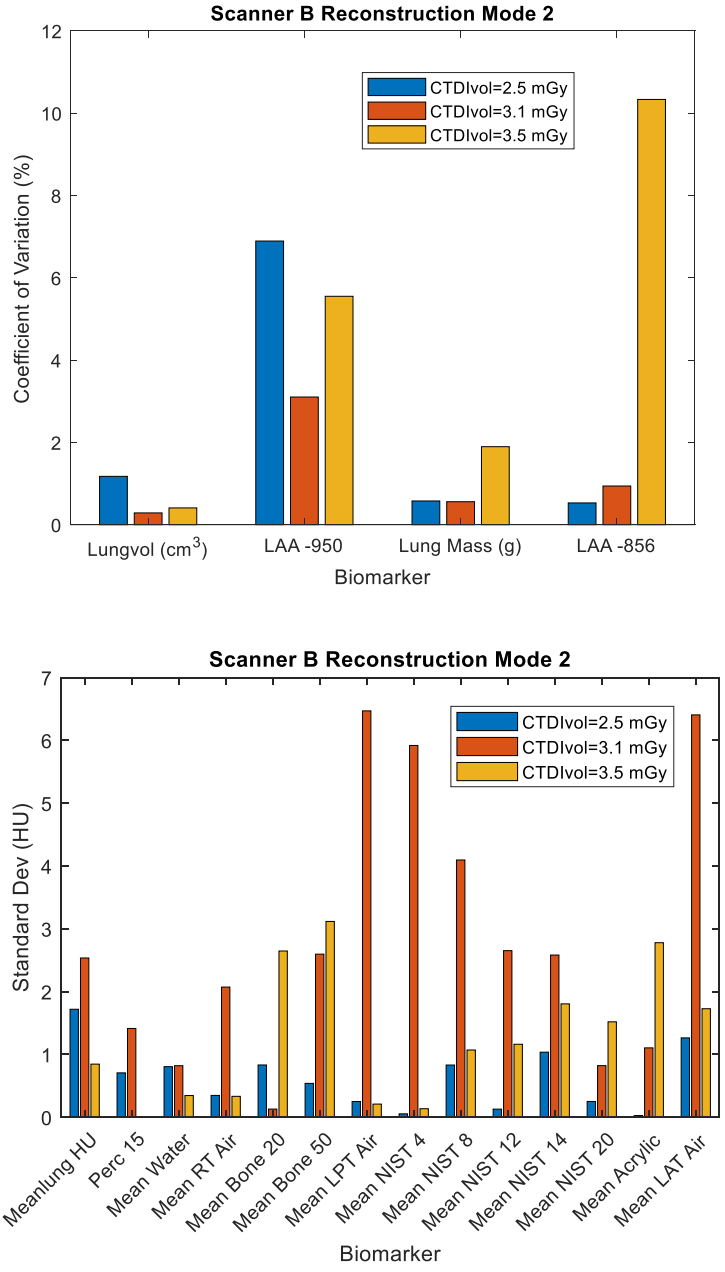

Figure 7: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 2 of Scanner B

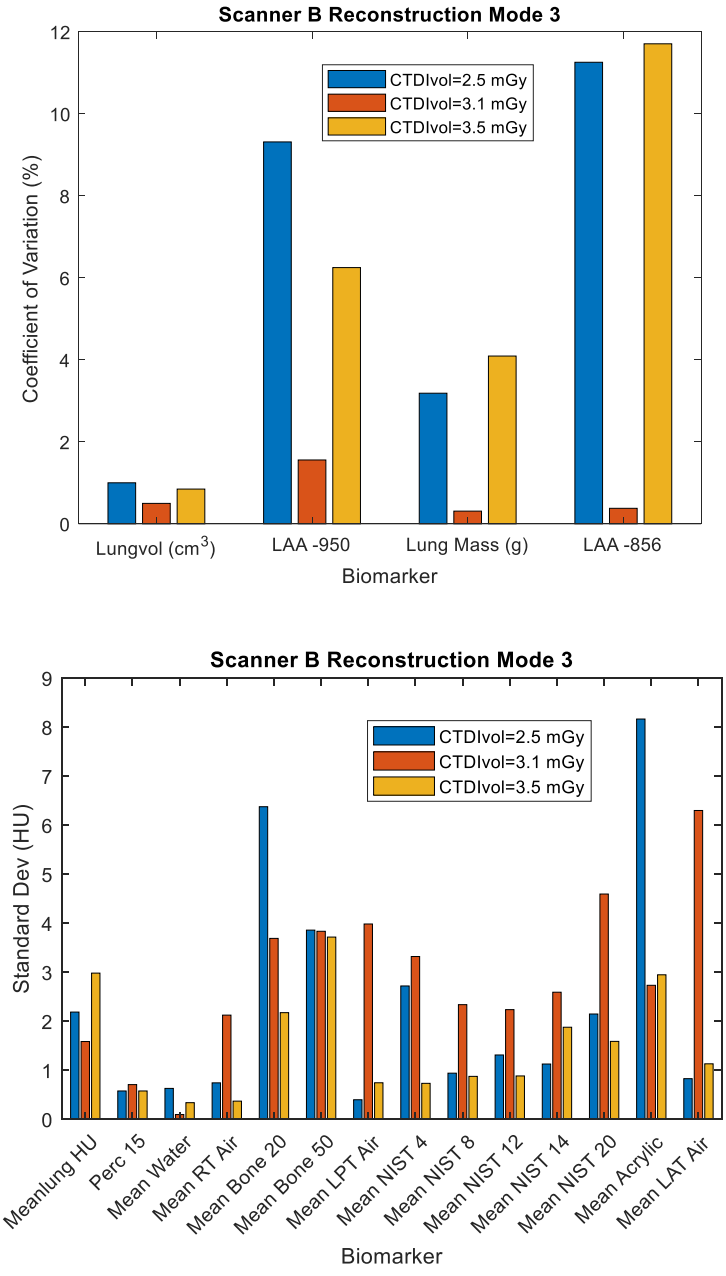

Figure 8: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 3 of Scanner B

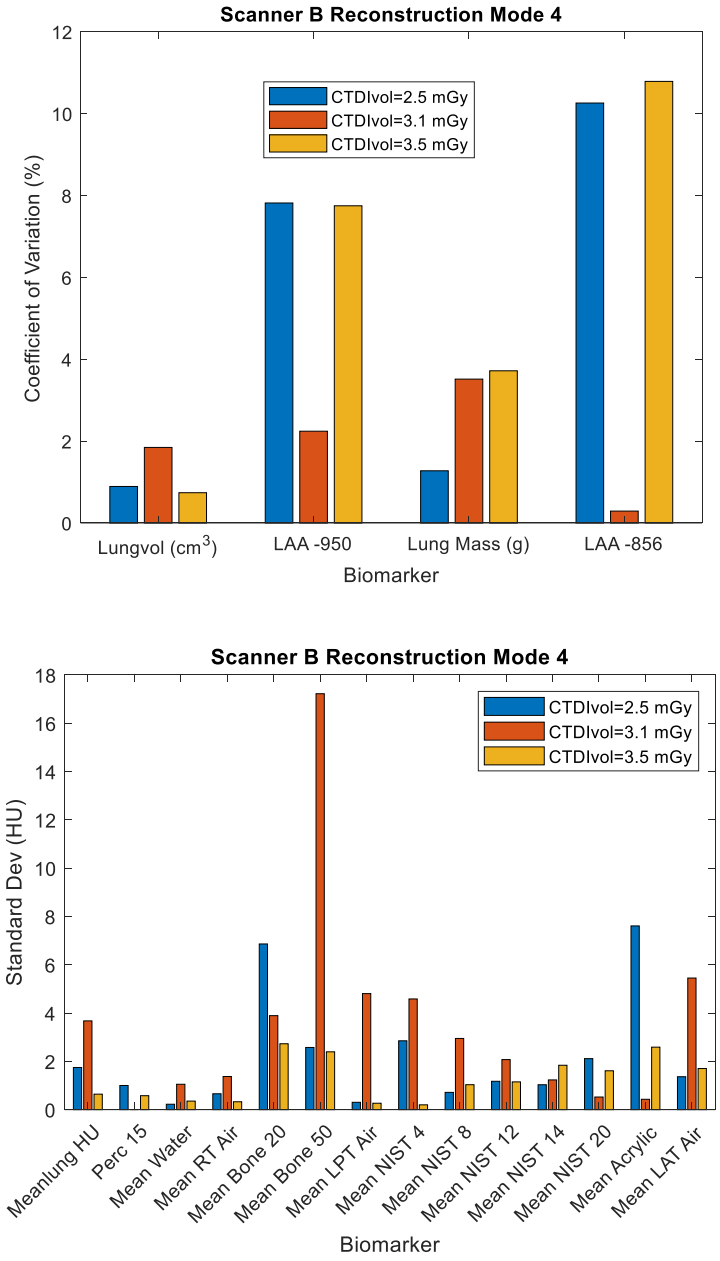

Figure 9: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 4 of Scanner B

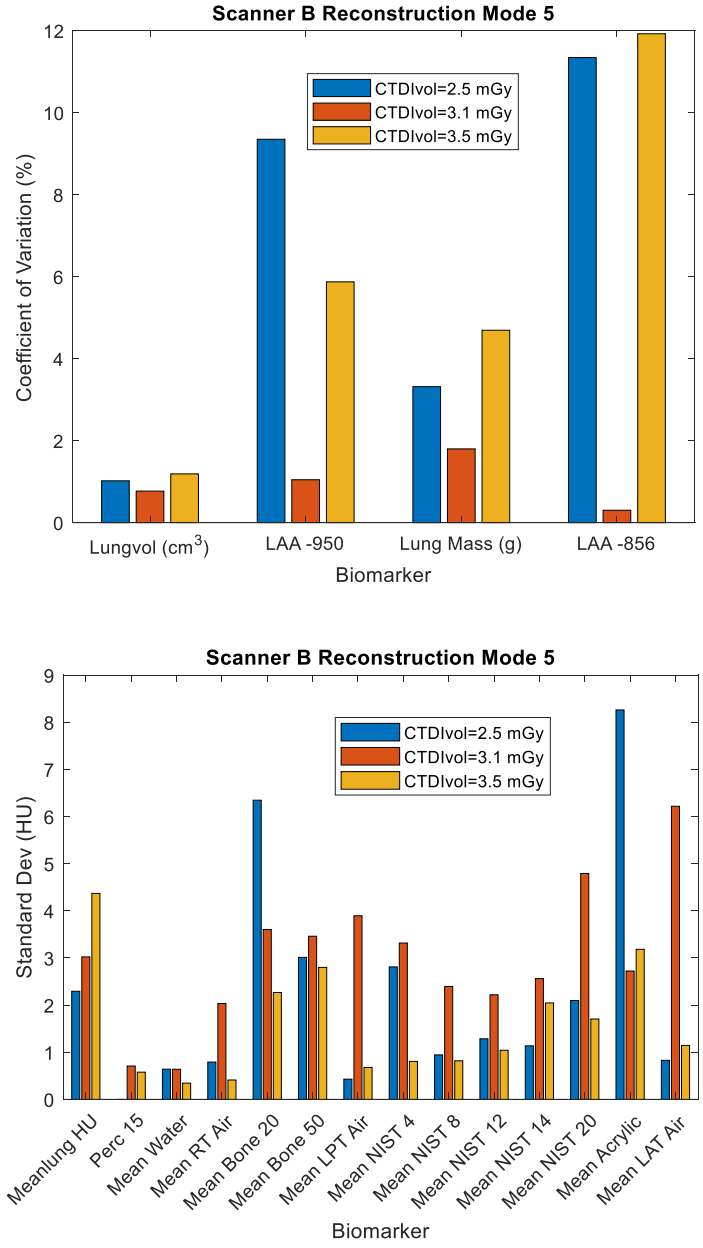

Figure 10: Intra-scan variability of non-HU (left) and HU (right) unit biomarkers conditioned to CTDIvol, for Reconstruction Mode 5 of Scanner B

2.3.2 Inter-Scan Variability

In order to examine biomarker measurement variability across different scanners, the mean value of each biomarker was plotted as a function of CTDI, with each of the 10 scanners superimposed on each other in each plot (49 figures in total). However, no statistically meaningful trend was found to be present across the plots (relating either the biomarker to a particular scanner or CTDIvol value), so only a portion of the plots are shown below in Figure 11 (for demonstration purposes).

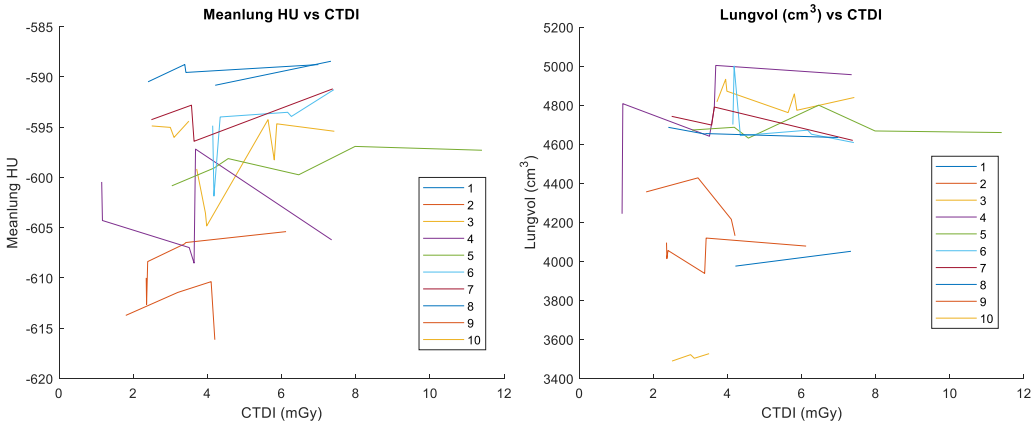

Figure 11: A subset of the 49 plots generated in the inter-scanner variability studies, from which no generalizable conclusion regarding trends between biomarkers and scanners could be drawn.

In addition, variability calculations for each scanner per biomarker across all imaging parameters used in the dataset were performed. Three different scanners which a great range of levels of variability with respect to one another (Scanners A, B, and C) were selected for this demonstration. The exact parameter space (with corresponding results in Figure 12) for each scanner can be seen in Table 4 below:

Table 4: Parameter space found in Scanners A, B, and C

	CTDIvol range (mGy)	kVp	Tube Current (mAs)	Voxel Size (mm^3)	Reconstruction Diameter	Number of Modes of Reconstruction
Scanner A	1.733 – 4.711	120	51 - 139	0.1728	260	4
Scanner B	2.5 -3.5	120	84-114	0.1287	259.765	5
Scanner C	3.04-11.40	120	80-300	0.1934	260	1

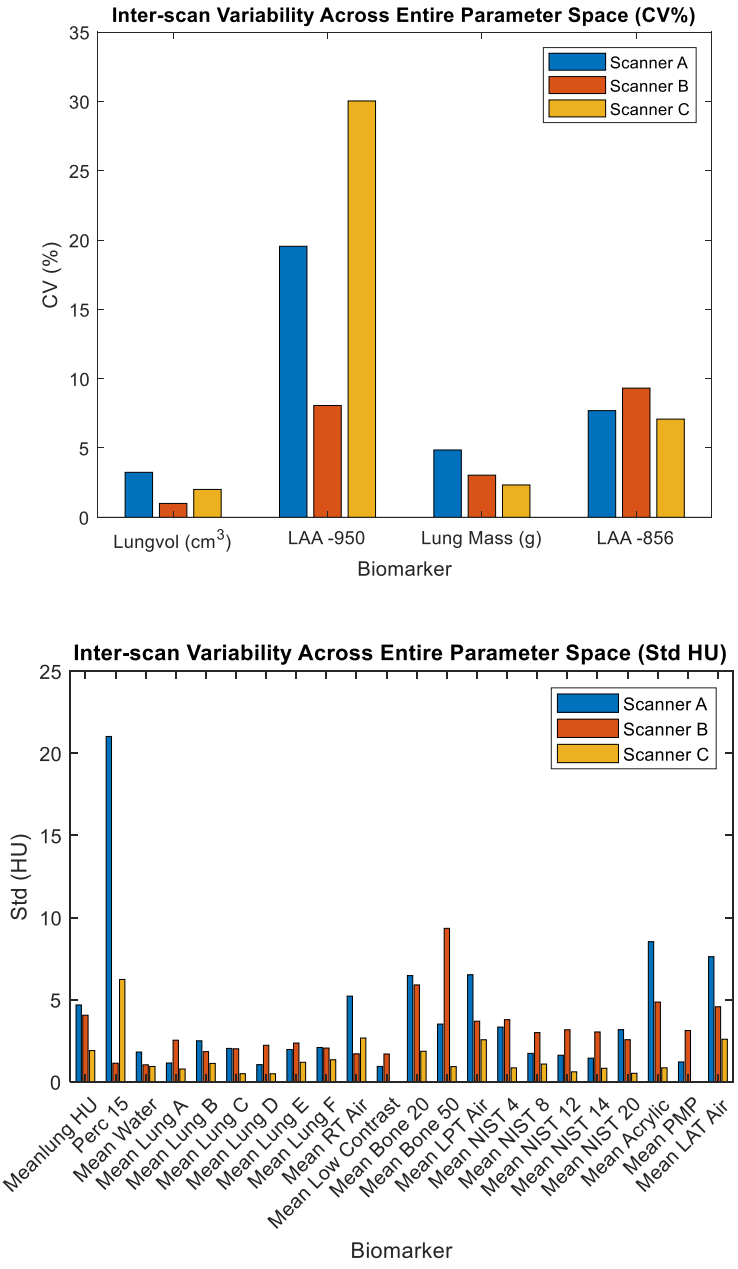

Figure 12: Variability of non-HU unit (left) and HU unit (right) biomarkers for each scanner across all imaging parameters used in the dataset.

2.4 Discussion

Regarding the results of intra-scanner variability (Figures 3 through 10), it can be seen that with respect to lung biomarkers, LAA -950 and LAA -856 tend to show the largest CV, often towering over other biomarkers at anywhere between 6% CV to 50% CV. One reason this could be is that very few lung HU voxels are as low as -950 and -856 (most are well above this), so since close to 0% of voxels are in this neighborhood small deviations from 0% could play bigger roles in terms of variability computations.

Generally speaking, all the other biomarkers showed smaller levels of intra-scan variability in terms of absolute number (HU biomarker usually had less than 10 HU standard deviation and non-HU biomarkers generally had a CV less than 5%).

Theoretically speaking, one would expect to see the general trend that as CTDI increases, variability across images tends to decrease. However, because each bar in these figures corresponds to three images and only a couple of different CTDI values (the range of CTDI values in Scanner B in Figures 6 through 10 are particularly small), these expected trends are not observed. Perhaps with a larger dataset, these expected trends would become more apparent.

Regarding the inter-scanner variability measurements in Figure 11, it can be seen that there is no clear consistent trend present between each biomarker and CTDI concerning which particular scanner was used, and that any potential corrections would not be as simple as a simple bias correction or a scanner-specific vertical shift. Perhaps

with more data points available (ex. more CTDIvol values used) or with a wider range of scanners, an overall trend could be more easily identified in future studies.

Looking at Figure 12, it can be concluded that the scanner model used does have an effect on the measured biomarkers, as for example Scanner C generally has a lower variability (HU biomarkers generally have 4 HU standard deviation and non-HU biomarkers generally have less than 10% CV) compared to the other two scanners. However, because the bars in Figure 12 represent aggregate variability measurements across the entire parameter space, it follows that scanners with broader ranges of variability in their imaging and reconstruction parameters will have a broader range of measured variability, so the above observation may be an artifact of this (ex. according to Table 3, Scanner C has only 1 reconstruction mode compared to the 4 and 5 reconstruction modes seen in Scanners A and B). Just like with the intra-scanner results, however, the variability for the LAA -950 and LAA -856 biomarkers seem to be the highest here, with the majority of the other biomarkers having variability less than either 10% CV or 10 HU standard deviation.

2.5 Conclusion

CT imaging biomarker variability has been demonstrated and analyzed in the context of real CT image data from a variety of scanners and imaging parameters though the usage of a clinically-relevant chest phantom. It has been shown that variability can be scanner, parameter, and biomarker specific and that analysis across different scanners

needs investigation in future research. Essentially, these variability findings may provide insight into how to harmonize these images in an effort to reduce measured variability for more accurate future quantifications in clinical contexts.

3. Validation of a Scanner-Specific CT Simulator

The second phase of this study was concerned with validating the accuracy of DukeSim in the context of Virtual Imaging Trials (VITs) for lung imaging applications. In brief, this was done by generating a computational version of the customized Kyoto Kagaku chest phantom, simulating this virtual phantom, and then comparing the resultant biomarkers from the simulation to the biomarkers from the corresponding real CT image with matching imaging and reconstruction parameters. The information presented in this chapter is planned to be submitted to Medical Physics. The contents of this chapter (text, tables, and figures) were largely adapted from that planned submission.

3.1 Introduction

The assessment and optimization of CT images for quantitative analysis traditionally involves working with real patient data, which can be challenging, expensive, and ground-truth limited [31]. With a growing number of CT scans performed every year, these limitations will only become more significant as time passes.

Alternatively, these assessments can be performed using realistic virtual simulations of the CT imaging process. These simulations can be conducted using Virtual Imaging Trials (VITs), which involves using computational models of the patient and imaging systems as opposed to using real patients imaged with clinical scanners.

The general idea of a VIT is shown below in Figure 13. VITs can overcome some of the challenges associated with conventional imaging trials. Because they are performed via computer simulation, they are easily replicable and cost efficient [8]. Additionally, VITs include the ground truth as the exact structural anatomy and material composition of the patient model is known [32]. Because VITs are easily replicable, they enable multiple image acquisitions of virtual patients without typical issues regarding patient data privacy and unnecessary radiation exposure [8].

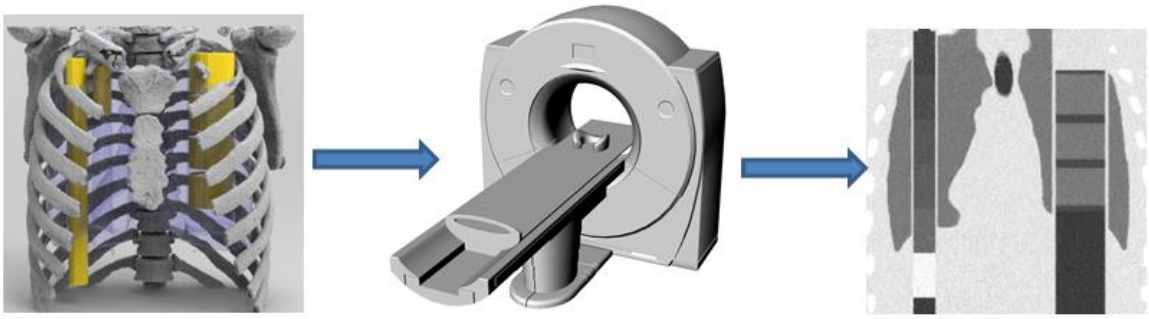

Figure 13: A depiction of a Virtual Imaging Trial (VIT) platform. A computer-generated human phantom is shown depicting a virtual patient (left) which is virtually scanned (middle) and a coronal view corresponding to the simulation result (right) using Window level -600 and Width 1500.

For virtual trials in CT, a major component is realistic simulation of the CT image acquisition process. DukeSim is a rapid and scanner-specific imaging simulator that produces realistic CT projection images given a high-resolution virtual phantom combining ray-tracing and Monte Carlo techniques [33-35]. DukeSim has been validated against experimental data acquired from various scanners and imaging conditions using cylindrical phantoms with various diameters using measurements of image contrast,

noise magnitude, noise texture, and spatial resolution [33, 36]. However, simple cylindrical-shaped phantoms are not seen in a clinical setting, thereby indicating a need for further investigations on the accuracy and reliability of DukeSim virtual simulations for more clinically relevant phantoms. For lung imaging applications in particular, a CT simulation needs to be validated against a corresponding anthropomorphic chest phantom and this must be done using more clinically relevant image quality metrics [37].

The purpose of this portion of the study was to validate the accuracy of DukeSim for lung imaging trials by simulating images of a computational version of an anthropomorphic chest phantom and comparing them to real images acquired with clinical scanners. These comparisons were done in terms of clinically relevant lung imaging biomarkers.

3.2 Methods

3.2.1 Chest Phantom

Similar to phase one of this study, a customized (University of Iowa Designed) version of the commercially available physical Kyoto Kagaku Multipurpose Chest Phantom N1 "LUNGMAN" (PH-1) was used [25-27]. Again, three vertical tubes were placed into the lung models, with one 59 mm diameter tube passing through the right lung and two 28 mm diameter tubes passing through the posterior and anterior portions of the left lung model. Similar to phase one, there were three configurations of tube

inserts used. The lung samples were fixed porcine lung cored samples from an inflated and an over-inflated pair of lungs and standardized foams acquired from the National Institute of Standards (NIST) were used [28]. Visuals of these setups for the chest phantom can be seen in Figure 1 and Table 1 of Chapter 2.

3.2.2 Real CT Data

The Kyoto-Kagaku anthropomorphic chest phantom was imaged with two clinical CT scanners (Siemens SOMATOM Definition Flash (n=15 images) and Siemens SOMATOM Force (n=13 images)) at 120 kVp at the University of Iowa. Both scanners in this dataset used either iterative approaches to reconstruction or weighted filtered back projections (wFBP) [38]. An overview of the 28 real CT images and relevant imaging parameters are listed in Table 5 below.

Table 5: CT scanners (rows) used to image the physical Kyoto-Kagaku Chest Phantom and imaging conditions (columns) utilized.

	Number of Images	CTDIvol Range (mGy)	kV	Reconstruction Algorithm	Reconstruction Kernel	Pixel Spacing (mm) x (mm)	Slice Thickness (mm)
Siemens SOMATOM Definition Flash	15	3.71- 7.42	120	SAFIRE Strength 3 & 5 or wFBP	Q30f or B35f	0.5078 x 0.5078	0.75
Siemens SOMATOM Force	13	1.15-7.35	120	ADMIRE Strength 3 & 5 or wFBP	Qr40d or Bf40d	0.5078 x 0.5078 or 0.9766 x 0.9766	0.75

3.2.3 Virtual Imaging Trial (VIT)

Acquisitions replicating the same phantom and image conditions were simulated. A computational model for the Kyoto-Kagaku chest phantom was created in a voxelized format with a 0.25 mm voxel size. To do this, a real CT scan of the physical Kyoto-Kagaku phantom was manually segmented using an open-source software (Seg3D [39]) to delineate masks representing various regions in the phantom (e.g.,. lung, thoracic bones, inserts, etc.). A mid-level region of this mask is shown below in Figure 14. The elemental compositions and volumetric densities (obtained either from vendor-specific information or an x-ray library database [40]) of each of the segments were used to compute their corresponding linear attenuation coefficients (μ).

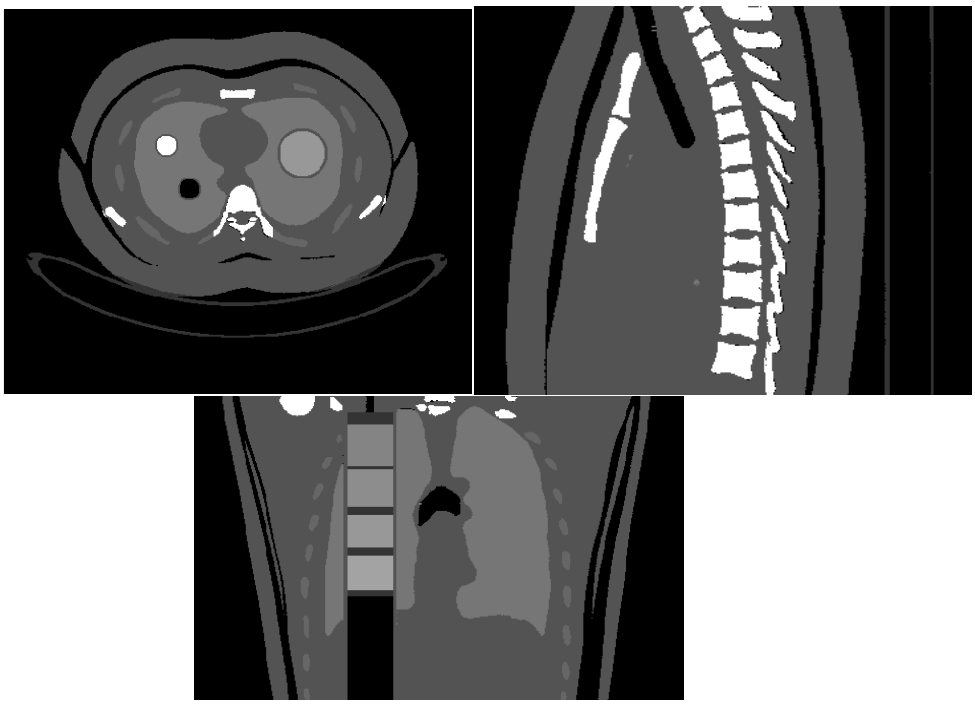

Figure 14: Computational Chest Phantom seen from an axial (top left), sagittal (top right), and coronal (bottom) view. Each shade on represents a different material

For the Low Contrast material (which had an unknown elemental composition), a similar liver tissue was utilized. The density of the synthetic lungs (a urethane foam made from a urethane based resin) was found by tuning the HU values in the simulated images to find a reasonable estimate. This was corroborated by computing the ratio of the lung mass and volume across multiple real CT images of the Kyoto phantom.

The above computational phantom, corresponding material information, and scanner specifications were inputted to DukeSim to generate projection data under the same imaging conditions used in the real acquisitions for each image in the dataset. The simulated projection data were reconstructed using a vendor-specific reconstruction software (ReconCT, version 15.0.53098.0) using the same reconstruction algorithm and imaging conditions.

3.2.4 Biomarker Analysis

Comparisons between the real (1) and simulated (2) images were made by computing various clinically relevant lung imaging biomarkers: lung density(average HU values in the lungs) , LAA -950 (percentage of lung voxels with an HU ≤ -950, which relates to airflow obstruction), LAA -856 (percentage of lung voxels with HU ≤ -856 a measure of gas trapping), Perc 15 (HU value at the 15th percentile of all lung voxels, used for measuring emphysema progression), lung mass, and lung volume [29]. A more in depth discussion of each of these biomarkers can be found in Table 3 of chapter 2.

Additionally, the mean and standard deviation of the HU values was computed in each sample region of each of the three tubes for Configurations 1, 2, and 3.

Differences in computed biomarkers were computed as a relative percent error defined as follows:

$$Percent\ Error = 100 * \frac{|Real - Simulated|}{Real}$$

in which Real is the computed biomarker in the real CT scan (1) and Simulated (2) is the computed biomarker in the reconstruction of the DukeSim simulated image. For LAA -950 and LAA -856, absolute percent differences were computed as these measures are natively expressed as a percentage.

Additionally, a two-sample difference of means test was conducted to evaluate the closeness of each simulated measurement to its corresponding real data [41]. In particular, a Welch's t-test (which assumes unequal variances for unlinked groups [42]) was performed. The null hypothesis assumed that the difference between the means between real and simulated for each biomarker was zero, whereas the alternative hypothesis was that the difference between the two was non-zero.

3.2.5 MTF Bead Slice Sensitivity Analysis

As mentioned in Figure 1, Configuration 3 contains one very dense (~ 15 g/cm3) tungsten carbide bead surrounded by low contrast. In situations like these in which there is such a high gradient, a windmill artifact is likely to appear which will distort the

reconstructed image [43]. To study this artifact, slice sensitivity profiles were developed for each scanner. These profiles depict normalized HU value as a function of vertical slice position [35].

3.3 Results

Representative reconstructions of the simulated images alongside the corresponding real CT image can be seen below in Figure 15 for both scanners. Qualitatively, the simulated images have similar visual realism to the real ones.

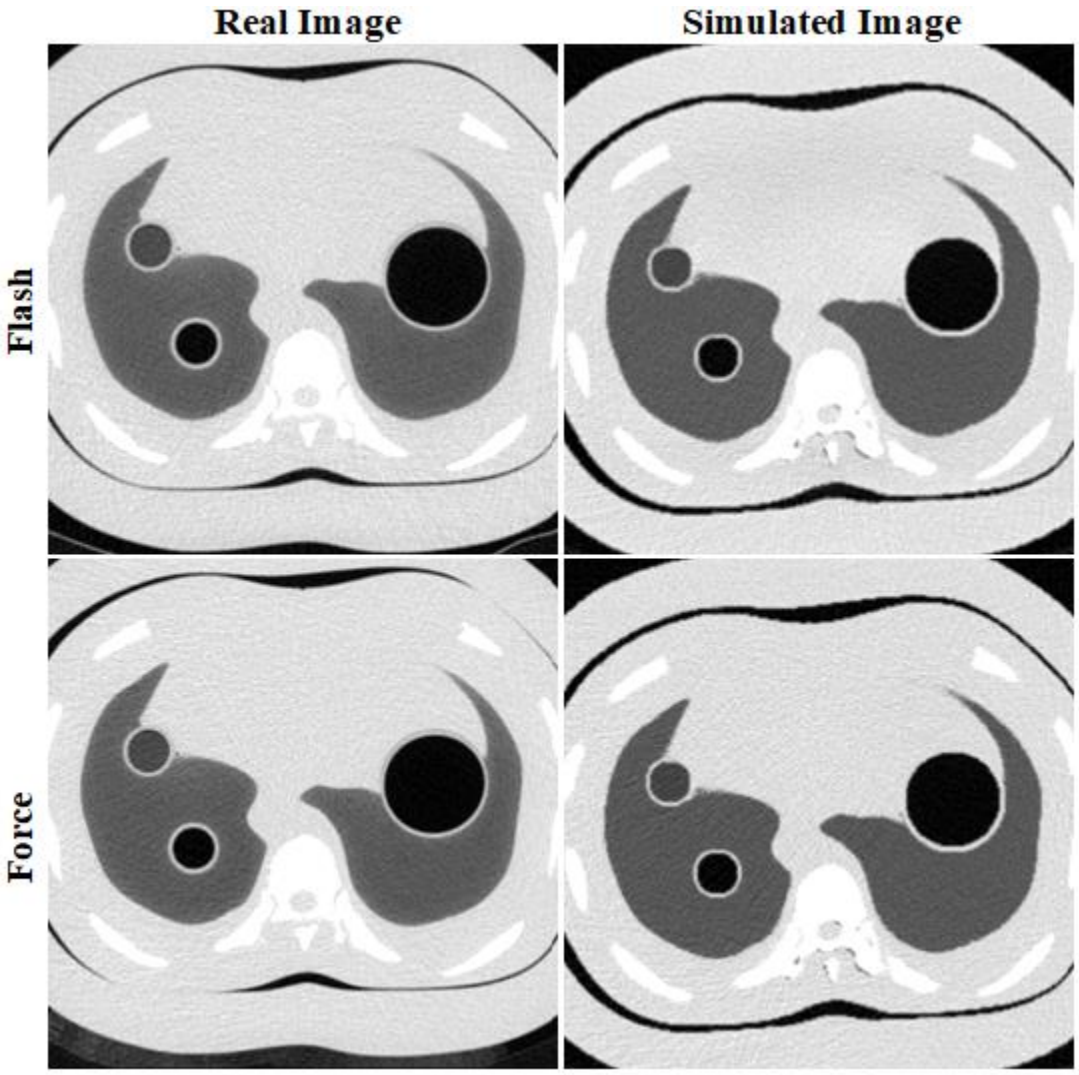

Figure 15: Low-level axial slice of Real and Simulated Images shown at Window level -437.5 and Width 1173. The top left image is the Real CT Image from Siemens SOMATOM Definition Flash and the top right is the corresponding simulated image. The bottom left is the Real CT Image from Siemens SOMATOM Force and the bottom right is the corresponding simulated image.

The quantitative comparisons for both scanner models are shown in Figures 16 and 17. Each bar corresponds to a particular biomarkers averaged across all images acquired from that scanner with the error bars corresponding to their standard deviation. The relative errors for the biomarkers between the real and simulated data were relatively small, with an average of only 3.84% error (range 0.187% to 18.269%). The error bars for some biomarkers overlap between real (red) and blue (simulated) data. The Force scanner generally had a higher magnitude for HU biomarkers across both real and simulated data compared to Flash, demonstrating that our DukeSim platform is sensitive to the differences between the scanner models.

Note that the "Water" insert results are not included since in the actual phantom the insert had extra materials which were unknown. In addition, the Lung Samples "A" through "F" were not evaluated since the inserts had textured materials that were unknown.

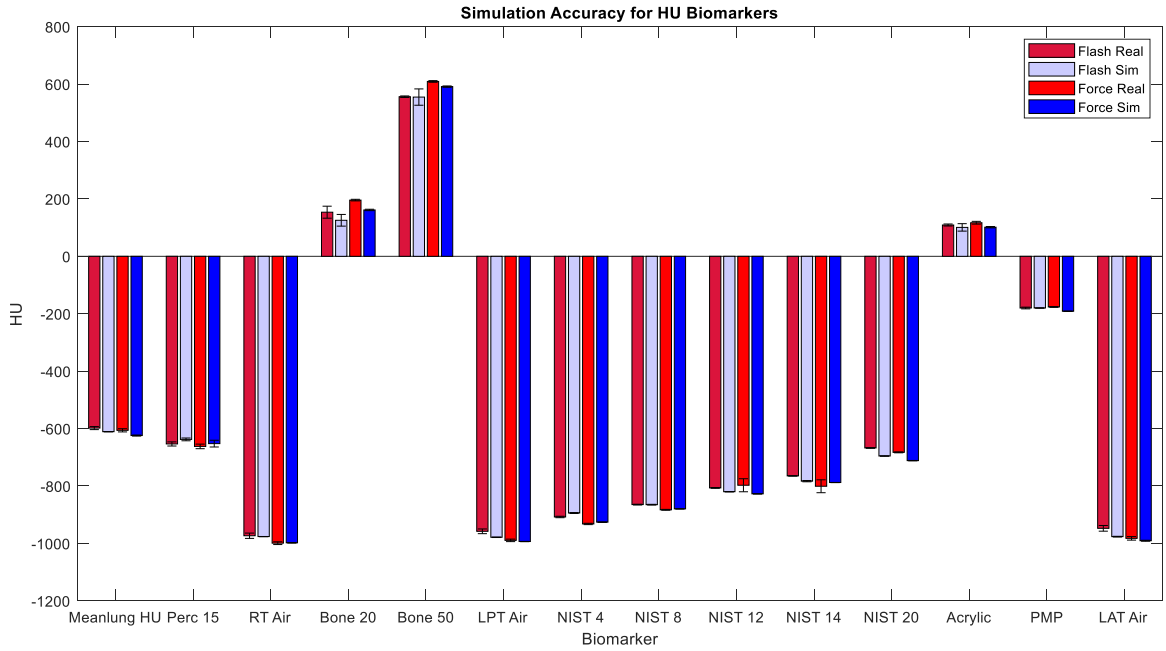

Figure 16: Comparison of the Biomarker in Siemens SOMATOM Definition Flash and Siemens SOMATOM Force for HU unit biomarkers. RT = Right tube, LPT = Left Posterior Tube, LAT = Left Anterior Tube

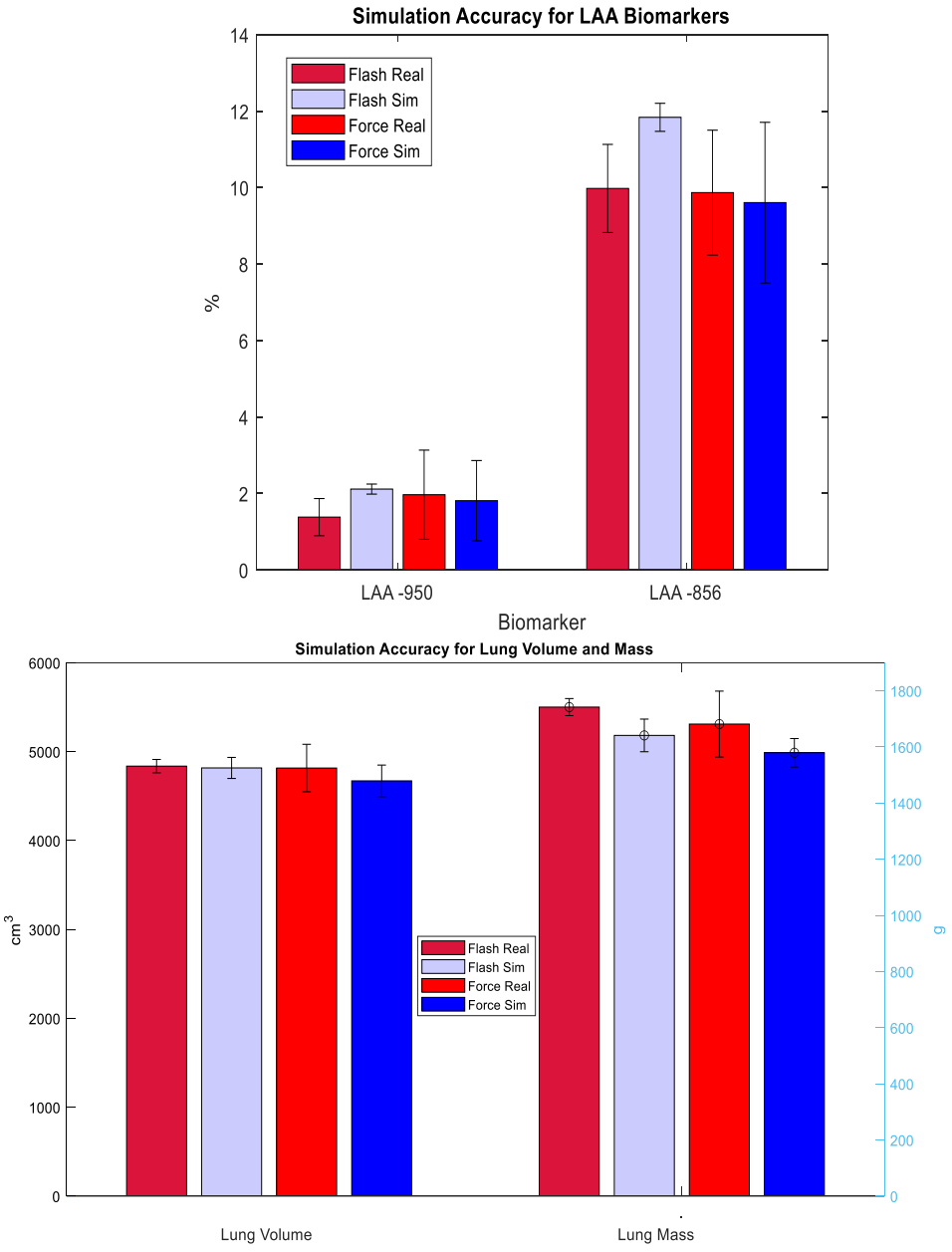

Figure 17: Comparison of the Biomarker in Siemens SOMATOM Definition Flash and Siemens SOMATOM Force for non-HU unit biomarkers: Percentage (top) and mass & volume (bottom). RT = Right tube, LPT = Left Posterior Tube, LAT = Left Anterior Tube

Following the results of Figures 16 and 17, the p-values for the two-tailed Welsch's t-test are shown in Table 6, demonstrating the relative closeness of the simulated results against real data. The biomarkers of the lung volume, RT Air, Bone 50, NIST 8, and PMP were less than 0.05 for the Siemens SOMATOM Definition Flash scanner and the biomarkers of the lung volume, LAA -950, LAA -856, RT Air, and NIST 14 were less than 0.05 for the Siemens SOMATOM Force scanner (in total only 28% of all 36 biomarkers).

Table 6: P-value results of Welsch's t-test

	Siemens Flash	Siemens Force
Mean Lung HU	2.45E-07	1.35E-08
Lung Volume (cm^3)	0.58223395	0.119923
LAA -950	3.65E-05	0.722341
Perc 15	2.32E-07	0.022376
Lung Mass (g)	6.31E-06	0.010711
LAA -856	1.63E-05	7.26E-01
Right Tube Air	0.2352481	0.367203
Bone 20	0.00083979	2.33E-21
Bone 50	0.85840501	7.24E-15
Left Posterior Tube Air	1.02E-07	0.006071
NIST 4	3.45E-15	1.63E-07
NIST 8	0.56108116	2.10E-06
NIST 12	3.21E-19	0.000457
NIST 14	9.93E-17	0.05884
NIST 20	8.57E-26	1.02E-19
Acrylic	0.02910012	5.96E-09
PMP	0.99970897	1.07E-06
Left Anterior Tube Air	1.58E-08	0.000362

Below in Figure 18 can be seen the slice sensitivity profiles for both scanners measured from both real and simulated data for one sample image. The sharp peak seen in each plot is caused by the windmill artifact caused by the high contrast from the very dense tungsten carbide bead. Each plot was produced by finding the slice sensitivity

profile (maximum HU voxel across each slice in the z-direction) and then normalizing from zero to one.

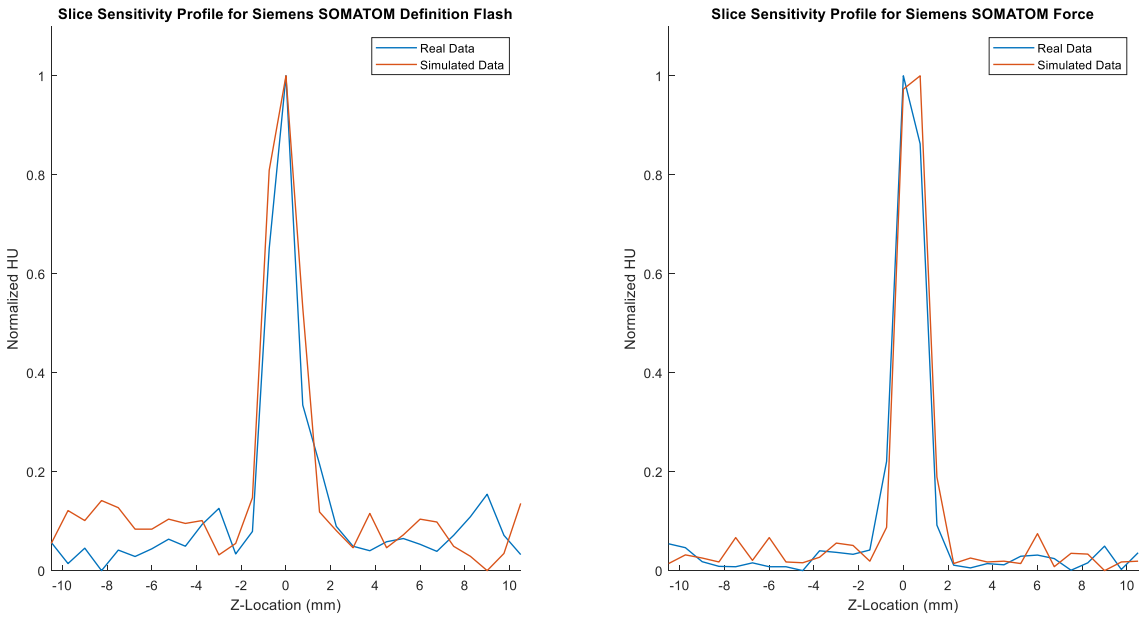

Figure 18: Slice Sensitivity profiles for each scanner comparing real and simulated data for one sample image

There is strong overlap of the peaks in the curves in Figure 18 above. To quantify each of these curves, the full width at half maximum (FWHM), expressed in mm, was computed. Table 7 below summarizes the FWHM results, with the errors of 0.4 and 0.15 mm for the Flash and Force scanners, respectively.

Table 7: Summary of the FWHM characteristics of the Slice Sensitivity Profiles. Each entry is the mean across all relevant images using std dev as the uncertainty.

	Real CT Data	Simulated Data
Siemens SOMATOM Definition Flash	1.46± 0.05 mm	1.85 ± 0.16 mm
Siemens SOMATOM Force	1.46± 0.07 mm	1.61± 0.04 mm

Hence, in terms of Slice Sensitivity, the Siemens SOMATOM Force scanner displayed more consistent behavior between the real and simulated results.

3.4 Discussion

This study validated the accuracy and reliability of DukeSim using an anthropomorphic phantom in using clinically relevant biomarkers. The measured biomarkers were highly close to the real measurements (3.8% on average) across the imaging conditions and the scanner models that were utilized in this study. The accuracy and sensitivity of DukeSim to various imaging settings, parameters, and biomarkers enable efficient and comprehensive virtual imaging studies for various imaging applications.

While the differences between the real and simulated measurements were generally small (3.8% on average). There are various potential reasons for the differences between the real and simulated measurements. For the density-based biomarkers, the accurate knowledge of the materials' density and elemental composition is crucial. While we knew those values for most materials, the values for the manufactured versions might be slightly different, resulting into differences in the final HU values. Further, there are various simulation modules that affect simulation accuracy (e.g., detector model, approximation for the primary and scatter signal, beam hardening correction). While our simulation modules are state-of-the-art, they have some inherent

approximation and errors associated with them. It should be noted that there is also an inherent variability across different scanners of the same model. With the data not available in this study, future studies may quantify whether our simulation errors are within the range of scanner-to-scanner variability.

In terms of the slice sensitivity, we observed FWHM errors of 0.15 and 0.4 mm. The tungsten carbide bead was approximated as one voxel (0.25 mm), which may affect the accuracy of slice sensitivity measurements. The 0.25 mm voxel size has shown to be an effective resolution for density-based imaging applications. However, for high-resolution tasks (e.g., airway measurements), the simulations would be more accurate with smaller voxel sizes which make the overall simulations more computationally expensive. Generally though, the results of this study seen with MTF bead seem to be in line with the results of previous studies examining the windmill artifact [35, 43, 44].

3.5 Conclusion

Through this study, it has been shown that DukeSim does hold validity in the context of virtual imaging trials (VIT) using clinically relevant biomarkers and anthropomorphic phantoms. Because it has been validated, VIT technology can be applied to further diseases and cases related to CT imaging to give scientists and clinicians a more efficient, replicable, and robust method of evaluating clinical disease without the risk of unnecessary radiation exposures commonly seen with traditional imaging methods.

4. A Pilot Virtual Imaging Trial (VIT) Study: Variability of Imaging Biomarkers in Simulated CT Image Data

The third phase of this study involved applying a VIT to demonstrate the utility of a validated CT simulator by simulating images of the chest phantom with different tube currents and reconstruction kernels to see how this would affect the measurement of clinically relevant biomarkers. The information presented in this chapter is planned to be submitted to Medical Physics. The contents of this chapter (text, tables, and figures) were largely adapted from that planned submission.

4.1 Introduction

Virtual Imaging Trials of CT simulations through DukeSim have been shown to be reasonably valid, as demonstrated in Chapter 3. With this in hand, variability in biomarker measurements can now be examined in the context of virtual simulations. Analysis of variability of measurements has been done before for real CT image data [15-18], but this study shall seek to do so in a virtual context. The accuracy of these virtual simulations can also be assessed. Again, the main advantage s of doing these variability studies virtually is that they are relatively efficient and avoid unnecessary radiation exposure. A depiction of the VIT performed in this part of the study can be seen in Figure 13 in Chapter 3.

4.2 Methods

A computational version of the customized Kyoto Kagaku chest phantom was used. Like before, three vertical tubes containing experimental sample inserts were utilized. For simplicity, however, only Configuration 1 (see Figure 1 in Chapter 2) was simulated.

Similar to phase two of this study, the computational phantom was created in a voxelized format with a 0.25 mm voxel size by manually segmenting (using Seg3D [39], an open source software) a real CT scan of the Kyoto Kagaku chest phantom. Masks representing various materials in the phantom were compiled together. The elemental compositions and volumetric densities of each of the segments from phase two of this study were reused, in order to compute the corresponding linear attenuation coefficients. Views of the final computational phantom can be seen in Figure 14 of Chapter 3.

For this particular application, virtual simulations of the Siemens SOMATOM Force CT scanner were performed at 120 kV and a pitch of 1.0. Three different tube current levels were used: 20 mAs, 100 mAs, and 300 mAs. All reconstructions were performed using iterative ADMIRE methods of strength 3 with utilizing vendor-specific software (using ReconCT, version 15.0.53098.0). The reconstruction kernel used varied, either Qr32f, Qr40f, or Qr69f. The number in each of these kernels corresponds to improvements in sharpness with Qr69f being the most sharp and Qr32f being the least.

The pixel size was held at 0.5078 x 0.5078 mm with a slice thickness of 0.75 mm. Based on previous work [45], it is known that these changes in sharpness will impact image and noise quality. The same biomarkers as phase one and phase two will be used for this study: mean and standard deviation of the lung HU, LAA -950, LAA -856, Perc15, lung mass, lung volume, and the mean and standard deviation of the HU values in each sample region of the three tubes. The corresponding differences in biomarker measurements across different tube currents and reconstruction kernels will be observed as a measure of biomarker variability in the context of CT lung imaging.

The accuracy of these simulations was measured with respect to a known ground truth. For Siemens SOMATOM Force simulations, the ground truth HU values can be computed from the linear attenuation coefficients (μ) corresponding to an effective energy of 67 kV using the following relationship for HU as a function of μ:

$$HU(\mu) = 1000\frac{\mu - \mu_{Water}}{\mu_{Water} - \mu_{Air}}$$

where μ represents the linear attenuation coefficient at 67 kV for some material.

4.3 Results

Shown below in Figure 19 are reconstructed images of the combinations of the dose levels (20, 100, and 300 mAs) and reconstruction kernels (Qr32f, Qr40f, and Qr69f)

used in this study.

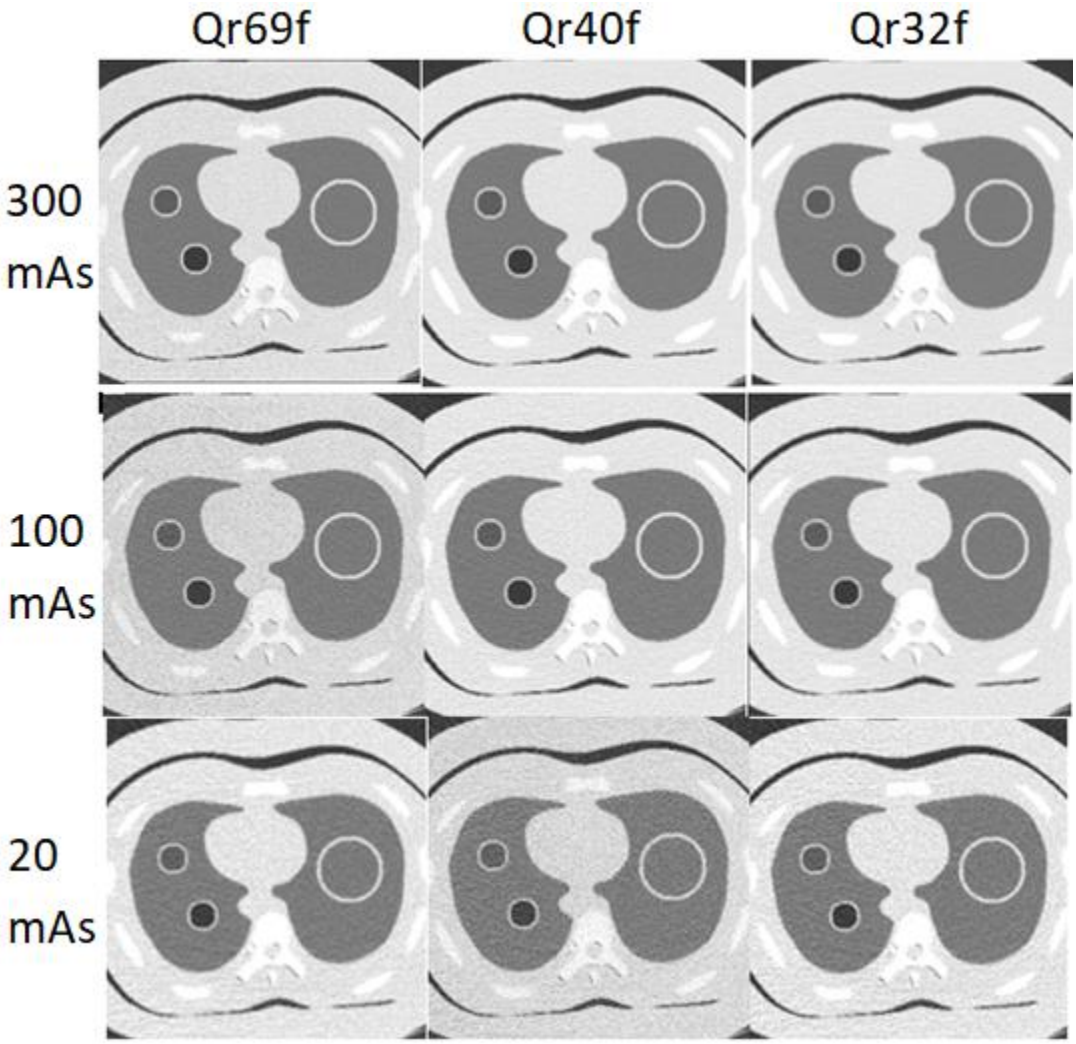

Figure 19: Mid-level axial views of all combinations of tube current and reconstruction kernels used in this phase of the study. The sharper kernels and lower dose images yield to noisier renditions.

Shown in Figures 20 through 22 below are plots of the errors for each biomarker with respect to the corresponding ground truth value, across the dose levels and reconstruction kernels. In general, the errors increased for the lower dose levels and

sharper kernels. Under the conditioned that were studied, the results demonstrated that kernels have higher impact on the biomarker accuracy compared to the tube current. For the low-attenuating objects, the measured values overestimated the ground truth values, whereas the high-attenuating objects overestimated the ground values, both of them likely due to the effects of scatter and beam hardening. The error for the air tube at the right and left lungs were also different, demonstrating that these errors depend on the location and shape of the imaged object.

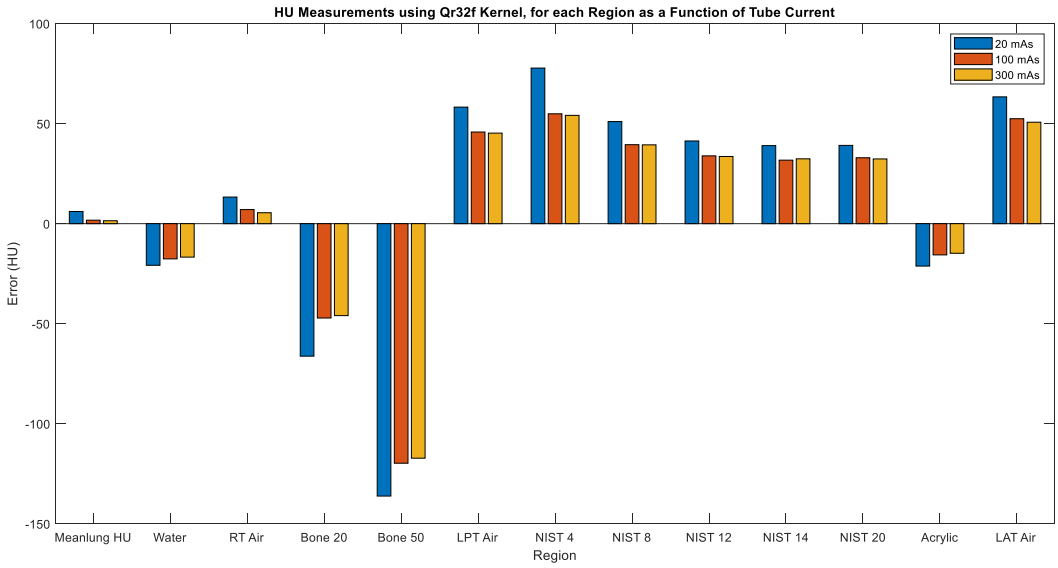

Figure 20: HU Biomarker errors for each region as a function of tube current using reconstruction kernel Qr32f

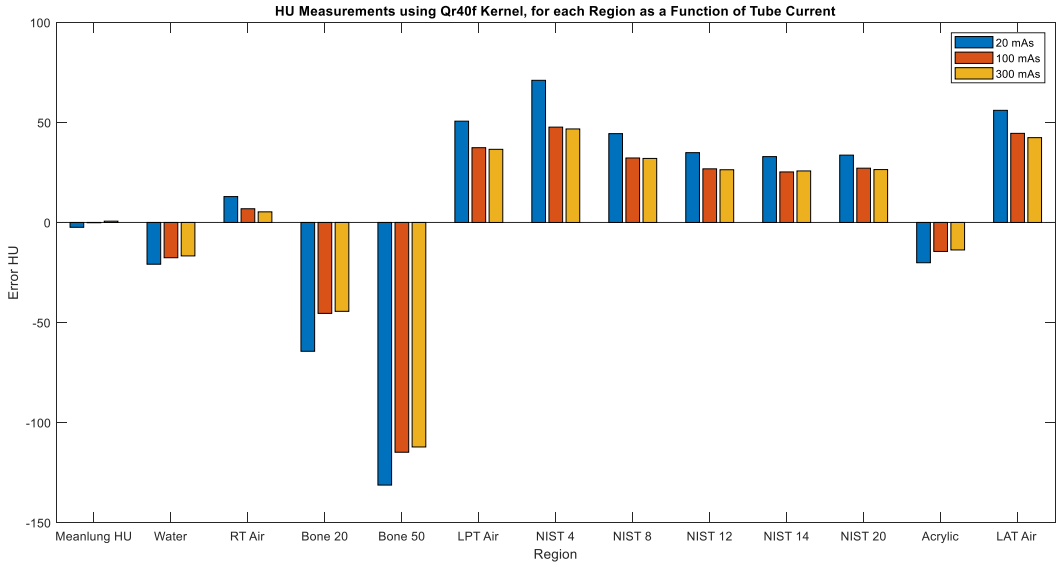

Figure 21: HU Biomarker errors for each region as a function of tube current using reconstruction kernel Qr40f

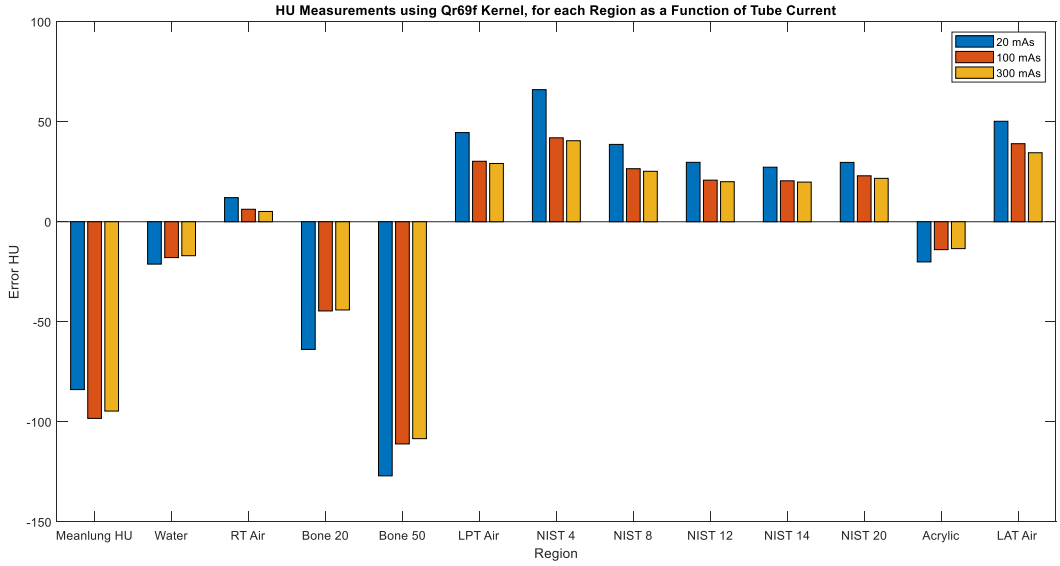

Figure 22: HU Biomarker errors for each region as a function of tube current using reconstruction kernel Qr69f

Figure 23 below shows a heat map visualizing the effects of tube currents and reconstruction kernels on the corresponding accuracy of the biomarker measurements for lung volume and lung mass biomarkers. Higher tube currents and softer kernels had lower errors in measuring these biomarkers. In particular, sharper kernels yielded errors as large as 797 to 3785 cm^3 and lung mass errors of 800 to 1751 g. Lower tube currents yielded errors in lung volume as large as 66 to 3785 cm^3 and lung mass errors of 170 to 1751 g. For the sharpest kernel (Qr69f), the images were so noisy that the segmentation algorithm failed to segment some of the lung regions, resulting into highly large errors in the lung volume and lung mass biomarkers.

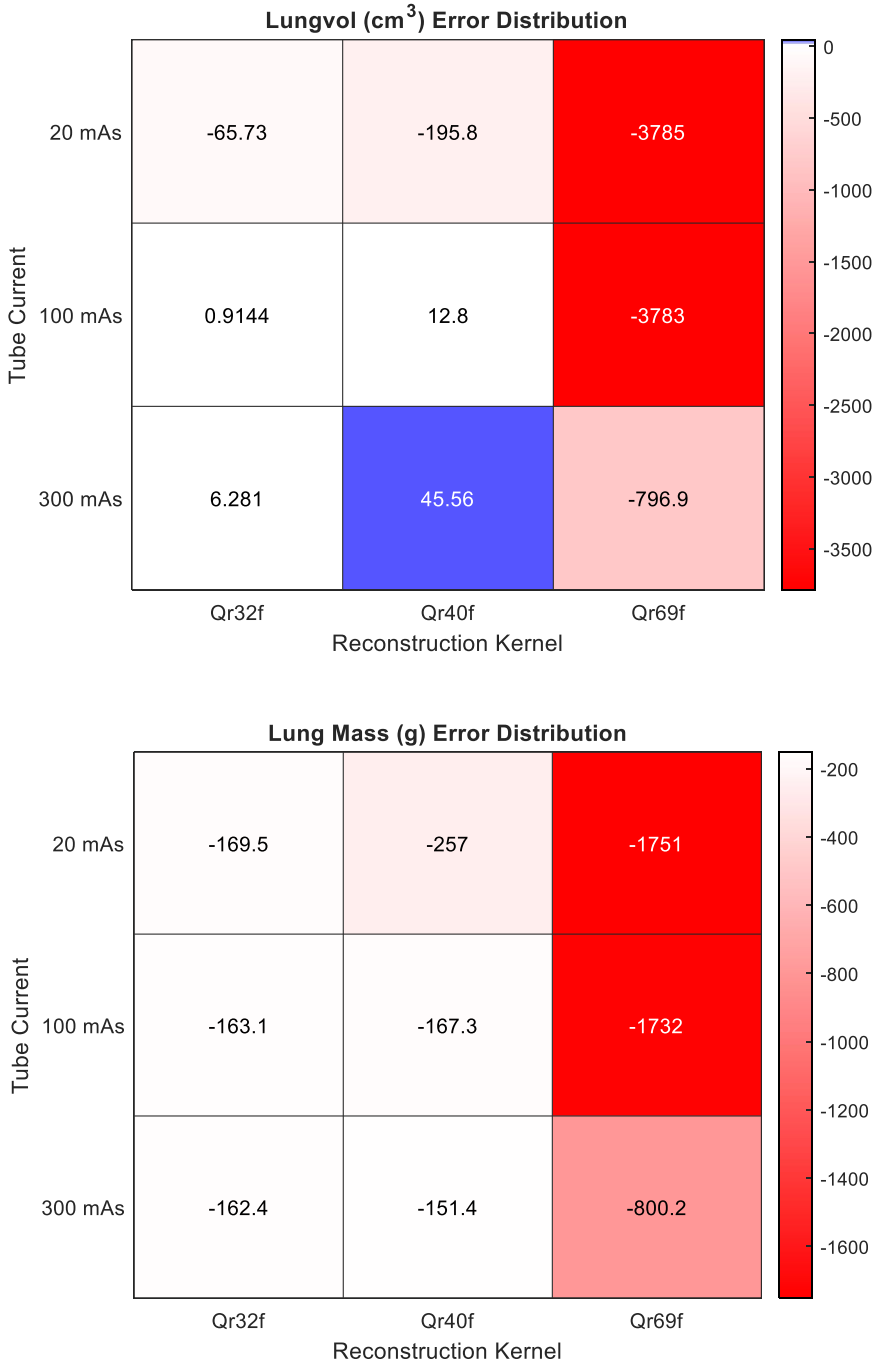

Figure 23: Heatmaps of the errors for the remaining biomarkers for demonstrating variability in a virtual context

4.4 Discussion

The intention of this study was to examine the reliability and variability of DukeSim through virtual CT imaging trials using different parameters. The pilot VIT study suggested that lower tube current and smoother kernels yield to more accurate quantifications. This is likely due to the reduced noise that is achieved with these conditions. Previous research is in line with this study, as others have found that after a certain point CT image noise greatly increases for even small changes in sharpness [45]. The pilot study also showed that the reconstruction kernels have significant effects on imaging biomarker quantifications, highlighting the significance and need for optimum protocol selections for quantitative imaging studies.

In the case of the sharpest kernel (Qr69f), the Otsu based segmentation algorithm failed to properly segment the whole lung, resulting into a high difference between the measured biomarkers and their corresponding ground truth. These results demonstrate that the biomarker accuracy depends on both imaging conditions and image processing algorithms (in this case segmentation method). This emphasizes the significance and need to explore optimum image processing algorithms towards accurate and robust CT quantifications.

4.5 Conclusion

The big take-away from this study is that varying imaging parameters and reconstruction approaches in the context of virtual simulations yields meaningful trends that further enhance the utility of VITs and open the doors for further studies in the realm of VITs. Additionally, this study in VITs has shown the value of being able to rapidly perform trials in some new imaging condition that has not been previously been examined before.

5. Conclusion

This study sought to demonstrate variability in clinically relevant biomarkers in real CT image data using a life-like chest phantom. Furthermore, it sought to validate a CT simulator (DukeSim) in this clinically relevant context and also understand how variation of certain imaging and reconstruction parameters affected biomarker measurements in virtual contexts.

It was found in regards to real CT image data that different scanners and biomarkers displayed different levels of intra-scan variability and that there was no simple trend concerning inter-scan variability. DukeSim was reasonably well validated for this anthropomorphic phantom, in that most simulated biomarkers were relatively close to their corresponding values in the real image. Lastly, it was found that sharper kernels and lower tube currents tend to reduce the accuracy of biomarker measurements in the virtual context. All three of these studies open the doors to future CT studies concerning accuracy and variability of images (in both real and virtual settings). In particular, VIT technology has a huge potential to be applicable to future study of relevant diseases, in part due to its advantages of relative efficiency and lack of radiation exposure. In the long term, these advances in CT technology have a huge scope to positively impact patient health outcomes in medicine.

References

[1] I. S. f. C. Tomography. "Half A Century In CT: How Computed Tomography Has Evolved." International Society for Computed Tomography. isct.org/computed-tomography-blog/2017/2/10/half-a-century-in-ct-how-computed-tomography-has-evolved (accessed 2021).

[2] H. H. Publishing. "Radiation risk from medical imaging." Harvard Health Publishing. health.harvard.edu/cancer/radiation-risk-from-medical-imaging (accessed 2021).

[3] M. Bosch de Basea *et al.*, "Trends and patterns in the use of computed tomography in children and young adults in Catalonia — results from the EPI-CT study," *Pediatric Radiology,* vol. 46, no. 1, pp. 119-129, 2016/01/01 2016, doi: 10.1007/s00247-015-3434-5.

[4] J. Hansen and A. G. Jurik, "Analysis of Current Practice of CT examinations," *Acta Oncologica,* vol. 48, no. 2, pp. 295-301, 2009/01/01 2009, doi: 10.1080/02841860802266714.

[5] D. Birenbaum, L. W. Bancroft, and G. J. Felsberg, "Imaging in acute stroke," (in eng), *West J Emerg Med,* vol. 12, no. 1, pp. 67-76, 2011. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21694755
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3088377/.

[6] M. G. M. Hunink and G. S. Gazelle, "CT screening: a trade-off of risks, benefits, and costs," (in eng), *J Clin Invest,* vol. 111, no. 11, pp. 1612-1619, 2003, doi: 10.1172/JCI18842.

[7] W. H. O. (WHO). "The top 10 causes of death." World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed 10-20-2021, 2021).

[8] E. Abadi *et al.*, "Virtual clinical trials in medical imaging: a review," (in eng), *J Med Imaging (Bellingham),* vol. 7, no. 4, p. 042805, Jul 2020, doi: 10.1117/1.Jmi.7.4.042805.

[9] R. Obuchowicz, M. Oszust, and A. Piorkowski, "Interobserver variability in quality assessment of magnetic resonance images," *BMC Medical Imaging,* vol. 20, no. 1, p. 109, 2020/09/22 2020, doi: 10.1186/s12880-020-00505-z.

[10] L. Basson *et al.*, "Chest Magnetic Resonance Imaging Decreases Inter-observer Variability of Gross Target Volume for Lung Tumors," (in English), *Frontiers in Oncology,* Original Research vol. 9, no. 690, 2019-August-13 2019, doi: 10.3389/fonc.2019.00690.

[11] A. Hagiwara, S. Fujita, Y. Ohno, and S. Aoki, "Variability and Standardization of Quantitative Imaging: Monoparametric to Multiparametric Quantification, Radiomics, and Artificial Intelligence," *Investigative Radiology,* vol. 55, no. 9, pp. 601-616, 2020, doi: 10.1097/rli.0000000000000666.

[12] H. Hatabu *et al.*, "Expanding Applications of Pulmonary MRI in the Clinical Evaluation of Lung Disorders: Fleischner Society Position Paper," *Radiology,* vol. 297, no. 2, pp. 286-301, 2020, doi: 10.1148/radiol.2020201138.

[13] M. Mohammadshahi, M. Alipouri Sakha, A. Esfandiari, M. Shirvani, and A. Akbari Sari, "Cost Effectiveness of Mobile versus Fixed Computed Tomography and Magnetic Resonance Imaging: A Systematic Review," (in eng), *Iran J Public Health,* vol. 48, no. 8, pp. 1418-1427, 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32292724
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7145907/.

[14] S.-H. Lee, H.-h. Cho, H. Y. Lee, and H. Park, "Clinical impact of variability on CT radiomics and suggestions for suitable feature selection: a focus on lung cancer," *Cancer Imaging,* vol. 19, no. 1, p. 54, 2019/07/26 2019, doi: 10.1186/s40644-019-0239-z.

[15] M. Liang *et al.*, "Variation in Screening CT–Detected Nodule Volumetry as a Function of Size," *American Journal of Roentgenology,* vol. 209, no. 2, pp. 304-308, 2017/08/01 2017, doi: 10.2214/AJR.16.17159.

[16] G. R. Oxnard *et al.*, "Variability of Lung Tumor Measurements on Repeat Computed Tomography Scans Taken Within 15 Minutes," *Journal of Clinical Oncology,* vol. 29, no. 23, pp. 3114-3119, 2011, doi: 10.1200/jco.2010.33.7071.

[17] J. A. Oliver, M. Budzevich, G. G. Zhang, T. J. Dilling, K. Latifi, and E. G. Moros, "Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer," *Translational Oncology,* vol. 8, no. 6, pp. 524-534, 2015/12/01/ 2015, doi: https://doi.org/10.1016/j.tranon.2015.11.013.

[18] S. J. v. Riel *et al.*, "Observer Variability for Classification of Pulmonary Nodules on Low-Dose CT Images and Its Effect on Nodule Management," *Radiology,* vol. 277, no. 3, pp. 863-871, 2015, doi: 10.1148/radiol.2015142700.

[19] R. B. Ger *et al.*, "Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies," *Scientific Reports,* vol. 8, no. 1, p. 13047, 2018/08/29 2018, doi: 10.1038/s41598-018-31509-z.

[20] S. Rajaraman, S. Sornapudi, P. O. Alderson, L. R. Folio, and S. K. Antani, "Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs," *PLOS ONE,* vol. 15, no. 11, p. e0242301, 2020, doi: 10.1371/journal.pone.0242301.

[21] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, "Variability and reproducibility in deep learning for medical image segmentation," *Scientific Reports,* vol. 10, no. 1, p. 13724, 2020/08/13 2020, doi: 10.1038/s41598-020-69920-0.

[22] T. P. Szczykutowicz, A. DuPlissis, and P. J. Pickhardt, "Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT," (in eng), *AJR Am J Roentgenol,* vol. 208, no. 5, pp. 1064-1072, May 2017, doi: 10.2214/ajr.16.17215.

[23] K. E. Stull, M. L. Tise, Z. Ali, and D. R. Fowler, "Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images," (in eng), *Forensic Sci Int,* vol. 238, pp. 133-40, May 2014, doi: 10.1016/j.forsciint.2014.03.005.

[24] B. M. Trotta, A. V. Stolin, M. B. Williams, S. B. Gay, A. S. Brody, and T. A. Altes, "Characterization of the relation between CT technical parameters and accuracy of quantification of lung attenuation on quantitative chest CT," (in eng), *AJR Am J Roentgenol,* vol. 188, no. 6, pp. 1683-90, Jun 2007, doi: 10.2214/ajr.06.1153.

[25] K. Murata. "Multipurpose Chest Phantom N1 "LUNGMAN"." Kyoto Kagaku. (accessed 10-12-2021, 2021).

[26] X. Xie *et al.*, "Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study," (in eng), *Eur Radiol,* vol. 23, no. 1, pp. 139-47, Jan 2013, doi: 10.1007/s00330-012-2570-7.

[27] T. Gomi, M. Nakajima, H. Fujiwara, and T. Umeda, "Comparison of chest dual-energy subtraction digital tomosynthesis imaging and dual-energy subtraction radiography to detect simulated pulmonary nodules with and without calcifications a phantom study," (in eng), *Acad Radiol,* vol. 18, no. 2, pp. 191-6, Feb 2011, doi: 10.1016/j.acra.2010.09.021.

[28] Z. H. Levine *et al.*, "A low-cost density reference phantom for computed tomography," *Med Phys,* vol. 36, no. 2, pp. 286-288, 2009, doi: https://doi.org/10.1118/1.3049596.

[29] S. P. Bhatt *et al.*, "Imaging Advances in Chronic Obstructive Pulmonary Disease. Insights from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) Study," (in eng), *Am J Respir Crit Care Med,* vol. 199, no. 3, pp. 286-301, 2019, doi: 10.1164/rccm.201807-1351SO.

[30] R. Ospina and F. Marmolejo-Ramos, "Performance of Some Estimators of Relative Variability," (in English), *Frontiers in Applied Mathematics and Statistics,*

Original Research vol. 5, no. 43, 2019-August-21 2019, doi: 10.3389/fams.2019.00043.

[31] H. L. Fred, "Drawbacks and limitations of computed tomography: views from a medical educator," (in eng), *Tex Heart Inst J,* vol. 31, no. 4, pp. 345-348, 2004. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15745283 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC548232/.

[32] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," (in eng), *Med Phys,* vol. 37, no. 9, pp. 4902-4915, 2010, doi: 10.1118/1.3480985.

[33] E. Abadi, B. Harrawood, S. Sharma, A. Kapadia, W. P. Segars, and E. Samei, "DukeSim: A Realistic, Rapid, and Scanner-Specific Simulation Framework in Computed Tomography," (in eng), *IEEE Trans Med Imaging,* vol. 38, no. 6, pp. 1457-1465, Jun 2019, doi: 10.1109/tmi.2018.2886530.

[34] G. Jadick, E. Abadi, B. Harrawood, S. Sharma, W. P. Segars, and E. Samei, "A scanner-specific framework for simulating CT images with tube current modulation," (in eng), *Phys Med Biol,* vol. 66, no. 18, Sep 13 2021, doi: 10.1088/1361-6560/ac2269.

[35] E. Abadi *et al.*, "Development of a scanner-specific simulation framework for photon-counting computed tomography," *Biomedical physics & engineering express,* vol. 5, no. 5, p. 055008, 2019.

[36] J. Giavanna, A. Ehsan, H. Brian, S. Shobhit, W. P. Segars, and S. Ehsan, "A framework to simulate CT images with tube current modulation," in *Proc.SPIE*, 2021, vol. 11595, doi: 10.1117/12.2580983. [Online]. Available: https://doi.org/10.1117/12.2580983

[37] Y. Cheng *et al.*, "Validation of algorithmic CT image quality metrics with preferences of radiologists," (in eng), *Med Phys,* vol. 46, no. 11, pp. 4837-4846, Nov 2019, doi: 10.1002/mp.13795.

[38] I. Dalehaug, K. Bolstad, D. Aadnevik, S. Flataboe, and H. E. S. Pettersen, "Admire vs. Safire: Objective comparison of CT reconstruction algorithms and their noise properties," *arXiv: Medical Physics,* 2017.

[39] *Seg3D*. (2016). Scientific Computing and Imaging Institute (SCI). Accessed: 2021. [Online]. Available: http://www.seg3d.org

[40] A. Brunetti, M. Sanchez del Rio, B. Golosio, A. Simionovici, and A. Somogyi, "A library for X-ray–matter interaction cross sections for X-ray fluorescence applications," *Spectrochimica Acta Part B: Atomic Spectroscopy,* vol. 59, no. 10, pp. 1725-1731, 2004/10/08/ 2004, doi: https://doi.org/10.1016/j.sab.2004.03.014.

[41] P. Schober and T. R. Vetter, "Two-Sample Unpaired t Tests in Medical Research," *Anesthesia & Analgesia,* vol. 129, no. 4, 2019. [Online]. Available: https://journals.lww.com/anesthesia-analgesia/Fulltext/2019/10000/Two_Sample_Unpaired_t_Tests_in_Medical_Research.3.aspx.

[42] Z. Lu and K.-H. Yuan, "Welch's t test," 2010, pp. 1620-1623.

[43] M. Silver, K. Taguchi, I. Hein, B. Chiang, M. Kazama, and I. Mori, *Windmill artifact in multislice helical CT* (Medical Imaging 2003). SPIE, 2003.

[44] K. Brown and S. Žabic, *Method for reducing windmill artifacts in multislice CT images* (SPIE Medical Imaging). SPIE, 2011.

[45] K. Eldevik, W. Nordhøy, and A. Skretting, "Relationship between sharpness and noise in CT images reconstructed with different kernels," (in eng), *Radiat Prot Dosimetry,* vol. 139, no. 1-3, pp. 430-3, Apr-May 2010, doi: 10.1093/rpd/ncq063.