Improving Dose-Response Correlations for Locally Advanced NSCLC Patients Treated with IMRT or PSPT

Yulun He

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Oncology Commons, Radiation Medicine Commons, and the Radiology Commons

Recommended Citation
He, Yulun, "Improving Dose-Response Correlations for Locally Advanced NSCLC Patients Treated with IMRT or PSPT" (2023). The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access). 1253.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1253

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digcommons@library.tmc.edu.
Improving Dose-Response Correlations for Locally Advanced NSCLC Patients Treated with IMRT or PSPT

by

Yulun He, B.S.

APPROVED:

Kristy Brock, Ph.D.
Advisory Professor

Radhe Mohan, Ph.D.
Advisory Professor

Zhongxing Liao, M.D.

Carol Wu, M.D.

Laurence Court, Ph.D.

Carlos Cardenas, Ph.D.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences
Improving Dose-Response Correlations for Locally Advanced NSCLC Patients Treated with IMRT or PSPT

A Dissertation

Presented to the Faculty of

The University of Texas
MD Anderson Cancer Center UTHealth
Graduate School of Biomedical Sciences

in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

by

Yulun He, B.S.
Houston, Texas
May 2023
Acknowledgements

First, I wish to express my sincerest gratitude to UTHealth and MD Anderson for the support and the vast array of resources that have made the past six years the most fruitful of my life, and I am proud to be representing UT in the next chapter of my career.

I would like to acknowledge my advisor Dr. Kristy Brock, for the unmatched mentorship that you provided. With your guidance, you shaped me into a problem solver and critical thinker, and I appreciate your frequent questions that you asked back to me, which helped me think outside the box and delve deeper into the complexities of medical physics. I will forever remember our last-minute meetings and extended paper discussions. Pursuing my PhD under your mentorship has been one of the best decisions I have made in my academic career.

I want to express my sincere gratitude to my co-advisor Dr. Radhe Mohan. I knew I wanted to work with you from the moment you shared your cell number with me during my research tutorial. Throughout my academic journey, you have been a consistent source of guidance and support. Your willingness to answer my questions, no matter how trivial they may have seemed, has been invaluable. Thank you for your consistent guidance and for inspiring me to stay curious. Your contributions to the field of medical physics are truly invaluable, and I feel fortunate to have had the opportunity to work with you.

To the rest of my advisory committee: Dr. Laurence Court, Dr. Carlos Cardenas, Dr. Carol Wu, and Dr. Zhongxing Liao, I am grateful for your support and guidance throughout my dissertation process. I appreciate the constructive feedback you provided throughout my dissertation process, challenging me to be a better scientist and helping me to improve the quality of my work. Your expertise and insights have been crucial in shaping my research. I would also like to extend my gratitude to each member individually. Dr. Court, thank you for always being there to offer advice on graduate school and career path. Your mentorship has
been invaluable. Dr. Cardenas, thank you for your continued interest in staying on my committee after your departure, and I hope you can continue to grow your research program successfully at UAB. Dr. Wu, I appreciate your expertise and patience in helping me with my image diagnosis and contouring requests. Dr. Liao, thank you for allowing me to shadow you during clinical consultations and sharing your knowledge to help me understand the clinical operations from the physician's standpoint. Lastly, I’d like to express my appreciation for the committee’s patience and flexibility in scheduling our meetings. I recognize that your time is valuable, and I am grateful for the effort you made to accommodate my schedule. It has been an honor to be advised by such an outstanding group of clinician scientists.

To the rest of the MD Anderson faculty and researchers that I have collaborated with, your contributions and expertise have also been instrumental to the success of this project. I firmly believe that it is because of your dedication and hard work that our institution remains the leader in the oncology field.

To my former program director Dr. Bud Wendt, I am especially grateful for the frequent check-ins and advice you offered during my early years in the program. Your feedback and insights helped me to navigate the challenges of graduate school. I also want to thank you for your counseling and encouragement when I doubted myself. You have made a lasting impact on the success of many students, and we will always be grateful for your dedication and support. I wish you a smooth retirement, and, though it may be bittersweet, I hope that cleaning out your office will be an easy process.

To my current program director Dr. Rebecca Howell, I would like to thank you for answering all my questions about the recruitment process of our graduate program seven years ago. I am also grateful for your assistance with the practice runs for my residency interviews: your feedback helped me to feel more confident and become better prepared for this important process. It is evident to me that our graduate program is in good hands with you at the helm, and I am confident that it will continue to thrive and evolve in exciting ways.
To my program classmates and lab members, I appreciate all the stimulating intellectual discussions, lighthearted moments, and adventures we shared. Knowing you and collaborating with all of you has been an enriching experience. Many of you have become lifelong friends, and I cherish the meaningful connections we have developed. Thank you for being an integral part of my personal and professional growth.

To my partner Athena, as I reflect on my academic journey, I could not have accomplished what I have without you by my side. Your support has been a constant source of strength and inspiration, and for that, I am truly grateful. I treasure the memories we have created together and the laughs and cries we shared. I appreciate your companionship, and I look forward to what the future holds for us.

Lastly, to my parents, as I reach this significant milestone in my academic journey and career, I am filled with extreme gratitude for all that you have done for me. Words can hardly express the depth of my feelings, but I want to convey how much your unwavering love and support have meant to me. You taught me the values of gratitude, strength, and curiosity, which have been essential in shaping who I am today, and I am forever grateful for the sacrifices you have made to help me reach this point of my life. Mom and dad, this work is for you, and I hope that it makes you proud. Thank you again for everything.

As I am finishing this section with a smile, thank you all from the bottom of my heart, Yulun.
Abstract

Improving Dose-Response Correlations for Locally Advanced NSCLC Patients Treated with IMRT or PSPT

Yulun He, B.S.
Advisory Professors: Radhe Mohan, Ph.D. & Kristy Brock, Ph.D.

The standard of care for locally advanced non-small cell lung cancer (NSCLC) is concurrent chemo-radiotherapy. Despite recent advancements in radiation delivery methods, the median survival time of NSCLC patients remains below 28 months. Higher tumor dose has been found to increase survival but also a higher rate of radiation pneumonitis (RP) that affects breathing capability. In fear of such toxicity, less-aggressive treatment plans are often clinically preferred, leading to metastasis and recurrence. Therefore, accurate RP prediction is crucial to ensure tumor coverage to improve treatment outcome. Current models have associated RP with increased dose but with limited accuracy as they lack spatial correlation between accurate dose representation and quantitative RP representation. These models represent lung tissue damage with radiation dose distribution planned pre-treatment, which assumes a fixed patient geometry and inevitably renders imprecise dose delivery due to intra-fractional breathing motion and inter-fractional anatomy response. Additionally, current models employ whole-lung dose metrics as the contributing factor to RP as a qualitative, binary outcome but these global dose metrics discard microscopic, voxel-(3D pixel)-level information and prevent spatial correlations with quantitative RP representation.

To tackle these limitations, we developed advanced deformable image registration (DIR) techniques that registered corresponding anatomical voxels between images for tracking and accumulating dose throughout treatment. DIR also enabled voxel-level dose-
response correlation when CT image density change (IDC) was used to quantify RP. We hypothesized that more accurate estimates of biologically effective dose distributions actually delivered, achieved through (a) dose accumulation using deformable registration of weekly 4DCT images acquired over the course or radiotherapy and (b) the incorporation of variable relative biological effectiveness (RBE), would lead to statistically and clinically significant improvement in the correlation of RP with biologically effective dose distributions.

Our work resulted in a robust intra-4DCT and inter-4DCT DIR workflow, with the accuracy meeting AAPM TG-132 recommendations for clinical implementation of DIR. The automated DIR workflow allowed us to develop a fully automated 4DCT-based dose accumulation pipeline in RayStation (RaySearch Laboratories, Stockholm, Sweden). With a sample of 67 IMRT patients, our results showed that the accumulated dose was statistically different than the planned dose across the entire cohort with an average MLD increase of ~1 Gy and clinically different for individual patients where 16% resulted in difference in the score of the normal tissue complication probability (NTCP) using an established, clinically used model, which could qualify the patients for treatment planning re-evaluation. Lastly, we associated dose difference with accuracy difference by establishing and comparing voxel-level dose-IDC correlations and concluded that the accumulated dose better described the localized damage, thereby a closer representation of the delivered dose. Using the same dose-response correlation strategy, we plotted the dose-IDC relationships for both photon patients (N = 51) and proton patients (N = 67), we measured the variable proton RBE values to be 3.07–1.27 from 9–52 Gy proton voxels. With the measured RBE values, we fitted an established variable proton RBE model with pseudo-$R^2$ of 0.98. Therefore, our results led to statistically and clinically significant improvement in the correlation of RP with accumulated and biologically effective dose distributions and demonstrated the potential of incorporating the effect of anatomical change and biological damage in RP prediction models.
Chapter 2: Central Hypothesis, Specific Aims, and Dissertation Organization ............29
  2.1 Central Hypothesis ....................................................................................................29
  2.2 Specific Aims .............................................................................................................29
  2.3 Dissertation Organization ...........................................................................................30

Chapter 3: Identifying Reference Image for Inter-4DCT DIR ........................................32
  3.1 Abstract .....................................................................................................................32
  3.2 Introduction ................................................................................................................34
  3.3 Methods and Materials ...............................................................................................36
    3.3.1 Patient Data ........................................................................................................36
    3.3.2 Hybrid Intensity-Based DIR .................................................................................36
    3.3.3 Quantitative Metrics for DIR Accuracy Assessment ............................................37
      3.3.3.a Dice Similarity Coefficient .............................................................................37
      3.3.3.b Target Registration Error ...............................................................................37
    3.3.4 Automatic Landmark Identification Method .......................................................38
    3.3.5 Landmark-Based DIR Accuracy Measurement ....................................................40
      3.3.5.a Comparison of Geometric Accuracy between AVG-AVG DIR and Phase-Phase DIRs ..............................................................................................................40
      3.3.5.b Effect of Breathing Motion .............................................................................40
      3.3.5.c Clinical Impact of DIR Geometric Accuracy on Dose Mapping ......................40
  3.4 Results ......................................................................................................................41
    3.4.1 Validation of Automatic Landmark Identification Workflow ...............................41
    3.4.2 AVG-AVG DIR Compared with Phase-Phase DIRs .............................................42
    3.4.3 Geometric Impact of Landmark Motion on DIR TRE ............................................43
    3.4.4 Clinical Impact of DIR Geometric Accuracy on Dose Mapping ............................44
  3.5 Discussion .................................................................................................................47
  3.6 Conclusion .................................................................................................................51

Chapter 4: Parameter Optimization for Biomechanical Model-Based DIR ......................52
  4.1 Abstract .....................................................................................................................52
  4.2 Introduction ................................................................................................................53
  4.3 Materials and Methods ...............................................................................................54
    4.3.1 Patient data .........................................................................................................54
    4.3.2 Biomechanical model-based registration algorithm in RayStation .....................56
    4.3.3 Parameters for boundary conditions ....................................................................57
    4.3.4 Evaluation ...........................................................................................................58
    4.3.5 Statistical Analysis ...............................................................................................58
  4.4 Results ......................................................................................................................58
    4.4.1 Accuracy .............................................................................................................61
    4.4.2 Robustness, Efficiency, and Complexity ..............................................................62
  4.5 Discussion .................................................................................................................63
  4.6 Conclusion .................................................................................................................66
List of Illustrations

Figure 1. Percentage of all estimated cancer deaths in both sexes in the world in 2019...... 1
Figure 2. Axial view of a contrast-enhanced computed tomography (CECT) ............... 4
Figure 3. Axial view of a overlaid positron emission tomography (PET) and computed
tomography (CT) with the tumor shown in the bright region ........................................... 5
Figure 4. Coronal view of the end-inhalation phase (T0), mid-ventilation phase (T3), end-
exhalation phase (T5), and the average-intensity image (AVG) of a 4-dimensional computed
tomography (4DCT) ............................................................................................................. 9
Figure 5. The percent depth dose (PDD) curves for proton (cyan) and X-ray (dashed). ......12
Figure 6. Correlated pneumonitis probability against mean lung dose from multiple
institutions overlaid on the same plot ..................................................................................15
Figure 7. Classification of radiographic changes after stereotactic body radiotherapy (SBRT).
............................................................................................................................................20
Figure 8. Fusion view of a rigid image registration example.. ...........................................22
Figure 9. Fusion view of a deformable image registration (DIR). .................................23
Figure 10. Target registration error (TRE) .......................................................................27
Figure 11. Manual landmark identification process to validate the automatic landmark
workflow .................................................................................................................................39
Figure 12. Target registration error (TRE) differences between deformable image
registration (DIR) across average-intensity images and DIR across phases for each patient.
............................................................................................................................................42
Figure 13. Boxplots of mean target registration error (TRE) of the 28 patients. ...........43
Figure 14. Deformed doses (left panels) and their difference (right panel) on an axial slice of
the average-intensity image (AVG) of the mid-treatment week for a sample patient.........44
Figure 15. 3D rendering of landmarks in liver (left) and lungs (right) ............................55
Figure 16. Schema for obtaining the 4D planned dose (4DP) (green), the 3D accumulated dose (3DA) (orange), and the 4D accumulated dose (4DA) (red). .............................................72

Figure 17. The effect of intra- and inter-fractional anatomical changes to the mean lung dose (MLD) and the normal tissue complication probability (NTCP) at the organ level.........................78

Figure 18. Boxplots of volumetric overlap between the 3D planned dose (3DP) and the 4D accumulated dose (4DA) for isodose intervals of 5-Gy increments.................................79

Figure 19. Cumulative histogram for voxels of the entire cohort showing the distribution of absolute dose difference between the 3D planned dose (3DP) and the 4D accumulated dose (4DA). ..........................................................................................................................80

Figure 20. Voxel-level dose-image density change (IDC) modeling results: .......................106

Figure 23. The created segmentation model in RayStation. ..............................................107
List of Tables

Table 1. Planning information for the ipsilateral lung of each analyzed patient. .................45
Table 2. BM-DIR results for liver images. ............................................................................60
Table 3. BM-DIR results for lung images. ............................................................................61
Table 4. Tumor location and dose prescription for the selected patients.........................70
Table 1. RBE modelling result. ..........................................................................................95
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%CV</td>
<td>percent coefficient of variation</td>
</tr>
<tr>
<td>3D</td>
<td>3-dimensional</td>
</tr>
<tr>
<td>3DA</td>
<td>3-dimensional accumulated dose</td>
</tr>
<tr>
<td>3DP</td>
<td>3-dimensional planned dose</td>
</tr>
<tr>
<td>4D</td>
<td>4-dimensional</td>
</tr>
<tr>
<td>4DA</td>
<td>4-dimensional accumulated dose</td>
</tr>
<tr>
<td>4DCT</td>
<td>4-dimensional computed tomography</td>
</tr>
<tr>
<td>4DP</td>
<td>4-dimensional planned dose</td>
</tr>
<tr>
<td>AAPM</td>
<td>American Association of Physicists in Medicine</td>
</tr>
<tr>
<td>ANACONDA</td>
<td>ANAtomically CONstrained Deformation Algorithm</td>
</tr>
<tr>
<td>AVG</td>
<td>average intensity image</td>
</tr>
<tr>
<td>BH</td>
<td>breath-hold</td>
</tr>
<tr>
<td>CBCT</td>
<td>cone-beam computed tomography</td>
</tr>
<tr>
<td>CECT</td>
<td>contrast-enhanced computed tomography</td>
</tr>
<tr>
<td>CT</td>
<td>computed tomography</td>
</tr>
<tr>
<td>CTV</td>
<td>clinical target volume</td>
</tr>
<tr>
<td>DECT</td>
<td>dual-energy computed tomography</td>
</tr>
<tr>
<td>DIR</td>
<td>deformable image registration</td>
</tr>
<tr>
<td>DL</td>
<td>deep learning</td>
</tr>
<tr>
<td>$D_{\text{max}}$</td>
<td>maximum dose</td>
</tr>
<tr>
<td>$D_{\text{mean}}$</td>
<td>mean dose</td>
</tr>
<tr>
<td>$D_p$</td>
<td>photon dose</td>
</tr>
<tr>
<td>$D_x$</td>
<td>proton physical dose</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>DSC</td>
<td>Dice similarity coefficient</td>
</tr>
<tr>
<td>DVF</td>
<td>deformation vector field</td>
</tr>
<tr>
<td>DVH</td>
<td>dose volume histogram</td>
</tr>
<tr>
<td>FB</td>
<td>free breathing</td>
</tr>
<tr>
<td>FEA</td>
<td>finite element analysis</td>
</tr>
<tr>
<td>FEM</td>
<td>finite element model</td>
</tr>
<tr>
<td>GTV</td>
<td>gross target volume</td>
</tr>
<tr>
<td>HD</td>
<td>Hausdorff distance</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield units</td>
</tr>
<tr>
<td>IDC</td>
<td>image density change</td>
</tr>
<tr>
<td>IGRT</td>
<td>image-guided radiation therapy</td>
</tr>
<tr>
<td>IMRT</td>
<td>intensity-modulated radiation (photon) therapy</td>
</tr>
<tr>
<td>IRB</td>
<td>institutional review board</td>
</tr>
<tr>
<td>ITV</td>
<td>internal target volume</td>
</tr>
<tr>
<td>LET</td>
<td>linear energy transfer</td>
</tr>
<tr>
<td>LET_d</td>
<td>dose-averaged linear energy transfer</td>
</tr>
<tr>
<td>linac</td>
<td>linear accelerator</td>
</tr>
<tr>
<td>MLC</td>
<td>multi-leaf collimator</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>NSCLC</td>
<td>non-small–cell lung cancer</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PSPT</td>
<td>passive scattering proton therapy</td>
</tr>
<tr>
<td>PTV</td>
<td>planning target volume</td>
</tr>
<tr>
<td>RBE</td>
<td>relative biological effectiveness</td>
</tr>
<tr>
<td>RIR</td>
<td>rigid image registration</td>
</tr>
<tr>
<td>ROI</td>
<td>region of interest</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>RTOG</td>
<td>The Radiation Therapy Oncology Group</td>
</tr>
<tr>
<td>STSC</td>
<td>small-cell lung cancer</td>
</tr>
<tr>
<td>T0</td>
<td>0% breathing phase</td>
</tr>
<tr>
<td>T3</td>
<td>30% breathing phase</td>
</tr>
<tr>
<td>T5</td>
<td>50% breathing phase</td>
</tr>
<tr>
<td>TRE</td>
<td>target registration error</td>
</tr>
<tr>
<td>TPS</td>
<td>treatment planning system</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Lung Cancer

Lung cancer is the second most common cancer worldwide with an estimated diagnosis of more than 2 million new cases annually (Figure 1) [1]. The World Health Organization reported that lung cancer is associated with the most cancer deaths among both men and women, leading to approximately 1.8 million fatalities worldwide in 2020 [2]. Lung cancer patients can present symptoms including cough, chest pain, and fatigue [3]. The shared symptoms between lung cancer and chronic respiratory conditions such as chronic obstructive pulmonary disease contribute to delayed diagnosis, which subsequently leads to the high mortality rate [4].

Figure 1. Percentage of all estimated cancer deaths in both sexes in the world in 2019

1.1.2 Classification and Staging

Lung cancer is broadly classified as small-cell lung cancer (SCLC) (15% of total diagnosis) or non–small-cell lung cancer (NSCLC) (85% of total diagnosis). NSCLC is made up by 40% with adenocarcinoma, 30% with squamous cell carcinoma, 15% with large cell carcinoma, and others [5].
The most accepted staging system for NSCLC is the American Joint Commission on Cancer (AJCC) TNM system, which stands for Tumor, Nodes, and Metastasis [6]. The T factor describes the size and extent of the primary tumor: it ranges from T1, indicating a small, confined tumor, to T4, indicating a larger tumor that has invaded nearby structures such as the chest wall or diaphragm. The N factor describes the extent of cancer spread in lymph nodes: it ranges from N0, indicating no nodal involvement, to N3, indicating extensive spread to multiple lymph nodes in the contralateral mediastinum and/or neck. Lastly, the M factor describes whether the cancer has spread to other parts of the body: it is either M0, indicating no distant metastasis, or M1, indicating the presence of distant metastasis in organs such as the liver, brain, or bones.

Combining the T, N, and M factors results in a stage grouping, which ranges from stage I (the least advanced) to stage IV (the most advanced). Detailed staging criteria can be referenced from the International Association for the Study of Lung Cancer (IASLC) [7]. In general, stage I indicates a small, localized tumor with no spread to lymph nodes or distant organs. Stages II and III indicate larger tumors and/or spread to nearby lymph nodes. Stage IV indicates extensive tumor growth with metastasis to other parts of the body. Upon diagnosis, the distribution of stages I–IV respectively are 26%, 8%, 28%, and 38% [8]. Most patients are diagnosed at advanced stages (e.g., locally advanced) because NSCLC may not produce any symptoms in the early stages. By the time symptoms appear, they can be mistaken for other chronic lung disease, and the cancer has already spread to other parts of the body, which contributes to the low survival [9]. Locally advanced NSCLC (LA-NSCLC) is a subtype of NSCLC where the tumor has extended beyond the primary site but has not yet metastasized to distant sites [10]. This dissertation project will focus on studying LA-NSCLC cases.
1.1.3 Diagnosis

Common diagnosis of the disease and staging includes anatomical imaging, functional imaging, and biopsy.

1.1.3.a Anatomical Imaging

X-ray and computed tomography (CT) are commonly used for initial diagnosis of lung cancer. Both techniques are non-invasive and widely available, and they utilize the attenuation capability of the organ (bone higher, soft tissue lower) to the photon beam. X-Ray provides 2-dimensional images with overlaid anatomical information, thus limited in detecting small tumors. However, X-ray technology has advanced from film-based to digital radiographic systems and has the advantage of low cost and fast turnaround time. Dual-energy X-ray technology evolved to provide improved differentiation of tissue composition by using two different energy levels of X-ray [11]. X-rays can also be used to diagnose lung cancer using volumetric CT [12]. CT employs a rotating X-ray source and detector array to acquire cross-sectional images of the human body. These images are then reconstructed using computer algorithms to generate 3-dimensional anatomical information. Compared to X-ray, CT provides enhanced contrast and image resolution of sub-millimeters.

Recent advancement in CT technology includes dual-energy CT (DECT) that quantifies material composition with the difference in material attenuation and the known changes in attenuation when acquired with different energy spectra [13]. DECT can be used to remove bones from the image to help evaluate tumors or other abnormalities. Contrast-enhanced CT (CECT) utilizes contrast agents injected through perfusion to highlight blood vessels and tumor angiogenesis. CECT of the chest can serve as a standard imaging technique for the initial evaluation of patients suspected of lung cancer [14]. CECT can also be used to monitor treatment response.
structures. However, anatomical imaging alone is limited in identifying advance-staged NSCLC tumors that can be less confined, near soft tissue, and confounded by co-morbidities such as inflammation that may also appear in similar shape as compared to tumors. The addition of functional imaging such as positron emission tomography (PET) can help define the extent of the primary tumor and identify metastasis and nodal involvement.

1.1.3.b Functional Imaging
1.1.3.c Biopsy

Biopsy remains the gold standard for lung cancer staging. It can determine whether the lesion is malignant or benign with an accuracy of 73%–90% [16]. Due to the invasive nature of biopsy, pneumothorax can occur during the process which decreases diagnostic accuracy [17]. Advancement such as fluoroscopy-guided biopsy provided real-time visualization of the needle tip and the site of the lesion which can achieve 100% diagnostic accuracy for lesions larger than 11 mm in diameter [18].

Upon diagnosis and staging of NSCLC, its treatment diversifies and includes surgery, chemotherapy, radiation therapy, and immunotherapy [10].

1.1.4.a Surgery

Surgery is a common treatment technique for stage I and II NSCLC with a curative intent. Surgery’s curative rate for stage I and II is 60%–80% and 40%–50%, respectively. The specific surgical modality includes segmentectomy, lobectomy, and pneumonectomy, depending on the size and location of the tumor [10]. Pre-operative pulmonary-function testing such as ventilation (air exchange) and perfusion (blood flow) are conducted to
determine if the patient can tolerate surgery [19].

1.1.4.b Ablation

Radiofrequency ablation (RFA) can be used as the primary treatment for stage I NSCLC patients who are not surgical candidates. RFA is usually guided by CT, and it uses a probe that generates electromagnetic energy to impart thermal damage. It was originally developed as the primary treatment for liver tumors but has been recently adopted for lung [20]. However, the 5-year survival rate for patients treated with RFA are approximately 27% [21], and the major complications of RFA include pneumothorax (i.e., collapsed lung), pleural effusion (i.e., buildup of fluid between the tissue layers that line the lungs and chest cavity), and intrapulmonary hemorrhage [10].

On the opposite end the thermal spectrum, cryoablation is a technique that uses extremely low temperatures to destroy abnormal tissue can also be used as an alternative to traditional surgical approaches. Similar to RFA, cryoablation may also cause healthy tissue damage and bleeding [22].

1.1.4.c Chemotherapy

In chemotherapy, drugs are administered orally or through intra-vascular injection. These drugs target rapidly dividing cells such as cancer cells but also normal cells such as hair follicles, thus causing the common side effects such as hair loss. Chemotherapy is commonly combined with other treatment techniques such as surgery and radiotherapy where it can be used before the primary treatment to shrink tumor before treatment (neoadjuvant). Chemotherapy can also be used as an adjuvant therapy to the primary treatment (e.g., surgery) where the drugs can reach the remaining or metastatic cancer cells that has evaded the primary treatment [23, 24]. Combining chemotherapy and radiotherapy (discussed in Section 1.1.4.c) can enhance the sensitivity of cancer cells to treatment, resulting in improved effectiveness compared to either treatment used alone. The
combination of paclitaxel and platinum-based drugs such as carboplatin is a commonly prescribed for chemo-radiotherapy for NSCLC [25].

1.1.4.d Immunotherapy

Immunotherapy has emerged as a promising treatment option for lung cancer, particularly advanced stage disease. Unlike conventional chemotherapy, immunotherapy harnesses the patient’s own immune system to identify and attack cancer cells. Immune checkpoint inhibitors such as Pembrolizumab have been shown to improve median progression-free survival from 6.0 months for patients treated with chemotherapy to 10.3 months [26].

1.1.4.e Radiotherapy

External beam radiotherapy (RT) is non-invasive and has replaced surgery in early staged cancer. RT delivers mega-voltage radiation beam of photons or other particles such as electrons to ionize and destroy tumor cells by damaging their DNA. Concurrent chemo-radiotherapy is the standard of care for LA-NSCLC and is also used in palliative RT to shrink tumor and control symptoms [10].

1.1.4.e.i Imaging for Treatment Planning

A personalized treatment plan is first developed for each patient to maximize tumor control and minimize healthy tissue damage. The planning process begins with imaging studies such as PET/CT that localize the tumor(s) and define their extent. Lung tumors are affected by the natural breathing motion. To address this, respiratory gating techniques such as breath hold (BH) can be utilized if the tumor moves more than a clinically set threshold during breathing (e.g., 5 mm in the inferior-superior direction) [27]. The deep inspiration BH technique involves coaching the patient to take a deep breath and hold it at the same level throughout treatment. The patient breathes through a spirometer, which monitors their respiration level by measuring the volume of inhaled and exhaled air. Prior to treatment, a
spirometry training session is conducted to determine the patient's breathing parameters, such as tidal volume and vital capacity [28].

However, the BH technique can be challenging especially for advance-staged patients with severe comorbidities that prevent extended periods of BH during each treatment session (i.e., fraction). In these situations, patients are treated with the free-breathing (FB) technique that is planned with 4-dimensional computed tomography (4DCT). A 4DCT is respiration-correlated over a full breathing cycle, consisting of a series of 3D CTs binned to represent various breathing stages along a full cycle [29]. 4CDTs are commonly divided into 10 phases starting from T0 (or T0%) as the end-inhalation phase, to T3 (or T30%) as the mid-breathing phase, to T5 (or T50%) as the end-exhalation phase, and eventually to T9 before the start of the next cycle to T0. Capturing lung anatomy at different time points in the normal breathing cycle allows clinicians to identify the extent of tumor motion. In addition, a maximum intensity projection image (MIP) is reconstructed by taking the highest intensity of any voxel (3-dimensional pixel) during the full breathing cycle, which can be used to define the tumor boundaries as the gross tumor volume (GTV). GTV can be extended to account for disease extension, and this extended volume is called the clinical target volume (CTV). Internal target volume (ITV) is CTV plus any additional tissue to account for variations in the location, size, or shape of the tumor during and between treatments. Finally, the planning target volume (PTV) is determined by adding an extra margin around the ITV to ensure that the tumor receives the appropriate dose (i.e., 95% of PTV is typically planned to receive the prescribed dose (e.g., 66 Gy)) [30]. Current clinical standard calculates dose on the average-intensity image (AVG) which is created by computing the arithmetical mean of the pixel intensity values of all breathing phases [31].
1.1.4.e.i Beam Modulation

The general goal of RT is to conform the high dose distribution to the PTV while limiting dose elsewhere. An early technique for tissue sparing is 3-dimensional conformal radiation therapy (3D-CRT) that utilizes the anatomical information provided by the 3D planning image (e.g., AVG). Dose conformation can be achieved by shaping the multi-leaf collimator (MLC) to the tumor in the ‘beams-eye view’ and using multiple beams directed at the tumor from different angles. Intensity-modulated radiation therapy (IMRT) is an advanced version of 3D-CRT that uses computer algorithms to ‘inversely’ determine the radiation beam settings (e.g., number of beams and beam angles) based on the clinical objectives in dose constraints (e.g., dose volume histogram (DVH) metrics) to the target and normal tissue [32]. IMRT has been shown to increase tumor dose and decrease damage to surrounding tissue in NSCLC [33].

1.1.4.e.ii Image Guidance

While the goal of treatment planning is to accurately plan the radiation delivery, the goal of image guidance is to consistently deliver radiation according to the treatment plan [34]. The treatment plan is initially optimized on a static anatomy, but patient anatomy may vary from fraction to fraction and needs to be aligned to the anatomy originally used in treatment planning. The alignment process can be achieved with on-board planar daily kV and/or MV X-ray images. However, planar imaging can be inadequate in localizing lung
tumor and its surrounding soft tissue. The development of on-board volumetric imaging such as cone-beam CT (CBCT) allows detailed volumetric visualization of the target region directly on the linear accelerator [35]. To generate a CBCT, the X-ray source emits cone-shaped X-ray beams that pass through the patient's body, and the attenuated X-rays are detected by the on-board imager from different angles to form a 3D image. Similar to 4DCTs, 4D-CBCTs can also be used to track tumor in real-time while the patient is still in the treatment position to adjust treatment plans accordingly [36].

Newer image guidance techniques have emerged such as ring-based gantries to allow for faster scanning and treatment time [37] and magnetic resonance imaging (MRI)-based linear accretor to leverage the superior soft tissue contrast provided by the MRI images for a clearer view of the tumor and surrounding critical structures [38].

1.1.4.e.iii Fractionation and Dose Prescription

Conventional RT delivers dose of 1.8–2 Gy per fraction (fx) due to the radiobiological effects of radiation on tumor and healthy cells. Fractionation relies on the differences in radiation sensitivity between tumor cells and normal tissue cells. Tumor cells are generally more sensitive to radiation than normal tissue. Dividing the total radiation dose into smaller fractions allows for the normal tissue cells to repair themselves between treatments while keeping the tumor cells under control. This allows for a higher total radiation dose to be delivered to the tumor over the course of several treatments while minimizing the risk of damage to normal tissues [39].

Stereotactic body RT (SBRT) utilizes hypofractionation regimen and delivers higher dose per fraction in fewer fractions. During each fraction, SBRT delivers precisely ablative dose to the tumor while minimizing toxicities [40]. SBRT can replace the need for surgery and reduce the time and inconvenience of treatment for patients. Due to the higher fractional dose, SBRT requires more precise imaging and planning than conventionally fractionated
IMRT [41]. SBRT is commonly used for stage I and II NSCLC that is medically inoperable [42]. For these patients, the RTOG 0813 trial compared the efficacy of SBRT and conventionally fractionated RT in non-operable NSCLC patients. The trial concluded two-year local control rates as high as 87.9% and progression free survival of 54.5% [43].

Another potential advancement in RT for lung cancer is dose escalation. For stage-III NSCLC, the commonly accepted radiation therapy dose of 60–63 Gy was established by the RTOG 7301 trial [44]. The dose of 60 Gy was chosen to optimize clinical factors such as local control. However, many studies such as RTOG 01173 [45] reported an increase in survival with higher than conventional dose (e.g., 74Gy). These findings led to RTOG 0617 [46] that compared 74 Gy concurrent chemo-radiotherapy and 60 Gy concurrent chemo-radiotherapy, the result of which showed no improvement and potentially worse survival with 74Gy than 60Gy. Although treatment-related deaths were more common in the dose escalation arm, the comparison did not reach statistical significance. Nevertheless, the standard prescription has been kept unchanged, and further clinical trials are needed to confirm the benefit of dose escalation. These findings suggest the potential benefit of highly conformal tumor coverage provided by dose tracking in adaptive radiotherapy [47] and using particle therapy such as protons for the increased biological damage [48].

1.1.4.e.iv Proton Therapy

Photon-beam RT remains the most common modality to treat NSCLC, but proton-beam RT is being increasingly utilized [49]. Protons are heavy charged particles. As protons traverse through tissue, they slowly deposit energy which ionizes nearby atoms. However, as protons slow down, due to their mass, they have increased time to ionize tumor cells to create many more double strand-breaks in the DNA, making the tumor unable to repair itself or grow new cells [48]. Such a phenomenon of losing most of their energy at their track end is called the Bragg peak [50]. Therefore, protons can achieve the same cell kill with less dose
compared to photons, and such increased ability of damage is represented with relative biological effectiveness (RBE), which is clinically applied as 1.1 [48]. Because of the sharp increase in tumor cell kill in the Bragg peak region, a spread-out Bragg peak (SOBP) can be created by staggering Bragg peaks with various depths from protons with different energies to uniformly cover the length of the target while sparing tissue beyond the target [51]. Therefore, compared to photon therapy, proton therapy has the advantage of maintaining tumor damage while minimizing damage to the surrounding normal tissue [48]. Proton therapy can be particularly beneficial for cancers near critical structures such as heart to potentially lower cardiac toxicities [52] and in pediatric patients to reduce secondary cancer risks [53].

![Figure 5. The percent depth dose (PDD) curves for proton (cyan) and X-ray (dashed). The tumor (red dot) is located at the Bragg peak location. Picture source: http://www.avoplc.com/en-gb/LIGHT/The-Potential-of-Proton-Therapy](http://www.avoplc.com/en-gb/LIGHT/The-Potential-of-Proton-Therapy)

### 1.2 Radiation-Induced Lung Toxicities

Despite significant advances in RT, median survival time of patients with LA-NSCLC remains less than 28 months [54]. Rate of loco-reoccurrence increases as primary stage increases (5–19% for stage I, 11–27% for stage II, and 24–40% for stage IIIA) [55]. The low
survival rates may be due in part to lack of tumor coverage with the full prescription dose in
fear of the dose-limiting treatment-related toxicities. When the prescribed tumor dose is
constrained by the concern of treatment-induced toxicity, microscopic disease may not
receive enough radiation to be destroyed. As a result, these cells can survive and potentially
spread to other parts of the body, such as the brain, forming distant metastases. [23, 56].

Radiation pneumonitis (RP) and pulmonary fibrosis (PF) are the two main types of
radiation-induced lung toxicities. Both RP and PF can be graded using Common
Terminology Criteria for Adverse Events (CTCAE) [57] and are correlated with increased
tissue density in computed tomography (CT) images (i.e., appearing brighter on grayscale).
RP is an acute toxicity that manifests 6 weeks – 6 months post RT. It degrades breathing
capability and causes respiratory failure [58]. Moderate to severe RP occurs in up to 30% of
irradiated patients [59]. RP grading is defined subjectively by grades, where grade 1 is
asymptomatic with radiographic change, grades 2–4 are symptomatic with worsening effect
on daily life (incidence ranges 15–40% [60]), and grade 5 is death. Currently, steroids are
prescribed to only alleviate symptoms rather than progression of RP while causing side
effects such as infection from immune suppression and hyperglycemia [61]. Increased tumor
dose has been linked to increased survival but also a higher rate of RP [62]. PF is a
permanent replacement of normal functioning lung tissue with scar tissue and may lead to
life-threatening respiratory failure. Unlike RP, PF is a latent effect that develops starting 6
months post RT and stabilizes after 2 years, and it almost always occurs after RT [63, 64].

1.2.1 NTCP Models

Normal Tissue Complication Probability (NTCP) models are clinically used to
describe the probability of a specific endpoint such as RP. NTCP models aim to distill
dosimetric and anatomic information down to a single risk measure. The Lyman-Kutcher-
Burman (LKB) model is a widely accepted model to predict RP [65, 66]. The model equation is given as follows:

\[
NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{D_{eff} - TD_{S0}}{mTD_{S0}}} e^{-\frac{x^2}{2}} dx \tag{1}
\]

\[
D_{eff} = \left( \sum v_i D_i \right)^n \tag{2}
\]

where NTCP is the probability of a specific endpoint such as grade ≥ 2 RP, the \(TD_{S0}\) parameter represents the dose at which 50% patients develop the endpoint, the \(m\) parameter describes the slope of the dose-response curve, i.e., how quickly the probability of the endpoint increases with increasing dose, and the \(D_{eff}\) term is the dose that, if given uniformly to the entire volume, is as effective as the actual nonuniform dose in leading to the same NTCP. \(D_{eff}\) is often referred to as the uniform equivalent dose (EUD), and its composition is described in equation (2): \(v_i\) is the volume factor (i.e., percentage of total volume) that corresponds to dose level \(D_i\), and \(n\) is the volume effect where \(D_{eff}\) becomes mean lung dose (MLD) for \(n = 1\) and root mean squared dose for \(n = 0.5\) where sub-volumes receiving higher doses carry a higher weight in contributing to the endpoint. A recent analysis conducted by Tucker et al. returned \(TD_{S0}\), \(m\), and \(n\) as 34.80 Gy, 0.22, and 0.5, respectively based on NSCLC patients treated with conventionally fractionated IMRT at MD Anderson Cancer Center [67].

1.2.2 Toxicity Prediction

Accurate RP predicting methods can help clinicians optimize dosimetric coverage of the tumor while limiting toxicities, thus improving treatment outcomes. Early identification of RP, such as changes seen in radiographic imaging, could allow clinicians make better treatment strategy [68]. Radiation dose has been consistently reported as the most prominent contributor to RP [62, 69–71]. Specifically, DVH-based metrics such as MLD and lung volume receiving more than 20 Gy (\(V_{20}\)) have been consistently documented to be
strongly correlative with RP [72–74]. Despite this pressing need, NSCLC patients still suffer from a lack of accurate toxicity prediction methods [75].

![Diagram showing symptomatic pneumonitis probability against mean lung dose from multiple institutions overlaid on the same plot. The dashed lines represent the average trend (red) and the confidence intervals (cyan).](image)

**Figure 6.** Correlated pneumonitis probability against mean lung dose from multiple institutions overlaid on the same plot. The dashed lines represent the average trend (red) and the confidence intervals (cyan). Source: Marks et al. [76]

One limitation is the usage of parameters derived from the planned dose distribution, which may not reflect the actual delivered dose. As previously introduced, for FB treatments, radiation dose is planned based on the CT that represents patient anatomy of the initial time point. Therefore, the planned dose is computed under the assumption of a static patient geometry, not fully considering the intra-fractional breathing motion during beam delivery or inter-fractional anatomical change throughout the treatment course such as tumor response. As a result, such planned dose may not accurately represent the delivered dose [77].

DVH-based metrics such as MLD and $V_{20}$ are macroscopic metrics that are overly generalized and lack spatial features, and their measurement is confounded by the moving
lung volume treated during free-breathing therapies. Due to non-uniform dose coverage in the lung, different dose distributions may result in the same values in these metrics. Therefore, relying on DVH metrics loses spatial information and may inherently render the predictions inaccurate.[76] Current normal tissue complication probability (NTCP) models are limited to predicting the symptomatic performance of the entire lung using scorings such as CTCAE. If physicians were able to examine a regional lung on the voxel-level, they could better tailor the treatment plans to avoid complications in certain difficult regions, especially in the lower lung that experiences the most breathing motion.

1.3 Uncertainties in Radiotherapy
1.3.1 Breathing Motion

As introduced, BH or 4DCT-based planning techniques have been developed to mitigate breathing motion. However, several studies have analyzed the difference between planned dose based on a static image such as AVG and the planned dose including breathing motion and concluded the latter can provide more accurate description of the delivered dose.

Glide-Hurst et al. investigated the feasibility of using AVG-based planned dose (dose calculated on the AVG is copied onto each phase image and summed back on to the average image) vs phase-based planned dose (dose was recalculated on each phase and summed on to the AVG with the same deformations used in the AVG-based summed dose) for SBRT plans [77]. The difference in the average dose of the normal lung (excluding GTV) was found to be less than ~1% and the $V_{20}$ was less than ~4%. The study was limited to 4 cases. Valdes et al. analyzed 10 SBRT patients with stage I–II lung cancer and showed that the phase image-based planned dose (dose recalculated on each breathing phase then accumulated to one single phase) is more accurate than the dose directly calculated on AVG [78]. The difference in MLD and $V_{20}$ between the two planned doses was $-0.14 \pm 0.38$
Gy and -0.15 ± 0.67 Gy, respectively. The authors estimated that the use of 4DCT-based planned dose was more accurate and could benefit the evaluation of normal tissue complication probabilities. In both studies, only early staged tumors were analyzed so the dose deformations might not be as complex for LA-NSCLC patients.

1.3.2 Longitudinal Response

The aforementioned studies laid a great foundation for the feasibility of 4DCT-based planned dose. However, these planned doses are assumed to be consistent patient geometry, thereby not incorporating the anatomical changes throughout treatment. In other words, these studies doses do not account for tumor response and changes in breathing pattern that occur over the course of radiotherapy. For standard-fractionated RT schedules, patients receive 4–7 weeks of treatment during which time the tumor and its surrounding normal tissue can change in shape/size to distort the dose deposition under the same beam setup. Therefore, without incorporating the doses reflected by the anatomy changes throughout treatment course, the assumption that the planned dose represents the delivered dose is inevitably invalid.

1.3.2.a Dose Accumulation

To quantify the effect of longitudinal response, dose delivered throughout the treatment can be accumulated, and such accumulated dose can be compared to the planned dose. In dose accumulation, dose delivered to patients at each intermediate timestamp is recomputed using longitudinal images such as weekly 4DCTs. These recomputed doses are then mapped/summed to a common, reference image. Retrospectively, dose accumulation can be used to provide a potentially better estimate of the delivered dose than the planned dose which is based on pre-treatment patient anatomy. However, many dose accumulation studies were limited to early staged tumors and the size of the studies were limited by a small patient cohort due to the labor-intensive nature of dose accumulation [79–82].
In addition, when the studies observed a dosimetric difference between the accumulated dose and the planned dose, they concluded the accumulated dose was more accurate in estimating the delivered dose. However, such conclusion was not validated through association with evidence from clinical response. Toxicity records such as RP (e.g., CTCAE) could potentially be used as clinical evidence but 28% of RP diagnoses can be confounded by comorbidities [62]. A quantifiable clinical metric is needed for evaluation. Voxel-level image density has is a candidate to demonstrate localized radiographic-based damage [83, 84]. Therefore, one aim of this project is to accumulate dose and compare the accumulated dose to the planned dose and compare their correlations to clinical response.

1.3.3 Proton Therapy Biological Damage

Dose accumulation can potentially return a more accurate representation of the delivered dose. However, our knowledge of the dose delivery on the biological level remains limited for patients treated with proton therapy: it is yet to be proven that the normal tissue sparing of protons translates into reduced toxicity compared to photon therapy for NSCLC patients. A randomized phase II clinical trial was conducted at MD Anderson Cancer Center to evaluate the difference in efficacy between photon RT and proton RT [85]. The results did not demonstrate the anticipated superior normal tissue sparing of protons. This suggested our limited understanding of proton’s ability to cause biological damage compared to photon—the RBE. Current consensus among clinical usage is a constant RBE of 1.1 [48]. Many in vitro studies suggest a much higher number near track end [86–88]. However, there is a lack of study directly using clinical evidence rather than cells. Knowing the exact RBE values inside patient anatomy would assist in understanding the actual biological damage, thus helping provide more accurate toxicity studies correlations.
1.4 Feasibility of Radiographic Change as the Clinical Endpoint

Many studies have investigated localized radiographic-based dose-response relationships [83, 84, 89, 90]. Underwood et al. compared the pre- to post-treatment lung CT image density change (IDC) for 20 chest wall patients treated with photon or proton (N=10 for each group), and the authors observed a higher level of IDC caused by the same amount of photon and proton dose, with the ratio reaching 4.7, indicating RBEs exceeding 1.1 [89]. Bernchou et al. studied the time evolution of IDC in the normal lung after IMRT for 20 patients with NSCLC who had regular CT scans before and after the treatment [83]. By plotting the voxel-level dose-IDC curves for follow-up images taken at various time points after RT, the authors observed the greatest IDC occurring in the first few months (<6 months) after treatment before IDC decreases and stabilizes. The IDC difference for early vs. late follow-up time points aligns well with the acute and latent effect of RP and PF, suggesting the feasibility of IDC as a clinical endpoint for RBE. Begosh-Mayne et al. compared the accuracy of four models in predicting lung response as represented by IDC and showed that the LKB NTCP model was the most accurate in predicting both early and the overall radiation effects.

Dahele et al. analyzed 68 patients treated with SBRT [91]. As shown in Figure 7, the authors defined four categories for image density change, in addition to no density change, for pneumonitis: diffuse (at least 5 cm in maximum diameter and which contained more than 50% abnormal lung) consolidation, patchy consolidation, diffuse ground-glass opacities (GGOs), or patchy GGO. Fibrosis was defined as occurring at soon as 6 months post treatment with categories of “modified conventional pattern” of fibrosis (characterized by consolidation, volume loss, and bronchiectasis GGO), mass-like fibrosis (local consolidation in the tumor region but larger than the original tumor), “scar-like fibrosis” (linear opacity in the tumor region with volume loss), besides no density increase. Their follow-up study found that, in short-term, 46% lesions showed no evidence of density increase and that the
incidence for patchy consolidation, diffuse consolidation, diffuse GGO, and patchy GGO were respectively 24%, 16%, 7%, and 6%. Late CT changes were observed for nearly all lesions (67/68; 99%), and the onsite of CT density changes at 6, 12, 24, and 36 months was 56%, 73%, 87%, and 99%, respectively.

Figure 7. Classification of radiographic changes after stereotactic body radiotherapy (SBRT). A: Acute radiological pneumonitis within 6 months post treatment. GGO: ground glass opacity. B: Late radiological fibrosis 6 months post treatment. Source: Dahele et al. [91]

Therefore, these studies suggest the feasibility of IDC as 1) the clinical evidence of radiation damage to be associated with different dose representations to compare their accuracy in representing the delivered dose and 2) the alternative clinical endpoint for RBE instead of congenic survival that is \textit{in vitro}. 
1.5 Image Registration

Dose propagation for dose accumulation and IDC calculation relies on registering the associated images [92]. Image registration is the process of establishing voxel-voxel correspondence between images, through which, dose deposited in the same anatomical voxel can be tracked and accumulated. The registration process returns a transformation that optimally aligns the target image (also known as secondary or moving image) to the same image space with the reference image (also known as primary or fixed image). The registration is optimized using similarity metric (i.e., level of agreement between the registered images) [77]. Categories of image registration include rigid, affine, and deformable.

1.5.1 Rigid and Affine Image Registration

Rigid image registration (RIR) restricts the image mapping to translation and/or rotation. RIR is an optimization problem that maximizes the similarity measurement between the reference image and the target image. The result of RIR is defined by a 4 x 4 transformation matrix that represents the translation and rotation components along each of the x, y, and z axis. The transformation matrix can then be applied to bring the target image into alignment with the reference image in the same image space.
In the treatment planning process, RIR can be used to propagate contours and overlay the functional uptake presented in the PET/CT with the planning CT for planner to define the target region. In patient setups of IGRT, daily images are rigidly aligned to the planning CT to replicate the patient position in the planning exam.

Rigid registration is a special version of affine registration which includes scaling and shearing in addition to the translation and rotation of rigid registration. Similar to rigid, affine also applies a single global transformation to bring the images to alignment.

However, for patient anatomy, tumor and normal tissue may change in shape and size during breathing (intra-fraction) or from day-to-day (inter-fraction). In these situations, simple rigid registration becomes inadequate to maintain the voxel correspondence. A non-rigid (i.e., deformable) approach can allow for more anatomically realistic correspondence between the two images. Nevertheless, RIR acts as the initialization step for deformable image registration (DIR).

1.5.2 Deformable Image Registration

DIR non-rigidly warps images in the presence of anatomical changes induced by breathing motion, tissue response, surgery, etc. [93]. Optimization of DIR algorithms also aims to maximize the similarity measurement between the reference and target images, and
the DIR result is represented by a deformation vector field (DVF). The optimization process is often regularized (i.e., penalized) by anatomically impossible deformations such as folding and invertibility.

![Fusion view of a deformable image registration (DIR).](image)

**Figure 9.** Fusion view of a deformable image registration (DIR). The orange and blue regions are the unaligned regions from the reference and moving images, respectively. Similarity metrics can be based on intensity information or geometric features.

Intensity-based algorithms (e.g., Demon’s algorithm [94]) strive to match the image intensity of all (or patches) of the image voxels through mathematical solutions such as (normalized) mutual information, sum of squared differences [95] and cross-correlation [96]. Intensity-based algorithms are widely applied, but they do not guarantee an anatomically reasonable deformation when intensity information is not preserved, e.g., in presence of tumor response and radiographic based tissue response. Geometric feature-based algorithms do not depend on image intensity information and instead rely on landmarks such as surfaces defined by organ contours. These algorithms can be more robust than intensity-based algorithms in the presence of response. However, landmark placement can be labor-intensive, and feature-based algorithms may not guarantee an accurate alignment outside the defined landmarks, i.e., ignoring anatomical distortions in regions outside these structures.

Therefore, it is important to use the appropriate algorithm for the specific image registration scenario. For example, intensity-based algorithms can be optimal for intra-4DCT...
registration: i.e., between the phases of the same 4DCT without anatomical response.

Subsequently, all phase doses can be mapped and summed to the reference phase to obtain the 4D dose. For registration of longitudinal lung CTs, the registration algorithm may be affected by the type of tumor response observed. Tumor shrinkage can be broadly classified into two types: elastic and inelastic shrinkage [93]. In elastic shrinkage, the tumor pulls the surrounding normal structures inward as it shrinks, thereby preserving the intensity information of the normal structure. In inelastic shrinkage, the tumor regresses and reveals previously hidden normal structures, often leaving behind microscopic extensions, resulting in loss of intensity preservation. We have observed that inelastic shrinkage is more common in LA-NSCLC tumors. Therefore, geometric feature-based algorithms may be optimal for longitudinal registrations such as between the corresponding phases (or AVGs) of two weekly 4DCTs and between the planning 4DCT and the follow-up CT for response assessment.

Nevertheless, algorithms employing a hybrid approach by combining both intensity information and geometric features can provide accurate registration while being anatomically plausible. The ANACONDA algorithm is a commercially available hybrid intensity-based algorithm [97]. The ANACONDA algorithm was shown to provide robust deformations with publicly available image datasets DIR-LAB (data source: https://med.emory.edu/departments/radiation-oncology/research-laboratories/deformable-image-registration/index.html) [98] and POPI (data source: http://www.creatis.insa-lyon.fr/rio/popi-model/) [99]. The algorithm measures the image similarity with correlation coefficient, and a focus region using lung contours is recommended to focus on the intensity information inside the lung when computing the image similarity metric. The regularization process includes minimizing the Dirichlet energy of the DVF, detecting inverted elements at all grid points as negative determinant of the Jacobian matrix, and preventing severe artifacts with large shape deviations of regions of interest (ROIs). The MORFEUS algorithm
is a commercially available biomechanical model-based algorithm [100]. The algorithm subdivides the ROI into triangulated finite elements and simulates organ deformation based on the assigned linear-elastic properties (Poisson’s ratio and Young’s Modulus). A ‘sliding’ boundary condition between the lung surfaces of the registered images simulates the lung breathing motion along the rib cages. The displacements of the finite element nodes inside the organs are calculated using finite element analysis. The generated DVF from the commercial solution can be expanded by externally aligning landmarks defined at the internal vessel bifurcations and including these correspondences as additional internal boundary conditions. For longitudinal lung CTs, the DVF generated from the algorithm can be modified to align lung vessels based on their intensity information, and such hybrid approach has been proven to provide robust results [101]. Detailed explanation of the ANACONDA and MORFEUS algorithms and their applications will be discussed in
Chapter 3: Identifying Reference Image for Inter-4DCT DIR and Chapter 4: Parameter Optimization for Biomechanical Model-Based DIR.

1.5.3 Accuracy Evaluation

DIR is an inherently ill-proposed solution, and it is crucial to quantify the accuracy of the registration.
Chapter 3: Identifying Reference Image for Inter-4DCT DIR and Chapter 4: Parameter Optimization for Biomechanical Model-Based DIR will discuss the methods to quantify the accuracy and robustness of the image registration workflow. AAPM TG-132 [102] provides a detailed guideline on the metrics and tests acquired to evaluate DIR for clinical implantation. These metrics can generally be divided into surface-based vs volumetric-based.

Dice similarity coefficient (DSC) and Hausdorff distance (HD) are two commonly used surface-based metrics. However, since DSC measures the relative volume, for larger organs such as lung, it is not difficult for algorithms to result in high DSC. Additionally, when a boundary condition is applied at the lung surface (e.g., MORFEUS algorithm), it inherently aligns the lung boundary, which forces a high DSC and low HD. However, the internal alignment is crucial for lung dose accumulation, which is not captured by DSC or HD. For this project, we will focus primarily on the target registration error (TRE) that measures the Euclidian distance (e.g., in mm) between the coordinates of the intended location and the actual location of the corresponding point after registration. We empathized on using TRE to measure the internal alignment so we could evaluate and validate the accuracy of voxel-level dose tracking in our dose accumulation study. The schematic is explained in Figure 10. A lower TRE indicates a more accurate voxel-voxel correspondence.

![Figure 10](image.png)

Figure 10. Two corresponding points (black x) in image 1 (target) and image 2 (reference) and the green ‘x’ mark as the mapped location from image 1 to image 2. Target registration
error (TRE) measures the distance between the green ‘x’ mark and the original ‘x’ mark in image 2. Figure taken from Cohen et al. [103].

Accurate dose accumulation requires accurate DIR workflow. However, there remains a need for accuracy assessment for inter-fx DIR for dose mapping. Samavati et al. studied the effect of DIR in 4DCT-based dose accumulation over the course of treatment for 10 SBRT cases and concluded that a maximum of 2.5 Gy difference in the minimum dose delivered to 0.5 cm³ in the tumor were seen when comparing different DIR algorithms [67]. The authors also concluded that dose heterogeneity potentially has significant impact on accumulated dose distribution when mapping contours. However, for this proposed dissertation project, we are focusing on the normal lung tissue in LA-NSCLC patients with larger, more complicated tumor structures. Therefore, quantifying the DIR accuracy used in inter-4DCT dose accumulation will be a focus in our work.

1.6 Summary

The optimal treatment of LA-NSCLC and prediction of radiation-induced lung toxicities continues to present challenges in clinical practice. This dissertation centers on quantifying the accumulation of dose and the biological damage in the normal lung tissue from photon and proton RT. With accurate image mapping and dose tracking enabled by advanced DIR algorithms, we are aiming to gain novel perspectives on the actual amount of damage inflicted, which could aid in making informed clinical decisions.
Chapter 2: Central Hypothesis, Specific Aims, and Dissertation Organization

The long-term goal of the dissertation research is to improve photon and proton NSCLC treatment outcomes by reducing radiation-induced toxicities through a dose-based toxicity prediction model. This model will improve the prediction of radiation-induced lung toxicities, leading to improvements in clinical treatment design. For this specific project, the objective is to fill the knowledge gaps of representing radiation dose and biological damage. I am evaluating our current assumptions in the accuracy of planned dose distribution and fixed RBE of 1.1 for proton patients. To achieve these goals, I am developing and quantifying advanced DIR techniques to track image and dose voxels, and such DIR workflow facilitated robust dose accumulation and response assessment for proton RBE modelling.

2.1 Central Hypothesis

I hypothesize that, for LA-NSCLC patients, dose accumulation using weekly 4DCTs acquired over the course of RT improves the accuracy of planned dose and variable proton dose-RBE relationship in the normal lung can be measured via radiographic change as the clinical endpoint with at least 0.90 goodness of fit. To test the hypothesis, I am retrospectively analyzing imaging and planning data of LA-NSCLC patients enrolled on a clinical trial at MD Anderson that compared efficacies of photon (conventional) and proton radiotherapies.

2.2 Specific Aims

Aim 1. Develop and validate a DIR workflow for 4DCT-based dose accumulation.

Goal: Determine the reference image for inter-4DCT DIR and the optimal parameters for biomechanical model-based DIR algorithm using such reference image.

Study 1-1: Define the reference image for longitudinal 4DCT-based DIR.

Study 1-2: Develop the optimal parameters for BM-DIR
Aim 2. Automate the dose accumulation workflow and evaluate the dosimetric and clinical impact as compared to the planned dose.

**Goal:** Using the DIR workflow defined in Aim 1, accumulate dose using longitudinal 4DCTs and compare the result with planned dose.

**Hypothesis:** At least 15% of patients demonstrate a 10% NTCP difference between planned dose and the accumulated dose and the accumulated dose more accurately describes the planned dose.

**Study 2-2:** organ/region/voxel-level dose comparison

**Study 2-3:** associate dose difference to the accuracy difference through dose-response correlations


**Goal:** Derive and validate variable RBE values using CT image density change

**Hypothesis:** The RBE model will return goodness of fit that achieves 0.9.

**Study 3-1:** Model an established RBE model on the RBE derived from clinical evidence of response.

2.3 Dissertation Organization

This dissertation's main body of work is Chapters 3-6.
Chapter 3: Identifying Reference Image for Inter-4DCT DIR and Chapter 4: Parameter Optimization for Biomechanical Model-Based DIR address Specific Aims 1-1 and 1-2 and are respectively entitled “Identifying Reference Image for Inter-4DCT DIR” and “Parameter Optimization for Biomechanical Model-Based DIR”. Chapter 5: Evaluate Dosimetric and Accuracy Differences between Accumulated and Planned Doses addresses Specific Aim 2 and is entitled “Evaluate Dosimetric and Accuracy Differences between Accumulated and Planned Doses”. Chapter 6: Variable Proton RBE Modelling Using Clinical Radiographic Evidence addresses Specific Aim 3 and is entitled “Variable Proton RBE Modelling Using Clinical Radiographic Evidence”. After the main body of work, a discussion follows in Chapter 7: Discussion, and References are included at the end.
Chapter 3: Identifying Reference Image for Inter-4DCT DIR

This chapter is based upon:


Permission policy of Wiley Periodicals LLC: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

3.1 Abstract

Purpose: Re-planning for four-dimensional computed tomography (4DCT)-based lung adaptive radiotherapy commonly requires deformable dose mapping between the planning average-intensity image (AVG) and the newly acquired AVG. However, such AVG-AVG deformable image registration (DIR) lacks accuracy assessment. The current work quantified and compared geometric accuracies of AVG-AVG DIR and corresponding phase-phase DIRs, and subsequently investigated the clinical impact of such AVG-AVG DIR on deformable dose mapping.

Methods and Materials: Hybrid intensity-based AVG-AVG and phase-phase DIRs were performed between the planning and mid-treatment 4DCTs of 28 non-small cell lung cancer patients. An automated landmark identification algorithm detected vessel bifurcation pairs in both lungs. Target registration error (TRE) of these landmark pairs was calculated for both DIR types. The correlation between TRE and respiratory-induced landmark motion in the planning 4DCT was analyzed. Global and local dose metrics were used to assess the clinical implications of AVG-AVG deformable dose mapping with both DIR types.

Results: TRE of AVG-AVG and phase-phase DIRs averaged 3.2±1.0mm and 2.6±0.8mm, respectively (p<0.001). Using AVG-AVG DIR, TREs for landmarks with <10mm motion
averaged 2.9±2.0mm, compared to 3.1±1.9mm for the remaining landmarks ($p<0.01$). Comparatively, no significant difference was demonstrated for phase-phase DIRs.

Dosimetrically, no significant difference in global dose metrics was observed between doses mapped with AVG-AVG DIR and the phase-phase DIR, but a positive linear relationship existed ($p=0.04$) between the TRE of AVG-AVG DIR and local dose difference.

**Conclusions:** When the region of interest experiences <10mm respiratory-induced motion, AVG-AVG DIR may provide sufficient geometric accuracy; conversely, extra attention is warranted, and phase-phase DIR is recommended. Dosimetrically, the differences in geometric accuracy between AVG-AVG and phase-phase DIRs did not impact global lung-based metrics. However, as more localized dose metrics are needed for toxicity assessment, phase-phase DIR may be required as its lower mean TRE improved voxel-based dosimetry.
3.2 Introduction

Recent advances in radiotherapy (RT) have enabled highly conformal treatment plans. However, over the treatment course, patient’s breathing pattern can change, and the tumor and normal tissue can change in volume, shape, and position in response to the treatment.[104] As a result, the original treatment plan may not be optimal to deliver the prescribed dose, potentially leading to tumor underdosing and healthy tissue overdosing. To account for these changes over the treatment course, the treatment plan must be adapted to the new anatomy as previous studies have demonstrated improved clinical outcomes with lung adaptive RT.[105, 106] The goal of adaptive RT is to maintain target coverage and normal tissue sparing by re-optimizing the treatment plan, based on the most recent imaging data that reflect setup differences, patient anatomy changes, and tumor response.[107, 108] An essential part of adaptive RT workflow is to consider the dose received in the initially completed fractions in the dose optimization based on the newly acquired planning image.[93]

In many RT practices for non-small cell lung cancer (NSCLC), four-dimensional computed tomography (4DCT) is used for treatment planning to account for respiratory-induced motion during beam delivery.[109] Composed of 3-dimensional CT images acquired over full breathing cycles, a 4DCT is usually binned into 10 breathing phases, ranging from end-inhalation phase (T0) to mid-ventilation phase (T3) to end-exhalation phase (T5).[110] An average-intensity image (AVG) can also be created by averaging pixel intensity values of all breathing phases of a 4DCT, and many lung RT practices calculate the radiation dose for treatment planning using AVG because it is composed of the average density seen at each voxel location,[85, 111, 112] which allows the delineation of the internal target volume, comprising of individual target volumes at all phases.[113–115] The current NRG lung template mandates that dose calculation is performed on AVG for free breathing lung cases.[116]
During plan adaptation, calculation of the composite dose requires the dose distribution on the AVG of the original planning 4DCT to be mapped onto the dose distribution on the AVG of the newly acquired 4DCT.[117] This dose mapping can be achieved with deformable image registration (DIR) between the associated AVGs (AVG-AVG DIR).

Intuitively speaking, accurate DIR is the prerequisite for accurate dose deformation for plan adaptation, and for 4DCT-based NSCLC treatment planning, it is important to quantify the uncertainty in DIR between the primary and adaptive 4DCTs. Although numerous studies have evaluated the accuracy of intra-4DCT DIR (e.g. between T0 and T5).[118–122] very limited number of studies investigated the accuracy of inter-4DCT (phase-phase) DIR to analyze the effect of longitudinal anatomic changes over the course of RT.[101] The RTOG 1106 trial implemented AVG-based planning for free-breathing treatments and rigidly registered the original and mid-treatment AVGs for dose propagation because the uncertainty of AVG-AVG DIR was unknown.[123] Therefore, the current work evaluated the accuracy of AVG-AVG DIR to provide the uncertainty within this adaptive treatment planning workflow for NSCLC.

In this study, the geometric accuracy of AVG-AVG DIR was quantified using corresponding anatomic landmarks and was subsequently compared with the geometric accuracy of DIR between corresponding phase pairs of the 4DCTs. The accuracy of AVG-AVG DIR and phase-phase DIR was also evaluated with respect to landmark motion observed for each patient during standard respiratory motion at the planning stage. Finally, by evaluating the dosimetric impact of such geometric accuracy, implications of AVG-AVG deformable dose mapping for adaptive re-planning were investigated and the associated clinical guidelines were provided.
3.3 Methods and Materials

3.3.1 Patient Data

The original planning 4DCT and a mid-treatment 4DCT (approximately four weeks into treatment to mimic an adaptive RT setting) were retrospectively evaluated for 28 randomly selected patients, who were previously treated with intensity-modulated RT under an Institutional Review Board–approved randomized clinical trial for locally advanced NSCLC.[85] Weekly 4DCTs, 6-8 per patient, were acquired during the treatment course for motion and target re-assessment. These 4DCTs were acquired with voxel size of 0.98x0.98x2.5 mm³ using the Discovery CT system (General Electric Healthcare, Waukesha, WI) with 120kVp, operated in cine mode. Patient surface motion was monitored by Varian’s Real-time Position Management Respiratory Gating (Varian Medical Systems, Palo Alto, CA).[124]

3.3.2 Hybrid Intensity-Based DIR

The DIR algorithm used for this evaluation was ANACONDA (ANAtomically CONstrained Deformation Algorithm), a hybrid intensity–based DIR algorithm commercially available in RayStation v9 (RaySearch Laboratories, Stockholm, Sweden).[118] The image registration process is as follows. Prior to image deformation, reference and target images are rigidly registered based on image similarity measured over all voxels that are enclosed by the external body contours in both images. The optimization process of ANACONDA is based on image similarity as measured by a correlation coefficient and is solved by a quasi-Newton algorithm.[118] This process is regularized by minimizing the Dirichlet energy of the generated deformation vector field (DVF). DVF smoothness is maintained by penalizing large shape deviations of regions of interest (ROIs) defined in the reference image, and invertibility is checked by the determinant of the Jacobian. When the optimization process is
constrained by user-delineated organ contours in both reference and target images, such geometric information makes the ANACONDA method hybrid.

In the current work, rigid registration was first established between the planning AVG and the mid-treatment AVG, focusing on bone and tumor regions and discarding rotations to mimic the daily kilovolt alignment. With ANACONDA, T0, T3, T5, and AVG of the mid-treatment 4DCT were registered to their counterparts of the planning 4DCT (as the reference image). Left and right lung boundaries were manually contoured in each image pair to guide the DVF as controlling ROIs. The rigid and deformable registrations were systematically performed through RayStation scripting. For each registration, the results were qualitatively assessed via image fusion of the deformed target image and the reference image with a focus on the bronchial and vascular alignment, per recommendation of the American Association of Physicists in Medicine (AAPM) Task Group 132.[102]

3.3.3 Quantitative Metrics for DIR Accuracy Assessment

3.3.3.a Dice Similarity Coefficient

AAPM Task Group 132 recommends metrics including Dice similarity coefficient (DSC) and target registration error (TRE) for validation of DIR accuracy. DSC is a measure of overlap between the ROI in the reference image and the same ROI in the deformed target image:

\[ DSC = \frac{2 \cdot (V_{\text{deformed}} \cap V_{\text{reference}})}{V_{\text{deformed}} + V_{\text{reference}}} \]  

(3)

where \( V_{\text{deformed}} \) is the volume of the deformed ROI and \( V_{\text{reference}} \) is the volume of the reference ROI.[125] DSCs of lung contours across all DIRs were obtained.

3.3.3.b Target Registration Error

Unlike DSC, which focuses on the alignment of the organ contour, TRE addresses the internal alignment of images. TRE is defined as the three-dimensional Euclidian distance between a landmark’s position in the deformed target image and its location in the reference
image:

$$TRE = \sqrt{(x_{deformed} - x_{reference})^2 + (y_{deformed} - y_{reference})^2 + (z_{deformed} - z_{reference})^2}$$ (4)

where $x$, $y$, and $z$ are the Cartesian coordinates in the reference image space. In the current work, vessel bifurcations were used as landmarks because of their abundance in the lungs and high contrast against air-like lung tissue. Landmark pairs were identified on corresponding phases (T0, T3, and T5) of the planning 4DCT and the mid-treatment 4DCT. These landmark pairs were overlaid onto the AVGs of the corresponding 4DCTs to assess the accuracy of AVG-AVG DIR, given that the AVG comprises all phases and exists in the same image space with phases of the same 4DCT. As a result, TRE of each landmark pair was computed for AVG-AVG DIR and phase-phase DIRs (T0-T0 DIR, T3-T3 DIR, or T5-T5 DIR).

3.3.4 Automatic Landmark Identification Method

Manual landmark identification is a cumbersome process that could introduce uncertainties such as inter-observer variability. For the current work, the following in-house fully-automatic landmark identification workflow was used.[126] By thresholding the reference and target images, vessels on these images were automatically segmented, through which the centerlines could be extracted and the bifurcations on the centerlines detected using the neighbors’ count. The segmented vessels on the reference and target images were then registered via a separate intensity-based Demons DIR algorithm,[94] after which bifurcations less than 4 mm apart after the deformation were considered landmark pairs. Landmarks in both lungs were automatically identified on T0 pairs, T3 pairs, and T5 pairs. This workflow has been previously validated against 10 pairs of T0/T5 lung 4DCTs, each with 300 manually identified landmark pairs (DIR-Lab, http//www.dir-lab.com) and 10 pre-/post-RT pairs of liver contrast-enhanced CTs with 5 manual landmarks each.
In the current work, the automatic workflow was also validated against manually identified landmarks in the ipsilateral lung from a random subset of 10 patients. Compared with the contralateral lung, the ipsilateral lung could inherently exhibit larger anatomic changes between the two 4DCTs owing to tumor response. For the validation, 16 anatomic landmark pairs were manually identified at vessel bifurcations in the ipsilateral lung in T0 pairs. These landmarks were uniformly distributed along the superior-inferior direction of the lung to potentially cover a wide range of respiratory motion exhibited by different regions of the lung.[127, 128] The same 16 landmark pairs were then re-identified in the T3 pairs and T5 pairs. Each landmark (position) was then directly copied to the AVG pairs, the same way as in the automatic workflow. In total, for each 4DCT set, 16 landmarks were identified in each of the three phases, and three variations of these 16 landmark positions (48 total) were identified in the AVG. The above process is described in Figure 1. Linear regression was used to compare mean TRE values from each phase-phase DIR of manual landmarks against the mean TRE values of automatic landmarks.

Figure 11. Manual landmark identification process to validate the automatic landmark workflow. The planning four-dimensional computed tomography (4DCT) and the mid-
treatment 4DCT are shown. Coronal views of the end-inhalation phase (T0), mid-ventilation phase (T3), and end-exhalation phase (T5), as well as the average-intensity image (AVG), are shown in the left column. Simplified cartoons of the two 4DCTs are shown on the right. Deformable image registrations (DIRs) were established between corresponding phases and AVGS of the two 4DCTs. A landmark at the lower vessel bifurcation in the right lung was identified with different colors in different phases. The landmark’s locations in these phases were transferred onto the AVG of each corresponding 4DCT.

3.3.5 Landmark-Based DIR Accuracy Measurement

3.3.5.a Comparison of Geometric Accuracy between AVG-AVG DIR and Phase-Phase DIRs

The TRE of each landmark pair from T0-T0 DIR, T3-T3 DIR, T5-T5 DIR, and AVG-AVG DIR was obtained for all 28 patients. Paired Student’s t-test was used to compare TRE of AVG-AVG DIR and TRE of T0-T0 DIR, based on T0-T0 landmark pairs, and such comparison was repeated for T3-T3 landmark pairs and T5-T5 landmark pairs. The same statistical testing was also used for comparisons among the phase-phase DIRs.

3.3.5.b Effect of Breathing Motion

For each patient, respiratory-induced landmark breathing motion was represented by the displacement of landmarks under hybrid intensity-based DIR between T5 (reference exam) and T0 of the planning 4DCT, with both lungs as controlling ROI. Assuming TREs of landmark pairs in the AVG pair and corresponding phase pairs follow a Gaussian distribution, we used linear regression to quantify the effect of landmark breathing motion on TRE for AVG-AVG DIR and T5-T5 DIR. A more simplified comparison was made in which these landmarks were divided using a cutoff of motion magnitude of 10 mm, and a two-sample unequal variance t-test was used to compare the average TRE values for landmarks that showed at least 10 mm motion and those that showed less than 10 mm motion, for both AVG-AVG DIR and T5-T5 DIR.

3.3.5.c Clinical Impact of DIR Geometric Accuracy on Dose Mapping

Hybrid intensity-based DIR was established between AVGS and between T5s (from planning to mid-treatment) with both lungs as controlling ROIs. Accordingly, the original planned dose on AVG, calculated with RayStation’s default uniform dose grid of 3mm, was
deformed to the mid-treatment 4DCT to simulate adaptive RT with both AVG-AVG DIR and T5-T5 DIR (warranted by AVG and phase existing in the same image space). For the ipsilateral lung, multiple linear regression was performed to correlate AVG-AVG DIR TRE, along with several planning metrics, to the absolute difference in mean lung dose (MLD) between the planned dose deformed with AVG-AVG DIR vs. with T5-T5 DIR, and the absolute difference in the volume of lung receiving at least 20Gy ($V_{20}$) between the planned dose deformed with AVG-AVG DIR vs. with T5-T5 DIR, as clinical endpoints. The planning metrics include gross tumor volume (GTV) motion in the superior-inferior direction, GTV volume inside lung, GTV center of gravity to diaphragm, diaphragm breathing motion in the superior-inferior direction, GTV dose homogeneity index (DHI) (concept proposed by Ding et al. [129]), DHI of normal lung (lung excluding GTV), and percent of primary tumor volume (PTV) in normal lung. In addition, the dose discrepancy on a voxel/sub-regional level was evaluated using the percentage of landmarks on T5 that had at least 2Gy absolute dose difference when the dose distribution was mapped using AVG-AVG DIR vs. T5-T5 DIR.

### 3.4 Results

AVG-AVG DIR and phase-phase DIR were successfully performed for all patients. On average, 654 ± 162, 603 ± 186, and 606 ± 194 landmarks were identified on each patient’s T0, T3, and T5 pairs, respectively. The average DSC of left and right lung contours combined was 0.95 ± 0.02 for AVG-AVG DIR and 0.93 ± 0.03 for all phase-phase DIRs. AAPM Task Group 132 considers DSC of 0.80-0.90 satisfactory within the contouring uncertainty of the structures[102]. Therefore, according to this metric, the DIRs were deemed successful.

#### 3.4.1 Validation of Automatic Landmark Identification Workflow

For 10 randomly selected patients, sixteen landmarks were manually identified where they were uniformly distributed along the superior-inferior direction within the
ipsilateral lung. TREs of these landmarks were compared against those of the automatic landmarks for each phase-phase DIR (Figure 12). With a slope of greater than 1.0, the automatic method consistently reported a larger TRE compared with the manual method (average TRE $2.5 \pm 0.9$ mm for automatic compared with $2.4 \pm 1.7$ mm for manual). However, the coefficient of determination was 0.8, and the averaged difference between the two methods was 0.1 mm, substantially smaller than the largest voxel dimension (2.5 mm). In addition, the paired Student’s $t$-test showed no statistical significance ($p = 0.55$), indicating that the automatic method was an acceptable substitute for the manual method.

![Figure 12](image)

Figure 12. Target registration error (TRE) differences between deformable image registration (DIR) across average-intensity images and DIR across phases for each patient. Each data point represents a labeled patient. The result of linear regression is $TRE_{\text{auto}} = 0.5 \times TRE_{\text{manual}} + 1.4$. Coefficient of determination is 0.81.

### 3.4.2 AVG-AVG DIR Compared with Phase-Phase DIRs

As shown in Figure 13, the mean TRE of T0 landmark pairs was $3.6 \pm 1.1$ mm for AVG-AVG DIR compared with $2.8 \pm 0.8$ mm for T0-T0 DIR ($p < 0.001$). The mean TRE of T3
landmark pairs was 3.0 ± 0.9 mm for AVG-AVG DIR compared with 2.6 ± 0.8 mm for T3-T3 DIR (p < 0.001). The mean TRE of T5 landmark pairs was 3.0 ± 0.8 mm for AVG-AVG DIR compared with 2.5 ± 0.8 mm for T5-T5 DIR (p < 0.001). In total, AVG-AVG DIR resulted in a mean TRE of 3.2 ± 1.0 mm compared with 2.6 ± 0.8 mm for phase-phase DIRs (p < 0.001). For all patients except one, TRE for AVG-AVG DIR was higher than TRE for phase-phase DIRs (for such patient, the TRE for AVG-AVG DIR was 0.07 mm lower than that for T0-T0 DIR and 0.03 mm lower than that for T3-T3 DIR).

Figure 13. Boxplots of mean target registration error (TRE) of the 28 patients. Each color pair of boxplots represents the mean TRE of phase pairs for the corresponding phase-phase DIR and for AVG-AVG DIR. The standard deviation of TRE ranged from 1.0 to 3.4 mm (not shown). *DIR: deformable image registration, AVG: average-intensity image, T0: end-inhalation phase, T3: mid-ventilation phase, T5: end-exhalation phase

3.4.3 Geometric Impact of Landmark Motion on DIR TRE

Respiratory-induced landmark motion was represented by the displacement of T5 landmarks under T5-T0 DIR of the planning 4DCT. Linear regressions between landmark motion and TRE for AVG-AVG DIR as well as for T5-T5 DIR both yielded coefficient of
determination values of less than 0.1. When comparing landmarks grouped using a cutoff of 10 mm motion, for AVG-AVG DIR, the mean TRE of T5 landmarks with less than 10 mm motion was 2.9 ± 2.0 mm compared with 3.1 ± 1.9 mm for T5 landmarks with at least 10 mm motion (p < 0.001), whereas for phase-phase DIR, the mean TRE of T5 landmarks with less than 10 mm motion was 2.5 ± 2.0 mm compared with 2.5 ± 1.9 mm for T5 landmarks with at least 10 mm motion (p = 0.30). Therefore, landmark pairs with at least 10 mm motion had significantly larger TREs than those with less than 10 mm motion in AVG-AVG DIR, which was not observed for phase-phase DIR.

3.4.4 Clinical Impact of DIR Geometric Accuracy on Dose Mapping

Figure 14 shows for patient #14 the comparison of the deformed planned dose with AVG-AVG DIR vs. T5-T5 DIR. This patient carried the largest absolute difference in MLD between the two deformed doses.

![Figure 14](image)

Figure 14. This figure shows, for patient 14, deformed doses (left panels) and their difference (right panel) on an axial slice of the average-intensity image (AVG) of the mid-treatment week. The upper-left and lower-left panels show the deformed planned dose with AVG-AVG DIR and with T5-T5 DIR, respectively. *DIR: deformable image registration, T5: end-exhalation phase.

Ipsilateral lung planning information for all 28 patients is shown in Table 1. Multilinear regression resulted in non-significant correlations (p > 0.05) when correlating AVG-AVG DIR TRE as well as clinical planning information (columns 3-9) to the difference in
global clinical metrics of MLD and $V_{20}$ between doses mapped with AVG-AVG DIR versus T5-T5 DIR (columns 11 and 12). However, when comparing these two mapped doses on a sub-regional level, both TRE of AVG-AVG DIR and GTV motion reached statistical significance ($p = 0.04$ and $p = 0.01$, respectively), and percent of PTV in normal lung achieved $p = 0.06$, when correlated against the metric: portion of T5 landmarks that had at least 2 Gy absolute dose difference ($\geq 2\text{Gy}\%$) (column 10). For this local dose metric, 12 patients achieved at least 10% (at least 10% of total T5 landmarks had at least 2 Gy absolute dose difference). Eleven of these twelve patients met at least one of the following conditions: TRE larger than 3.0 mm (recommended dose grid size, per AAPM Task Group 132[102]), GTV motion larger than 5.0 mm (AAPM Task Group 76[130] recommends that respiratory management techniques be considered when tumor motion is larger than 5.0 mm), or percent of PTV in normal lung larger than 20%, achieving sensitivity of 0.92. Only 3 of the 16 patients with less than 10% of $\geq 2\text{Gy}\%$ demonstrated any of these three criteria, achieving a specificity of 0.81.

Table 1. Planning information for the ipsilateral lung of each analyzed patient.

| Patient Characteristics columns: | Diaph motion: the largest distance between two voxels on the diaphragm between end-exhalation phase and end-inhalation phase along the superior-inferior direction. GTV motion: the distance between the center of gravity of the GTVs in end-exhalation phase and end-inhalation phase along the superior-inferior direction. GTV to Diaph: the distance between the center of gravity of the GTV to the diaphragm along the superior-inferior direction. GTV vol: volume of GTV inside normal lung (lung volume excluding GTV). PTV%: percent PTV volume in normal lung. |
| Dose Metrics columns: | $|\Delta MLD|$: absolute difference in MLD between the planned dose deformed with AVG-AVG DIR vs. with T5-T5 DIR $|\Delta V_{20}|$: absolute difference in $V_{20}$ between the planned dose deformed with AVG-AVG DIR vs. with T5-T5 DIR $\geq 2\text{Gy}\%$: Portion of T5 landmarks with $\geq 2$ Gy absolute dose difference between doses mapped with AVG-AVG DIR vs T5-T5 DIR |

| Patient | AVG- AVG DIR TRE [mm] | Diaph motion [mm] | GTV motion [mm] | GTV to Diaph [mm] | GTV vol [cm$^3$] | PTV% | GTV DHI | DHI of normal lung | $|\Delta MLD|$ [cGy] | $|\Delta V_{20}|$ ≥2Gy% |
|---------|------------------------|-------------------|-----------------|------------------|----------------|------|--------|-------------------|---------------|------------------|
| 1       | 2.0                    | 16.0              | 4.4             | 80               | 9              | 13.4 | 3.5    | 0.9               | 7             | 0.1              | 2.1             |
| 2       | 2.3                    | 7.6               | 1.5             | 79               | 32             | 11.0 | 4.2    | 0.9               | 7             | 0.0              | 9.6             |
| 3       | 1.8                    | 5.1               | 1.7             | 73               | 25             | 8.4  | 2.4    | 0.9               | 20            | 0.3              | 0.0             |
| 4       | 2.2                    | 19.5              | 4.0             | 124              | 5              | 3.1  | 2.6    | 0.5               | 7             | 0.4              | 4.8             |
| 5       | 2.6                    | 10.4              | 6.2             | 20               | 19             | 13.8 | 2.9    | 0.8               | 55            | 0.8              | 16.5            |
| 6       | 2.4                    | 10.2              | 0.8             | 56               | 90             | 18.1 | 1.8    | 1.0               | 15            | 0.1              | 0.4             |
| 7       | 2.7                    | 5.5               | 0.6             | 85               | 9              | 10.4 | 2.2    | 0.9               | 37            | 0.6              | 6.3             |
| 8       | 5.7                    | 4.5               | 0.9             | 76               | 21             | 5.5  | 3.5    | 0.5               | 5             | 0.1              | 2.1             |
| 9       | 1.7                    | 14.7              | 0.6             | 35               | 29             | 12.4 | 2.4    | 1.0               | 20            | 0.3              | 3.2             |
| 10      | 3.1                    | 27.3              | 3.7             | 27               | 180            | 8.6  | 4.6    | 0.9               | 80            | 2.0              | 16.6            |
| 11      | 2.1                    | 8.3               | 1.4             | 19               | 12             | 12.1 | 2.5    | 0.9               | 11            | 0.1              | 4.3             |
| 12      | 3.1                    | 5.5               | 1.9             | 32               | 193            | 24.1 | 4.0    | 1.0               | 16            | 0.1              | 6.0             |
| 13      | 1.9                    | 26.5              | 0.9             | 81               | 8              | 7.1  | 2.6    | 0.7               | 1             | 0.0              | 2.1             |
| 14      | 4.8                    | 5.0               | 0.5             | 12               | 18             | 7.3  | 2.6    | 0.8               | 84            | 1.6              | 15.5            |
| 15      | 2.2                    | 6.2               | 4.0             | 90               | 1              | 8.1  | 2.7    | 0.8               | 28            | 0.9              | 17.8            |
| 16      | 4.7                    | 9.3               | 0.4             | 93               | 363            | 13.4 | 3.0    | 1.0               | 18            | 0.2              | 12.2            |
| 17      | 2.7                    | 10.4              | 1.7             | 80               | 192            | 23.4 | 4.1    | 0.9               | 12            | 0.1              | 10.2            |
| 18      | 3.8                    | 14                | 3.6             | 112              | 235            | 12.6 | 4.2    | 1.0               | 32            | 0.1              | 11.8            |
| 19      | 2.4                    | 1.4               | 2.8             | 25               | 53             | 16.8 | 1.7    | 0.8               | 39            | 0.5              | 5.3             |
| 20      | 5.0                    | 9.6               | 3.0             | 75               | 28             | 8.1  | 4.5    | 0.8               | 2             | 0.2              | 11.7            |
| 21      | 2.3                    | 1.0               | 0.6             | 90               | 21             | 7.7  | 2.5    | 0.8               | 1             | 0.0              | 0.9             |
| 22      | 4.3                    | 3.6               | 1.8             | 94               | 35             | 12.3 | 2.2    | 0.9               | 21            | 0.5              | 10.7            |
| 23      | 2.4                    | 10.6              | 1.0             | 101              | 9              | 8.6  | 6.3    | 1.0               | 38            | 0.6              | 8.8             |
| 24      | 5.4                    | 7.8               | 4.7             | 51               | 23             | 4.7  | 3.3    | 0.5               | 21            | 0.2              | 21.6            |
| 25      | 2.3                    | 17.7              | 2.2             | 72               | 41             | 13.9 | 2.9    | 0.9               | 55            | 0.8              | 7.4             |
| 26      | 3.3                    | 28.3              | 7.7             | 90               | 33             | 5.3  | 2.4    | 0.6               | 48            | 1.0              | 8.2             |
| 27      | 1.9                    | 14.2              | 8.6             | 55               | 19             | 11.2 | 3.6    | 0.9               | 5             | 0.0              | 10.9            |
| 28      | 4.6                    | 15.1              | 2.9             | 81               | 53             | 18.5 | 3.3    | 0.9               | 58            | 0.7              | 10.7            |
3.5 Discussion

In the current work, we evaluated the accuracy of AVG-AVG DIR and compared it with that of DIR between corresponding phases (T0, T3, and T5) for 4DCT-based treatment planning in NSCLC. Based on imaging data and landmark pairs from 28 NSCLC patients, our findings indicated that AVG-AVG DIR has larger uncertainty, i.e., inferior accuracy, compared with phase-phase DIR, and that breathing-induced motion degrades the accuracy of AVG-AVG DIR more than that of phase-phase DIR.

AAPM Task Group 132 recommends that the target TRE should be less than the maximum image voxel size.[102] Therefore, given that 4DCTs used in the current work were 2.5 mm thick, T5-T5 DIR achieved the AAPM-recommended target TRE, with mean TRE of 2.5 ± 0.8 mm, and thus can be deemed clinically acceptable. For phases typically subject to larger breathing artifacts,[131] T3-T3 DIR and T0-T0 DIR almost met the target TRE, with 2.6 ± 0.8 mm and 2.8 ± 0.8 mm, respectively. However, with an overall mean TRE of 3.2 ± 1.0 mm, AVG-AVG DIR did not meet the uncertainty recommendation, with more than half of the lateral resolution (0.98 mm) over the recommended threshold. In addition, our analysis of the effect of motion showed that large landmark respiratory motion (i.e., at least 10 mm) affects AVG-AVG DIR more than phase-phase DIR. In summary, phase-phase DIR was more robust against breathing motion, which was not achieved by the AVG-AVG DIR. Therefore, per AAPM Task Group 132, additional uncertainty should be assumed, depending on the clinical protocol, if AVG-AVG DIR is used for dose mapping. This is especially true for patients who breathe with a relatively large magnitude, which can potentially introduce a large amount of uncertainty, as well as for anatomic structures that typically have a large motion, such as tumors close to the diaphragm.

Geometrically robust points, which are necessary for TRE, require a geometrically robust image. AVG is constructed from averaging the pixel intensities of all phases over the
breathing cycle. As a result, anatomy on AVGs is blurred due to breathing motion and thus carries inherent uncertainty in representing the true anatomic shapes, which makes AVG not geometrically robust. Practically speaking, AVG is commonly used as the planning image for dose calculation because it captures the entire tumor movement under respiratory motion.[111, 132–134] This helps to avoid overdosing normal tissue near the tumor and under-dosing the tumor itself, which can happen with a smooth dose intensity map around the tumor. Although AVG-AVG DIR is the more straightforward choice for clinicians, extra care must be taken when using AVG-AVG DIR to perform dose mapping when adaptive planning is required, given the small but potentially clinically significant uncertainty. Furthermore, considering the effect of TRE on dose metrics, a TRE reduction of 1.6 mm for phase-phase DIR used to deform and propagate the planning dose has been shown to clinically affect decision-making in stereotactic body RT treatment planning.[82] These findings support the potential clinical benefit of using phase-phase DIR rather than AVG-AVG DIR.

The major strength of the current work was the large number of landmark pairs made available by the automatic landmark identification workflow (average of 600+ per DIR). As a result, this work contributed to closing the gap on evaluating inter-4DCT DIR in the presence of anatomy changes through internal landmark points. In addition to using landmarks to determine TRE as the internal metric, we also used DSC of the left and right lung contours as the external metric. Because the hybrid intensity-based DIR method was constrained by matching the lung contours on the reference and target images, DSC values would naturally be large. In addition, the DSC can quantify how well the organ boundary matches, but it does not guarantee accurate modeling of internal volume.[135] Therefore, using DSC as the sole metric to represent DIR accuracy would be inadequate, and consequently DSC was reported to supplement TRE.
In terms of landmark identification on AVG, the same anatomic landmarks were detected on the three phase images on both 4DCTs and then were directly used for AVG-AVG DIR. The presented method that transfers landmarks defined in the breathing phases onto the AVG of the same 4DCT enabled representation of the actual position of these landmarks (the average of a landmark’s locations in breathing phases) in AVG, which is difficult to identify directly, especially in regions that experience large respiratory-induced motion.

In our validation of automated landmarks, automatic landmarks achieved larger TREs than manual landmarks. This is potentially due to landmarks identified by automatic workflows systematically produce worse DIR, so identifying landmarks throughout the entire lung would on average produce larger TREs. This result could also be attributed to a lack of one-one correspondence between automatic landmarks and manual landmarks and the limited number of manual landmark pairs identified. A small set of landmark points has been shown to be insufficient for calculating TRE because this misrepresents the actual spatial accuracy.[136] In fact, only 16 landmark pairs were chosen manually per patient due to the cumbersome nature of manual landmark identification, thus potentially limiting the TRE accuracy based on these landmarks.

The difference in DIR accuracy among the three phases (T0, T3, and T5) could be attributed to differences in the magnitude of breathing artifacts. At T0, the full inhale phase, the patient changes breathing direction, so this phase is most prone to breathing artifacts due to diaphragm motion. In addition, the inhale position tends to vary from breathing cycle to breathing cycle. In contrast, in T5, the patient spends a longer proportion of the breathing phase in the exhale position, thus leading to the fewest motion artifacts; vessels especially near the diaphragm are not “washed out” and can be more accurately identified. Therefore, T5-T5 DIR can provide the highest registration accuracy among phase-phase DIRs. This can be applied in four-dimensional dose calculation, with phase dose mapped and summed
on T5 as the gold-standard dose calculation method for a 4DCT.[78] Therefore, our findings can support the investigation of the differences between T5-based four-dimensional planned dose and accumulated dose using T5-based four-dimensional weekly doses with T5-T5 DIR. Retrospective toxicity studies on a voxel or regional level are also warranted, to achieve the most accurate representation possible for the delivered dose.

In the dosimetric study, as the landmark TRE increased, the probability of the dose to an image voxel in the planning AVG being mapped to a different anatomical location in the re-planning AVG also increased. Such effect was pronounced in the high dose-gradient region near the GTV where adjacent image voxels shared large dose differences. It was magnified when the GTV motion was large, which further increased TRE of AVG-AVG DIR. However, their effect was lessened on global clinical metrics such as mean lung dose, where these voxel-level dose differences cancel each other out. For instance, patient 24 had AVG-AVG DIR TRE of 5.4 mm (largest among all patients) and achieved 21.6% in the previously defined metric ≥2Gy%, but only achieved 21 cGy and 0.2 in absolute MLD difference and V20 difference, respectively. On the contrary, patient 8 carried AVG-AVG DIR TRE of 5.7 mm but only achieved 2.1% in ≥2Gy% due to low GTV motion (0.9 mm) and low PTV in normal lung (5.5%). These results are similar to those concluded by Hardcastle et al.[137] where larger registration error in high dose gradient regions caused larger dosimetric uncertainty. These observations demonstrated that the clinical impact of DIR geometric accuracy, as represented by TRE, must be interpreted in the context of the patient and dosimetric parameters. Therefore, when dose is mapped with AVG-AVG DIR vs. phase-phase (T5-T5) DIR, global clinical metrics are unlikely to differ significantly. However, if local doses are evaluated based on the mapped dose, phase-based mapping should be considered.

Lastly, to provide context in the clinical adaptive workflow, the following three simplified workflows in the order of increasing complexity can be considered. 1) 3D
treatment planning using AVG-AVG DIR: original planned dose is calculated on AVG of planning 4DCT; at adaptation, the fraction-corrected planned dose is deformed to the AVG of the newly acquired 4DCT with AVG-AVG DIR; the adapted dose is created based on the deformed dose on the new AVG. 2) Three-dimensional treatment planning using phase-phase DIR: this is an alternative to 1) where a specific phase-phase DIR is performed and used to map the dose calculated on the original AVG onto the secondary AVG for better accuracy. 3) Full four-dimensional treatment planning using phase-phase DIR: dose is calculated on each phase of the planning 4DCT, deformed to a reference phase, and summed over the breathing phases; the fraction-corrected 4D planned dose on a specified phase is deformed to the corresponding specified phase of the newly acquired 4DCT with phase-phase DIR; the adapted dose is created based on the deformed dose.[77]

3.6 Conclusion

In the current work, TRE was used to quantify the geometric uncertainty of DIRs between corresponding AVGs and corresponding phases for 4DCT-based lung adaptive RT for 28 NSCLC patients. When the region of interest has respiratory-induced motion less than 10 mm, AVG-AVG DIR may provide sufficient clinical accuracy; however, when motion is at least 10 mm, extra attention is warranted and phase-phase DIR, especially T5-T5 DIR, is recommended. Dosimetrically, the geometric accuracy of AVG-AVG DIR has not been shown to significantly impact global lung-based clinical metrics. However, phase-phase DIR may be required for advanced toxicity correlation studies that utilize local dose metrics (e.g., dose used in voxel-based analyses).
Chapter 4: Parameter Optimization for Biomechanical Model-Based DIR

This chapter is based upon:


Permission policy of Wiley Periodicals LLC: The author is hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process.

4.1 Abstract

Background: Successful generation of biomechanical model-based deformable image registration (BM-DIR) relies on user-defined parameters that dictate surface mesh quality. The trial-and-error process to determine the optimal parameters can be labor intensive and hinder DIR efficiency and clinical workflow.

Purpose: To identify optimal parameters in surface mesh generation as boundary conditions for a BM-DIR in longitudinal liver and lung CT images to facilitate streamlined image registration processes.

Methods: Contrast-enhanced CT images of 29 colorectal liver cancer patients and end-exhale 4DCT images of 26 locally advanced non-small cell lung cancer patients were collected. Different combinations of parameters that determine the triangle mesh quality (voxel side length and triangle edge length) were investigated. The quality of DIRs generated using these parameters was evaluated with metrics for geometric accuracy, robustness, and efficiency. Metrics for geometric accuracy included target registration error (TRE) of internal vessel bifurcations, dice similar coefficient (DSC), mean distance to agreement (MDA), Hausdorff Distance (HD) for organ contours, and number of vertices in the triangle mesh. AAPM Task Group 132 was used to ensure parameters met TRE, DSC, MDA recommendations before the comparison among the parameters. Robustness was
evaluated as the success rate of DIR generation, and efficiency was evaluated as the total
time to generate boundary conditions and compute finite element analysis.

**Results:** Voxel side length of 0.2 cm and triangle edge length of 3 were found to be the
optimal parameters for both liver and lung, with success rate of 1.00 and 0.98 and average
DIR computation time of 100 and 143 seconds, respectively. For this combination, the
average TRE, DSC, MDA, and HD were 0.38-0.40 mm, 0.96-0.97, 0.09-0.12 mm, and 0.87-
1.17 mm, respectively.

**Conclusion:** The optimal parameters were found for the analyzed patients. The decision-
making process described in this study serves as a recommendation for BM-DIR algorithms
to be used for liver and lung. These parameters can facilitate consistence in the evaluation
of published studies and more widespread utilization of BM-DIR in clinical practice.

**4.2 Introduction**

Biomechanical model-based deformable image registrations (BM-DIRs) can aid
image-guided radiation therapy in registering liver and lung anatomies, allowing contour
propagation, tumor tracking, dose accumulation, and response assessment.[92, 119, 138,
139] They have been reported to provide improved accuracy and more realistic deformation
compared to intensity-based DIRs.[101, 140–142] However, biomechanical models estimate
a dense deformation vector field (DVF) by necessitating organ-specific boundary conditions
applied on organ contours to estimate the finite element model (FEM) representing the
patient anatomy before performing the finite element analysis (FEA).[143] which can be
more time consuming than classical intensity-based DIR algorithms.[144] Additionally, user-
defined parameters such as boundary condition and mesh resolution dictaFEM quality
and eventually affect the topology and success of the generated DVF. As a result, the
current implementation of BM-DIR is limited by uncertainties in model parameters which
further increases registration time and vulnerability. From our clinical research experience,
DIR failure can occur in approximately 15% cases. This disrupts the image registration workflow and especially hinders the efficiency of inter-fractional dose mapping where the DIRs sharing the same reference image would consistently fail. Therefore, identifying parameters that consistently produced accurate registrations is crucial towards a more comprehensive integration of BM-DIRs into clinical workflow, especially when longitudinal registrations are involved such as adaptive radiation therapy and image/dose deformations for response assessment.[139, 145–147]

A BM-DIR algorithm (Morfeas) has been implemented in a commercial treatment planning system (TPS), (RayStation, RaySearch Laboratories, Stockholm, Sweden).[148] In this study, we investigated the different parameters in surface mesh generation that influence the boundary conditions in the estimation of a FEM-based DVF for longitudinal liver and lung images. The robustness and efficiency of the DIR process using these parameters for liver and lung anatomies were compared to find the optimal parameters.

4.3 Materials and Methods

4.3.1 Patient data

For liver registrations, pre- and post-treatment contrast-enhanced breath-hold CT images of 29 colorectal liver cancer patients were obtained under a retrospective institutional review board (IRB)-approved protocol. These patients were previously treated with microwave ablation to treat colorectal liver metastasis. The median image voxel size was 0.74 x 0.74 x 2.5 mm³ (minimum: 0.56 x 0.56 x 2 mm³, maximum: 0.88 x 0.88 x 5 mm³). For each patient, the liver was manually contoured by a graduate student under the guidance and review of a board-certified interventional radiologist (BO).
For lung registrations, four-dimensional CT (4DCT) images of 26 locally advanced non-small cell lung cancer patients without major atelectasis or substantial tumor shrinkage were acquired under a retrospective IRB-approved protocol. These patients were previously treated with concurrent chemo-radiotherapy on a prospective clinical trial. End-exhalation phases of the planning and mid-treatment 4DCT were used as the reference and target images, respectively. Image resolution across all images was 0.98 x 0.98 x 2.5 mm$^3$. Both left and right lungs were contoured by a graduate student under the guidance and reviewed by a board-certified radiation oncologist (ZL).

To assess volumetric alignment, target registration error (TRE) was quantified using landmarks at vasculature branching points inside liver and lung anatomy. Liver landmarks were manually identified with an average of 4 per patient (0 to 29). The number of liver landmarks was limited by the number of vessels, varying contrast stages of the analyzed CTs, and underlying diseases. On the other hand, lung landmarks were identified using a validated, in-house, automatic landmark identification algorithm with an average of 191 per patient (111 to 343)\cite{149}. The magnitude of deformation was assessed by computing the TRE immediately following rigid registration, which averaged 5.7 mm, with maximum TRE values per patient of 7.4 to 21.1 mm. Figure 15 shows renderings of landmarks within organs from sample patients in RayStation.

![Figure 15. 3D rendering of landmarks in liver (left) and lungs (right)](image-url)
4.3.2 Biomechanical model-based registration algorithm in RayStation

The BM-DIR implementation in the commercial TPS (RayStation 10B, Stockholm, Sweden) consists of two steps: finite element model (FEM) generation and finite element analysis (FEA). In the first step, organ and body drawn polygon-sliced contours in both reference and target images are converted into voxel masks of the same spacing (Figure 15 (a)). Triangular surface meshes are generated using a defined triangular length derived from the down-sampled voxel size of the voxel masks in the reference image, exerting a smoothing effect to the overall mesh surface reducing the number of sharp edges (Figure 15 (b)). The reference triangle meshes are then rigidly transformed to the target image and adapted to the corresponding voxel mask. Distances from the edge of the voxel mask are used to drive the cost function in the adaptation process.

In the second step, the reference triangular surface meshes are converted to tetrahedral volumetric meshes (i.e. FEM) using a finite element generator Netgen/NGSolve. [150] Using the FEM representation of the reference patient anatomy and the point-to-point boundary conditions, an FEA is performed to estimate the displacements of the FEM nodes inside the organs, not constrained by the boundary conditions.

RayStation allows two types of relationship between organs within the FEM. The “fixed” interface attaches the surface points of the organ to the body nodes in the FEM, allowing a continuous deformation over the organ surface where the reference mesh vertices are mapped to the corresponding target mesh vertices – a point-to-point correspondence. In the case of a liver registration, the nodes outside the liver will follow the boundary condition direction.[148] The “sliding” interface models a frictionless contact surface between the inner and external organ elements where the nodes directly outside the inner elements do not follow along. In the case of a lung registration where the sliding
interface model is used, the FEM will allow discontinuity between the nodes at the interface of the lung and chest wall.[151]

The complexity of the triangular surface mesh, as determined by user defined parameters, directly affects the generation of the FEM. If FEM generation fails, the user must recreate surface meshes using different parameters until the algorithm succeeds, hindering efficiency of the workflow. At the same time, the triangular surface mesh should retain the original organ shape. The following section therefore introduces the investigated parameters and the metrics that reflect the robustness and efficiency of the DIR process using these parameters.

4.3.3 Parameters for boundary conditions

In RayStation 10B, different combinations of two parameters: voxel side length and triangle edge length were tested to generate the surface meshes and boundary conditions required for the biomechanical-model based DIR. The voxel side length refers to the voxel size for the voxel masks. The triangle edge length is the approximate triangle edge length expressed in number of voxels in the voxel mask.

Parameters for liver registrations were first investigated. Voxel side lengths of 0.15, 0.2, and 0.25 cm were each investigated in combination with triangle edge lengths of 1, 2, and 3. For lung, the tested parameters were narrowed down to four parameters that well-performed for liver per the evaluation metrics mentioned in the following section: voxel side lengths of 0.2 cm and 0.25 cm were each investigated in combination with triangle edge lengths of 2 and 3.

All tests were performed in a Research version of RayStation (10B DTK, v10.1.110.138) running on Windows 2016 Server edition equipped with 6 Tesla T4 15Gb GPU.
4.3.4 Evaluation

The resulting BM-DIR created from different parameter combinations were first evaluated based on geometric accuracy of the registration, per recommendations of the American Association of Physicists in Medicine (AAPM) Task Group 132 (TG-132).[102] The surface alignment was evaluated for each organ contour (i.e. liver, left lung, or right lung) using Dice similarity coefficient (DSC), mean distance to agreement (MDA), and Hausdorff distance (HD) between the deformed reference and the target organ representations. Achieving the TG-132 recommendations was identified as the most important goal.

Second, robustness was evaluated based on the average success rate (i.e. ability to create a DIR: pass/fail). Robustness was identified as the next highest priority as modest increases in time were deemed worth ensuring that the algorithm would complete the DIR. Third, efficiency was evaluated based on the total time to generate the DIR (from surface mesh to DVF generation). Together, these ensure minimum user intervention in a streamlined workflow (e.g. image-guided clinical workflow). Finally, FEM complexity was evaluated for each organ contour using the number of surface mesh vertices.

4.3.5 Statistical Analysis

Performance of the different biomechanical-model based DIRs were compared per set of parameters and evaluation metrics with paired comparisons using the paired student t-test, wherein a comparison was considered significant if the p value was strictly inferior to 0.05.

4.4 Results

Table 2 and Table 3 reported the mean (SD) performance of the BM-DIR using accuracy (mean and max TRE, DSC, MDA, and HD), robustness (success rate), efficiency
(computation time), and FEM complexity for different boundary condition parameters for liver and lung images, respectively.
Table 2. BM-DIR results for liver images. Bolded row represents the optimal parameters and their metrics for each function. *TRE: target registration error; DSC: Dice similarity coefficient; DTA: distance to agreement; HD: Hausdorff distance.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.56</td>
<td>625 ± 98</td>
<td>0.35 ± 0.16</td>
<td>0.49 ± 0.32</td>
<td>0.98 ± 0.00</td>
<td>0.08 ± 0.01</td>
<td>0.93 ± 0.28</td>
<td>19529 ± 2750</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.93</td>
<td>277 ± 143</td>
<td>0.41 ± 0.28</td>
<td>0.53 ± 0.39</td>
<td>0.97 ± 0.00</td>
<td>0.08 ± 0.01</td>
<td>1.05 ± 0.35</td>
<td>8964 ± 1732</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.00</td>
<td>173 ± 78</td>
<td>0.39 ± 0.25</td>
<td>0.52 ± 0.35</td>
<td>0.97 ± 0.00</td>
<td>0.09 ± 0.01</td>
<td>1.06 ± 0.28</td>
<td>5254 ± 996</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.54</td>
<td>475 ± 136</td>
<td>0.48 ± 0.27</td>
<td>0.66 ± 0.37</td>
<td>0.97 ± 0.00</td>
<td>0.10 ± 0.01</td>
<td>0.96 ± 0.22</td>
<td>11496 ± 1741</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.89</td>
<td>187 ± 134</td>
<td>0.40 ± 0.22</td>
<td>0.55 ± 0.34</td>
<td>0.97 ± 0.01</td>
<td>0.11 ± 0.01</td>
<td>1.10 ± 0.27</td>
<td>4938 ± 969</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.00</td>
<td>100 ± 53</td>
<td>0.38 ± 0.18</td>
<td>0.55 ± 0.35</td>
<td>0.97 ± 0.01</td>
<td>0.12 ± 0.02</td>
<td>1.17 ± 0.32</td>
<td>2936 ± 574</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.64</td>
<td>341 ± 192</td>
<td>0.50 ± 0.32</td>
<td>0.63 ± 0.39</td>
<td>0.97 ± 0.00</td>
<td>0.11 ± 0.01</td>
<td>0.99 ± 0.27</td>
<td>7431 ± 1643</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.89</td>
<td>161 ± 157</td>
<td>0.46 ± 0.22</td>
<td>0.60 ± 0.31</td>
<td>0.96 ± 0.01</td>
<td>0.13 ± 0.01</td>
<td>1.12 ± 0.35</td>
<td>3098 ± 636</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.00</td>
<td>84 ± 46</td>
<td>0.42 ± 0.20</td>
<td>0.55 ± 0.29</td>
<td>0.96 ± 0.01</td>
<td>0.14 ± 0.02</td>
<td>1.18 ± 0.29</td>
<td>1850 ± 364</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. BM-DIR results for lung images. Bolded row represents the optimal parameters and their metrics. Mean and max TRE were calculated for both lungs combined. For DSC, mean DTA, HD, and number of vertices, results of each left and right lung were shown in the left and right half cells, respectively.

* TRE: target registration error; DSC: Dice similarity coefficient; DTA: distance to agreement; HD: Hausdorff distance.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Lung</td>
<td>0.2</td>
<td>2</td>
<td>0.92</td>
<td>249 ± 136</td>
<td>0.41 ± 0.09</td>
<td>1.09 ±0.28</td>
<td>0.97 ±0.01</td>
<td>0.09 ±0.01</td>
<td>0.84 ±0.38</td>
<td>5222 ± 996</td>
</tr>
<tr>
<td>Right Lung</td>
<td>0.2</td>
<td>3</td>
<td>0.96</td>
<td>143 ± 100</td>
<td>0.40 ±0.09</td>
<td>1.07 ±0.26</td>
<td>0.96 ±0.01</td>
<td>0.09 ±0.01</td>
<td>0.87 ±0.40</td>
<td>5218 ± 629</td>
</tr>
<tr>
<td>Left Lung</td>
<td>0.25</td>
<td>2</td>
<td>0.96</td>
<td>159 ± 102</td>
<td>0.41 ±0.09</td>
<td>1.08 ±0.27</td>
<td>0.95 ±0.01</td>
<td>0.11 ±0.01</td>
<td>0.91 ±0.44</td>
<td>3321 ± 616</td>
</tr>
<tr>
<td>Right Lung</td>
<td>0.25</td>
<td>3</td>
<td>0.88</td>
<td>155 ± 166</td>
<td>0.40 ±0.07</td>
<td>1.02 ±0.23</td>
<td>0.95 ±0.01</td>
<td>0.12 ±0.01</td>
<td>0.94 ±0.40</td>
<td>1930 ± 337</td>
</tr>
</tbody>
</table>

4.4.1 Accuracy

AAPM TG-132 recommends that TRE and MDA be within the maximum image voxel dimension. The diagonal dimension (longest vertex-vertex distance) of liver and lung image voxel was 2.71 (2.15-5.15) mm and 2.86 mm, respectively. The parameters for liver
and lung both returned mean TREs exceeding the respective tolerance but were comparable to the mean TRE of original publication of the DIR algorithm using inhale-exhale liver and lung CT images as the benchmark of 4.4 ± 2.0 mm. The 3 parameters with the lowest TRE were recommended for further consideration based on this test, which achieved a TRE of less than 4.0 mm. The largest MDA among liver parameters was 1.4 mm, and the largest MDA among lung parameters was 1.2 mm (left lung) and 1.1 mm (right lung).

The task group recommends a DSC of 0.80-0.90, which is the typical inter and intra-observer variability in contouring.[102] The lowest DSC was 0.96 for liver and 0.95 and 0.96 for left and right lungs, respectively. Both MDA and DSC results were based on comparing the target contour with the mapped reference contour. All liver and lung parameters satisfied MDA and DSC recommendations.

4.4.2 Robustness, Efficiency, and Complexity

Triangle edge length of 1 resulted in 54-64% success rate and was therefore not considered a clinically robust solution. 100% success was found in three parameter combinations: voxel side lengths of 0.15, 0.2, and 0.25 cm each combined with triangle edge length of 3. However, 0.15 cm side length on average required more than double the time to compute FEA for the remaining two (both $p < 0.001$), thereby not clinically advantageous. In addition, compared to voxel side length of 0.25 cm, voxel side length of 0.2 cm resulted in no significant difference for MDA, HD, or DSC, a decrease in efficiency (longer FEA computation time, $p < 0.05$), but an improvement in TRE ($p < 0.05$). As a result, the optimal parameter combination was determined to be a side length of 0.2 cm and edge length of 3 cm.

For lung, all parameter combinations experienced failures from the Netgen FEA generator, the third-party software component implemented for volumetric tetrahedral
generation. However, voxel side length of 0.25 cm + triangle edge length of 3 succeeded in less than 90% of total cases and was therefore not considered ideal for a clinically robust solution. Subsequently, DIR computation time was evaluated for the remaining parameters, and voxel side length of 0.2 cm + triangle edge length of 2 on average required more than one minute more than the remaining parameters, thereby not clinically advantageous. Comparing the remaining two parameter combinations: voxel side length of 0.2 cm + triangle edge length of 3 vs voxel side length of 0.25 cm + triangle edge length of 2, the former on average yielded insignificantly superior TRE metrics and HD but significantly superior DSC and MDA ($p < 0.01$), despite not clinically significant. However, the former yielded coarser meshes (number of vertices) indicating inferior FEM complexity ($p < 0.05$). Despite the similar quality of FEM and DIR between these two parameters, the optimal parameter was determined to be voxel side length of 0.2 cm with triangle edge length of 3 due to its slightly outperforming surface alignment. However, voxel side length of 0.25 cm with triangle edge length of 2 is recommended as a viable option.

4.5 Discussion

This study investigated different parameter combinations used to generate the surface mesh for the BM-DIR implemented in a commercial TPS, using pre- and post-ablation liver contrast-enhanced CT and planning and mid-treatment radiotherapy lung 4DCT images. The success rate, efficiency, and accuracy of the generated DIR using these parameters were evaluated, and the optimal parameters for each organ were found. The decision-making process described in this study serves as a recommendation for the RayStation biomechanical-model based DIR algorithm and a potential reference for other biomechanical DIR systems. These parameters were subsequently verified on, 16 head-and-neck cancer patients and 16 prostate cancer patients randomly selected from an
existing IRB protocol. All cases had biomechanical DIR successfully completed using the optimized parameters for the parotid glands[152] and prostate[153].

First and foremost, the DIR should strive to meet volumetric and surface accuracy tolerances recommended by AAPM TG-132.[102] Volumetrically, the optimal parameters yielded comparable TRE results compared to the original Morfeus publication[148] but did not meet the TG-132 tolerance, with the TRE exceeding the voxel dimension by 1.1 mm. However, TG-132 did note that these metrics were a target goal and may not be met with current DIR algorithms, and when not met, the uncertainty should be included in the clinical process. Research into adding additional boundary conditions at internal vasculatures to the standard Morfeus DVF demonstrated a significant reduction in TRE for both lung and liver longitudinal CTs.[101, 140] For surface alignment, DSC and MDA were met by all remaining parameters. These metrics would naturally be satisfactory as boundary conditions required matching organ mesh surfaces between the reference and target images.

Once a consistent accuracy was ensured, robustness (success rate) was the next criteria for the selection process as it ensures minimum user input needed, which is critical for automated clinical workflows. Even for parameters achieving superior accuracy, a lower success rate would indicate a higher chance for potential failure, necessitating user manual interventions to optimize parameter settings (e.g. voxel side length of 0.2 cm with triangle edge length of 3 for lung) or create the boundary conditions in the user interface on a patient-specific bases. This would be especially burdensome for heavy offline dose accumulation workflows to monitor delivered dose during online adaptive radiotherapy or image-guided ablation that relies on robust DIRs.[141, 154] In addition, parameters that are accurate and robust alone would not guarantee a clinically feasible solution unless they are reasonably efficient. Therefore, for parameters with similar accuracy and robustness, those that require substantially longer computational time were excluded to ensure robust and fast
DIRs. Even though the results from this study showed a potential positive relationship between efficiency and robustness, parameters with exceptional accuracy but a long computational time would not be recommended.

The last criteria, mesh complexity was subsequently considered because a coarse mesh would not provide enough vertices to maintain a low surface accuracy (MDA and HD) while a substantially complex mesh might be burdensome to solve the finite element problem, which negatively impacts efficiency and even robustness. This can be seen from the liver result where decreasing the triangle edge length from 3 to 1 for the same voxel side length drastically increased the number of vertices but required significantly longer processing time. Therefore, finding the balance of accuracy, robustness, efficiency, and complexity is the essence of parameter optimization. The parameter selection process can be automated by configuring a scoring system where different metrics are weighted to find the highest-scoring parameter sets. It is important to note that users can adapt the ordering of the metrics according to their own clinical needs: online adaptive radiotherapy may require robustness and efficiency weighed highest while hypo-fractionated treatments may require accuracy weighed highest.

In this study, there were two primary sources of failure in the tested biomechanical-model based DIR algorithm: mesh generation failure and DIR generation failure. Mesh generation failure can be due to low number of smoothing iterations to achieve the desired fine mesh structures. Note that these failures are not due to the magnitude of deformation between the reference and target images but the complexity of structures being modeled and the quality of the generated mesh as determined by the parameters. DIR generation failure can be due to intersecting triangles in the meshes that rendered solving the FEA impossible. As observed in the study, DIR generation failure can also come from Netgen failing to convert the reference triangular meshes to tetrahedral volumetric meshes, for which the failure is not reproduceable and the source is currently under investigation by
RaySearch. Further work is required to seek parameters that achieve potential 100% success rate. For the one case that failed for voxel side length of 0.2 cm and triangle edge length of 3, the remaining three parameter combinations completed without failing, which encouraged the selection of an alternative parameter combination (voxel side length of 0.25 cm and triangle edge length of 2) as a temporary solution when the optimal parameter fails. Rerunning with the same parameters could also resolve the failure but it is a less optimal solution.

4.6 Conclusion

In this study, several parameter settings for boundary condition generation of the biomechanical-model based DIR algorithm in a commercial TPS have been evaluated based on accuracy, robustness, and efficiency. The optimal parameters were found for liver CT and lung CT DIR to be voxel side length of 0.2 cm and triangle edge length of 3. Improved robustness and efficiency may encourage more widespread usage of biomechanical model-based DIR algorithms in clinical practice, such as adaptive radiotherapy workflows and retrospective dose accumulation studies. The consistent use of the same reported optimal parameter by users across institutions can also facilitate standardized evaluation of published studies.
Chapter 5: Evaluate Dosimetric and Accuracy Differences between Accumulated and Planned Doses

In Revision: International Journal of Radiation Oncology - Biology - Physics

5.1 Sample Size Justification

Dose accumulation is a labor-intensive process. To justify the sample size needed for the aim, a preliminary study with six IMRT patients was conducted to compare phase-based accumulated dose and the AVG-based planned dose. These accumulated doses were propagated from the exhale phase to the AVG of the planning 4DCT to be compared against the planned dose on the same image. Specifically, we compared the region that received more than 80 Gy in the normal, ipsilateral lung. On average, these regions had volumes of 48 cm³ and 22 cm³ for the accumulated dose and the planned dose, respectively. Volumetric overlaps between the regions from the two distributions had an average score of 0.37 ± 0.20. The null hypothesis was that the two dose distributions are the same, for which we set the criteria of dice score to be at least 0.9. Therefore, for a 2-tailed test and 0.05 alpha, 26 cases were needed to have a power of 80% in demonstrating a significant difference in dose distributions between the phase-based weekly dose accumulation and average-image based planned dose.

We developed an automated approach utilizing the scripting capabilities of RayStation treatment planning system (RaySearch Laboratories, Stockholm, Sweden). The end of the thesis will provide some general tips on scripting in RayStation.

5.2 Abstract

Purpose: To investigate the dosimetric and clinical impacts of 4-dimensional computed tomography (4DCT)-based longitudinal dose accumulation in locally advanced non-small-
cell lung cancer patients treated with standard-fractionated intensity-modulated radiotherapy (IMRT).

**Materials and Methods:** Sixty-seven patients were retrospectively selected from a randomized clinical trial. Their original IMRT plan, planning and verification 4DCTs, and ~4-month post-treatment PET/CTs were imported into a commercial treatment planning system. Two deformable image registration (DIR) algorithms were implemented for dose accumulation, and their accuracies were assessed. The planned and accumulated doses computed using average-intensity images or phase images were compared. At the organ level, mean lung dose (MLD) and normal-tissue complication probability (NTCP) for grade ≥ 2 radiation pneumonitis were compared. At the region level, mean dose to lung subsections and the volumetric overlap between isodose intervals were compared. At the voxel level, the accuracy in estimating the delivered dose was compared by evaluating the fit of a dose vs. radiographic image density change (IDC) model. The dose-IDC model fit was also compared for sub-cohorts based on the magnitude of NTCP difference (|ΔNTCP|) between planned and accumulated doses.

**Results:** DIR accuracy was quantified, and the uncertainty was considered for the voxel-based analysis. Compared to planned doses, accumulated doses on average resulted in < 1-Gy lung dose increase and < 2% NTCP increase (up to 8.2 Gy and 18.8% for a patient, respectively). Volumetric overlap of isodose intervals between the planned and accumulated dose distributions ranged 0.01–0.93. Voxel-level dose-IDC models demonstrated a fit improvement from planned dose to accumulated dose (pseudo-R² increased 0.0023) and a further improvement for patients with ≥ 2% |ΔNTCP| vs. for patients with < 2% |ΔNTCP|.

**Conclusions:** With a relatively large cohort, robust image registrations, multi-level metric comparisons, and radiographic image-based evidence, we demonstrated that dose accumulation more accurately represents the delivered dose and can be especially beneficial for patients with greater longitudinal response.
5.3 Introduction

The standard of care for locally advanced non–small-cell lung cancer (LA-NSCLC) patients includes radiotherapy. Despite the treatment’s curative intent, these patients suffer from low 5-year survival rate.[7] Current clinical practice in external-beam radiotherapy calculates the delivered dose at the planning stage by assuming stationary anatomy throughout the course of treatment. However, patient anatomy often changes in volume and shape, which affects the treatment response of the target and the organs at risk.[155] Having an accurate representation of the delivered dose may help clinicians to predict toxicity and develop a better understanding of dose-dependent predictors of response.

Intra-fractional and inter-fractional anatomical changes have been demonstrated to alter dose deposition in the tumor and normal tissue.[155] For the intra-fractional anatomical change, four-dimensional computed tomography (4DCT) is routinely used in treatment planning to quantify motion for use in defining the planning target volume,[29] and studies have found that by incorporating breathing motion, the phase-based 4D planned dose is more accurate than the average-intensity image (AVG)–based 3D planned dose.[77, 78, 156–159] The effect of inter-fraction anatomical changes can be realized with longitudinal dose accumulation based on repeat volumetric imaging over the course of treatment. Previous studies have accumulated dose using cone-beam computed tomography (CBCT) in small patient cohorts (< 30 patients).[79–81] These studies focused on DVH-based organ-level metrics that lack spatial features; however, increasing evidence has suggested clinical benefits of region/voxel-level metrics.[160, 161] Therefore, true validation of the dosimetric accuracy improvement of deformable image registration (DIR)-based dose accumulation compared to the planned dose in delivered dose representation has been limited. To date, no study has evaluated improvements in dosimetric accuracy using models of dose vs. radiographic evidence of response to prove the superior accuracy of the accumulated dose over the planned dose.
In this study, we hypothesized that the accumulated dose is superior in accuracy than the planned dose in representing the delivered dose, with an improved correlation to radiographic treatment response. To test this, we automated a 4DCT-based deformable dose accumulation workflow for standard-fractionated IMRT and retrospectively investigated the dosimetric differences between the planned dose and the accumulated dose on the organ, region, and voxel levels. We assessed the voxel-level dose-radiographic treatment response relationships to determine the clinical implications of dose accumulation. We employed state-of-the-art DIR techniques to ensure accurate dose and image mappings and addressed the uncertainties involved in dose accumulation and image-based response assessment.

5.4 Methods and Materials

5.4.1 Patient Data

Data on 67 LA-NSCLC patients were curated under an IRB-approved retrospective protocol and imported into a commercial treatment planning system (TPS) (RaySearch Laboratories, Stockholm, Sweden) for processing. Patients were originally treated under free-breathing conditions with standard-fractionated IMRT in a prospective clinical trial.[85] Table 4 lists the tumor locations and dose prescriptions of the cohort. Patients had a single treatment plan and an average of 6 (range, 4–8) weekly 4DCTs to quantify response over the course of radiotherapy. The net tumor volume change was 23.0 ± 70.0 cm³ over the course of treatment.

Table 4. Tumor location and dose prescription for the selected patients.

<table>
<thead>
<tr>
<th>Tumor location</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper lobes (L/R)</td>
<td>45</td>
</tr>
<tr>
<td>Right middle lobe</td>
<td>6</td>
</tr>
<tr>
<td>Mediastinum (nodal)</td>
<td>5</td>
</tr>
<tr>
<td>Lower lobes (L/R)</td>
<td>11</td>
</tr>
</tbody>
</table>
For a subset of 60 patients, image response was assessed using a non-contrasted free-breathing PET/CT acquired 4.0 (1.1–8.2) months after treatment to measure acute inflammatory response. The voxel sizes for 4DCT and PET/CT were 0.98 x 0.98 x 2.5 mm³ and 0.98 x 0.98 x 3.3 mm³, respectively. All images were free from lobar atelectasis or pleural effusion with > 2-cm axial thickness to minimize lung volume imbalance for image registration accuracy. Both lungs were segmented using a commercially available 3D U-Net–based auto-segmentation algorithm as provided by the TPS. The algorithm was trained with in-house images that were manually edited by an experienced graduate student (YH) and checked by a board-certified radiologist (CW).

5.4.2 Dose Accumulation Workflow

A clinical dose grid of 3 x 3 x 3 mm³ was set for all dose calculations. Dose was accumulated in a 3-step process: recalculation, deformation, and summation. For clarity, the term “summation” corresponds to adding doses that share the same image frame, whereas the term “accumulation” is reserved for longitudinal dose mapping. We investigated 4 types of doses: the AVG–based 3D planned dose (3DP) as the standard of care, the phase-based 4D planned dose (4DP), the AVG–based 3D accumulated dose (3DA), and the phase-based 4D accumulated dose (4DA) as the most extensive dose accumulation. Their schema are shown in Figure 16.

Planned doses (3DP and 4DP) were recalculated based on the planning 4DCT. 3DP was obtained by recalculating the dose by the treatment plan on the AVG of the planning 4DCT, and 4DP was obtained by calculating the dose on each phase of the planning 4DCT.
(with equally distributed weights), mapping non–end-exhale phase (T5) dose distributions to T5 through intra-fractional DIRs (green arrows, Figure 16), and summing the deformed doses with the T5 dose.

To accumulate dose (3DA or 4DA), weekly doses (3D or 4D) were weighted by the number of fractions assigned to each 4DCT (the nearest fractions according to the treatment record) with respect to the total treated fractions. For 3DA, weekly 3D doses were deformed to the planning AVG and summed. For 4DA, weekly 4D doses (obtained in the same way as with 4DP) were deformed to the planning T5 and summed. We selected T5 as the reference image for dose propagation as recent studies showed that T5-T5 DIR maintains the most robust registration accuracy for inter-fractional registrations.[146]

Figure 16. Schema for obtaining the 4D planned dose (4DP) (green), the 3D accumulated dose (3DA) (orange), and the 4D accumulated dose (4DA) (red). Images shown are the AVG and the 0%/50%/90% breathing phases of the planning 4D computed tomography (4DCT) and the weekly verification 4DCTs. Arrows indicate the direction of dose mapping.
5.4.3 Image Registration Process and Accuracy

Prior to DIR, rigid alignments were performed in the TPS. The AVGs of the planning and weekly 4DCTs were registered with translation only and a focus on bony anatomy to mimic patient setup during original treatment. Subsequently, breathing phases of weekly 4DCTs were automatically registered to those of the planning 4DCT since images of each 4DCT shared the same image frame—this allowed dose recalculation for the phase images of weekly 4DCTs. For response assessment, the planning T5 was registered with the PET/CT.

Two deformable image registration (DIR) algorithms were implemented. An image intensity-based algorithm focusing on the lung region (ANACONDA) (commercially available in the TPS) was used for intra-fractional mappings.[118] A biomechanical model-based algorithm with sliding lung motions and internal vasculature alignment (vbcMORFEUS) was used for longitudinal mapping and response assessment by modifying a commercially available algorithm in the TPS.[101] The process for vbcMORFEUS was the following: 1) triangle mesh was generated with previously optimized parameters,[162] 2) boundary conditions for a sliding lung interface were established, and 3) the deformation vector field was exported from the TPS and modified by adding constraints on vessel alignment.

We validated geometric registration accuracy within the ipsilateral normal lung (excluding the gross tumor volume (GTV)). We implemented a previously validated automatic landmark identification algorithm to quantify the target registration error (TRE) at vasculature branching points,[149] to evaluate the accuracy for intra-fractional DIR (represented by T0-T5 DIR of the planning 4DCT), inter-fractional T5-T5 DIR and AVG-AVG DIR between the planning and the third weekly 4DCTs (both using landmark pairs identified on the T5s), and the T5-PET/CT DIR for response assessment. The American Association
of Physicists in Medicine Task Group 132 (AAPM TG 132) Report was employed to determine whether voxel-level accuracy was achieved for these DIRs.[102]

5.4.4 Organ-Level Dose and Response Metrics

The mean dose to the ipsilateral normal lung (MLD) was recomputed as the organ-level dose metric. Probability of grade ≥ 2 radiation pneumonitis was evaluated as the clinical metric using the Lyman–Kutcher–Burman (LKB) normal-tissue complication probability (NTCP) model (with dose distributions of the total normal lung). The model is expressed in equations (1) and (2) where[65, 66, 163]

\[
NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{D_{eff} - TD_{50}}{mTD_{50}}} e^{-\frac{x^2}{2}} dx
\]  

(1)

\[
D_{eff} = \left( \sum_i v_i D_{i} \right)^{\frac{1}{n}}
\]

(2)

parameters \( TD_{50} \), \( m \), and \( n \) were 34.80 Gy, 0.22, and 0.5, respectively, per Tucker et al. that shared the same patients with our study.[67] These patients were treated with standard fractionation, so no fraction-correction was implemented for \( D_i \). One-way ANOVA with Dunnett’s multiple comparisons test (with \( P = 0.05 \) as the threshold for statistical significance) was used to compare MLD and NTCP for each of 4DP, 3DA, and 4DA against 3DP. Histograms were created for 4DP, 3DA, and 4DA to demonstrate the distribution of the magnitude of MLD deviation (\(|\Delta \text{MLD}|\)) and NTCP deviation (\(|\Delta \text{NTCP}|\)) against 3DP. The number of patients with \(|\Delta \text{NTCP}| \geq 5\% \) or 2\% was quantified to identify patients with potentially clinically meaningful changes over the course of treatment. These thresholds were proposed by Langendijk et al. to guide the decision making of treatment modality in clinical trials.[164]
5.4.5 Region-Level Dose Metrics

The ipsilateral normal lung in the planning T5 was tri-partitioned into equal-volume subsections along the superior-inferior direction. Using the paired Student t-test, the mean dose to each subsection was compared between 3DP (evaluated using T5 contours) and 4DA. Further, the percentage volumetric overlap between isodose intervals were evaluated for 3DP and 4DA. These isodose intervals were defined by voxels of 5-Gy increments that spanned 0–80 Gy.

5.4.6 Voxel-Level Dose and Response Metrics

Dose grids of 3DP and 4DA were resampled to match the resolution of the planning T5. For each patient, we subtracted the dose maps between 3DP and 4DA for voxels in the ipsilateral normal lung to obtain the level of dose difference. Subsequently, we created a cumulative histogram for the magnitude of dose difference for the entire cohort.

To link the dosimetric difference to clinical impact, we calculated CT image density change (IDC) between the planning T5 and the PET/CT (after the PET/CT was deformed via vbcMORFEUS to the planning T5). Since 3DP was calculated on the planning AVG, we overlaid 3DP to the inherently registered planning T5. We considered voxels of the planning T5 that were at least 2 mm away from the lung and GTV boundaries to avoid partial-volume averaging and boundary effects. Similar to the isodose intervals in the regional analysis, these voxels were grouped into 5-Gy dose bins. Dose bins with fewer than 1 million voxels were excluded to ensure robust statistics—this defined the highest dose bin to be plotted. The average IDC and dose for voxels of each dose bin were obtained. Subsequently, voxel-level dose-IDC relationships were established by fitting a modified LKB NTCP model, where NTCP was expressed as the percent IDC against the maximum IDC, and $D_{eff}$ was represented by dose to each voxel. Non-linear least squares fitting from the Python SciPy
package (version 1.6.2) was used to obtain \( m \) and \( TD_{50} \). Pseudo-R\(^2\) in equation (3) was used to reflect the non-linear fit, where the residual IDC represented the difference between the modeled IDC and the measured IDC.

\[
R^2 = 1 - \frac{\text{variance(residual IDC)}}{\text{variance(measured IDC)}}
\]  

(3)

The voxel-level dose-IDC modeling was repeated for the sub-cohort of patients with \( \geq 2\% \) \( |\Delta\text{NTCP}| \) between 3DP and 4DA and again for patients with \( < 2\% \) \( |\Delta\text{NTCP}| \).

To ensure voxel-based analysis, if the accuracy of T5-PET/CT DIR did not meet the recommendations from AAPM TG 132,[102] the images and doses involved (i.e., planning T5, deformed PET/CT, 3DP, and 4DA) were simultaneously resampled by binning voxels uniformly along all three dimensions (e.g., 2 x 2 x 2 voxels) until the diagonal length of the “super” voxel was larger than the average TRE of T5-PET/CT DIR. Subsequently, the voxel-level dose-IDC modeling was repeated using the “super” voxels, and the model fit was compared with that of the original voxel dimension.

### 5.5 Results

#### 5.5.1 Image Registration Accuracy

The evaluated DIRs were the intra-fractional T0-T5 DIR (ANACONDA), the inter-fractional T5-T5 DIR and AVG-AVG DIR (vbcMORFEUS), and the longitudinal T5-PET/CT DIR (vbcMORFEUS). For T5-PET/CT DIR, the algorithm was modified to include bifurcations on thinner vessels to compensate for the complex response in the follow-up PET/CTs, and the TRE evaluation was limited to cases with > 20 identified landmarks. The average TRE and number of identified landmarks were 2.3 ± 1.8 mm and 62 ± 27 (T0-T5 DIR), 2.8 ± 1.9 mm and 60 ± 26 (T5-T5 DIR), 3.2 ± 1.8 mm and 60 ± 26 (AVG-AVG DIR), and 5.6 ± 2.2 mm and 61 ± 43 (T5-PET/CT DIR). The AAPM TG 132 Report recommended the registration uncertainty to be within the longest vertex-vertex distance of the
images,[102] which was 2.9 mm as the diagonal dimension of the analyzed 4DCTs. Both the
T0-T5 DIR and the T5-T5 DIR satisfied this recommendation, indicating that the phase-
based 4DP and 4DA achieved voxel-level accuracy. TRE for AVG-AVG DIR indicated
slightly lower dose mapping accuracy compared to T5-T5 DIR. To meet the
recommendation for T5-PET/CT DIR, we resampled images and doses involved in the
voxel-level dose response analysis by binning every two voxels along all three dimensions
into “super” voxels with diagonal dimension of 5.7 mm (1.95 x 1.95 x 5.0 mm³).

5.5.2 Organ-Level Dose and Response Metrics

Figure 17 shows the effect of intra- and inter-fractional anatomical changes at the
organ level. Figure 17(a) and Figure 17(b) demonstrate the MLD and NTCP, respectively for
all four doses: 3DP, representing standard of care, had MLD of 30.6 ± 7.7 Gy and NTCP of
28.5% ± 16.2%; 4DP, accounting for intra-fractional anatomical change, had MLD of 30.3 ±
7.7 Gy and NTCP of 27.4% ± 15.7%; 3DA, accounting for inter-fractional anatomical
change, had MLD of 31.4 ± 8.0 Gy and NTCP of 30.4% ± 17.6%; 4DA, accounting for both
intra- and inter-fractional anatomical changes, had MLD of 31.3 ± 8.0 Gy and NTCP of
29.7% ± 17.3%. Figure 17(c) and Figure 17(d) show the frequency distribution for |ΔMLD|
and |ΔNTCP|, respectively, for each of 4DP, 3DA, and 4DA compared to 3DP. The largest
MLD increase from 3DP to 4DA was 8.2 Gy, and the NTCP increased 18.8% for this patient,
and more patients with larger deviations were seen for 3DA and 4DA than 4DP. The number
of patients with ≥ 5% or 2% |ΔNTCP| between 3DP and 4DA was 11 or 30 (16% or 45% of
the cohort).
Figure 17. The effect of intra- and inter-fractional anatomical changes to the mean lung dose (MLD) and the normal tissue complication probability (NTCP) at the organ level. Whiskers span the 10th-90th percentile. Top sub-plots: boxplots for average MLD and NTCP for the 3D planned dose (3DP), the 4D planned dose (4DP), the 3D accumulated dose (3DA), and the 4D accumulated dose (4DA) for the studied cohort. The NTCP comparison between 3DP and 4DA resulted in $P = 0.08$. Middle and bottom sub-plots: histograms for $|\Delta\text{MLD}|$ and $|\Delta\text{NTCP}|$ for each of 4DP, 3DA, and 4DA against 3DP.
5.5.3 Region-Level Dose Metrics

Compared to 3DP, 4DA resulted in higher mean dose of the superior and middle subsections (47.4 Gy vs. 48.0 Gy and 34.8 Gy vs. 35.4 Gy, respectively, both $P < 0.01$) and lower mean dose of the inferior subsection (13.8 Gy vs. 13.6 Gy, $P = 0.73$). Boxplots in Figure 18 illustrate that the volumetric overlap of 5-Gy isodose intervals between 3DP and 4DA ranged 0.01–0.93.

![Volumetric Overlays of Isodose Intervals between 3DP and 4DA](image)

Figure 18. Boxplots of volumetric overlap between the 3D planned dose (3DP) and the 4D accumulated dose (4DA) for isodose intervals of 5-Gy increments. The triangles next to each boxplot represent the average volume of each isodose interval under both 3DP and 4DA. Due to differences in dose prescriptions, the number of patients that contributed to the last 2 boxplots were labeled above each boxplot.

5.5.4 Voxel-Level Dose and Response Metrics

For individual patients, more than one-fifth of the cohort (14/67) had ≥ 20% voxels experiencing absolute dose difference of ≥ 5 Gy between dose distributions of 3DP and 4DA. Figure 19 shows the ‘dose-difference’ volume histogram for the comparison between 3DP and 4DA for the entire cohort. Approximately 35% and 15% voxels received absolute dose difference of ≥ 2 Gy and 5 Gy, respectively.
Figure 19. Cumulative histogram for voxels of the entire cohort showing the distribution of absolute dose difference between the 3D planned dose (3DP) and the 4D accumulated dose (4DA). Dose difference of > 20 Gy was excluded in the plot.

The voxel-level dose-IDC model fit for the original voxel resolution is shown in Figure 20(a). The average number of voxels in each dose bin was 1.8e+6 for both doses (3DP: 1.0e+6–7.5e+6, 4DA: 1.1e+6–7.6e+6). Compared to 3DP, 4DA resulted in a higher $T_{D50}$ (dose causing 50% of maximum IDC), a higher mean IDC for each dose bin ($P < 0.01$ except for 10-15 Gy with $P = 0.08$), and a modest, but consistent, improvement in model fit (pseudo-$R^2$ increase of 0.0023). Figure 20(c) and Figure 20(d) show that the model fit for patients with $\geq 2\% \Delta NTCP$ had larger improvement in pseudo-$R^2$ (0.0118) than for patients with $< 2\% \Delta NTCP$ (0.0007)—i.e., those with larger longitudinal anatomical changes vs. those with less changes. Figure 20(b) shows that 2x2x2-binned “super” voxels maintained comparable fitted parameters and pseudo-$R^2$ compared to the original voxel dimension in Figure 20(a), demonstrating the results from the original voxel resolution was not degraded by accounting for the larger TRE.
5.5 Discussion

For LA-NSCLC patients treated with standard-fractionated IMRT, we quantified the dosimetric differences between the planned and accumulated dose distributions by independently evaluating the effect of intra-fractional and inter-fractional anatomical changes. We determined the clinical impact of such dose differences by correlating the analyzed doses with radiographic image-based outcomes.

Figure 20. Voxel-level dose-image density change (IDC) modeling results: the mean image density change (IDC) (left y-axis), mean dose (x-axis), and the normalized IDC against the maximum IDC (right y-axis) for voxels in each dose bin. The fitted plots are represented by dashed lines. TD50 represents dose that causes 50% of maximum IDC. (a) & (b): Fitted models using original sized voxels and “super” voxels of the entire cohort, respectively. (c) & (d): Fitted models for patients with ≥ and < 2% |ΔNTCP| (magnitude of change in the normal tissue complication probability) between the 3D planned dose (3DP) and the 4D accumulated dose (4DA), respectively.
We employed two DIR algorithms: ANACONDA for intra-fractional dose mapping and vbcMORFEUS for inter-fractional dose mapping and longitudinal image mapping for response analysis. Intra-fractional image mapping does not involve tissue response, which warranted ANACONDA as an accurate, robust, and fast method. For inter-fractional image registrations, lung tumors can respond 1) inelastically, where the tumor boundary erodes away while normal lung tissue stays intact, 2) elastically, where normal lung tissue follows the tumor boundary, indicating no microscopic disease, or 3) both inelastically and elastically.[93, 165] Intensity-driven DIR algorithms may force align the tumor boundary between the registered images to maximize the image similarity metric. In contrast, biomechanical model-based DIR algorithms are independent of image intensity, and the internal boundary conditions drive the alignment of structural similarities that are maintained in each image. It was observed in our sample that most tumors responded (at least partially) inelastically. Therefore, vbcMORFEUS would provide a more realistic registration. In addition, for elastically responding tumors, not force-aligning the tumor boundary was a conservative approach to accumulate lung dose—target dose to lung voxels adjacent to shrunken tumor was mapped to tumor voxels in the planning image, so the high dose was not accounted for the accumulated lung dose, since only the normal lung was considered for dose metrics. Accuracy of these DIRs were validated by automated TREs and compared against the recommendation of AAPM TG 132. To maintain voxel-level accuracy for dose-response analysis, we binned the voxels of the images and doses involved into “super” voxels to compensate for the high TRE for T5-PET/CT DIR.

We leveraged the commercially available scripting capability of the TPS to perform auto-segmentation, image registration, dose mapping, and dose summation to achieve, to our knowledge, the most comprehensive dose evaluation for the largest patient sample to date. Among these doses, 3DP represents the planned dose calculated on the AVG and is widely used as the clinical standard for treatment planning. In contrast, 4DP explicitly uses
the anatomy of a full breathing cycle, and the difference between 3DP and 4DP has been well-documented in the literature.[77, 78] Compared to 3DP, 3DA resulted in a higher MLD, indicating the effect of inter-fractional anatomical change: e.g., fewer photons were attenuated by smaller tumor volume and, in turn, deposited more energy in the normal lung. The higher MLD of 4DA compared to 3DP reflected the net result of the interplay of intra- and inter-fractional anatomical changes. Neither the MLD nor NTCP difference was clinically significant in the population-based analysis. However, a substantial portion (16–45%, depending on criteria used) of the cohort individually resulted in |ΔNTCP| that could trigger a change in the treatment planning re-evaluation per Langendijk et al., suggesting the benefit of dose accumulation on a patient-by-patients basis for assessment of complication risk. Therefore, automated dose accumulation workflows can help identify dose deviation that triggers re-planning.

Among the geometric subsections of the ipsilateral normal lung, the highest mean dose in the superior subsection reflected that the majority of the cohort had superiorly located tumors. Both superior and middle subsections experienced increased mean dose while their inferior counterpart experienced decreased mean dose—the interplay of intra- and inter-fractional anatomical changes was especially pronounced in the inferior subsection. For the volumetric comparison of isodose intervals, the < 50% average overlap for the mid-level dose ranges shown in Figure 18 indicated low agreement between the planned and accumulated doses. In contrast, isodose intervals for low and high dose ranges resulted in relatively better agreement, potentially due to increase in the compared volumes. Therefore, evaluating the dose metrics at a regional level may improve our understanding of how longitudinal anatomical changes affect different lung sub-regions. Future studies should consider applying automatic lobe segmentation to investigate the dosimetric and functional impact of anatomical changes to different lobes, especially the inferior lobes with increased sensitivity—a multi-institutional clinical trial.[70, 166–168]
We utilized IDC from planning to post-treatment to quantify radiographic response rather than subjective, qualitative symptomatic scorings.[169] Similar methods have been documented in the literature.[83, 84, 89, 170, 171] However, we addressed the DIR accuracy for voxel-to-voxel correspondence to obtain IDC and used non-contrasted PET/CTs to avoid artificial IDC increase in voxels near vessels. Studies have demonstrated a time-dependency of IDC where IDC peaked at around 6 months for acute response (e.g., pneumonitis) and then decreased to a plateau at around 12 months as a permanent response (e.g., fibrosis).[83, 170, 171] The PET/CTs from our study were, on average, obtained ~4 months after treatment, representing mostly acute inflammatory response. Future studies can include additional longitudinal images to examine the time effect, but the number of available images could be limited due to loss of patient contact for follow-ups.

We used a modified NTCP model (i.e., a sigmoid curve) instead of a linear model, as IDC theoretically cannot increase indefinitely. However, the model had to be modified so that each voxel represented itself as a sample, independent of the location and patient. As a result, \( D_{\text{eff}} \) became the dose to voxel itself. With the relative IDC against the maximum IDC as the endpoint, the model calculated the likelihood that a voxel would achieve the maximum radiographic response. The modest, but consistent improvement in model fit from the accumulated dose, compared to the planned dose, is evidence, especially for the sub-cohort with larger anatomical changes, that the accumulated dose more accurately represents radiation damage than the planned dose. Given this improvement, the ability to robustly automate the dose accumulation process, and the substantial differences observed in individual cases with larger response, we recommend using dose accumulation to represent the delivered dose when performing radiation-induced toxicity analysis and to determine the composite dose for re-irradiation or adaptive re-planning. However, the increased voxel-mapping sensitivity in regions of high dose and large dose gradient calls for
caution for lung tissue near tumor. Nevertheless, the observed dosimetric benefits of dose accumulation should be further evaluated in a multi-institutional prospective trial.

Our findings aligned well with those of previous studies. Wang et al. accumulated dose for a hypo-fractionated regimen for stage III patients (N = 27) using 5 CBCTs for a total of 17 fractions per patient, and observed a limited increase in MLD (median increase < 5%).[81] Luo et al. grouped 3-dimensional conformal radiotherapy and IMRT (N = 24, each cohort size unknown) and found no significant differences in organ-level dose metrics in lung, unless patients had a large target response.[80] Ren et al. combined doses from the original plan and the new plan after 20th fraction (N = 30) and found a positive relationship between the difference in the planned–accumulated doses and the lung volume change.[79] Alam et al. focused on the esophagus dose for LA-NSCLC (N = 11) and found that the median accumulated mean esophagus dose was higher than the planned dose (24 Gy vs 21 Gy, P < 0.01).[172]

Our study benefited from a large patient cohort, the component-wise investigation of the impact of breathing motion and longitudinal response, region/voxel-level dose/response analysis, and the correlation of dose to the clinical, image-based response. However, one limitation is the lack of daily imaging, such as CBCT, to more frequently represent the inter-fractional anatomical changes. However, we believe that the number of weekly 4DCTs was sufficient to reflect anatomical changes throughout treatment and that 4DCTs had the advantage of realizing the intra-fractional anatomical change and full field-of-view compared to CBCT.

The studied patients were enrolled in a clinical trial that compared photon vs proton therapies, and proton dose accumulation is the focus of ongoing research. We hypothesize a more significant dose difference due to protons’ increased sensitivity to anatomical changes.
5.6 Conclusions

We performed 4DCT-based longitudinal deformable dose accumulation for LA-NSCLC patients treated with standard-fractionated IMRT and compared the accumulated dose with the planned dose. Population-based comparison returned statistically significant dose deviations yet modest differences in the predicted pneumonitis rate. Individually however, a substantial portion of the cohort (16–45%, depending on criteria used) resulted in deviations in predicted pneumonitis rate that could trigger a change in the treatment planning re-evaluation. Additionally, clinical image-based evidence showed that accumulated dose improved the correlation to localized tissue damage than planned dose, especially for patients with larger deviations in predicted pneumonitis rate. Therefore, our findings demonstrated that dose accumulation more accurately estimates the dose delivered from the treatment course and can be particularly beneficial when patients experience large longitudinal anatomical changes.
6.1 Abstract

**Purpose:** To derive the proton variable relative biological effectiveness (RBE) from radiographic changes and validate it with two published empirical models for locally advanced non–small-cell lung cancer patients.

**Materials and Methods:** We retrospectively analyzed patients previously treated on a prospective randomized trial with standard-fractioned intensity-modulated photon therapy (n = 51) or passive scattering proton therapy (n=67). We recomputed, for photon patients, the planned dose (Dₓ) and, for proton patients, the physical planned dose (Dₚ) and the dose-averaged linear energy transfer (LETₐ). For each patient, we calculated the image density change (IDC) in the normal ipsilateral lung from the planning CT to a follow-up CT. Using IDC as the clinical evidence of response, we correlated voxel-level dose-IDC relationships by fitting a modified Lyman–Kutcher–Burman normal tissue complication probability (NTCP) model for both patient groups together. With the fitted NTCP models, we calculated Dₓ and Dₚ at each defined IDC level and the clinical RBE value as Dₓ/Dₚ. We validated the observation of variable RBE values by fitting two empirical models: McNamara *et al* and Wedenberg *et al*.

**Results:** The measured radiographic-based RBE values ranged from 3.0 to 1.2 for voxels that received proton physical doses ranging from 9.7 Gy to 55.6 Gy (corresponding to IDC levels of 20 HU to 115 HU, in increments of 5 HU). The McNamara and Wedenberg models returned the fitted parameter(s) of p₀ = 3.9, p₁ = 15.8, p₂ = 2.3, and p₃ = -0.3 and q = 13.0 with a pseudo-R² of 0.99 and 0.92, respectively.

**Conclusion:** Variable RBE values were modeled using voxel-based radiographic changes as the clinical endpoint and validated using established models. This is the first study to
demonstrate the potential clinical radiographic-based endpoint for RBE modelling to better understand proton biological dose.

6.2 Introduction

The standard of care for patients with locally advanced non–small-cell lung cancer (LA-NSCLC) includes radiation therapy (RT). However, RT patients experience a low 5-year survival rate [7]. Dose-limiting toxicity such as radiation pneumonitis (RP) may contribute to this low survival rate. Given concerns about toxicity, clinicians may prescribe sub-optimal dose coverage, enabling tumor cells to avoid lethal damage. The RTOG 0617 trial concluded that photon dose escalation increases tumor control but also increases toxicity [46]. Photon therapy is the conventional radiation modality, and its tumor conformality has improved as 3-dimensional conformal RT has evolved into intensity-modulated RT (IMRT). However, IMRT still delivers a considerable amount of radiation to surrounding healthy tissues (i.e., exit dose).

Proton therapy has emerged as an alternative radiation modality because it theoretically offers better conformal dose delivery than photon therapy, and thus a lower risk of toxicity [49]. Proton therapy’s superior dose conformality comes from protons stopping after delivering most of their energy at the tumor, through the Bragg peak phenomenon [50]. With this unique characteristic, proton therapy has the advantage of maintaining tumor dose while minimizing damage to the surrounding normal tissue [48]. However, studies have shown that the efficacy achieved with proton therapy is not significantly different than that achieved with photon therapy. A randomized phase II clinical trial aimed to compare the efficacy of PSPT with that of IMRT for LA-NSCLC [85]. The trial showed that although PSPT provided better conformity, with a reduced volume of lung receiving at least 5 or 10 Gy, it did not yield significantly lower rates of local failure (~11% for both IMRT and PSPT) or grade 3 RP (6.5% for IMRT vs. 10.5% for PSPT). Therefore, the theoretical tissue-sparing
advantage of proton therapy has not translated into improved clinical outcomes, and we have a limited understanding of the underlying biological damage done by protons in the lung.

Relative biological effectiveness (RBE) describes the difference in biological damage between protons and photons. It is defined as the ratio between the proton dose and photon dose contributing to the same biological endpoint. Based on an average of evidence, the current consensus for the clinical implementation of proton therapy is to use a constant RBE of 1.1 [173], so the physical proton dose is obtained by dividing the biological dose by 1.1. However, we questioned the validity of the clinically implemented fixed RBE of 1.1 because the clinical trial results did not confirm the anticipated superior normal tissue sparing promised by proton therapy [85]. Even though the fixed RBE–weighted lung dose was preserved by PSPT, the higher biological effectiveness of PSPT, which was not accounted for in treatment planning, could have contributed to the increased toxicity.

Increasing experimental and simulation evidence suggests RBE varies depending on dose [174–177]. However, these studies do not reflect the response of normal tissue inside patients. Therefore, measuring the RBE values directly in the patient could help us understand the clinical biological effects of protons. Therefore, we used the image density change (IDC) from the planning computed tomography (CT) to post-treatment CT as a directly quantifiable clinical endpoint for RBE. Although the IDC does not always indicate RP, it is still considered a grade 1 response according to the Common Terminology Criteria for Adverse Events [57]. In addition, an increasing number of studies has demonstrated the feasibility of localized radiography-based dose-response models [83, 84, 90, 178]. The common relationship is a positive sigmoidal relationship between the dose and radiographic change. Therefore, in the present study, we considered the clinical RBE to be the ratio between the photon and proton doses causing the same level of IDC. As we obtained changing RBE values with dose, we validated the observation by fitting an established RBE
model. Many phenomenological models have been developed using empirical data from \textit{in vitro} cell studies [176, 177]. In general, these models are based on a linear-quadratic model of cell killing, and common parameters include photon dose, proton physical dose, linear energy transfer (LET), alpha/beta ratio ($\alpha/\beta$), and their unique parameters to be fitted. Such models assume that LET, a macroscopic dosimetric parameter, is a main contributor to proton RBE [179]. As a proton slows towards the end of its track, LET is the main contributor to the increase in the proton RBE [180]. Due to the stopping of proton, LET dramatically increases, which increases RBE [176] whereas LET stays constant for photons.

The goals of the present study were to (1) leverage clinical response evidence of the variable relationship of RBE with dose using IDC as radiographic response and (2) validate this evidence using two established empirical RBE models. This work provides an innovative method in leveraging IDC as the clinically viable metric to better understand the proton biological damage in patients.

\section*{6.3 Methods}

\subsection*{6.3.1 Patient data}

Under an Institutional Review Board–approved protocol, we retrospectively acquired clinical data for patients with LA-NSCLC who previously received IMRT ($n = 51$) or passive scattering proton therapy (PSPT; $n=67$) to 60–74 Gy (RBE = 1.1 for PSPT) in a prospective randomized trial [85]. We recomputed the planned dose for IMRT patients ($D_p$) in RayStation 11B DTK (RaySearch Laboratories, Stockholm, Sweden). We recomputed the planned physical dose for PSPT patients ($D_p$) and dose-averaged LET ($LET_d$) with a clinically commissioned track-repeating Monte Carlo algorithm [181]. For each patient, we obtained contrast-free follow-up CT acquired approximately 4 months after the end of the course of radiation therapy. We selected this imaging timeline because it matched the period of clinically reported RP or, if no RP was reported, captured acute response in general.
6.3.2 Testing the feasibility of using image density change as the clinical endpoint for RBE

To ensure that the IDC reflected toxicity development, we randomly selected 20 patients who were clinically identified to have developed RP, and the RP regions were contoured by a board-certified radiologist (XXX). The average IDC for voxels in the RP region was compared with that for voxels in the non-RP region (i.e., the normal ipsilateral lung excluding the RP region). A significantly higher IDC for the voxels in the RP region would indicate that RP contributed more to IDC than did all other radiation responses (e.g., pneumonia, atelectasis) combined, confirming the association of IDC with toxicity.

6.3.3 Establishing Dose–Response Relationship

We used a biomechanical model–based deformable image registration algorithm that is commercially available in the RayStation treatment planning system [100] to deform the follow-up CT to the planning CT. To ensure voxel-level accuracy, we performed the following steps: 1) we generated a triangle mesh using parameters previously optimized for robustness and accuracy [162]; 2) we used boundary conditions for a sliding lung interface; and 3) we modified the treatment planning system–generated deformation vector field by adding constraints on vessel alignment [101]. The voxel-level correspondence enabled us to obtain the IDC in voxels of the normal ipsilateral lung. These voxels were at least 2 mm from the boundaries of lung and the primary tumor. All lung and tumor contours were reviewed by a board-certified radiologist (XXX). To establish the dose–IDC correspondence for each image voxel, we resampled $D_x$, $D_p$, and $LET_a$ maps to match the resolution of the planning CT (0.98 x 0.98 x 2.5 mm$^3$) using nearest-neighbor interpolation with SimpleITK v2.0.2 (http://simpleitk.org/). Using a previously validated automatic landmarks identification method, we quantified the target registration error (TRE) at vasculature branching points of the DIR to be 5.6 mm. AAPM TG-132 recommends that TRE should be less than the voxel
size (diagonal length) which is 2.9 mm for the images used in the study [102]. Therefore, to account for DIR uncertainty (mis-registration), all dose and image voxels were simultaneously binned along all three dimensions into “super” voxels of size 1.98 x 1.98 x 5.0 mm³.

We first grouped the lung voxels into dose bins at 5-Gy intervals and calculated the mean physical dose and mean IDC for the voxels belonging to each dose bin. To ensure statistical power, we excluded dose bins with fewer than 1 million voxels, which defined the highest dose bin to be plotted. Subsequently, we correlated voxel-level dose–IDC relationships by fitting the Lyman–Kutcher–Burman (LKB) normal-tissue complication probability (NTCP) model [65, 182]:

\[
NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{EUD-TD_{50}}{m\cdot TD_{50}}} e^{-\frac{x^2}{2}} dx
\]

(1)

with a change in the meaning of the parameters: \( NTCP \) becomes the normalized IDC (against the maximum IDC of any dose bin); the effective uniform dose (\( EUD \)) represents the mean dose of each dose bin, since each voxel represents itself as a sample, independent of the location and patient; \( TD_{50} \) represents the dose that would achieve half the maximum IDC (the global maximum for both groups); and \( m \) represents the speed at which the IDC increases with the increasing physical dose. We used non-linear least squares from the Python SciPy package (version 1.6.2) to fit the model and obtain \( m \) and \( TD_{50} \). To evaluate the goodness of fit, we obtained the pseudo-\( R^2 \) using the following equation:

\[
R^2 = 1 - \frac{\text{variance(residual IDC)}}{\text{variance(measured IDC)}}
\]

(3)

where the residual IDC represents the difference between the modeled IDC and the measured IDC.
6.3.4 Obtaining RBE values

After we plotted the fitted NTCP models for both the IMRT and PSPT groups, we graphically determined the IDC levels at which to compute $D_x$ and $D_p$. Specifically, we selected the lowest and highest IDC levels that both curves reached (i.e., both $D_x$ and $D_p$ could be associated given the IDC value). We then computed $D_x$ and $D_p$ with their respectively fitted NTCP models at each IDC level of 5-HU increments between the lowest and highest IDC levels. We then computed RBE values using the ratio of $D_x$ and $D_p$ corresponding to each IDC level.

6.3.5 Fitting RBE models

We then fitted McNamara [177] and Wedenberg [183] models using the measured RBE values, the corresponding $D_x$, $D_p$, and $\text{LET}_d$ values, and an $\alpha/\beta$ of 3 Gy for lung tissue [177]. $\text{LET}_d$ values were each estimated as the average $\text{LET}_d$ value for voxels receiving $D_p \pm 0.001$ Gy. We used the pseudo-$R^2$ calculated with equation (2) to describe how closely the model-fitted RBE values matched the measured RBE values.

6.4 Results

The voxel-level NTCP fit result is shown in Figure 21. The average number of voxels in each dose bin was $2.0e+6$ for IMRT ($1.3e+6$–$6.0e+6$) and $1.7e+06$ for PSPT ($1.3e+6$–$2.6e+6$). Voxels in PSPT patients demonstrated that dose had a more linear relationship with IDC according to the higher fitted $m$ parameter of 0.71 (i.e., a faster IDC increase for low-dose proton voxels than photon voxels). For all dose bins (except 62–67 Gy), the PSPT IDC was significantly higher than the IMRT IDC ($p < 0.001$).
To identify the lowest and highest IDC levels for RBE modeling, we compared the fitted curves between IMRT and PSPT. Because the PSPT curve was higher than the IMRT curve, we set the lowest IDC level at 20 HU based on the PSPT curve and set the highest IDC level at 115 HU based on the IMRT curve. The range of IDC levels, the inversely derived $D_x$ and $D_p$ corresponding to each IDC level, and the estimated LET$_d$ are given in Table 5. The McNamara and Wedenberg models returned the fitted parameter(s) of $p_0 = 3.9$, $p_1 = 15.8$, $p_2 = 2.3$, and $p_3 = -0.3$ and $q = 13.0$ with a pseudo-$R^2$ of 0.99 and 0.92, respectively.
Table 5. RBE modelling result. At each IDC (image density change) level, the ratio between the corresponding $D_x$ (photon dose) and $D_p$ (proton physical dose) returned the measured RBE. LET$_d$ (dose-averaged linear energy transfer) was estimated as the average LET$_d$ of voxels receiving $D_p \pm 0.001$Gy.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>28.8</td>
<td>9.8</td>
<td>5.1</td>
<td>3.0</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>25</td>
<td>31.3</td>
<td>12.8</td>
<td>4.1</td>
<td>2.5</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>30</td>
<td>33.5</td>
<td>15.4</td>
<td>4.1</td>
<td>2.2</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>35</td>
<td>35.6</td>
<td>17.9</td>
<td>4.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>40</td>
<td>37.5</td>
<td>20.2</td>
<td>3.8</td>
<td>1.9</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>45</td>
<td>39.3</td>
<td>22.4</td>
<td>4.1</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>50</td>
<td>41.1</td>
<td>24.5</td>
<td>4.2</td>
<td>1.7</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>55</td>
<td>42.8</td>
<td>26.6</td>
<td>3.6</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>60</td>
<td>44.5</td>
<td>28.7</td>
<td>4.1</td>
<td>1.6</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>65</td>
<td>46.2</td>
<td>30.7</td>
<td>3.0</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>70</td>
<td>47.9</td>
<td>32.7</td>
<td>3.4</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>75</td>
<td>49.6</td>
<td>34.7</td>
<td>4.0</td>
<td>1.4</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>80</td>
<td>51.3</td>
<td>36.8</td>
<td>3.9</td>
<td>1.4</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>85</td>
<td>53.1</td>
<td>39.0</td>
<td>4.2</td>
<td>1.4</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>90</td>
<td>54.9</td>
<td>41.2</td>
<td>3.5</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>95</td>
<td>56.9</td>
<td>43.6</td>
<td>3.5</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>100</td>
<td>59.0</td>
<td>46.1</td>
<td>3.2</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>105</td>
<td>61.3</td>
<td>48.9</td>
<td>3.8</td>
<td>1.3</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>110</td>
<td>63.9</td>
<td>52.0</td>
<td>3.6</td>
<td>1.2</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>115</td>
<td>66.9</td>
<td>55.6</td>
<td>3.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Testing for the association of IDC with clinical toxicity response revealed that RP voxels had an IDC of 451 ± 176 HU, whereas non-RP voxels had an IDC of 23 ± 74 HU (p < 0.001).

6.5 Discussion

Herein, we demonstrate a novel approach to model proton RBE using voxel-level clinically imaged response inside patients as the biological endpoint. Our results demonstrate that RBE varies depending on dose, in which RBE reached 3.0 for low-dose voxels and decreased to 1.2 for high-dose voxels. We validated the RBE values by fitting
two published empirical RBE models and achieved a pseudo-$R^2$ of 0.99 and 0.92, respectively. To our knowledge, this is the first study that shows clinical evidence of variable RBE relationships in normal lung tissue.

The difference in IDC between RP and non-RP voxels validated our usage of using image changes in post-treatment images as a clinical endpoint for RBE. In addition, several studies have suggested the feasibility of using IDC as the clinical endpoint for RBE modeling for proton therapy. For example, Peeler et al. analyzed post-treatment magnetic resonance images of pediatric patients with ependymoma and found that voxel-level image changes depended on increasing LET and dose [178]. Underwood et al. studied radiographic changes in postmastectomy chest wall CT images of 20 breast cancer patients (proton $n = 10$, photon $n = 10$) [89]. By analyzing the image change against the dose at the voxel level, they found that proton patients had a greater increase in both image density and IDC compared with photon patients. In these patients, the higher IDC per dose was associated with higher-grade abnormalities, which suggests clinical RBE values higher than 1.1.

After we obtained both the dose maps and the IDC maps, the next step was to determine a correlation so dose could be calculated given an IDC. Our decision to use the LKB NTCP model was inspired by Begosh-Mayne et al., who compared the accuracy of four NTCP models in correlating dose with IDC response on the binned voxel basis and showed that the LKB NTCP model was the most accurate in predicting both early and overall radiographic response [90]. The 0.95+ pseudo-$R^2$ for both the IMRT and PSPT patients validated the accuracy of the fitted LKB NTCP model. However, as shown in Figure 21, the high dose datapoints of the PSPT patients did not follow the sigmoidal curve. This may have been due in part to the difference in the patients’ dose prescriptions. The clinical trial in which the patients had been enrolled implemented a dose escalation to 74 Gy (RBE) instead of the conventional 60–66 Gy (RBE). As a result, voxels receiving 60 Gy (RBE) in patients with dose escalation could be physiologically different than voxels with the same
dose but in patients with conventional dose prescriptions. Physically tracking voxels—for example, by creating a map showing voxels’ proximity to the tumor surface—could provide more insight into the varying sensitivity of tissue voxels in biological dose damage.

After we measured the RBE values for the IDC levels, we validated these RBE values using two empirical models derived from existing cell experiments. Here, the aim was to use existing models to validate our clinically derived RBE values, not to determine the best model with which to validate our results. In future studies, researchers are encouraged to select models according to their preferences or develop full mechanistic models to describe the clinical endpoint of IDC. One limitation of the study was that the patients were not matched even though they were randomized in the clinical trial, as we tried to include as many patients as possible. As a result, RBE is currently derived from voxels of proton and photon patients that might inherently carry different response sensitivity. Nevertheless, future studies should focus on matching patients while enrolling a sufficient number to ensure strong statistical power.

6.6 Conclusion

Using CT-based IDC as the clinical endpoint, we obtained voxel-level variable RBE values in patients with LA-NSCLC previously treated with IMRT or PSPT. Therefore, our findings demonstrate the feasibility of proton RBE modeling based on clinically imaged evidence and warrant additional studies of similar approaches to explain clinical proton biological damage, using models focusing on clinical response as the endpoint.
Chapter 7: Discussion

7.1 Overall Summary

The presented project aims to improve the understanding of radiation damage through dose accumulation and proton biological dose modelling to help improve decision making based on toxicity predictions. If dose-limiting radiation-induced toxicities can be better predicted, dose distributions can be better optimized. Patients with lower toxicity risks could be treated with a higher tumor dose, potentially improving treatment outcomes and quality of life.

The phase-based accumulated dose (4DA) is the most extensive dose accumulation workflow that incorporates both breathing motion and longitudinal response. The first step to achieve such workflow is to establish a reference phase that provides the most robust longitudinal phase-phase DIR, which is our current gap in knowledge. This was achieved in Chapter 3. We analyzed the three most representative phases of the breathing cycle: T0, T3, and T5. We also compared these phase-phase DIRs with the AVG-AVG DIR because dose is conventionally planned on the AVG image and the AVG-AVG DIR accuracy needs to be quantified if we want to apply it clinically in dose mapping for plan adaptation purposes. Therefore, this work not only helped us decide on the reference phase for our dose accumulation workflow, but also served as the clinical guideline for AVG-based adaptive workflow.

Next, we needed to optimize our selection of DIR algorithms used for both intra-4DCT and inter-4DCT DIRs to ensure accuracy and efficiency. Since we utilize the RayStation TPS for its scripting capability to automate our dose accumulation, it was most
convenient to leverage the DIR algorithms directly available in RayStation: the intensity-based ANACONDA algorithm and the biomechanical model-based MORFEUS algorithm. We used ANACONDA for intra-4DCT DIR since studies have shown it is both accurate and efficient for registering breathing phases obtained in the same imaging session [97]. We decided to use the MORFEUS algorithm since most of the tumor response in these patients were inelastic or a mixture of elastic and inelastic, thereby not preserving image intensity. However, the success rate and accuracy of the MORFEUS algorithm is affected by the mesh generating parameters, which are selected by the user. When the algorithm fails, the only way to find a suitable parameter is through trial-and-error, which prevents automation. Therefore, we conducted the study in Chapter 4 to find the optimal (and alternative) parameter set that consistently produces accurate T5-T5 DIR. With the optimal reference phase and the parameter set for MORFEUS DIR algorithm identified, the workflow met the accuracy recommendation of AAPM TG-132 on clinical implementation of DIR, proving our hypothesis. Subsequently, we could streamline the dose accumulation pipeline.

In Chapter 5, our comparison of 3DP, 4DP, 3DA, and 4DA was comprehensive and individually evaluated the effect of breathing motion and the effect of longitudinal response. Although the DVH-based organ-level metrics showed minimal differences on the population basis, the differences in individual patients demonstrated the importance of individualized treatment and the potential of ART: with 16% patients found to experience the deviation (magnitude of difference) of NTCP score from planned dose distribution to the accumulated dose distribution, we proved the hypothesis of a clinical dose difference between the planned dose and the accumulated dose. The final step was to associate their dose difference to accuracy difference in representing the delivered dose. We established the dose-response correlations where the goodness of fit (pseudo-$R^2$) indicated how well each dose described localized damage, thereby their accuracy. The clinical radiographic based response modelling was inspired by Begosh-Mayne et al. which recorded that the LKB
NTCP model most accurately modeled the normalized IDC for short-term radiation effects (<6 months) on the voxel level. This method was implemented in our study where follow-up PET/CTs close to 4 months post RT were selected. The improvement in the pseudo-$R^2$, although modest, demonstrated that the accumulated dose indeed more closely estimates the delivered dose than the planned dose.

The dose-IDC correlation also served as the backbone of our proton RBE modelling work in Chapter 6. Since the concept of RBE is flexible with the clinical endpoint, we chose IDC as the endpoint. RBE was derived from proton and photon doses achieving the same level of IDC. As we measured the RBE values to be larger than 1.1, we decided to fit an established model published by McNamara et al. [177]. The fit result of pseudo-$R^2 = 0.98$ confirmed the hypothesis that the goodness of fit for the variable RBE model was more than 0.9.

Our results on the anatomical response and RBE effect of dose have the potential to help physicians tailor treatment plans to individual patients, based on the predicted toxicity and adapt the treatment plan according to anatomical response while maintaining sufficient tumor dose coverage. Ultimately, we hope to improve the quality of life and long-term survival for NSCLC patients.

7.2 Discussion on Specific Studies

7.2.1 Inter-4DCT DIR Accuracy

This work narrowed our knowledge gap in the effect of registration of 4DCT images acquired at different timepoints and accuracy in phase-phase and AVG-AVG DIR for IGRT. In most clinical practices, patient dose is calculated using the AVG while they are treated with free breathing. Therefore, the most clinically achievable dose mapping is through AVG-to-AVG mapping. RTOG 1106 [184] used AVG-based planning and adaptive treatment. Its protocol only allowed rigid registration between the AVG images due to the lack of accuracy
analysis of AVG-AVG DIR, which justified the clinical need to quantify the accuracy of AVG-AVG DIR.

Evaluating the results of the phase-phase comparisons, T3 and T5 did not demonstrate a statistically significant difference in TRE. We made the decision to use T5 since patients spend most of the time in the exhalation stage within a breathing cycle. Therefore, T5 is usually anatomically robust with the least breathing artifact. In addition, the response assessment required the registration with PET/CT. In our clinical practice of PET/CT acquisition, we use T5 for correction of static PET when misregistration between PET and CT occurs [185].

7.2.2 Parameter Optimization for Biomechanical Model-Based DIR Algorithm

As a collaboration with Dr. Brian Anderson, this work provided recommendations on the most robust parameter selection for BM-DIR in RayStation for lung and liver DIRs. We automated the mesh creation, finite element analysis, accuracy, efficiency, and robustness assessment through the scripting capability of RayStation, and assessed the DIR results following the recommendations of AAPM TG-132 [102].

The immediate goal of this work was to facilitate automated dose accumulation for lung and liver. Previously, the tetrahedral mesh generation in RayStation could fail without pointing to the exact area of failure, thus requiring trial and error on the parameters that initially generated the surface mesh. With parameters that consistently achieve anatomically robust and accurate BN-DIRs, we streamlined the inter-4DCT dose mapping.

The secondary goal was to develop appropriate BM-DIR working practices allowing integration into clinical workflows such as adaptive radiotherapy. For this goal, we have tested the optimal parameters on 16 head-and-neck patients and 16 prostate patients, and they also successfully generated meshes then DIRs for parotid glands and prostate, respectively. Furthermore, we expect the same parameters when extrapolating to other DIR
algorithms, since we are focusing on fundamental components of mesh generation of BM-DIR.

7.2.3 Dose Accumulation

We addressed the limitations in dose accumulation studies due to the lack of systematic review of the effect of breathing motion and longitudinal response. Most studies focused on the planning 4DCT [77, 78], which does not account for tumor response and changes in breathing motion that occur over radiotherapy, especially for advanced stage lung cancer patients. Studies that do investigate the effect of longitudinal response are limited by small sample size due to the labor-intensiveness of image and dose mapping, focus on organ-level DVH-based metrics such as MLD and V\textsubscript{20} that lack spatial features, or hypo-fractionated treatments that restrict the effect of response [79–81].

Besides the organ-level DVH metrics, we interrogated the dose evaluations at the region and voxel level. Lung is considered as a parallel organ for clinical dose constraints. However, it consists of different lobes, airways, and blood vessels. Moreover, increasing studies on the functional substructures of lung are showing the potential benefit of lung functional avoidance [186, 187]. The DVH-based MLD and V\textsubscript{20} are inevitably insufficient in describing the localized damage to the lung. Our results on the overlap of isodose regions confirmed that albeit the organ-level metrics are minimally different on the population basis, the underlying dose differences can be very dramatic. In DVH metrics, these sub structural dose differences are washed out by the cancelling effect between the breathing motion and the longitudinal effect.

To correlate dose differences with clinical evidence, we quantified response using IDC from pre- to post-treatment. This was obtained by registering the follow-up image to the planning image. Subsequently, the same response map for each patient can be associated with either the planned or the accumulated dose, and the difference in the dose-response
correlations indicated how well each dose representation describes the localized damage, thereby the delivered dose [49].

Results of the project showed significant differences between accumulated and planned dose distributions for certain patients (e.g. large tumor breathing motion and large tumor shrinkage throughout treatment), we showed the importance of using dose accumulation to represent the delivered dose rather than simply the planned dose, and such information can aid in the development of clinical dose accumulation strategies for adaptive radiotherapy, especially with treatment planning for re-irradiation. The results currently are only for the photon arm of the clinical trial, and I have been working on the dose accumulation for the proton arm based on Monte Carlo simulations. Future work can leverage the pipeline to evaluate these proton cases. We are expecting a more significant difference from the planned dose to the accumulated dose due to the proton range uncertainties.

In summary, the completed study was the first to analyze the difference between planned dose and accumulated dose for multiple 4DCTs. This ensures dose to be clinically used in toxicity prediction model is the closest representation of the actual delivered dose. Therefore, our 4DCT-based weekly dose accumulation contributed to the knowledge in the field by providing a more complete dose accumulation scheme.

7.2.3.a Defining Accuracy of Dose Accumulation

The quantification of dose accumulation accuracy can be challenging as it cannot be measured directly (e.g., through a dosimeter measure in deforming patient tissues to determine the true dose). However, we investigated how the geometric accuracy of image deformation interplays with dose calculation – using TRE to determine the accuracy of DIR. The intra-4DCT and inter-4DC (phase-phase) reported TRE of 2.3 ± 1.8 mm and 2.8 ± 1.9 mm, respectively, which means that lung voxels were deformed to within 3 mm radius of the
corresponding voxel. With a clinical dose grid of 3x3x3 mm³ for conventionally fractionated IMRT, the uncertainty of image deformation was on a similar scale with the size of dose calculation. Therefore, we addressed the concerns such as geometric accuracy of image deformation and changing mass when determining the accuracy of dose accumulation, especially when working with tumors that are heavily affected by breathing motion (e.g., tumors at diaphragm) and large shrinkage throughout the treatment course.

In addition, the changing anatomy over the course of treatment affected the definition of dose deposition. During image deformation, the tumor region can experience volumetric changes and deformation causing adjacent voxels to merge into one and where portions of tumor cells that used to exist in the voxel now are replaced by normal lung tissues when they are eliminated by radiation, making tumor dose accumulation difficult to define. Therefore, cases with responding tumors will have to be scrutinized when a high dose gradient is seen at the tumor boundary in the original plan dose distribution. If a tumor shrinks elastically where the surrounding lung tissue is ‘pulled’ centripetally, a boundary condition can be placed at the tumor boundary to preserve the bordering relationship. When the tumor is ‘eroding’ away as it responds to radiation, it reveals the underlying lung tissue that was covered by tumor rather than being pushed away by tumor. In such cases, a boundary condition at the tumor boundary would not be considered. However, with the help of Lusmeralis Almodóvar-Abreu (undergraduate student at the University of Puerto Rico who I mentored as a CPRIT summer experience student), we identified that most tumors regressed inelastically. We manually placed 4 landmarks at vessel bifurcations closest to the tumor at the planning 4DCT and a mid-treatment 4DCT. And we compared the distance from each landmark to the closest tumor surface. If distance between each landmark and tumor surface all increased, the tumor would have experienced a complete inelastic regression. If half the landmarks increased, the tumor would have experienced a mixture of elastic and inelastic regression. Both scenarios would require not placing the boundary
condition over the tumor, and we found that most of the tumors qualified for this situation. However, a more automated method in finding the 4 landmarks to determine the tumor shrinkage should be developed to facilitate personalized DIR method.

7.2.3.b Defining Clinical Significance of Dose Difference

The dose accumulation study improved the predictive power of current methods. Currently, NTCP scoring is the default clinically used indicator for toxicity prediction [74, 76], but it often lacks appropriate level of complexity. It assumes a sigmoid relationship between toxicity and independent variables such as dose, which might not be the universal solution. [26] In addition, such method assumes uniform irradiation of target volume, and the parameters come from interpolation or extrapolations from whole organ data or are based purely on the experience of the clinicians involved in the contouring [27].

However, considering the lack of more sophisticated complication prediction methods, even though NTCP scoring can be criticized for being overly simplified since it does not maintain the 3-dimensional dose distribution and is subject to human factors, it can still be used as a caliper to reasonably estimate the tendency of a patient to experience adverse events. McCulloch et al. studied the potential impact on NTCP models from dose accumulation for liver patients who received stereotactic radiation therapy and concluded a statistically significant difference between the NTCP models based on the two different dose representation methods. [28] Therefore, we designed the hypothesis such that if differences in lung NTCP based on accumulation dose vs planned dose (used by current methods) yielded > 5%, such difference can be considered clinically significant. This was previously proposed by Langendijk et al. to aid clinical decisions in re-planning to guide clinical trial strategies [164].
7.3 Auto-Segmentation for Lung

The nature of this project involves retrospective analysis of patient data from a clinical trial [85]. Lung segmentation is a crucial step to guide the image registration for dose accumulation and for IDC calculation using follow-up imaging. The dose-IDC correlation also requires lung boundary to be defined. With the sheer number of images to contour, it was important to utilize an auto-segmentation tool. However, the methods available at the start of the project were not robust enough in the presence of the complex lung tumor and the lung conditions such as pleural effusion which should be excluded and atelectasis which should be included in lung volume. These methods included the Atlas-based method within the RayStation TPS, and the publicly available deep-learning based algorithm that were trained with mostly healthy or early staged lungs [188]. Another available source was the algorithms that participated in the 2017 AAPM Thoracic Segmentation Challenge. However, the training set in the Challenge did not include pre-existing diseases such as pneumonia, which can be seen across the patients in the dataset used for the completed project. Figure 22 is an example of an evaluation of different auto-contouring algorithms in RayStation in the presence of pre-existing lung disease.

Figure 22. Comparison of different lung auto-segmentation methods in RayStation in the presence of pre-existing lung disease. The left lung contains opacity above the tumor and all auto-contouring methods (yellow: model-based, green: atlas-based, and blue: deep-learning-based algorithm trained with the same dataset of the AAPM Challenge) failed to correctly identify the actual lung boundary as depicted by the pink, manual. The right lung is disease-free, and contours from all segmentation methods align well. This shows the
importance for auto-contouring algorithms to consider cases with pre-existing lung conditions.

Therefore, I trained RayStation’s patch-based 3D U-NET algorithm using my manually curated lung contours for 100+ CT images, including most of the patient analyzed so that the remaining weekly images could be automatically contoured with prior knowledge. Therefore, no systematic testing of accuracy was conducted. The segmentation on the AVG and T5 of the planning 4DCT for dose comparisons (both images) and response assessment (T5 only) were further modified manually on top of the auto-segmentation result to ensure high precision for voxel-level dose and response analysis. The resulting algorithm is available in the research version of RayStation. This work was a collaboration with Dr. Bastien Rigaud.

Figure 23. The created segmentation model in RayStation.
7.4 Future Directions

7.4.1 CBCT-Based Dose Accumulation

We realize that the acquisition of weekly 4DCT images is not common for most clinical practices. Therefore, CBCT-based dose accumulation can be the most clinically viable way to investigate the effect of dose accumulation [189]. AVG-based accumulated dose (3DA) most closely mimics FB CBCT-based dose accumulation. However, the CBCT image equality can limit CBCT from being used as the new planning image for plan adaptation but recent hardware and software advancements in CBCT’s image quality may enable adaptive radiation therapy [190, 191].

Prospectively, dose accumulation can be used in adaptive radiotherapy (ART) to compute the composite dose which consists of the dose already delivered from the original plan and the dose to be delivered from the new plan. I hope our results encourage future prospective clinical trials to evaluate the clinical impact of using dose accumulation to adaptive the treatment plan vs not-adapting original treatment plan by comparing toxicity development in lung and other organ-at-risks such as heart and esophagus.

7.4.2 Incorporating Clinical Metrics/Record for A Comprehensive Toxicity Model

For our study, the clinical toxicity record such as symptomatic RP was defined and graded according to CTCAE v3.0 [57]. Despite the clear guideline of CTCAE, the criteria can be subjective, varying among individual treating physicians and confounded by underlying comorbidities. In the clinical trial, the outcomes review committee met to discuss each patient reported to have developed RP, and the final grading was decided by the outcomes review committee.

Ideally, for the dose accumulation study, a predictive model such as receiver operator characteristic (ROC) curve can be constructed and the area under the curve (AUC) value be compared between the NTCP values from the planned dose and the accumulated
dose using the toxicity record for symptomatic RP. This could further validate our result that the accumulated dose is more accurate than planned dose based on radiographic evidence. However, the number of available patients for the dose accumulation study was restricted, mostly due to accessibility of weekly 4DCT datasets in the clinical storage system which rendered the patients unqualified for the study. Only 10 patients out of the analyzed 67 patients developed symptomatic RP. Therefore, a much larger number of patients need to be analyzed. Furthermore, the probability of exceeding clinical dose constraints for organs-at-risks with and without consideration of dose accumulation can also be studied to demonstrate the necessity of having more accurate dose representations in the prediction model.

In addition to dose metrics, studies have demonstrated the improvement in toxicity prediction by incorporating clinical metrics. Krafft et al. discovered that by including clinical data and dose information with radiomics features extracted from the normal lung volume on pre-treatment CTs, pneumonitis predictability as represented with the AUC score, increased from 0.51 (no image features) to 0.68 [192]. They used Common Terminology for Criteria for Adverse Events v3.0. Castillo et al. studied 100 NSCLC patients that underwent FDG PET/CT prior to RT and used SUV\textsubscript{95} information of PET and achieved AUC of 0.78 (95% CI = 0.69 – 0.87) in predicting symptomatic RP using CTCAE v4.0 [69]. Wang et al. studied predicting capability of MLD excluding (GTV, PTV, CTV) and concluded highest AUC for MLD was 0.707 to predict RP grade 2+ based on CTCAE v3.0 [72].

The ultimate goal is to create the most robust RP prediction model based on an extensive dose accumulation that will ensure better representation of the delivered dose, coupled with clinical metrics such as age and smoking status. We hope that eventually the developed toxicity prediction strategy from the proposed project can be deployed in the clinic, which can help improve NSCLC treatment and toxicity management. This will also
require prospective clinical trials, which can be supported by the automated workflows developed and validated in this thesis.
General RayStation Scripting Advice

The project benefited from the scripting/automation functionality of the RayStation treatment planning software. Here is some general advice on how to script in RayStation:

1. RayStation is based on python, so learning Python syntax is a must.
2. The ‘record’ function is a direct and simple method to know the desired script.
3. Reference the user manual, the scripting manual, and other manuals.
4. There are also sample scripts that can be found in the same network folder where RayStation is installed.
5. Create a toolbox with functions that you often use so you can keep using the same function in different scripts. When the function needs to be changed, you only need to change it in the toolbox once rather than changing it in every script that uses it.
6. Ask your colleague and RaySearch support.
7. Lastly but most importantly: just because the script ran without any hiccups, doesn’t mean the result is guaranteed to be physically possible! Always QA the process and check the results visually in the user interface before analyzing the results!
References


head and neck cancers. International Journal of Radiation Oncology*Biology*Physics

40. Chua GWY, Chua KLM (2019) Which patients benefit most from stereotactic body
radiotherapy or surgery in medically operable non-small cell lung cancer? An in-depth
look at patient characteristics on both sides of the debate. Thorac Cancer 10:1857.
https://doi.org/10.1111/1759-7714.13160

Advanced Lung Cancer: A Simple Question With a Complicated Answer. JAMA Oncol

radiotherapy – is volumetric modulated arc therapy the answer? J Med Radiat Sci

43. Bezjak A, Paulus R, Gaspar LE, Timmerman RD, Straube WL, Ryan WF, Garces YI,
Pu AT, Singh AK, Videtic GM, McGarry RC, Iyengar P, Pantarotto JR, Urbanic JJ,
Sun AY, Daly ME, Grills IS, Sperduto P, Normolle DP, Bradley JD, Choy H (2019)
Safety and Efficacy of a Five-Fraction Stereotactic Body Radiotherapy Schedule for

Standard-dose versus high-dose conformal radiotherapy with concurrent and
consolidation carboplatin plus paclitaxel with or without cetuximab for patients with
stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two
2045(14)71207-0


48. Paganetti H (2017) Proton Beam Therapy


https://doi.org/10.1016/j.prro.2017.03.005

59. Hope AJ, Naqa I el, Bradley JD, Vivic M, Lindsay PE, Bosch WR, Purdy JA, Deasy
JO (2004) Radiation pneumonitis/fibrosis risk based on dosimetric, clinical, and
https://doi.org/10.1016/J.IJROBP.2004.06.150

(Basel) 10

Bremer M, Dickgreber N (2016) Inhalative steroids as an individual treatment in
symptomatic lung cancer patients with radiation pneumonitis grade II after

62. Marks LB, Bentzen SM, Deasy JO, Kong F-MS, Bradley JD, Vogelius IS, el Naqa I,
dose-volume effects in the lung (QUANTEC: ORGAN-SPECIFIC PAPER). Int J
Radiat Oncol Biol Phys. https://doi.org/10.1016/j.ijrobp.2009.06.091

63. Evans W, Leucutia T (1925) Intrathoracic Changes Induced by Heavy Radiation.
American Journal of Roentgenology Radium Therapy 13:203–220

64. Libshitz H (1993) Radiation Changes in the Lung

65. Lyman JT (1985) Complication probability as assessed from dose-volume

66. Kutcher GJ, Burman C (1989) Calculation of complication probability factors for non-
uniform normal tissue irradiation: The effective volume method. International Journal
of Radiation Oncology*Biology*Physics 16:1623–1630. https://doi.org/10.1016/0360-
3016(89)90972-3


concurrent chemotherapy and three-dimensional conformal radiotherapy (3D-CRT).

https://doi.org/10.1016/j.ijrobp.2006.07.1337

https://doi.org/10.1016/j.ijrobp.2009.06.091

https://doi.org/10.1016/S0360-3016(02)03986-X


https://doi.org/10.1016/j.ijrobp.2009.07.1754


https://doi.org/10.1016/J.SEMRADONC.2019.02.002


11:1271–1294. https://doi.org/10.1142/S0218001497000597


https://doi.org/10.1088/0031-9155/54/7/001


https://doi.org/10.1016/j.prro.2018.08.005

https://doi.org/10.1016/j.ijrobp.2005.05.045

https://doi.org/10.1186/s13014-017-0907-8


https://doi.org/10.1118/1.4894702

https://doi.org/10.1088/0031-9155/60/8/3359


134. Wijesooriya K Importance of 4D simulation, Planning, and delivery


patients. Scientific Reports 2017 7:1 7:1–8. https://doi.org/10.1038/s41598-017-07586-x


176. Chen Y, Ahmad S EMPIRICAL MODEL ESTIMATION OF RELATIVE BIOLOGICAL EFFECTIVENESS FOR PROTON BEAM THERAPY †. https://doi.org/10.1093/rpd/ncr218


Vita

Yulun He was born and raised in China. In 2011, he arrived in the United States and continued his high school studies in Sayre School in Lexington, Kentucky. In 2013, he entered Worcester Polytechnic Institute in Worcester, Massachusetts and graduated in 2017 with a Bachelor of Science of a double major in Mechanical engineering and Physics. In Fall 2017, he began the pursuit of a Doctor of Philosophy in Medical Physics. In his free time, Yulun enjoys snowboarding, trying out new restaurants in Houston, and travelling around the world.