Approaches for Incorporating Patient-Specific Response in Radiation Therapy

by

Daniel Frank Polan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Martha M. Matuszak, Chair
Professor Marina A. Epelman
Professor John E. Foster
Professor Marc L. Kessler

Daniel F. Polan

polandan@umich.edu

ORCID iD: 0000-0002-8182-8832

Dedication

To those impacted by childhood cancer, and the caregivers and healthcare professionals who tirelessly work to care for and support them.

Acknowledgements

First, I would like to acknowledge and thank my dissertation committee, especially my PhD advisor, Professor Martha Matuszak. Throughout my education at the University of Michigan, Dr. Matuszak has strongly supported my growth as both a student in the Nuclear Engineering and Radiological Sciences Department and as a researcher in the Radiation Oncology Department. Additionally, thank you to Professor Marc Kessler who has been a fearless supporter of my career development in addition to being a wonderful mentor over the last few years. I also greatly appreciate the valuable support of Professor Marina Epelman, who has provided her expertise to better my research and personal development, and Professor John Foster for inspiring me to pursue a career of impactful science and meaningful outreach.

I would also like to thank my former research group, especially my first research advisor, Professor Kristy Brock, who motivated me as an undergraduate student to combine my skills in programming and problem solving with my desire to better the lives of those affected by cancer. Additionally, thank you to the support and camaraderie provided by my fellow trainees including Guillaume Cazoulat, Molly McCulloch, and Rocky Owen.

I also wish to share my sincere appreciation for the many individuals in the Radiation Oncology Department who have guided me throughout the years, including Randy Ten Haken, James Balter, Jean Moran, Choonik Lee, and Rojano Kashani, and Matt Schipper. Thank you to the research dosimetry group, including Robin Marsh, Janell Dow, Karen Vineberg, and David Jarema for assisting me on many projects and to Markus Varsta for assisting in software

development. Furthermore, thank you to the many physicians who provided clinical context and insight, including Ted Lawrence, Mary Feng, Shruti Jolly, and Caitlin Schonewolf.

Most importantly, I wish to thank my loving family and friends. My wife, Erin Spencer, has provided continuous support and without her I would not have made it this far. Thank you to my parents, Ed and Jeanne Polan, and to my brothers and sister for their unwavering love and support (even if they didn't exactly understand what I was doing). Lastly, thank you to my dog, Bunsen, for keeping me company during a year of working from home and for reminding me when it is time to take a walk.

This work is supported in part by the National Institutes of Health grant P01 CA059827 and Varian Medical Systems.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
3D-CRT	Three-Dimensional Conformal Radiation Therapy
4DCT	Four-Dimensional Computed Tomography
AAA	Analytical Anisotropic Algorithm
AP	Anterior-Posterior
API	Application Programming Interface
CBCT	Cone-beam Computed Tomography
$CE_{3+,2y}$	Cardiac Event within 2 years (grade 3 or higher)
CIs	Confidence Intervals
CRC	Colorectal Cancer
CT	Computed Tomography
$CVaR^{+}$	Conditional Value-at-Risk
DCE-MRI	Dynamic Contrast-Enhanced Magnetic Resonance Imaging
DIR	Deformable Image Registration
DOO	Dose-Only Optimization
Dose BCs	Dose Boundary Conditions
DSC	Dice Similarity Coefficient
DVC	Dose-Volume Constraints

DVF	Deformation Vector Field
DVH	Dose-Volume Histogram
EBRT	External Beam Radiation Therapy
EQD_2	Equivalent Dose in 2 Gy Fractions
ESAPI	Eclipse Scripting Application Programming Interface
FEA	Finite Element Analysis
FEM	Finite Element Models
GTV	Gross Tumor Volume
HCC	Hepatocellular Carcinoma
HN	Head and Neck
HNC	Head and Neck Cancer
HNQOL	Head and Neck Quality of Life
HRCTV	High-Risk Clinical Tumor Volume
IGRT	Image-Guided Radiation Therapy
IMRT	Intensity-Modulated Radiation Therapy
ℓEUD	Linearized Equivalent Uniform Dose
LR	Left-Right
$LRPFS_{2y}$	Local Regional Progression-Free Survival at 2 years
MCO	Multi-Criteria Optimization
MDA	Mean Distance to Agreement
MLC	Multileaf Collimator
MRI	Magnetic Resonance Imaging
MU	Monitoring Units

NSCLC	Non-Small Cell Lung Cancer
NTCP	Normal Tissue Complication Probability
OAR	Organ-at-risk
P-TTB	Prioritized Total Toxicity Burden
PC	Pharyngeal Constrictor
PET	Positron Emission Tomography
PG	Parotid Gland
PTV	Planning Target Volume
PUO	Priority Utility Optimization
PWL	Piecewise-Linear
QOL	Quality-of-Life
RE_{3+}	Radiation Esophagitis (grade 3 or higher)
$RILT_{3+}$	Radiation-Induced Lung Toxicity (grade 3 or higher)
RT	Radiation Therapy/Radiotherapy
SBRT	Stereotactic Body Radiation Therapy
SD	Standard Deviation
SI	Superior Inferior
TCP	Tumor-Control Probability
TPS	Treatment Planning System
TRE	Target Registration Error
TTB	Total Toxicity Burden
VaR	Value at Risk
VMAT	Volumetric Modulated Arc Therapy

Abstract

In an era of increasingly personalized medical care, radiation therapy (RT) continues to be a prominent and effective focal cancer treatment. However, the standard RT treatment process remains largely uniform, using a one-size-fits-all approach designed to maximize therapeutic benefit to patient populations typically stratified only by tumor type, anatomical location, and extent of the disease. However, additional factors within these populations may lead individual patients to respond differently to RT during and after treatment. Therefore, current RT treatment practices and processes are likely suboptimal for individual patients and could be better tailored to each patient by incorporating additional patient-specific information into the RT planning and treatment evaluation processes. To facilitate this, actionable treatment response models that account for patient variability must be developed and implemented. These models can be formulated though analysis of anatomical and functional changes observed within patient imaging and by evaluating treatment outcomes based on underlying biological factors in addition to the therapeutic intervention. This dissertation includes multiple investigations focused on developing and implementing patient-specific models to support personalized, evidence-based RT.

Advancements in anatomical and functional patient imaging have allowed for non-invasive evaluation and analysis of changes throughout the course of RT and follow-up care. Accurate spatial alignment of patient imaging is required to incorporate these new imaging techniques into the RT planning process, evaluate treatment response, and better correlate anatomical and functional changes to therapeutic radiation delivery. However, the ability to link and compare imaging studies before, during, and after treatment is often obscured by significant spatial and

volumetric variations caused by physiological and treatment-related factors. To address this, we conducted a series of studies on geometrically-based patient modeling to improve the accuracy of deformable image registration (DIR) and evaluate the clinical impact. First, an intensity-based DIR algorithm was characterized in the context of modern clinical imaging scenarios. Algorithm parameters were evaluated with respect to clinical accuracy metrics and dosimetric impact. Next, radiation dose, in combination with patient factors including tumor location and type, were used to biomechanically model longitudinal liver anatomy changes during RT and follow-up care. A previously developed biomechanical DIR algorithm was modified to incorporate the newly developed liver-response models and was shown to improve spatial and volumetric correlation.

Following studies in geometric patient modeling, we studied the direct incorporation of patient-specific response models into the RT planning process. Through a prioritized fluence optimization approach, we implemented the concept of utility-based planning where the optimization objective is to maximize the predicted value of overall treatment utility for a patient, defined by the probability of efficacy (e.g., local control) minus the weighted sum of toxicity probabilities. Implementation of the prioritized utility-based optimization strategy offers an intuitive approach to biological optimization in which planning trade-offs are explicitly optimized. The feasibility of this new planning approach was demonstrated on a cohort of non-small cell lung cancer patients and was shown to improve overall plan utility through situational tumor dose modification and normal tissue dose redistribution.

Chapter 1
Introduction

1.1 Radiation Therapy

Despite advances in the biological understanding of cancer and the development of new treatments, cancer continues to be the leading cause of death globally.[1] Furthermore, the burden of cancer incidence and mortality continues to grow rapidly worldwide, driven in part by aging populations and increasing cancer risk factors associated with globalization and developing economies.[1] Radiation therapy (RT) plays a central role in the management of cancer, both in the US and worldwide, with a majority of cancer patients benefiting from RT at some point during their care.[2,3] Over the last decade, technological advancements throughout the RT workflow, summarized in Figure 1.1, have aimed to increase the precision of RT, improve tumor control, and decrease toxicity with the underlying goal of delivering therapeutic radiation doses to tumors while attempting to avoid radiation damage to surrounding normal tissues. This includes advancements in pre-treatment imaging, where the incorporation of magnetic resonance imaging (MRI) and functional imaging, in addition to the standard-of-care computed tomography (CT) imaging, has provided physicians with improved visualization of tumor structures and organs, facilitating precise identification of target and normal tissue volumes for treatment planning.[4,5] Improvements in image registration tools, which provide the ability to spatially correlate data from multiple imaging sources, have contributed to the ability to delineate these structures, allowing for physicians to simultaneously use multiple imaging datasets to contour structures.[5] Additionally, accurate image registration has permitted improved flow of patient imaging data into the treatment

planning process, empowering treatment planning to be based on not only anatomical data, but also functional imaging data (e.g., functional avoidance planning).[6]

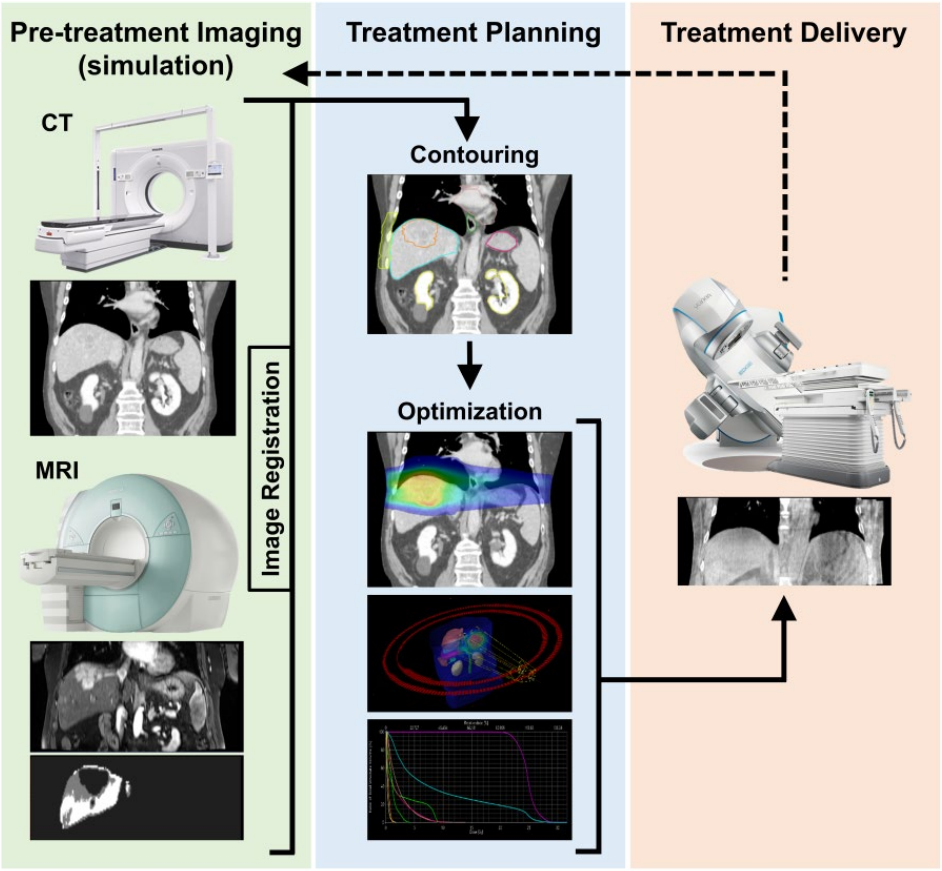


Figure 1.1. External beam radiation therapy treatment planning workflow including pre-treatment imaging, treatment planning, and treatment delivery. First a patient receives pre-treatment imaging. This is typically in the form of a CT scan, but also can include MR imaging and nuclear medicine studies to help visualize anatomical structures and quantify organ function for treatment planning. Then imaging data is correlated through image registration allowing for accurate contouring of tumor and normal tissue volumes across the multiple imaging modalities. Anatomical information is then used to create and optimize a personalized treatment plan (dose distribution) with the goal of delivering radiation to the tumor while avoiding surrounding normal tissues. Following the creation of an acceptable plan, radiation is delivered to the patient with assistance from image guidance. The process can be repeated if the patient responds to therapy or if anatomical changes occur during the course of treatment. Machine images from Philips Healthcare (Amsterdam, Netherlands), Siemens Healthineers (Erlangen, Germany), and Varian Medical System (Palo Alto, CA).

The impact of improved imaging continues to treatment delivery, where daily 3D image guidance through in room cone-beam CT (CBCT) decreases patient setup variations and, when combined with image registration, enables tracking of patient response and delivered dose.[7]

Similarly, in RT treatment planning, technological and engineering advancements in treatment delivery, along with advancements in computing, have made it possible to develop highly conformal intensity-modulated RT (IMRT) treatment plans (i.e., beam trajectories and fluence intensities) that can be accurately delivered to the patient.[8] Additionally, it is now feasible to iterate through the treatment planning process, adapting a patient's treatment plan to account for anatomical and biological changes that present during the course of RT.[9]

With these advancements and others increasing the precision in RT treatments, it has become evident that increased personalization may also be possible. This represents a potential paradigm shift in RT, which has been historically known for employing 'one-size-fits-all' approaches, such as using uniform prescription doses and applying population-level normal-tissue dose constraints in treatment planning. However, incorporating patient-specific data into RT poses many challenges, from understanding the accuracy and limitations of new technologies to determining how and when to act on information not previously available or considered. This thesis focuses on multiple approaches working towards incorporating patient-specific response into RT. This encompasses steps related to generating improved anatomical and physiological patient models and the ability to directly intervene on the basis of such models.

1.2 Deformable Image Registration

Patient management throughout oncology care increasingly relies upon imaging for non-invasive diagnosis, response assessment, and follow-up care.[10] Although advancements in imaging have improved care broadly across oncology, patient imaging is particularly important in RT, serving as an essential component for the accurate delivery of radiation.[11] Prior to treatment, imaging is used to visualize the extent of disease, delineate the position of the tumor and surrounding normal-tissue structures for treatment planning, calculate radiation dose, and assess

the functional capacity or distribution of radiation-sensitive organs.[12,13] This process may include gathering and integrating data from multiple imaging modalities to provide physicians increased insight into the patient and disease. For example, some modalities may be focused on improving the anatomical targeting accuracy of treatment by providing better visualization of the boundary or motion of the tumor (e.g., 4DCT, MRI), while others are focused on quantitatively estimating tumor cell metabolism or organ function (e.g., positron emission tomography (PET), Dynamic Contrast-Enhanced (DCE) MRI).[13,14]

Furthermore, advancements in the technical capabilities of radiation detectors and reconstruction techniques have enabled the use of imaging during treatment to improve the precision and accuracy of the actual treatment delivery, which can occur days after initial patient imaging and across multiple treatment sessions.[15,16] In addition to improving the spatial targeting of radiation delivery at the time of treatment, image-guided radiation therapy (IGRT) through on-board imaging modalities (e.g., CBCT, MRI) also provides the opportunity to visualize anatomical and physiological changes over the course of treatment. This enables clinicians to non-invasively monitor and track the tumor's response to radiation in addition to factors attributed to increased patient setup errors, such as patient weight loss, that can lead to decreased treatment efficacy and unnecessary radiation-induced toxicity. In the presence of anatomical or physiological changes that cause the treatment to deviate from the original, intended RT plan, there is a potential to intervene by adjusting or re-planning the RT treatment based on a more recent anatomical representation and clinical presentation of the patient.[17-20] Applications of these methods, referred to broadly as adaptive radiotherapy, have become increasingly popular in RT and have the potential to directly improve patient outcomes as a result of more personalized, precise, and accurate treatment. Numerous clinical trials (e.g., NCT01504815, NCT03416153, NCT04751747)

are currently underway to investigate the application and determine the effectiveness of adaptive RT for multiple anatomical locations and cancer types.

Another application for imaging data collected during IGRT is to longitudinally reconstruct the actual dose delivered to the patient over the course of treatment factoring in daily variations in patient setup and anatomy.[21] While gathering this information at the individual level can be useful in determining when and if to adapt a patient's treatment, it may also be useful in generating more accurate representations of patient dose (i.e., delivered dose compared to planned dose) for efficacy and toxicity modeling at the population level. Understanding the link between radiation and patient outcomes (e.g., tumor control, toxicity) is fundamental to the clinical success of RT. However, historical models relating dose to patient outcomes, such as tumor-control probability (TCP) and normal-tissue-complication probability (NTCP) models, have been based on the original, planned dose to a patient, despite the known presence of anatomical and patient-positioning variations throughout treatment. Multiple studies have demonstrated that a more accurate representation of radiation dose accumulated in tumors and normal tissues over the complete course of RT can be more predictive of patient outcomes than planned dose.[22,23] Furthermore, updated models incorporating the accumulated, or delivered, dose can impact clinical treatment decisions related to tumor dose selection and normal-tissue dose limits.[24]

With the wealth of imaging data now available in RT and the development of related applications, accurate spatial alignment of images acquired from various imaging modalities at different time points for a patient is a critical step to enable effective clinical use for both qualitative and quantitative assessments. The goal of image registration is to provide accurate spatial alignment of two image datasets by determining the spatial transformation that exists between the images.

Generally speaking, this transformation can either be represented by a rigid registration, where the transformation is limited to uniform translation and/or rotation, or a non-rigid (deformable) registration, where the transformation can vary in magnitude and direction across the image. While rigid registration has become a common tool for 2D and 3D data fusion in RT, such as its use for daily patient positioning, it is inherently limited by the degrees of freedom present within the transformation (up to six parameters, three for translation and three for rotation) and cannot accurately represent shape changes, volume changes, or distortion between structures within the two image datasets.[25] Determining the optimal set of parameters for a rigid registration is relatively easy and is largely considered a solved issue in RT. Alternatively, deformable image registration (DIR) has the ability to model complex, spatially-variant anatomical changes that occur within the human body between image acquisitions and over the course of treatment. However, solving for the optimal DIR between two images is computationally more challenging, with the number of parameters extending up to three times the number of voxels within the dataset (i.e., a deformation vector field (DVF) comprised of independent 3D translation vectors for each voxel). Although the concept of DIR is almost four decades old and many algorithms have been proposed, a renewed interest in this area of research has been generated by increased computational abilities, more broadly adopted IGRT, and the growing field of adaptive RT.[25-30] These solutions can be grouped into two general categories, intensity-based approaches (e.g., B-spline, Demons) and geometric approaches (e.g., Chamfer, biomechanical finite element), both of which are investigated in this thesis.[25] Figure 1.2 shows an example intensity-based deformable registration between lung images of the same patient at different time points.

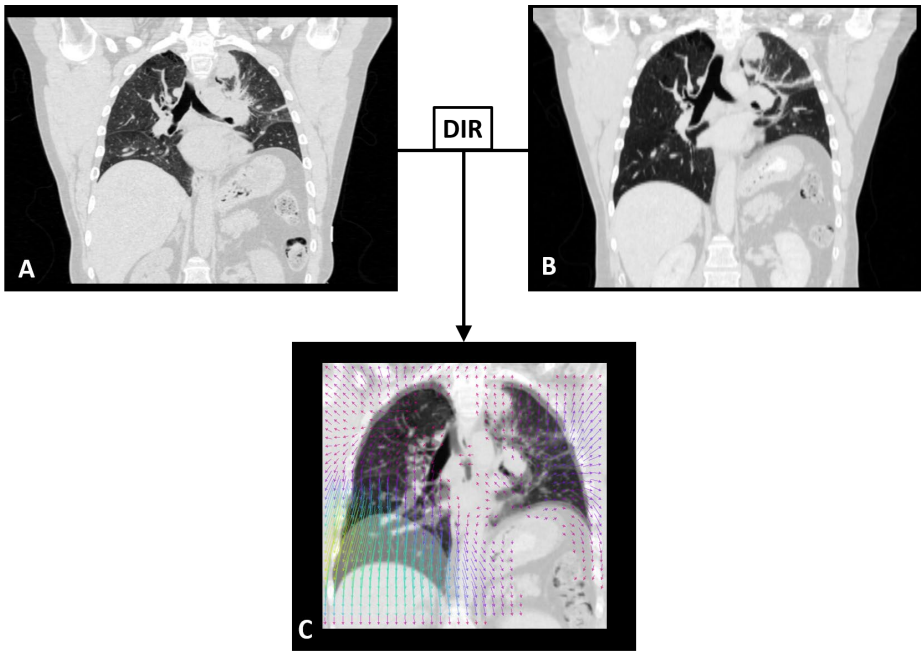


Figure 1.2. Example of intensity-based DIR for a lung patient with CT images acquired in different breathing states. Image A shows a deep exhale CT, Image B shows a breath-hold treatment planning CT, and Image C shows an overlay of the images with the DIR calculated DVF.

Intensity-based DIR methods use similarities between numerical greyscale information (e.g., Hounsfield units, MRI signal intensity) within two image datasets to generate a transformation spatially linking the images. The registration parameters composing the transformation can be based on global spline-based interpolation methods (thin-plate splines) where movement of control points influences deformation globally, local spline-based models (B-splines) where a weighted grid of basis functions influences the deformation locally, or optical flow models (e.g., demons) where local voxel motion is determined by intensity gradients. Regardless of the transformation model, registration parameters are optimized based on the measured similarity of spatially overlapping image intensities within the two images. Depending on the image types, this similarity metric is typically calculated using either the image intensities directly, in the form of normalized cross-correlation (for singular-modality DIR), or the joint

probability distribution functions of image intensity levels (i.e., voxels grouped based on similar intensities) in the form of mutual information (for multi-modality DIR).[25]

Geometric methods offer an alternative approach to using image intensity information to optimize and generate the transformation. First, corresponding point-based landmarks and/or anatomical surfaces are identified in the images. These features are then extracted from the images and are used to drive the registration through point or surface matching. In general, these methods also require an interpolation or re-sampling method, similar to those used in intensity-based methods, to convert the point and surface transformations back into a transformation that can be applied to the entire image. One method for preserving the physical integrity of the global transform is to use biomechanical finite element models (FEMs) of the deforming tissues in which realistic material properties, such as the stiffness and Poisson's ratio of the tissue, are assigned to each of the specific tissue types present in the image. An example DIR algorithm using this approach is shown in Figure 1.3. Although this may produce more physically plausible deformations when compared to intensity-based DIR methods, biomechanical methods require delineation of organs, tumors, or corresponding points prior to registration. Historically this has been a time-consuming manual task, but advancements in deep learning techniques for image segmentation could soon provide robust, automated pathways for segmentation and increase the adoption of biomechanical DIR in RT.[31]

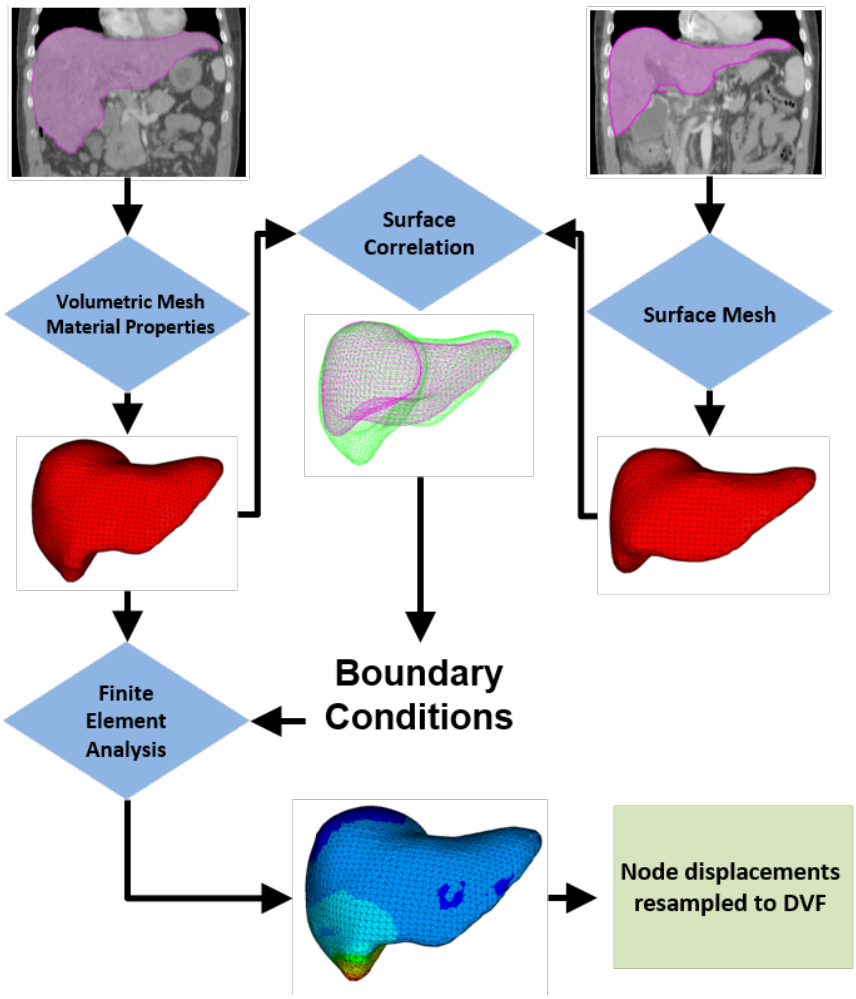


Figure 1.3. Example workflow of Morfeus, a FEM-based biomechanical DIR algorithm, for a liver patient. First the liver volumes are manually delineated and converted into 3D finite-element meshes with realistic material properties. Then, through surface correlation, boundary conditions are created and fed into a finite element analysis (FEA) solver to generate a DVF.

1.3 Treatment Plan Optimization

The overall goal of RT is to deliver a therapeutic dose of radiation to a tumor while avoiding radiation damage to nearby organs. In external beam RT (EBRT), the main type of RT that we focus on in this dissertation, it is not possible to irradiate cancer cells without simultaneously irradiating surrounding tissue. Although advancements in image guidance and delivery techniques have improved the precision and conformality of EBRT, radiation beams still transverse normal tissues before and after converging on the target structure. EBRT treatment plan

optimization aims to create a dose distribution, dependent on beam angles, collimator shapes, and fluence patterns, that maximizes radiation dose within the tumor to achieve local control and minimizes the dose within normal tissues to decrease treatment-related acute and late morbidities. This aim is embodied in the concept of the therapeutic window, shown in Figure 1.4, in which plan optimization aims to maximize the therapeutic window through increasing the probability of tumor control with awareness of the increasing probability of normal tissue toxicities at higher doses.

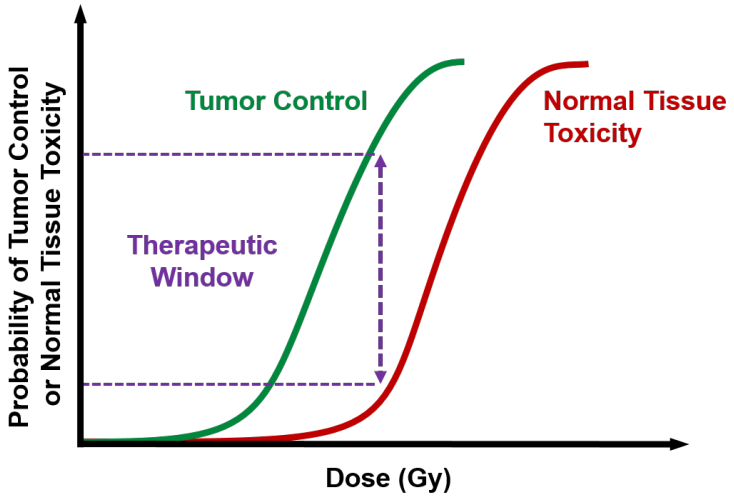


Figure 1.4. TCP and NTCP curves demonstrating the therapeutic window in radiation therapy. At any given radiation dose, there is a probability of tumor control and normal-tissue toxicity. The separation between tumor control and toxicity represents the therapeutic window (or therapeutic ratio).

In the mid-1990s, the emergence of 3D imaging, the development of 3D treatment planning systems, and computer-controlled treatment machines equipped with multileaf collimators (MLCs) ushered in an era of 3D conformal RT (3D-CRT).[32,33] The implementation of 3D-CRT was largely based on the concept of forward planning, in which appropriate beam angles and collimator shapes were first determined by a treatment planner.[34] Then, using a uniform or linearly varying fluence pattern, the planner calculates the required machine output necessary to achieve the prescription dose. This tended to be a manual, iterative process, requiring the planner to revisit

the selected beam angles and collimator shapes, then recalculate the machine output, to generate a treatment plan that satisfied the planning goals.[35]

Advancements in computing, particularly those related to dose calculation and optimization methods, led to the development of inverse planning, which is not only a more automated approach to generating treatment plans but also can improve normal tissue sparing while still delivering a therapeutic dose to the tumor. In inverse planning, treatment planning goals, typically represented by numerical criteria in the form of dose metrics, are combined into an objective function that is then optimized to generate a plan. The concept of inverse planning has allowed treatment planners to also explore the use of non-linear fluence patterns that were not conceivable in manual planning. This planning technique, referred to as intensity modulated radiation therapy (IMRT), has become common practice in RT, providing even greater conformal shaping of dose distributions than 3DCRT.[36]

In this dissertation, the concept of beamlet-based inverse planning is used to create personalized treatment plans. In this method, also referred to as fluence map optimization, the inverse optimization problem is solved for a distribution of beamlet weights that minimize the dose to normal tissue structures, defined by dose metrics included in the optimization objective function, subject to constraints on competing metrics of tumor dose and coverage. Beamlets are defined by segmenting the larger beams at each beam angle into smaller sub-beams whose optimal intensity, or weight, can be solved for during optimization. Typically, this requires beam directions to be defined prior to optimization so that the individual dose contributions of beamlets to discretized points in the patient, representing tumor and normal tissue volumes, can be calculated. A visual overview of this IMRT planning process is shown in Figure 1.5. Historically, the objective functions used in inverse IMRT planning have been defined using dose metrics as the optimization

criteria; however, these metrics are only surrogate measures of tumor and normal tissue dose response. Recently there has been increased interest in developing techniques that move beyond the use of simple dose metrics for planning criteria, and instead represent the inverse RT planning problem directly based on patient response and treatment outcomes.[37] Chapter 4 focuses on these techniques through the inclusion of patient-specific dose-response models in the inverse planning objective function.

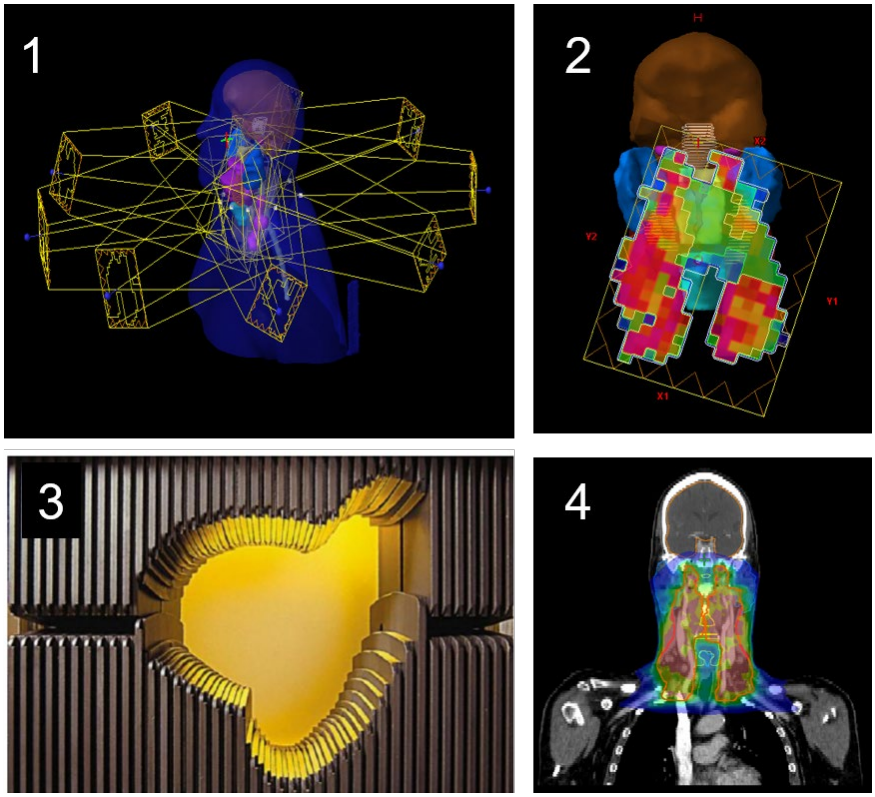


Figure 1.5. Visual overview of IMRT planning for a head and neck cancer patient. First, beams are arranged around the patient (Image 1) and dose calculation is performed to determine the contribution of smaller segmented beams, 'beamlets,' to points within the patient. Then a numerical optimization objective is defined and the beamlet-weight optimization problem is solved to determine the optimal fluence patterns (Image 2) for each beam. Optimal fluence is then turned into a physically deliverable plan through calculating patterns of MLC (Image 3) motion with the final dose distribution (Image 4) calculated after this motion is determined. Image 3 provided by Varian Medical System (Palo Alto, CA).

1.4 Goals and Organization of this Dissertation

The generation and incorporation of patient-specific response models in RT has the potential to improve disease control and lower toxicity rates. In part, the creation of more personalized response models requires accurate image registration to spatially correlate and analyze longitudinal changes in patient anatomy and organ function. Additionally, image registration is required to more accurately determine the delivered dose to patient, which could improve the accuracy and applicability of historical and newly-generated TCP and NTCP models. Once new or updated models are created, clinical interventions need to modified to allow for inclusion of patient-specific information in the treatment planning process. This thesis focuses on these aspects of personalized radiation therapy through (Ch. 2) quantitative characterization of an intensity-based DIR algorithm, (Ch. 3) development and evaluation of a DIR algorithm incorporating patient-specific dose-based response, and (Ch. 4) implementation of a patient-specific response models in RT treatment planning. The end of each chapter includes an "extensions" section which briefly describes additional studies related to the primary foci of this dissertation. This includes co-authored and ongoing studies that are based on work developed during the course of this research.

In Chapter 2, titled "Clinical Characterization of an Intensity-Based Deformable Image Registration," a comprehensive evaluation of a clinically implemented intensity-based DIR algorithm is presented using quantitative metrics of registration accuracy. The sensitivity of the DIR algorithm to varying input parameters was explored and we demonstrated that registration accuracy could be improved, but only in certain anatomical locations.

In Chapter 3, titled "Characterization and Implementation of Radiation-Induced Dose-Volume Liver Response for Biomechanical Deformable Image Registration," a novel

biomechanical modeling method for hepatic tissue including independent dose-driven deformation forces was developed and implemented on a series of primary and metastatic liver cancer patients. The proposed DIR technique demonstrated improved target registration and deformed segment volume accuracy for longitudinal patient modeling when compared with a previously developed biomechanical DIR algorithm.

In Chapter 4, titled "Incorporation of Patient-Specific Efficacy and Toxicity Estimates in Radiation Therapy Plan Optimization," a novel IMRT optimization strategy is introduced that directly incorporates patient-specific dose response models into the planning process. In this strategy, we integrate the concept of utility-based planning where the optimization objective is to maximize the predicted value of overall treatment utility, defined by the probability of efficacy minus the weighted sum of toxicity probabilities. We believe that implementation of the utility-based objective function offers an intuitive approach to biological optimization in which planning trade-offs are explicitly optimized. To demonstrate the feasibility of the approach, we apply the strategy to treatment planning for non-small cell lung cancer patients.

Chapter 2
Clinical Characterization of an Intensity-Based Deformable Image Registration Algorithm

Portions of this chapter are substantially equivalent to work the author previously submitted as an internal report to a commercial research partner as part of a co-development agreement.

2.1 Abstract

Image registration tools have become more commonly available in commercial radiotherapy software systems over the last decade. Despite this availability, the clinical adoption and use of these tools, particularly deformable image registration (DIR), remains limited in part due to the inability to quantitatively evaluate and characterize the performance of the system. Additionally, many clinically available DIR implementations do not incorporate the necessary tools required to robustly test user-controlled parameters or do not even allow users to adjust algorithm parameters. In this study, we aim to quantitatively evaluate an intensity-based DIR algorithm for numerous clinically relevant imaging situations, including single and multi-modality images. Additionally, through the creation of a new scripting interface, we characterize the performance of the algorithm across a range of input parameters to determine the sensitivity of registration accuracy to the algorithm. In this retrospective study, the DIR algorithm was evaluated for patients in each of the following anatomical locations: head and neck (HN), liver, male pelvis and female pelvis. HN DIRs were evaluated from planning-CT (pCT) to repeat-pCT and pCTs to daily CBCTs using dice similarity coefficients (DSC) and mean distance to agreement (MDA) of corresponding anatomical structures. Male pelvis DIRs were evaluated from pCT to CBCTs using

DSC, MDA, and target registration error (TRE) of implanted radiofrequency (RF) beacons within the prostate. Liver DIRs were evaluated from pMR to pCT using DSC and TRE of vessel bifurcations. Female pelvis DIRs were evaluated between fractionated brachytherapy MRIs using DSC and MDA of corresponding anatomical structures. Registration algorithm parameters, including those related to the image intensity histogram, deformation model, and optimization process, were tested on the HN and lung cases with both pCT and daily CBCT imaging. Evaluation of the algorithm demonstrated clinically acceptable DIR accuracy for lung pCT-CBCT. Near clinically acceptable accuracy was achieved for head and neck (pCT-pCT and pCT-CBCT), lung 4DCT, and GYN MR-MR. Clinically acceptable accuracy was not achieved in liver MR-CT and prostate CT-CBCT. Analysis of registration parameter sensitivity demonstrated that a large improvement in registration accuracy could be achieved in lung cases, but only a marginal and variable improvement could be achieved for head and neck cases. Overall this work demonstrates that although clinically acceptable DIR accuracy can be achieved in some imaging situations with class-level DIR algorithm parameter selection, refinement of the parameters for specific anatomical locations may significantly improve accuracy. However, for certain clinical applications, registration accuracy of intensity-based DIR is relatively insensitive to algorithm parameters and other DIR techniques should be considered.

2.2 Introduction

Throughout the cancer care continuum, technological advancements in minimally or non-invasive patient imaging have transformed the detection, treatment, and management of cancer. Improved detection, signal processing, and reconstruction techniques in computed tomography (CT), magnetic resonance (MR) imaging, and tomographic nuclear medicine imaging have improved the spatial accuracy of anatomical imaging, increased the ability to differentiate different

tissue types in the body, and have facilitated wider-spread adoption of functional, or physiological, imaging to guide treatment. With these advancements, modern radiation therapy has become an increasingly image-rich environment where the quantitative use of imaging to calculate dose, delineate structures, and position daily treatment is fundamental to the precise delivery of radiation.[38] In a modern RT clinic, patients may receive numerous pre-treatment imaging studies, including CT imaging for normal tissue contouring, dose calculation, and motion characterization, MR imaging for improved visualization of tumors in soft tissue and functional assessment, and nuclear medicine studies for characterization of tumorigenic activity and additional functional assessment.[39,40] Additionally, with the extensive use of image-guided radiation therapy (IGRT), driven by the development of in-room or on-board 3D imaging devices, patients may receive daily CT, cone-beam CT (CBCT), or MR imaging to guide precision delivery of radiation and track treatment response.[41-43] While individually each of these imaging techniques serves an important purpose in guiding treatment, the true potential of multimodality, multi-instance imaging in RT can only be realized when imaging datasets can be spatially linked together for quantitative and qualitative assessment.

Rigid registration of image datasets, through global translation and rotation parameters, provides one method of spatially correlating imaging data. However, this method is only sufficient for anatomical sites with rigid body motion (e.g., brain, bones) and is ill-suited for portions of the body that are subject to complex physiological motion (e.g., lungs, bowel) or motion caused by external forces such as gravity. Furthermore, longitudinal imaging assessments may demonstrate changes in morphology and anatomy caused by treatment interventions (e.g., tumor shrinkage) and/or body habitus changes unrelated to treatment.[47,48] Non-rigid, deformable image registration (DIR) offers a spatial alignment solution that has the potential to account for complex

physiological motion and changes in anatomy. While rigid-registration may be limited to 6 parameters to determine spatial alignment, DIR offers the ability for every voxel in an image dataset to move independently. In medical imaging, however, it is highly unlikely that every voxel within an imaging dataset moves independently, therefore multiple DIR algorithms have been proposed to create physically-plausible deformations.[46-51]

Despite the breadth of work invested in creating new DIR algorithms, the quantitative assessment of DIR accuracy has remained limited.[52] Even with commercial adoption of DIR algorithms, robust clinical evaluations of DIR implementations are sparse, in part due to the lack of proper quantitative evaluation tools within commercial software systems.[25] A recent report from the American Association of Physicists in Medicine (AAPM) has highlighted the need for increased clinical evaluation, quality assurance, and quality control of DIR tools.[25] Quantifying the sensitivity of the DIR algorithms to user-controlled input parameters remains even more limited because many implementations function as 'black-boxes' in which the user does not have the ability to clearly tune the algorithm parameters for specific clinical needs. Without a clear understanding of these parameters and their effect on the registration process, large variability in DIR accuracy has been noted between institutions, even if they are using the same commercial DIR product.[53,54] This study aims to provide a comprehensive clinical evaluation of an intensity-based DIR algorithm and test the effect of the DIR algorithm parameters on quantitative measures of registration accuracy.

Due to the quantity of data and analysis included in this chapter, this chapter is non-traditionally formatted so that specific methods and materials can be easily referenced while reading the results and discussion. An overview section of the methods and materials is provided to explain the common elements of the multiple smaller sub-studies. The grouped methods,

materials, results, and discussion sections for each of the sub-studies can be separated into two primary categories. First we tested and evaluated the DIR algorithm across multiple anatomical sites, then we evaluated the sensitivity of the algorithm to parameter inputs on a subset of the patients (HN and lung).

2.3 Methods and Materials Overview

2.3.1 DIR Algorithm

In this study, we test and evaluate the intensity-based DIR algorithm found commercially in the Velocity software package (Varian Medical System, Palo Alto, CA). Velocity utilizes a B-spline-based DIR model in which mutual information is used as the intensity-based similarity metric to optimize the image transformation. B-spline models are a popular method for determining DIR coordinate system transformations due to their ability to capture complex local deformations without influencing the global transformation and for their ability to inherently produce smooth, continuous DVFs.[55] This is attributed to modeling the deformation through sparse control points across the image, and then calculating the displacement at any voxel through interpolation of the nearby control points. Velocity also incorporates the concept of hierarchical, multi-resolution B-spline models, in which the deformation is first calculated with sparse control point spacing to resolve larger, global deformations. Then the deformation is refined using denser grids of control points to capture small, local deformations. For the intensity-based similarity metric, Velocity uses Mattes's formulation of mutual information, described in detail in Mattes et al.[56] Briefly, mutual information is calculated by measuring the entropy present in the joint probability distribution of voxels sampled between the two image datasets. As entropy is reduced, through optimization of the control point displacements, mutual information within the two images is increased, signaling more accurate spatial alignment of the images. Since the mutual information

metric is based on bins, or groupings, of voxels with similar image intensities, rather than the voxel intensities directly, it is possible to calculate meaningful mutual information metrics between images of differing modalities (e.g., CT and MR). This is particularly useful in RT, where multimodality pre-treatment imaging and daily image guidance has become more prevalent.

All registrations were performed in a development/evaluation build of Velocity version 4.0. Although this specific build includes additional features developed to aid in the evaluation process, the underlying DIR algorithm and implementation remains consistent with the clinical (commercially available) build of Velocity v3.2 and v4.0. In all cases, an automated rigid registration was performed prior to the deformable registration unless both image volumes were already in the same frame of reference (e.g., 4DCT). To test the sensitivity of this DIR algorithm to changes in the input parameters, a C# application programming interface (API) was jointly developed with Varian to provide access to registration parameters that are not currently exposed in the commercial user interface.

2.3.2 Evaluation Metrics

In this study, two structure-based metrics and one point-based metric were used to assess the quality of the DIRs. The structure metric, Dice similarity coefficient (DSC), provides a statistic for spatial overlap between two structures and has a range of 0 to 1 (0 = no overlap, 1 = complete overlap). In the absence of a manual contour reproducibility study, a DSC of 0.70 is generally considered clinically acceptable for moderately sized structures. Since DSC is dependent, in part, on the overall volume of the structure and does not differentiate between overlap of bulk volume (at the center of the structure) and overlap at the edge of the structure, the clinically acceptable DSC levels should be determined in the context of both the manual contour reproducibility and structure size.[25] For small structures, DSC becomes increasingly sensitive to minor errors between

contours, and therefore, it may be more useful to analyze registrations using the second structure metric, mean distance to agreement (MDA).[25] MDA provides the average distance between the boundaries of two structures. Generally, a MDA within a few voxels is clinically acceptable. For larger structures (e.g., liver, lungs), the structure-based metrics become less sensitive and it is preferable to use the point-based metrics such as target registration error (TRE).[25] TRE provides a measurement of the residual error within a registration and is calculated by comparing the known and transformed locations of stable anatomical landmarks, such as vessel bifurcations. Figure 2.1 shows visual representations of these three metrics, with the blue representing expected transformation result and red representing the registration result.

Figure 2.1. Visual representation of the three registration evaluation metrics where blue represents the ground truth and red represents DIR prediction. Abbreviations: Dice similarity coefficient (DSC), Mean Distance to Agreement (MDA), Target Registration Error (TRE).

2.3.3 Patient Data and Clinical DIR Evaluation

Anatomized data from four anatomical locations (head and neck (HN), lung, liver and pelvis) and multiple imaging modalities (planning CT, CBCT, MRI) were used for the analysis. Figure 2.2 shows example images for each of the anatomical and imaging situations considered. In the next section, each imaging scenario is described in further detail and is presented in proximity to the results and discussion. For each of these scenarios, the Velocity DIR algorithm was tested using options broadly available in the clinically released Velocity user interface (UI).

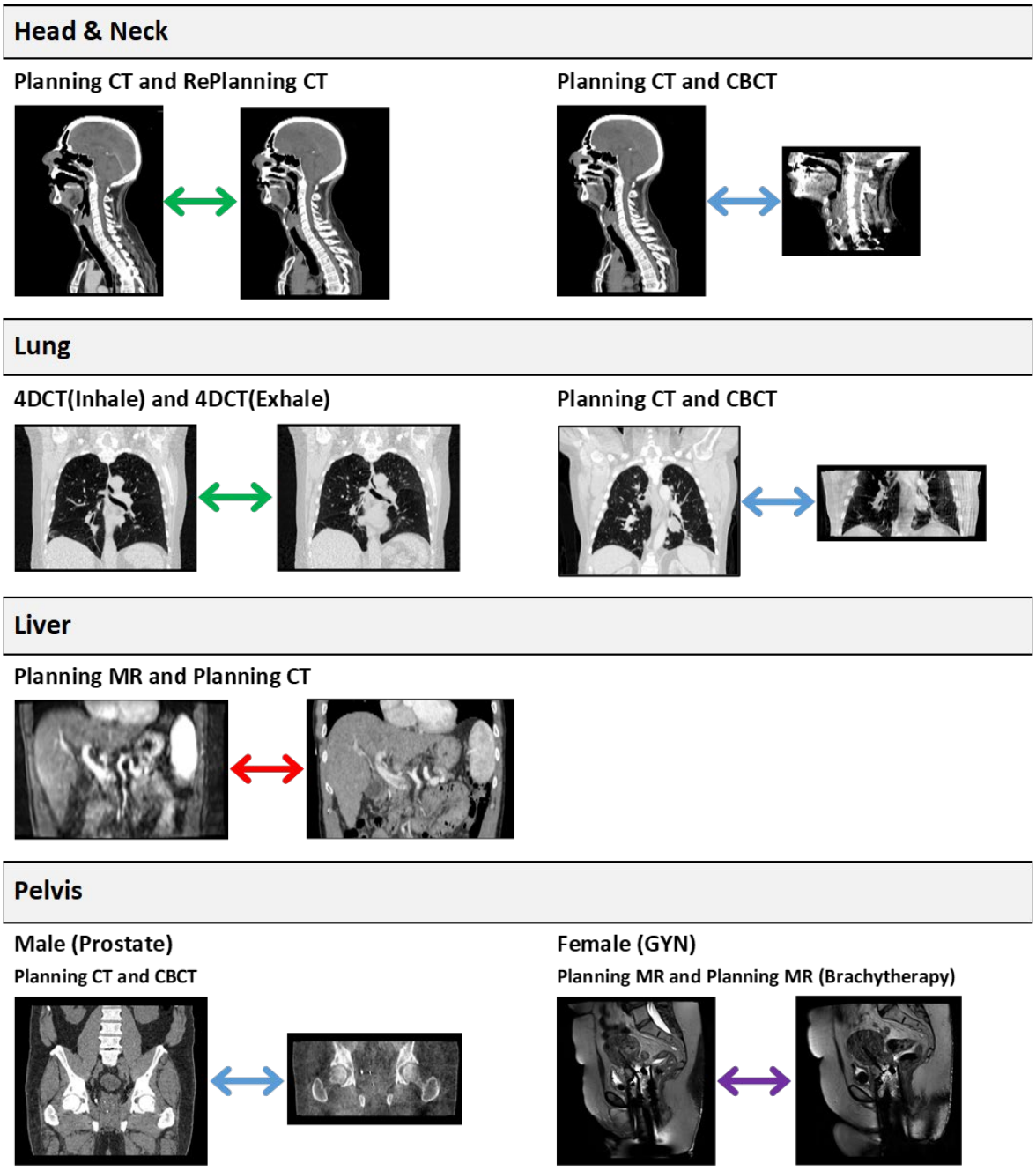


Figure 2.2. Example images from the anatomical sites and imaging situations used to evaluate the DIR algorithm. Green arrows indicate registrations between planning CT images, blue arrows indicate registrations between planning CT and CBCT images, the red arrow indicates registrations between planning MR and CT images, and the purple arrow indicates registrations between planning MR images.

2.3.4 DIR Parameter Sensitivity Analysis

The sensitivity of the DIR algorithm was tested using the HN and lung data from the initial clinical DIR Evaluation. This allowed for exploration of registration parameter sensitivity using all three quantitative registration evaluation metrics. A linear search for each registration parameter was performed, with parameters varied independently when possible. This was done as an attempt to determine the individual effect of each parameter on registration accuracy. When not possible, such as for number of multiresolution levels and grid spacing, preset values were compared in an attempt to determine the registration sensitivity to the joint parameters. Figure 2.3 shows all of the parameters available for adjusting in the Velocity C# API, some of which were previously adjustable in the UI. For registrations involving a CBCT image volume, the CBCT pre-processing filter was applied during the registration process unless otherwise noted.

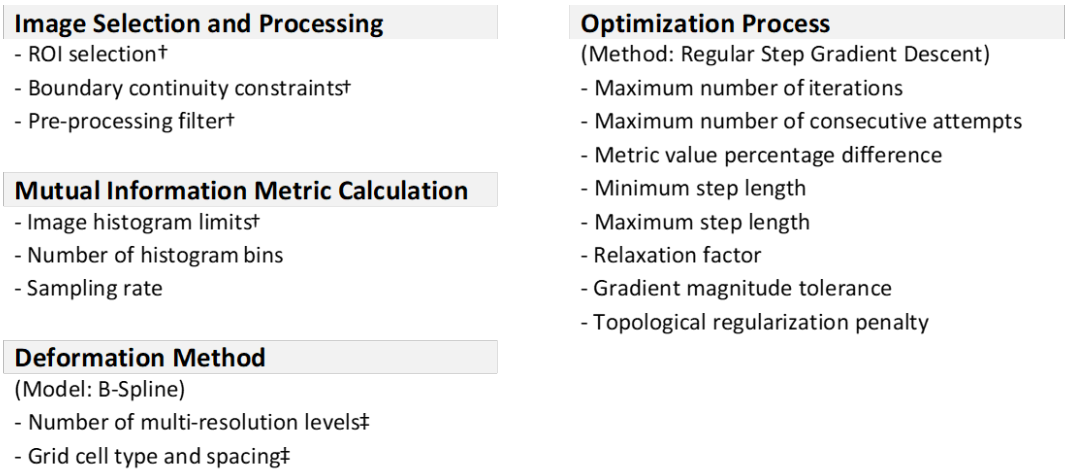


Figure 2.3. DIR algorithm parameters available in the C# API. † Denotes parameter has full adjustability in UI, ‡ Denotes parameter has preset adjustability in UI.

2.4 Methods, Materials, Results, and Discussion (per Sub-Study)

2.4.1 Head and Neck

Planning CT - to - Re-Planning CT

The registration of head and neck planning CTs is clinically useful for contour propagation and mapping doses (planned or accumulated) between treatment plans. Five patients treated at the University of Michigan for head and neck cancer were evaluated. Each of the five patients underwent mid-treatment re-planning as a result of anatomical changes throughout the patient's treatment course. For each re-plan, a new simulation (planning) CT was acquired and contoured for purposes of generating a new clinical treatment plan on the updated patient geometry. This allowed for contour-based evaluation of deformations between each patient's planning CTs. The 'Deformable Multi Pass' option was used for the registration with a manually positioned region-of-interest (ROI) encompassing all the structures of clinical interest. The quantitative evaluation utilized the structure-based metrics, DSC and MDA, for the following 13 normal tissue structures: brainstem, right and left cochlea, inferior and superior pharyngeal constrictor muscles, spinal cord, spinal canal (cervical region), esophagus, larynx, right (R) and left (L) parotid, right (R) and left (L) submandibular (submand) glands.

Figure 2.4 shows the evaluation results for this DIR application. In summary, the DIR performs consistently well in propagating the brainstem, spinal cord, spinal canal, esophagus, left and right parotid glands, and left and right submandibular glands (DSC > 0.7 and MDA < 2.0 mm). Small structures, including the cochlea and constrictors, have lower DSC (0.62–0.70), but the MDA is within a voxel (1.1–1.5 mm). Overall, these results demonstrate that the Velocity DIR algorithm is clinically acceptable in most cases for normal structure propagation between head and neck planning CTs. For patients with rapidly responding or growing lesions near organs at risk

(particularly of concern for HPV+ patients), the DIR may fail to provide locally accurate deformations.

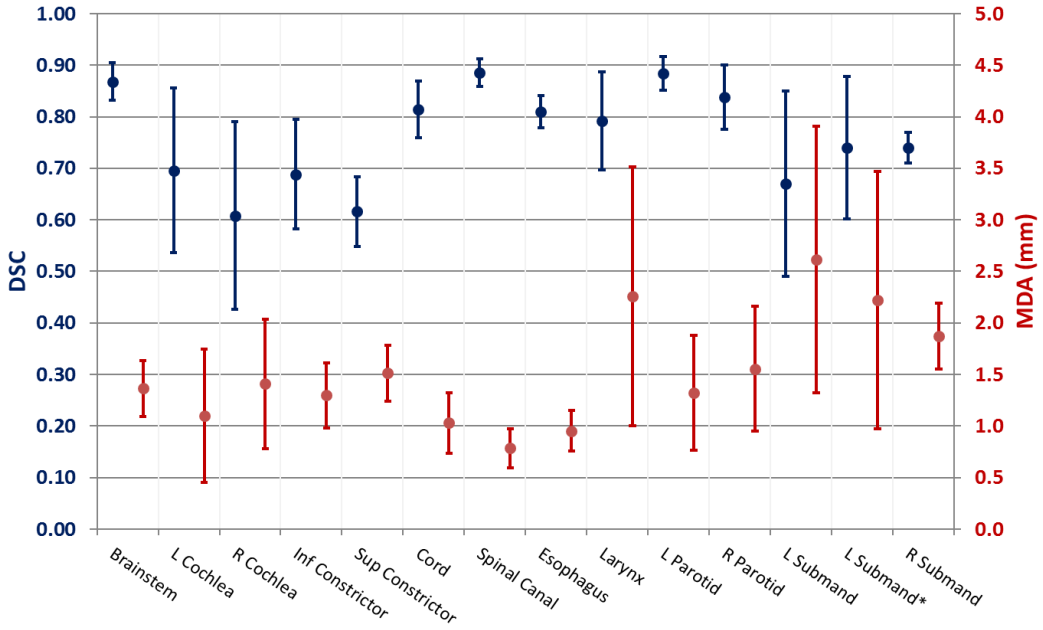


Figure 2.4. DSC and MDA results for DIRs between head and neck planning CTs. Submand* excludes a case with rapidly responding lesion near the left submandibular gland.

Planning CT - to - CBCT

Similar to the previous DIR application, the registration of a cone-beam CT (CBCT) to a patient's planning CTs is clinically useful for tracking patient response over the course of RT and for reconstructing delivery dose. The same five head and neck cancer patients used in the previously described evaluation of planning CTs were used for evaluation of deformations between planning CTs and daily CBCTs. For each of the five patients, three CBCTs were selected with time points spread throughout the treatment course to give a range of deformation magnitudes. Each of the selected CBCTs were contoured by a resident radiation oncologist with previous experience in head and neck contouring for treatment planning. An intra-observer contour variability study was performed since it is not standard clinical practice to contour on CBCTs. For

this study, the physician re-contoured five CBCTs (one per patient) in order to gather statistics on manual contouring variations. The structure base-statistics, DSC and MDA, from the manual contouring were then used as a reference for analyzing the DIR performance. Velocity's 'CBCT Corrected Deformable' registration option was used with an ROI placed at the boundary of the CBCT image volume. The following normal tissue structures were analyzed using the DSC and MDA structure metrics: inferior and superior pharyngeal constrictor muscles, spinal canal (cervical region), esophagus, larynx, right and left parotid, right and left submandibular glands.

Results of the intra-observer contour reproducibility study are shown in Figure 2.5. Based on these results, the clinically acceptable levels for the DIR performance were set at DSC ≥ 0.70 and MDA ≤ 2.0 mm.

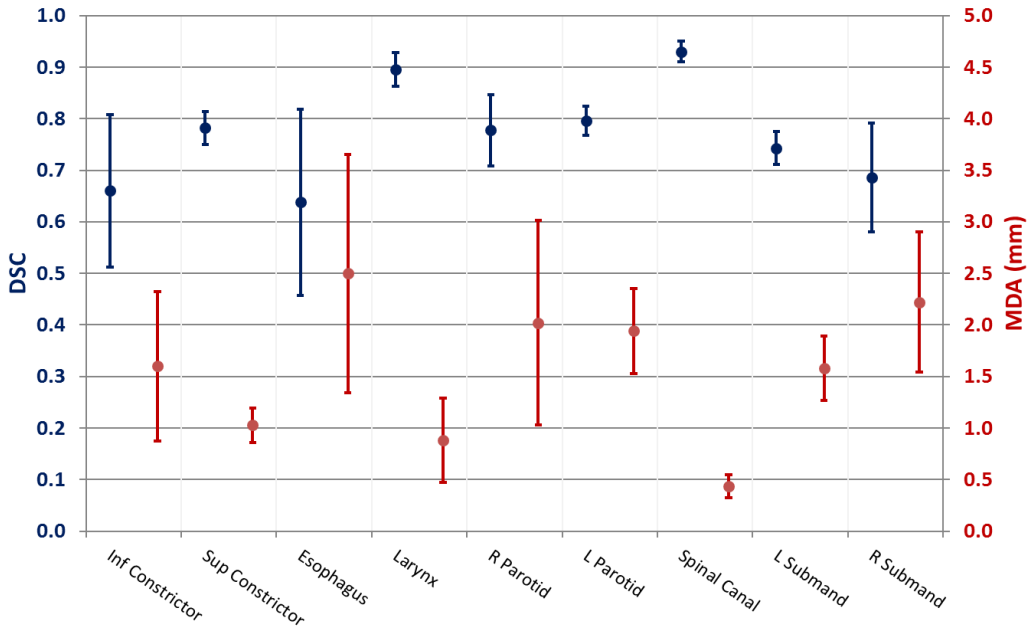


Figure 2.5. Intra-observer reproducibility of contouring normal tissue head and neck structures on CBCTs.

Results from the DIR evaluation are shown in Figure 2.6. The DIR was found to be clinically acceptable for the propagation of spinal canal, larynx and parotid glands (DSC 0.7–0.9) and pharyngeal constrictors (MDA < 2 mm). Although the spinal cord was not directly evaluated, the cervical region of the spinal canal was accurately propagated which suggests that evaluation of accumulated dose to the cord should also be accurate. Evaluation of the esophagus and submandibular glands demonstrated unacceptable DIR performance. However, the esophagus results should be taken in the context of the relatively large manual contour variation. The reduced DIR performance for propagating the submandibular glands is clinically concerning and could be caused by the proximity to the lower jaw, especially in the presence of CBCT photon starvation artifacts caused by metal implants.

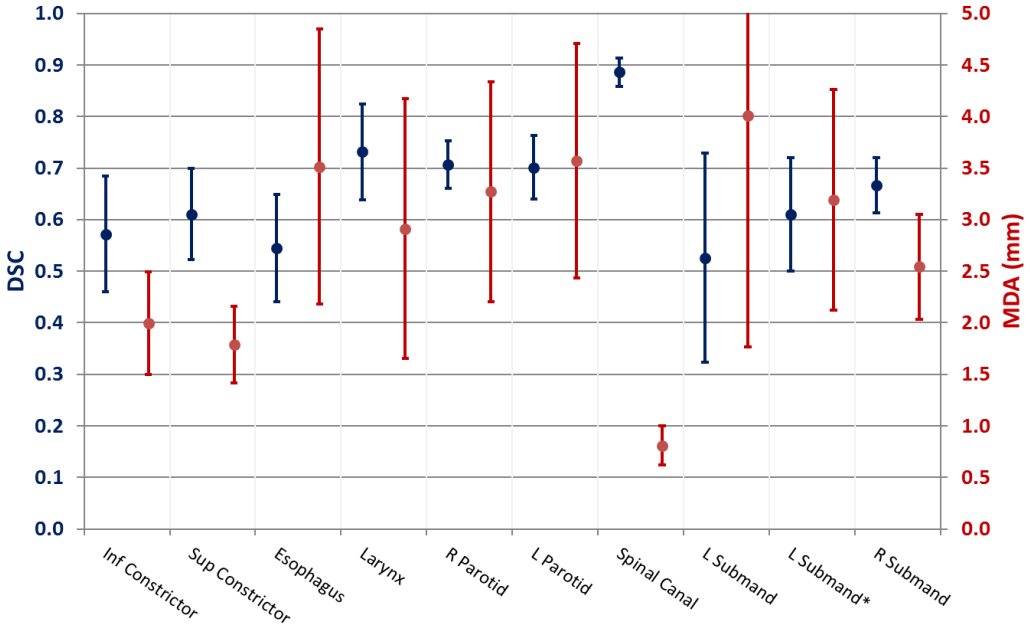


Figure 2.6. DSC and MDA results for DIRs between head and neck planning CTs and CBCTs. The L Submand* structure category excludes one patient with rapidly responding lesion in close proximity to the left submandibular gland which was found to contribute locally to large registration errors.

2.4.2 Lung

4DCT(Inhale) - to - 4DCT(Exhale)

Deformable registration between phases of a 4DCT can be useful in both treatment planning and estimation of functional parameters such as ventilation. A publicly available dataset (DIR-Lab) was used to evaluate this application of the DIR algorithm.[57,58] This dataset has been used throughout the image registration research community and often serves as benchmarking data for evaluation of new DIR algorithms. This dataset includes 10 cases (5 of which have 256x256 reduced field of view 4DCTs) with 300 manually selected landmarks on the extreme phases of the 4DCT. Although these cases offer a range of respiratory motion magnitude (4.0–15.16 mm), it has previously been shown that near voxel level DIR accuracy can be achieved across the cohort.

Initially for this evaluation, the 'Deformable Multi Pass' was compared to the 'Extended Deformable Multi Pass' to determine whether additional multiresolution steps improved the registration quality. These registrations were performed with a manually placed ROI encompassing the entire lung and with the 'Use contrast' setting turned on. The contrast was set at the preset lung value, [−1000, 200 HU]. As the shown in Figure 2.7, 'Deformable Multi Pass' gave an overall TRE vector of 4.5 mm (SD 2.3 mm), while the 'Extended Deformable Multi Pass' resulted in a TRE of 3.7 mm (SD 1.8 mm). The slight accuracy improvement with the extended option is likely a result of the finer resolution B-spline grids allowing for more flexibility in aligning small substructures within the lung.

Despite the improvement in registration accuracy when using the extended option, the average TRE using the preset contrast settings was determined to be clinically unacceptable on the basis of comparison to the known achievable accuracy when using other DIR algorithms. However, initial testing of image contrast setting in the UI demonstrated that the algorithm is

highly sensitive to this input parameter. For example, the results of using a contrast setting of [−1000, −242 HU] are shown in green on Figure 2.7. For these DIRs, the average TRE was found to be 2.2 mm (SD 0.2 mm). This demonstrates a significant improvement, but the error remains larger than clinically acceptable. It should also be noted that the ribs (chest wall) show residual misalignment and non-physical deformations when using these registration contrast settings.

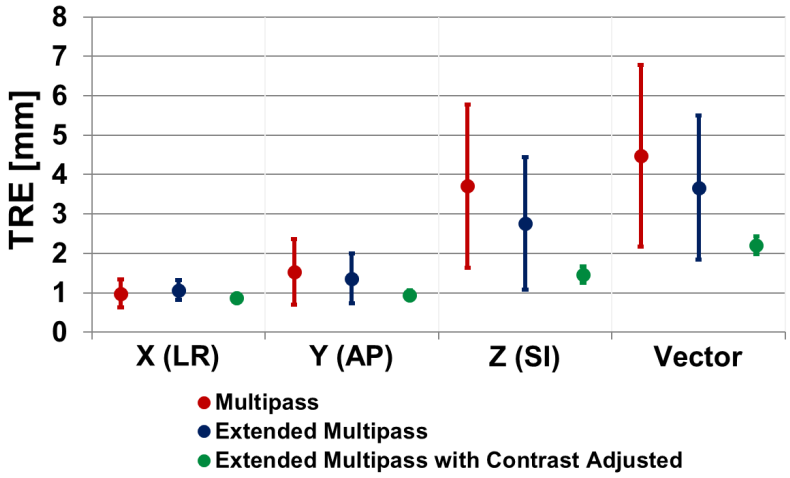


Figure 2.7. Mean population TRE for deformable registration of 4DCT extreme phases.

Planning CT - to - CBCT

Deformable registration of lung planning CTs to CBCTs was evaluated for the purposes of dose accumulation. Ten patients treated for lung cancer at the University of Michigan were selected for evaluation. 30 corresponding landmarks were manually selected on the planning CT and two CBCTs for each patient. A reproducibility study was performed to determine the error in manually selecting these points, and the average error was found to be sub-voxel (0.7 mm TRE vector). Therefore, in this evaluation, the goal for residual error following a deformable alignment was set at the voxel level (1.7 mm TRE vector). Deformations were performed using the 'CBCT Corrected Deformable' option with an ROI cropped to the vertical CBCT boundary and the lateral

lung boundaries. Although these breath-hold scans had limited respiratory deformation, preliminary testing found that multi-pass performed better than the single-pass option. For clinical evaluation, the 'Use Contrast' setting was turned on and the contrast was set at the lung preset. Figure 2.8 shows the TRE results of the deformable registration compared to the rigid registration. The deformable registration performs well, giving a mean TRE of 1.8 mm (SD 0.3 mm) with low patient variability (1.6–2.2 mm). This demonstrates that the Velocity algorithm is robust and provides near voxel level accuracy for this clinical use case.

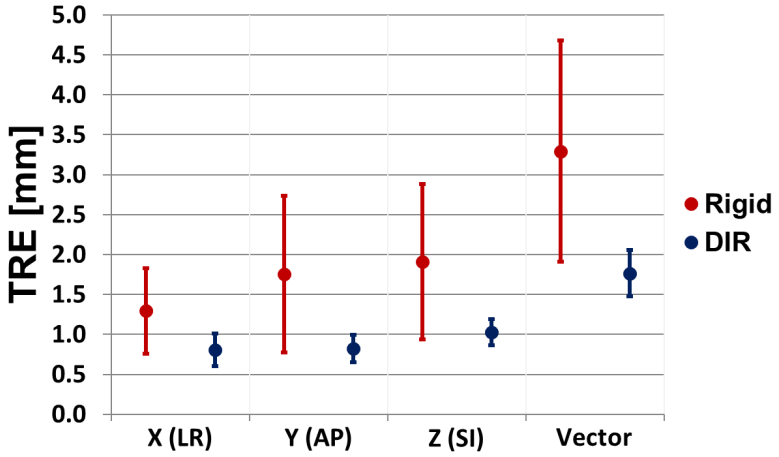


Figure 2.8. Mean population TRE for deformable registration of lung planning CTs to breath-hold CBCTs.

2.4.3 Liver

Planning MR - to - Planning CT

The usage of MR imaging in radiation oncology continues to increase through the acquisition of pre-treatment MRIs to aid in delineating tumors and, more recently, through MR-guided RT and functional avoidance treatment planning. Therefore, the ability to accurately register MR scans to CTs has become an important clinical use case for multimodality DIR algorithms. One common usage of MR imaging is for GTV delineation. Ten liver cancer patients

treated at the University of Michigan were used to evaluate the accuracy of the Velocity DIR algorithm in aligning pre-treatment MR imaging to a patient's treatment planning CT. Although the specific MR scan sequence varied between patients, all scans were contrast enhanced T1-weighted VIBE and selected by the Radiation Oncologist at the time of treatment planning for the ability to clearly delineate the boundary of the GTV. DIR accuracy was evaluated using TRE calculated using 10 manually selected corresponding vessel bifurcations identified on both patient scans. A reproducibility study was performed to evaluate the variability in manually selecting these landmarks. Based on the reselection of 10 points across the patient cohort, the variability was found to be 0.6 mm in-plane and 1.2 mm axially. Since this variability was found to be lower than half the average voxel size of the MR (approximately 2x2x3 mm), the voxel size was used as the limiting factor in evaluating TRE for these patients. Deformable registration utilized the 'MR Corrected Deformable' option and an ROI that encompassed the entire liver vertically and extended to the abdominal walls laterally. The 'Use Contrast' setting was turned off since there is not an intuitive value for contrast settings between MR and CT images.

Figure 2.9 shows the TRE results for this evaluation with a comparison between the Velocity automated rigid registration and deformable registration. The rigid registration resulted in an overall vector TRE of 4.9 mm (SD 1.9 mm) and the use of DIR did not substantially improve accuracy, with a TRE of 4.0 mm (SD 1.1 mm). Only 7 of 10 cases show improvement in TRE when using the deformable registration compared to the automated rigid registration, and only 5 of 10 cases show improvement over the manually-performed, clinically-used rigid registration. It should also be noted that one case showed incorrect rotation following rigid registration and required manual intervention prior to DIR. These results highlight the need for robust optimization of the registration inputs for this important DIR clinical use case.

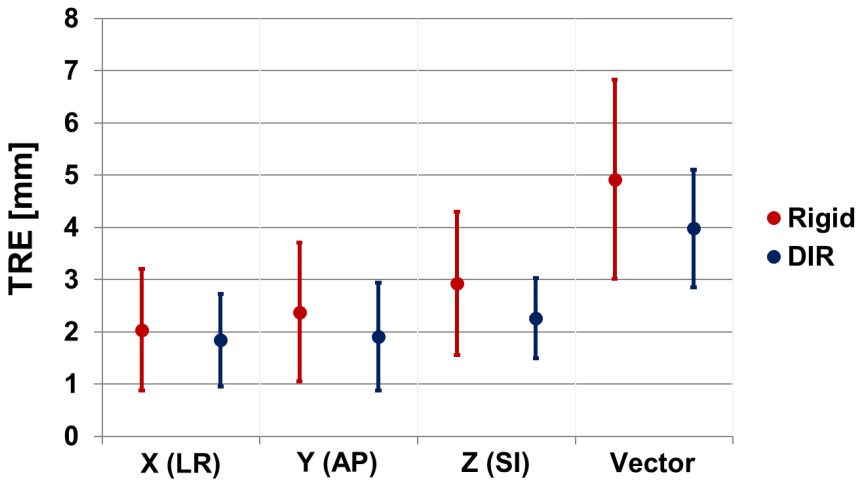


Figure 2.9. Mean population TRE for deformable registration of liver planning CT to planning MR.

2.4.4 Male Pelvis (Prostate)

Planning CT - to - CBCT

Deformable registration of daily CBCTs to planning CTs for prostate cancer treatments could be useful for the evaluation of delivered doses to organs at risk and target structures. Ten prostate cancer patients treated at the University of Michigan were used for this evaluation, with three treatment fractions used per patient. The bladder, prostate, rectum, and proximal femurs were contoured on each CBCT to allow for structure-based evaluation of the DIR. TRE was also used in the evaluation process by manually localizing three implanted RF beacons for each patient. Variability in manually selecting the centroid of the beacons was 0.4 mm in-plane and 0.3 mm axially. The deformable registrations were performed using the 'CBCT Corrected Deformable' with an ROI placed at the boundary of the CBCT so that the ROI position could in replicated automatically in the future.

Figure 2.10 shows the TRE results for registration application. The automated rigid registration performed prior to the deformable results in a vector TRE of 6.8 mm (SD 5.3 mm),

and the DIR gives a TRE of 4.1 mm (SD 3.5 mm). In one case, incorrect rotation within the auto rigid initialization caused a patient-specific TRE of 20 mm. This appears to be caused by a large bladder volume difference and changes in the spatial relationship between soft tissue and bony anatomy. Overall, the large average TRE and standard deviation could be the result of CBCT artifacts around the RF beacons. In the absence of post processing tools used to suppress this artifacts, using the DIR contrast setting may help to improve registration quality.

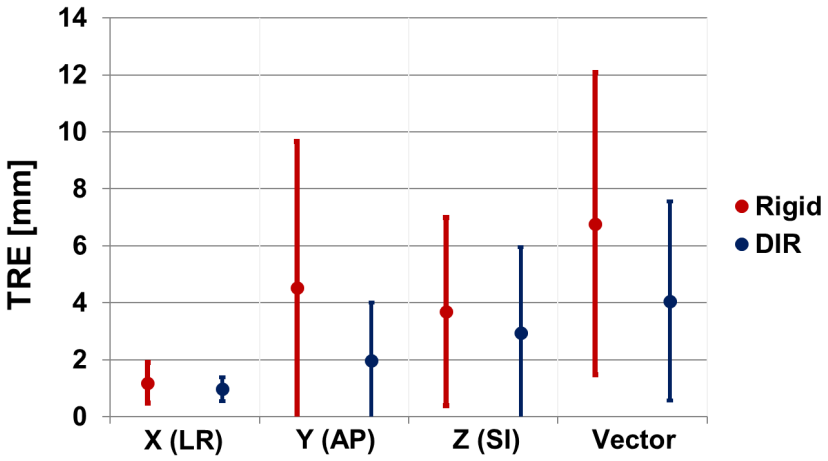


Figure 2.10. Mean population TRE for deformable registration of prostate planning CT to daily CBCT imaging.

Figure 2.11 shows the results of the structured-based evaluation using both DSC and MDA. Structure mapping for the femurs is consistently accurate with the average DSC > 0.90 and MDA < 2 mm. The bladder contour propagation is generally acceptable (DSC > 0.70) but is dependent on bladder volume changes between scans. When the fractional bladder volume change is less than 0.5, clinically acceptable alignment is achieved, but when the fractional bladder volume change is above 0.5, the DIR gives a large range of results (DSC 0.45–0.80). The average prostate alignment is near clinical acceptability (DSC = 0.69), but is also dependent on bladder volume. For cases

with large bladder volumes (> 350 cc), the prostate alignment tends to be worse (DSC 0.50–0.70), and DIR in these cases results in a large TRE range (1.4–8.4 mm). Structure mapping of the rectum was found to be near the clinically acceptable limit with an average DSC of 0.64 (SD 0.06).

Overall, structure mapping in the presence of moderate bladder volume change is clinically acceptable, but the large average implant TRE and inability to accurately map structure boundaries in proximity to the prostate (e.g., anterior rectum wall) demonstrate limitations of the current algorithm inputs for this application.

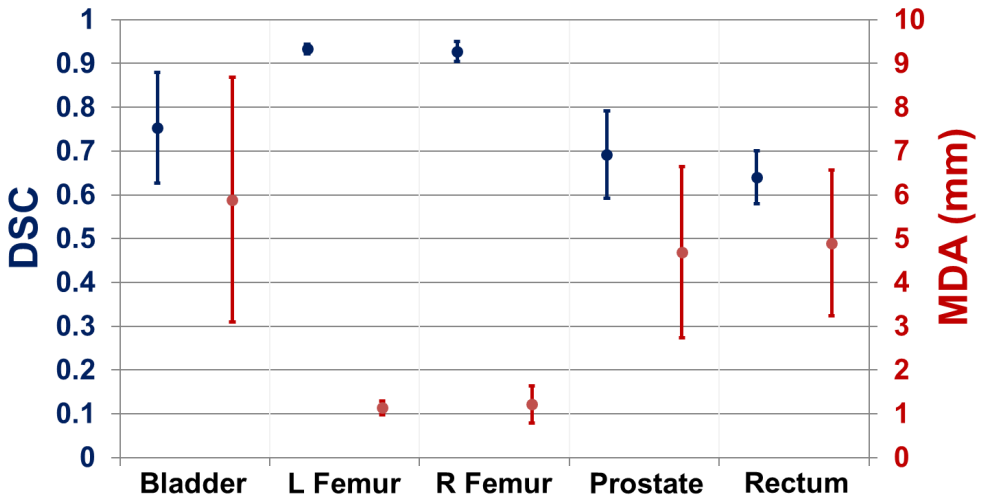


Figure 2.11. DSC and MDA results for DIRs between prostate planning CT to daily CBCT imaging.

2.4.5 Female Pelvis (Cervical Cancer)

Planning MR - to - Planning MR

Treatment planning for multi-fraction brachytherapy treatments, which is the standard of care for cervical cancer, requires significant clinical resources since each fraction requires new contours be drawn for daily plan optimization. Usage of a DIR algorithm to accurately propagate structures between treatment fractions would improve clinical workflow and decrease time

between patient imaging and treatment. To analyze this clinical use case, 10 patients treated at the University of Michigan for cervical cancer were evaluated. For each fraction, a T1-weighted MR scan was acquired and contoured for the purpose of treatment planning. Each patient was treated in five fractions; however, some fractions were excluded from this study on the basis of inadequate bladder draining despite use of a Foley catheter during imaging. In total, 43 fractions were used, with 33 fractions being analyzed for contour propagation since each patient required a single manually contoured fraction. The structure metrics DSC and MDA were used to evaluate the quality of the DIR for the following structures: bladder, sigmoid colon, high risk CTV (HRCTV). Although contouring on the MR images is standard clinical practice, variability was noticed in contouring the rectum and sigmoid. An intra-observer contour variability study showed that re-contouring gives an average DSC for the rectum and sigmoid of above 0.80. The MDA for re-contouring the rectum was 1.8 mm (SD 2.8), but average MDA for the sigmoid was larger at 3.1 mm (SD 5.1). Since the boundary of the rectum and sigmoid was difficult to determine, the union of the two structures (labeled C-S + R) was also analyzed.

Preliminary testing showed that the standard 'MR Corrected Deformable' registration option did not allow for enough deformation in the presence of bowel gas movement and bladder volume changes. Therefore, the 'Deformable Multi Pass' option was used with the MR fade correction manually applied prior to the registration. Although this can lead to highly-variable deformation fields that are unsuitable for voxel-based dose accumulation, it allows for improved structure propagation.

Figure 2.12 shows the results of the structure propagation using the DIR. For the bladder, HRCTV, and rectum, the structure propagation is clinically acceptable (DSC > 0.70). Although the sigmoid falls below the clinically acceptable limit, the union of the sigmoid and rectum is

clinically acceptable. Significant decrease in the HRCTV volume led to poor propagation results for the HRCTV in two cases. Similar to the male pelvis use case, the results are dependent on bladder volume changes between fractions. When the fractional bladder volume change is less than 0.5, clinically acceptable alignment is achieved, but when the fractional bladder volume change is above 0.5, the DIR accuracy is highly variable and case dependent (DSC 0.40–0.90). However, bladder volume change does not correlate well with the accuracy results of other structures. Variability in the results for the sigmoid and rectum appear to be caused by large changes in gas volume and spatial distribution.

Overall, structure mapping is clinically acceptable when using the multi pass deformable registration and manually applied MR fade correction. As previously noted, voxel-based dose accumulation would require further investigation when using this registration option

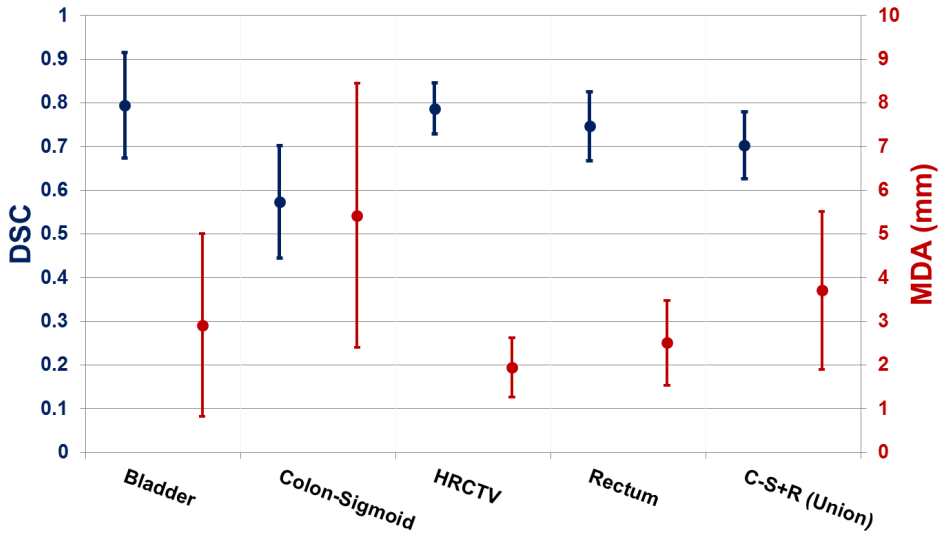


Figure 2.12. DSC and MDA results for DIRs between female pelvis MR imaging.

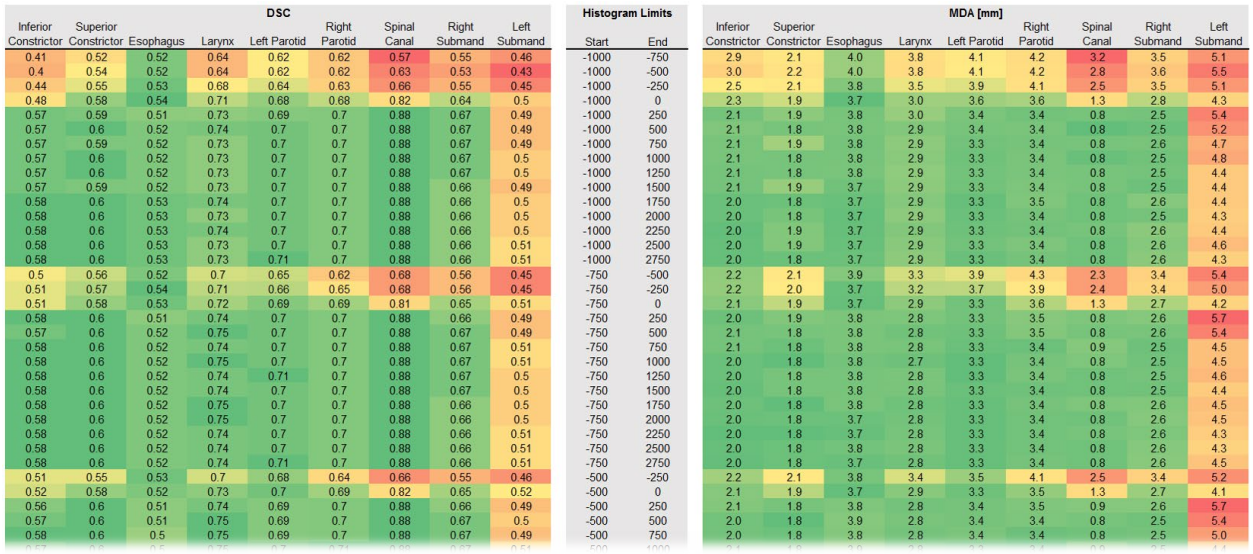
2.4.6 Image Histogram Limits

Image histogram limits are the minimum and maximum intensity values used when binning intensities together for use in the mutual information metric calculation. These limits, which are individually adjustable for each image volume, set the levels of the first and last bins within the intensity binning (down sampling) process. Information beyond these cutoffs is not discarded but instead grouped into two bins above and below the levels. By default, the algorithm uses the image minimum and maximum intensities for the histogram limits. Therefore, we refer to histogram limits as not being enabled by default since they are not user-controlled, although they are inherently present.

Although these levels can be individually adjusted for each image volume in a registration, this study used the same limits between volumes given that only CT to CT and CT to CBCT registrations were performed. This means that two, instead of four, parameters were varied across the registrations. A linear search of the parameter space was performed in increments of 250 starting at level of −1000.

For head and neck registrations, the histogram limits were varied using the preset deformable multipass registration parameters as a basis for other parameters. DSC and MDA were visualized as heat maps across all previously noted structures. Figure 2.13 shows an example heatmap produced for analysis of these registrations. Each row represents a specific set of registration parameters executed over the patient cohort with the population mean DSC and MDA reported. Overall, no significant improvement was obtained by adding and adjusting the histogram limits. No individual histogram limit pair produced the best registration accuracy across all of the analyzed structures. More specifically, although registration results varied across the multiple structures, structures near bony anatomy showed differing trends than those surrounded mainly by

soft tissue. It should be noted that the behavior of the registration metrics for the left submandibular gland is driven, in part, by a patient with a rapidly responding lesion.



DSC								
Inferior Constrictor	Superior Constrictor	Esophagus	Larynx	Left Parotid	Right Parotid	Spinal Canal	Right Submand	Left Submand
0.41	0.52	0.52	0.64	0.62	0.62	0.57	0.55	0.46
0.4	0.54	0.52	0.64	0.62	0.62	0.63	0.53	0.43
0.44	0.55	0.53	0.68	0.64	0.63	0.66	0.55	0.45
0.48	0.58	0.54	0.71	0.68	0.68	0.82	0.64	0.5
0.57	0.59	0.51	0.73	0.69	0.7	0.88	0.67	0.49
0.57	0.6	0.52	0.74	0.7	0.7	0.88	0.67	0.49
0.57	0.59	0.52	0.73	0.7	0.7	0.88	0.67	0.49
0.57	0.6	0.52	0.73	0.7	0.7	0.88	0.67	0.5
0.57	0.6	0.52	0.73	0.7	0.7	0.88	0.67	0.5
0.57	0.59	0.52	0.73	0.7	0.7	0.88	0.66	0.49
0.58	0.6	0.53	0.74	0.7	0.7	0.88	0.66	0.5
0.58	0.6	0.53	0.73	0.7	0.7	0.88	0.66	0.5
0.58	0.6	0.53	0.74	0.7	0.7	0.88	0.66	0.5
0.58	0.6	0.53	0.73	0.7	0.7	0.88	0.66	0.51
0.58	0.6	0.53	0.73	0.71	0.7	0.88	0.66	0.51
0.5	0.56	0.52	0.7	0.65	0.62	0.68	0.56	0.45
0.51	0.57	0.54	0.71	0.66	0.65	0.68	0.56	0.45
0.51	0.58	0.53	0.72	0.69	0.69	0.81	0.65	0.51
0.58	0.6	0.51	0.74	0.7	0.7	0.88	0.66	0.49
0.57	0.6	0.52	0.75	0.7	0.7	0.88	0.67	0.49
0.58	0.6	0.52	0.74	0.7	0.7	0.88	0.67	0.51
0.58	0.6	0.52	0.75	0.7	0.7	0.88	0.67	0.51
0.58	0.6	0.52	0.74	0.71	0.7	0.88	0.67	0.5
0.58	0.6	0.52	0.74	0.7	0.7	0.88	0.67	0.5
0.58	0.6	0.52	0.75	0.7	0.7	0.88	0.66	0.5
0.58	0.6	0.52	0.75	0.7	0.7	0.88	0.66	0.5
0.58	0.6	0.52	0.74	0.7	0.7	0.88	0.66	0.51
0.58	0.6	0.52	0.74	0.7	0.7	0.88	0.66	0.51
0.58	0.6	0.52	0.74	0.71	0.7	0.88	0.66	0.51
0.51	0.55	0.53	0.7	0.68	0.64	0.66	0.55	0.46
0.52	0.58	0.52	0.73	0.7	0.69	0.82	0.65	0.52
0.56	0.6	0.51	0.74	0.69	0.7	0.88	0.66	0.49
0.57	0.6	0.51	0.75	0.69	0.7	0.88	0.67	0.5
0.58	0.6	0.5	0.75	0.69	0.7	0.88	0.67	0.49

Histogram Limits	
Start	End
-1000	-750
-1000	-500
-1000	-250
-1000	0
-1000	250
-1000	500
-1000	750
-1000	1000
-1000	1250
-1000	1500
-1000	1750
-1000	2000
-1000	2250
-1000	2500
-1000	2750
-750	-500
-750	-250
-750	0
-750	250
-750	500
-750	750
-750	1000
-750	1250
-750	1500
-750	1750
-750	2000
-750	2250
-750	2500
-750	2750
-500	-250
-500	0
-500	250
-500	500
-500	750

MDA [mm]								
Inferior Constrictor	Superior Constrictor	Esophagus	Larynx	Left Parotid	Right Parotid	Spinal Canal	Right Submand	Left Submand
2.9	2.1	4.0	3.8	4.1	4.2	3.2	3.5	5.1
3.0	2.2	4.0	3.8	4.1	4.2	2.8	3.6	5.5
2.5	2.1	3.8	3.5	3.9	4.1	2.5	3.5	5.1
2.3	1.9	3.7	3.0	3.6	3.6	1.3	2.8	4.3
2.1	1.9	3.8	3.0	3.4	3.4	0.8	2.5	5.4
2.1	1.8	3.8	2.9	3.4	3.4	0.8	2.5	5.2
2.1	1.9	3.8	2.9	3.3	3.4	0.8	2.6	4.7
2.1	1.8	3.8	2.9	3.3	3.4	0.8	2.5	4.8
2.1	1.8	3.8	2.9	3.3	3.4	0.8	2.5	4.4
2.1	1.9	3.7	2.9	3.3	3.4	0.8	2.5	4.4
2.0	1.8	3.7	2.9	3.3	3.5	0.8	2.6	4.4
2.0	1.8	3.7	2.9	3.3	3.4	0.8	2.5	4.3
2.0	1.9	3.7	2.9	3.3	3.4	0.8	2.6	4.4
2.0	1.9	3.7	2.9	3.3	3.4	0.8	2.6	4.6
2.0	1.8	3.7	2.9	3.3	3.4	0.8	2.6	4.3
2.2	2.1	3.9	3.3	3.9	4.3	2.3	3.4	5.4
2.2	2.0	3.7	3.2	3.7	3.9	2.4	3.4	5.0
2.1	1.9	3.7	2.9	3.3	3.6	1.3	2.7	4.2
2.0	1.9	3.8	2.8	3.3	3.5	0.8	2.6	5.7
2.1	1.8	3.8	2.8	3.3	3.5	0.8	2.6	5.4
2.1	1.8	3.8	2.8	3.3	3.4	0.9	2.5	4.5
2.0	1.8	3.8	2.7	3.3	3.4	0.8	2.5	4.5
2.0	1.8	3.8	2.8	3.3	3.4	0.8	2.5	4.6
2.0	1.8	3.8	2.8	3.3	3.4	0.8	2.5	4.4
2.0	1.8	3.8	2.8	3.3	3.4	0.8	2.6	4.5
2.0	1.8	3.7	2.8	3.3	3.4	0.8	2.6	4.5
2.0	1.8	3.7	2.8	3.3	3.4	0.8	2.6	4.3
2.0	1.8	3.8	2.8	3.3	3.4	0.8	2.6	4.3
2.0	1.8	3.7	2.8	3.3	3.4	0.8	2.6	4.5
2.2	2.1	3.8	3.4	3.5	4.1	2.5	3.4	5.2
2.1	1.9	3.7	2.9	3.3	3.5	1.3	2.7	4.1
2.1	1.8	3.8	2.8	3.4	3.5	0.9	2.6	5.7
2.0	1.8	3.9	2.8	3.4	3.4	0.8	2.5	5.4
2.0	1.8	3.8	2.8	3.4	3.4	0.8	2.5	5.0

Figure 2.13. Example heat map of mean population DSC (left) and MDA (right) results for head and neck CT to CBCT registration with varying histogram limits (center). Heat scale represents worse (red) to better (green) registration accuracy.

Deformable multipass registration parameters were also used as a basis for analyzing lung registrations. TRE was visualized as a 3D plot of TRE with varying histogram limits on the x-y plane. Figure 2.14 shows the TRE across all 4DCT registrations. A clear minimum mean population TRE of 1.94 mm is found at histogram limits of −1000, −250. The behavior of the parameter space for these registrations is relatively smooth compared to that of CT to CBCT registrations, shown in Figure 2.15. For CT to CBCT registration, the minimum mean population TRE is less pronounced, with histogram limits from −1000, −250 to −1000, +2000 resulting in similar TRE. Similar to the 4DCT, the minimum mean TRE of 1.71 mm was found at limits of −1000, −250. Overall, lung registrations are extremely sensitive to the histogram limits and using

the limits of −1000, −250 results in registrations with near voxel level accuracy for both lung registration use cases.

Overall, adjusting the intensity limits has a significant impact on registrations with clear material/tissue type intensity distinctions, such as in the lung registrations. However, in registrations with less intensity-distinguished structures, adjusting the histogram limits is less helpful in achieving uniform accuracy improvement across multiple structures. In these situations, implementing the ability to discard information above or below the cutoff limits could help, particularly for soft tissue structures in close proximity to boney anatomy or air.

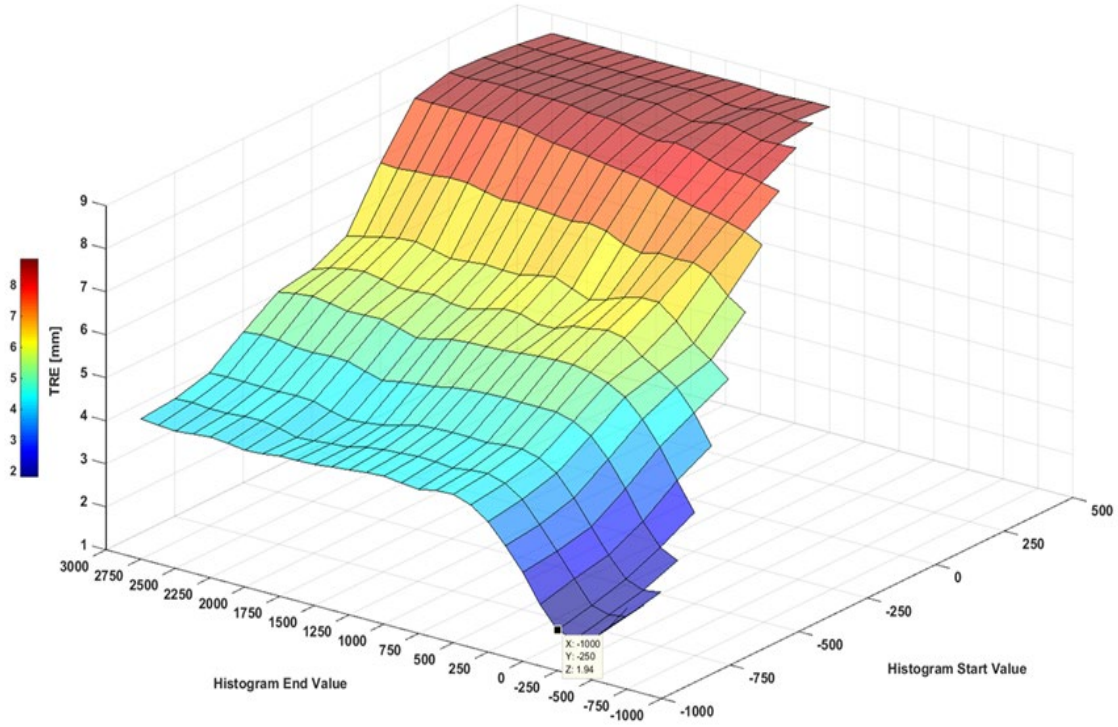


Figure 2.14. 3D visualization of mean population TRE results of lung 4DCT registration. A minimum TRE of 1.9 mm is achieved at a histogram minimum value of −1000 and maximum value of −250.

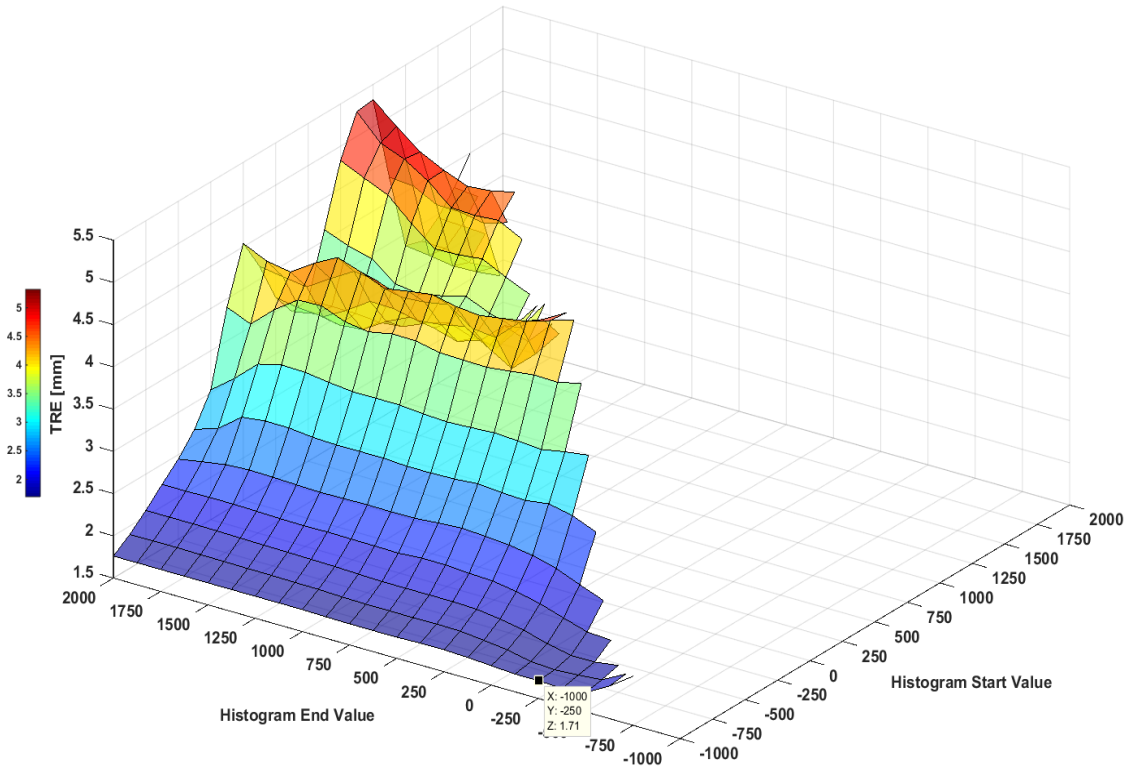


Figure 2.15. 3D visualization of mean population TRE results of lung CT to CBCT registration. A minimum TRE of 1.7 mm is achieved at a histogram minimum value of −1000 and maximum value of −250.

2.4.7 Number of Histogram Bins

The number of histogram bins defines how much the image intensity is down sampled prior to computing the mutual information metric. In general, too few bins can result in a registration that is unable to capture the relationship between the two image volumes; whereas too many bins can result in increased entropy within the joint histogram which could inhibit the mutual information metric from being accurately estimated. Velocity's default value for this parameter is 50 bins, with a minimum of 5 bins, as required by using Mattes mutual information. As previously noted, these histogram bins are distributed uniformly between the user defined histogram limits. Although not tested in this study, a separate number of histogram bins can be used for each B-spline resolution level.

For head and neck registrations, the deformable multipass registration settings were used with a varying number of histogram bins from 5 to 75. Figure 2.16 shows the mean population DSC and MDA for CT to CT registrations (CT to CBCT registrations behaved similarly). Highly variable registration results were found when less than 10 bins were used, with gradually improving registration accuracy between 10 and 20 bins. Beyond 20 bins, the registration metrics became increasingly stable, with the default of 50 bins residing in this stable region.

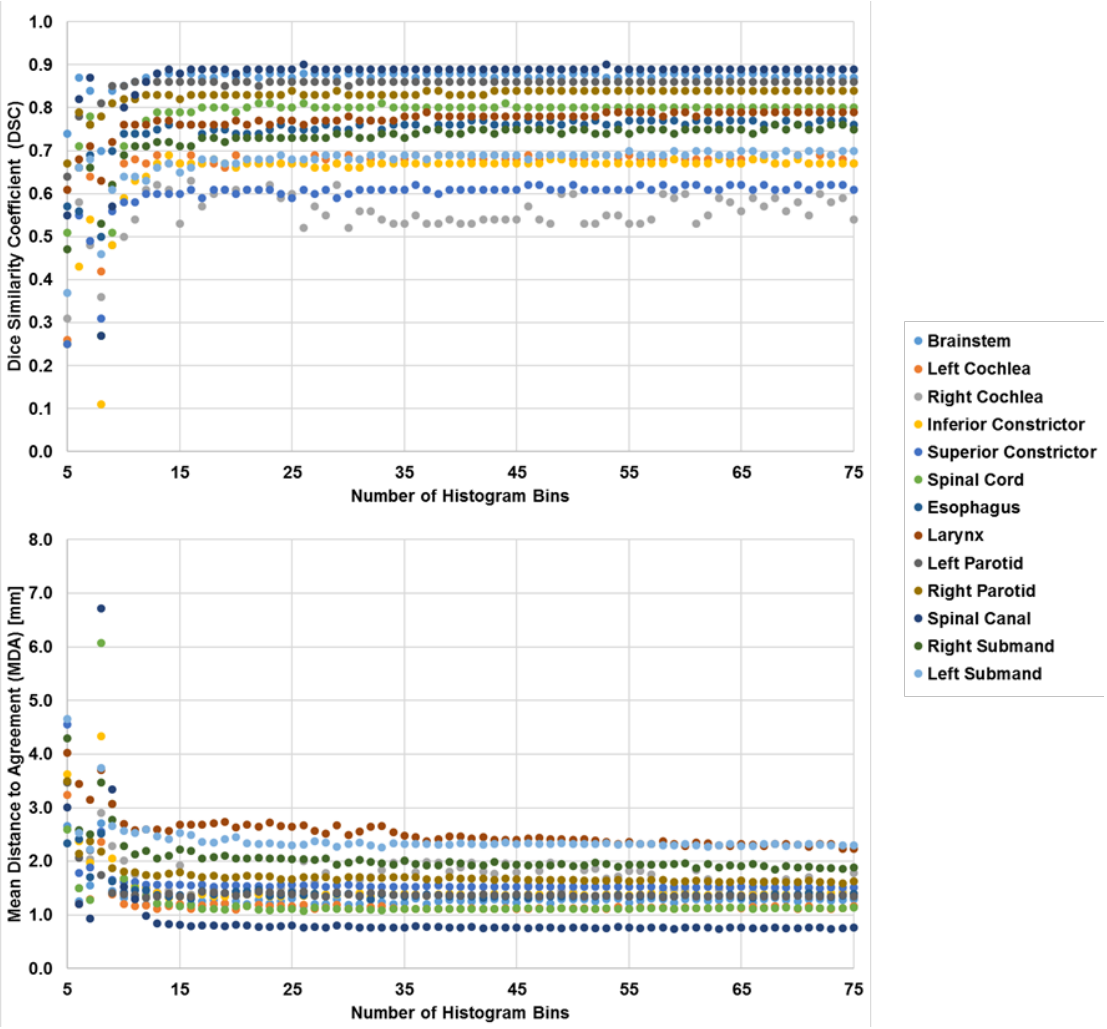


Figure 2.16. Mean population DSC and MDA of head and neck CT to CT registrations with varying number of histogram bins. The default value of 50 bins is located in the maximum DSC and minimum MDA plateau of the parameter space.

Analysis of lung registrations used the extended deformable multipass settings with the same range of bins, 5 to 75. Figure 2.17 shows that for lung 4DCT registrations, accuracy drastically improves from 5 to 10 bins and becomes stable after 20 bins. Once again, the default of 50 bins resided in this stable region. Lung and CT to CBCT registrations had similar results and demonstrated that only slight and variable improvement in accuracy can be achieved by varying the number of histogram bins from the default value. Additionally, reducing the number of histogram bins was found to have no appreciable impact on registration speed.

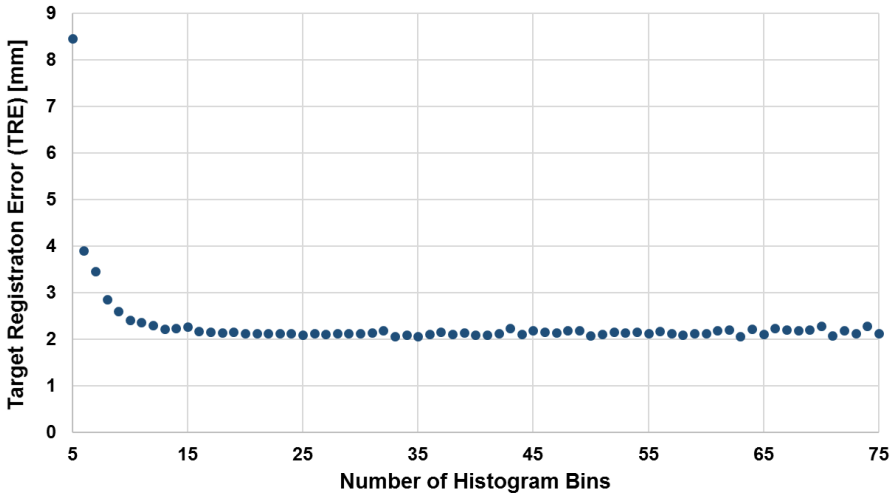


Figure 2.17. Mean population TRE results of lung 4DCT registration with varying number of histogram bins. The default value of 50 bins is located in the minimum TRE plateau of the parameter space.

In this analysis, the DIR algorithm does not appear to be sensitive to the number of histogram bins within the range surrounding the default value of 50. Decreasing the number of bins below 15 should be avoided, and appropriate values for the parameter appear to range from 35 to 75 (the maximum value tested). In registration applications involving MR image volumes, the number of histogram bins should be retested since the mutual information metric is likely to be noisier than the CT to CT and CT to CBCT registrations tested in this study. In the future, non-

uniform bin spacing may be worth investigating, although this may conflict with the method used to estimate the mutual information metric.

2.4.8 Sampling Rate

The sampling rate parameter controls the number of spatial samples randomly selected and used in the mutual information metric calculation. Generally, as the sampling rate is increased, registration results become more accurate at the cost of increased registration computation effort. Therefore, in an unsupervised registration environment without timing pressure, the sampling rate should be set at the lowest value that does not reduce registration accuracy. However, in a supervised environment, registration execution time may become a more important trade-off with registration accuracy. By default, Velocity uses a sampling rate of 20% across all registration options. In this study, the sampling rate was varied from 10% to 100% in increments of 10% to determine the effect on registration accuracy and computation time. All other registration parameters were set at the default values. For registrations involving a CBCT image volume, the CBCT preprocessing filter was applied.

Registration of head and neck CT to CBCT and CT to CT image volumes were performed using the Deformable Multipass preset parameters. Figure 2.18 shows the results of varying the sampling rate on DSC and MDA for head and neck CT to CBCT registrations. Marginal variation in DSC and minimal improvement in MDA were found with increasing sampling rate. A slight decrease in the variation in registration results across the patient cohort was also noted. Similar results were obtained for head and neck CT to CT registrations. Average registration computation time, from a sampling rate of 20% to 100%, increased 22% for CT to CBCT registrations and 47% for CT to CT registrations.

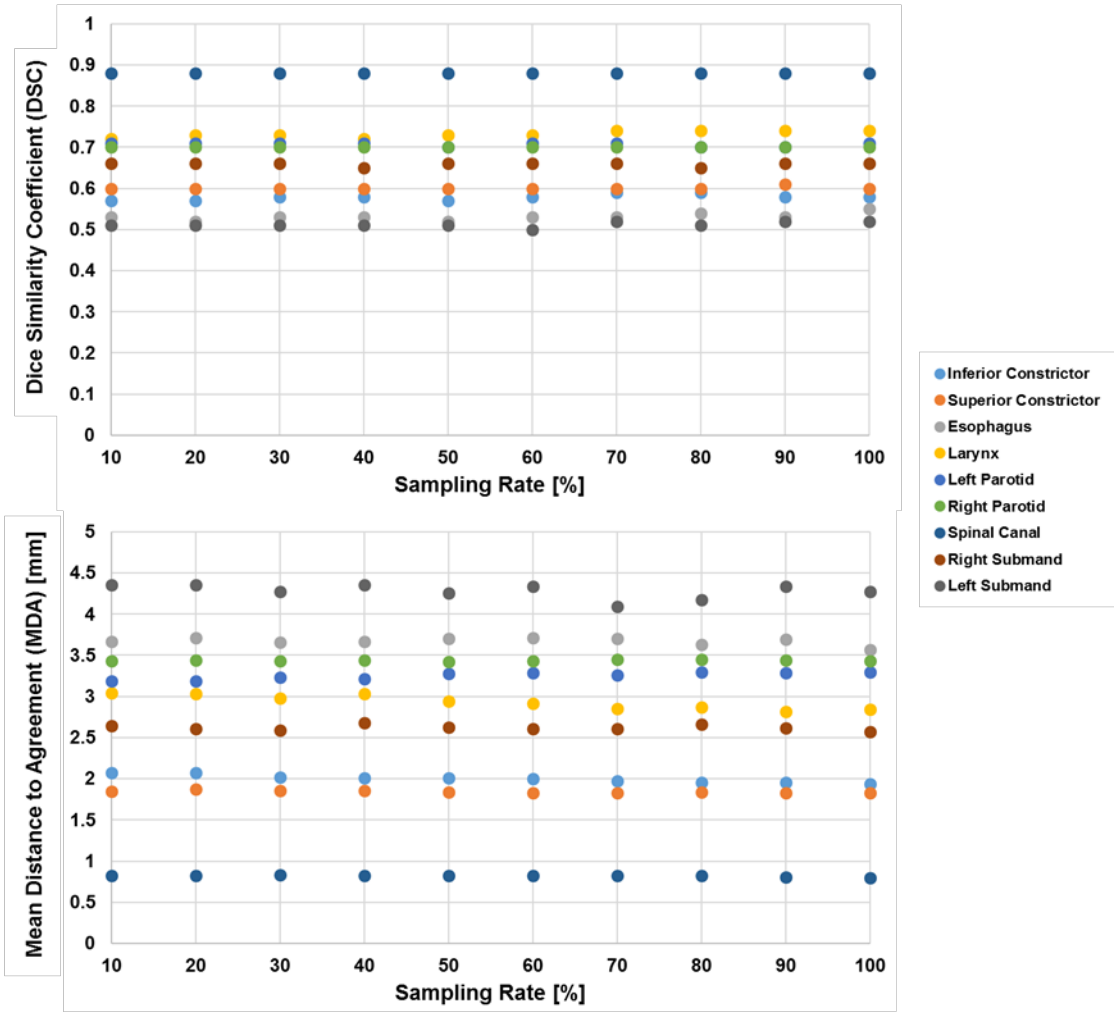


Figure 2.18. Mean population DSC and MDA of head and neck CT to CBCT with varying sampling rate. Minimal improvement in registration accuracy is noted at higher sampling rates.

For both lung CT to CBCT and 4DCT, the Extended Deformable Multipass preset parameters were used. Figure 2.19 shows the results of varying the sampling rate on lung TRE and relative computation time, normalized to 50% sampling, for both lung CT to CBCT and 4DCT. For CT to CBCT registration, a 11% improvement in TRE can be achieved by increasing the sampling rate to 100%, but the computation time almost doubles. For 4DCT registrations, a 19% improvement in TRE was found when increasing the sampling rate to 100%, but this corresponds

to a tenfold increase in computation time. For both registration cohorts, variation in the individual patient TRE was decreased with increasing sampling rate.

Based on this analysis, when using this DIR algorithm in an unsupervised registration environment, the sampling rate should be increased from 20% to 100%. Although this increases computational expense, it will lead to a more robust and accurate registration solution, particularly for registrations focused on aligning small structures. Additionally, in a dose accumulation application requiring voxel to voxel accuracy, rather than just contour matching accuracy, including each voxel in the mutual information metric should improve registration confidence at the substructure level.

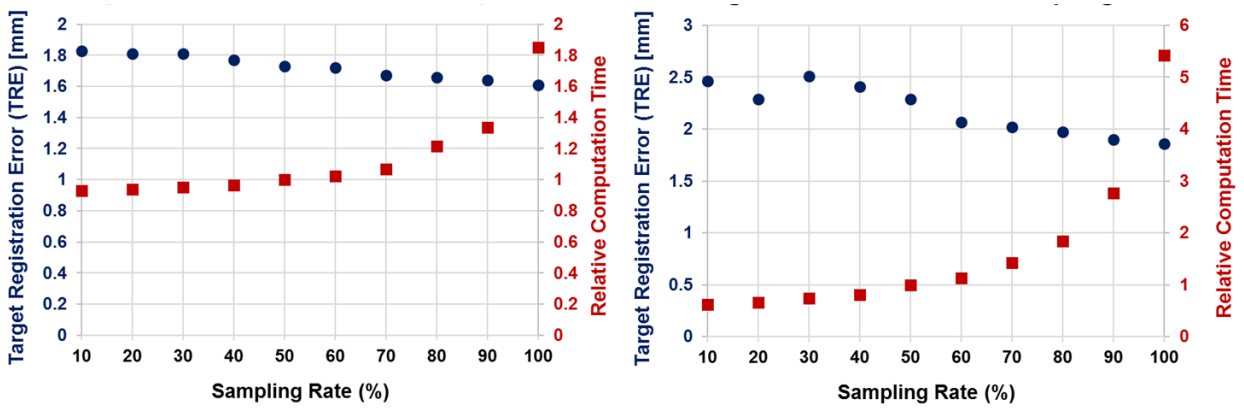


Figure 2.19. Mean population TRE and relative registration execution time for deformable registration of lung CT to CBCT (left) and 4DCT (right) with varying sampling rate. A slight improvement in TRE can be noted with increased computation time at higher sampling rates.

2.4.9 Number of Multiresolution Levels and Grid Spacing

Multiresolution levels and variable grid spacing parameters function to help the deformable registration avoid local maxima in the mutual information metric. Typically, hierarchical deformations are performed from coarse to fine grid spacing to aid in finding the best possible overall registration. In the Velocity UI, there are three preset values for this: Deformable or single

pass (1 resolution level), Deformable Multipass (3 resolution levels), and Extended Deformable Multipass (5 resolution levels). For each resolution level in these presets, the number of grid points in each direction is the same fixed value regardless of the physical size of the ROI. The number of grid points for each of the preset's resolution levels is (7), (5, 10, 15), and (5, 10, 15, 20, 25), respectively. Given the multitude of variability available for these parameters, we initially compared the three preset values and then varied the grid spacing in one of the preset options.

Comparison of the three presets was performed for the head and neck CT to CBCT, lung CT to CBCT, and lung 4DCT datasets, using histogram limits of −1000 and +2750 for head and neck cases and −1000 and −250 for the lung cases. Figure 2.20 shows the head and neck CT to CBCT comparison with an additional test of turning off the CBCT pre-processing filter when using the deformable multipass preset. Minimal variation in the evaluation metrics DSC and MDA was found between the individual presets, although the deformable Multipass had a slightly better overall performance. The CBCT pre-processing filter did not have a large effect on most structures with exception of the left submandibular gland in which improved results were found when not applying the pre-processing. For both lung datasets, deformable Multipass performed the best with the lowest population mean TRE and variation in the population mean. For lung CT to CBCT mean population TREs were 1.8 ± 0.9 mm, 1.7 ± 0.9 mm, 1.9 ± 1.0 m, and for lung 4DCT the were 2.8 ± 2.4 mm, 1.9 ± 1.5 mm, 2.1 ± 1.6 mm, respectively for each resolution preset.

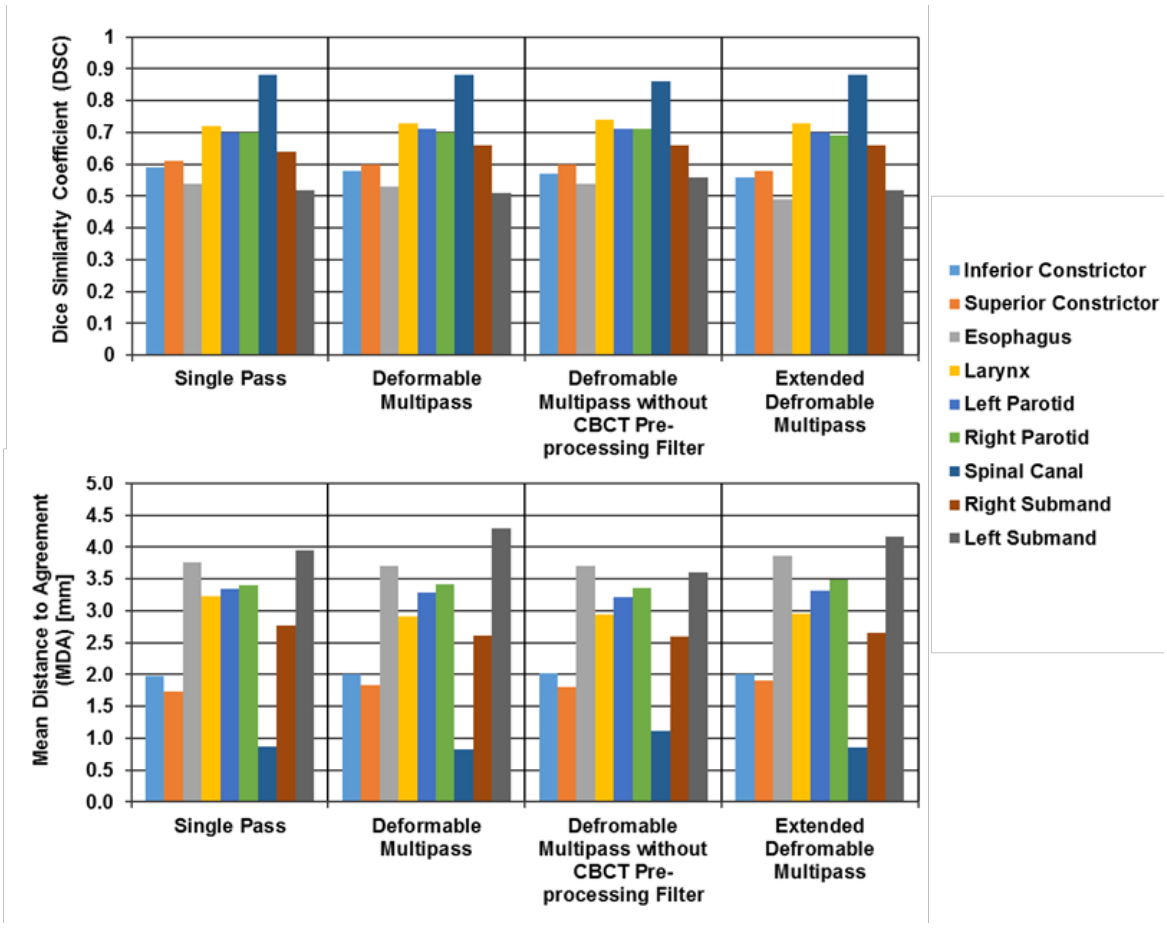


Figure 2.20. Mean population DSC and MDA of head and neck CT to CBCT with the three registration presets. Minimal variation in the metrics was noted, although Deformable Multipass had the best overall performance.

The grid spacing parameter was tested using the head and neck CT to CBCT dataset. Spacing was varied from x = 1 to x = 10 using the same multiresolution grid scheme of deformable multipass preset. This meant that the number of grid points at each level was the same in each direction and set as (1x, 2x, 3x), giving multiresolution deformations with the number of grid points ranging from (1, 2, 3) to (10, 20, 30). Figure 2.21 shows the results of varying this parameter. Using this multiresolution scheme, a grid spacing of (3, 6, 9) gave the best overall result. However, changes in the metrics for individual structures did not demonstrate a consistent trend.

Overall, when comparing the multiresolution presets, switching from a single resolution to hierarchal multi-resolution registration scheme did not significantly affect the registration

accuracy for either lung or head and neck applications. However, given the slight improvement in accuracy and robustness of hierarchal multi-resolution registrations, the option utilizing three resolution levels should likely be used. Deviating from the default grid density of (5, 10, 15) is unlikely to result in improved registration accuracy but future analysis should be performed on adjusting the grid density based on the physical size of the ROI.

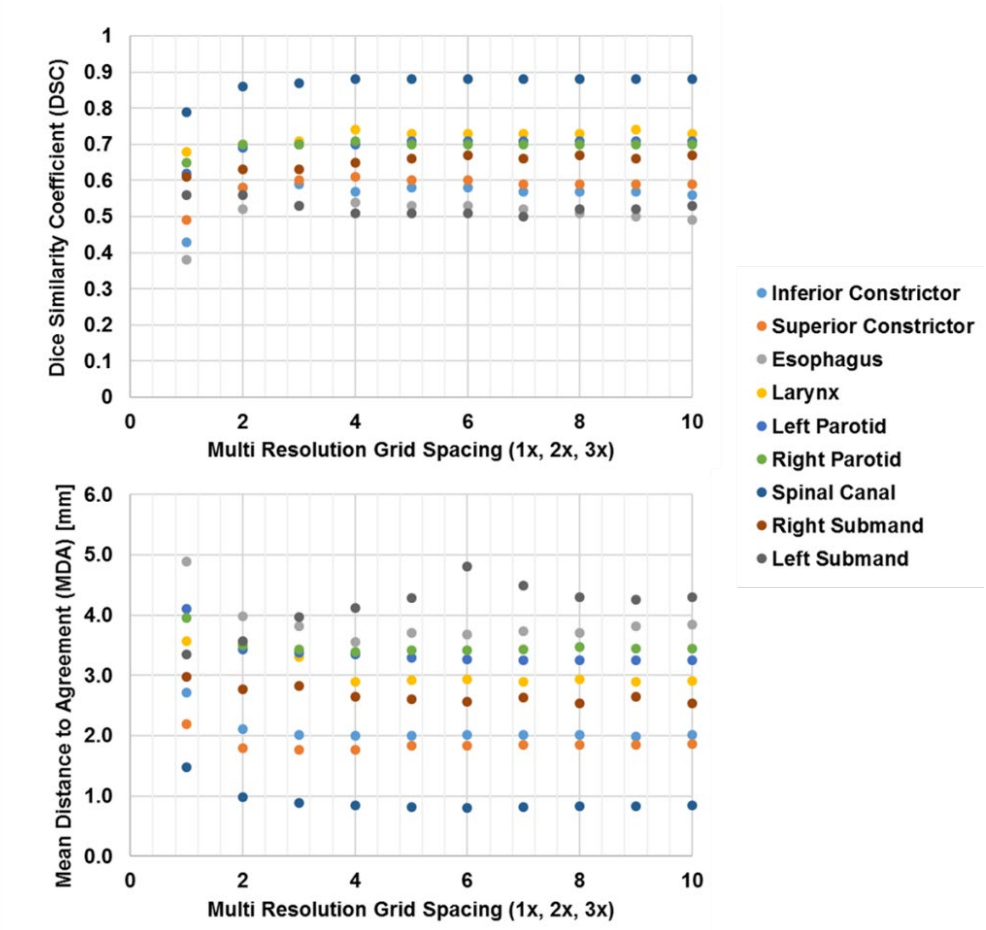


Figure 2.21. Mean population DSC and MDA of head and neck CT to CBCT with varying grid spacing. Results of changing the spacing depend on the evaluated structure, but in this multiresolution scheme, a spacing of (3,6,9) gave the best overall registration result.

2.4.10 Topological Regularization Constraint

Deformation constraints can be employed within DIR algorithms to limit extreme deformations and effectively smooth the deformation field. Commonly used constraints include the Jacobian determinant, which is used as a penalization term within the registration optimization process. The goal of incorporating this constraint is to create a more robust registration algorithm by limiting local volume change and image inversion/folding. Velocity has implemented a constraint into their algorithm that performs similarly to a Jacobian penalty without the significant computational cost of calculating the Jacobian determinate values across the vector field.[59] By default, this method is not used in the deformable registration process; however, the API allows activation of this method with user-adjustable constraint parameters.

This feature was tested using the head and neck CT to CBCT registrations with distance-limiting coefficients ranging from 0 to 0.95 in increments of 0.05. This tunable regularization parameter controls the penalization of differences between two adjacent deformation coefficients. At a value of 0, this corresponds to a quadratic roughness penalty in which all differences are penalized. As the value is increased, non-penalized (or allowable) differences between coefficients are also increased, and therefore, the deformation field roughness is penalized less. Although the coefficient could be changed independently in each orthogonal direction, the coefficient was only adjusted isotopically in this study. Figure 2.22 show the results of applying the regularization penalty with all other registration parameters matching the Deformable Multipass preset. Using a distance limiting coefficient of above 0.65 results in registrations that do not substantially differ from registrations without the regularization. Degradation of registration metrics is noted when applying regularization with a distance limiting coefficient below 0.10. However, employing the regularization penalty with a distance limiting coefficient between 0.25 and 0.40 results in similar

or improved mean registration metrics for most structures, particularly for the left submandibular gland which is affected largely by a single patient in the cohort that presented with a rapidly responding lesion. Additionally, variation in registration results across the patient cohort are noticeably decreased. This suggests that utilizing the regularization penalty with a distance limiting coefficient near 0.35 could result in slightly improved registrations. As a potentially more important consideration, this could result in a more robust and stable registration algorithm, which would be beneficial when deployed in an unsupervised registration process.

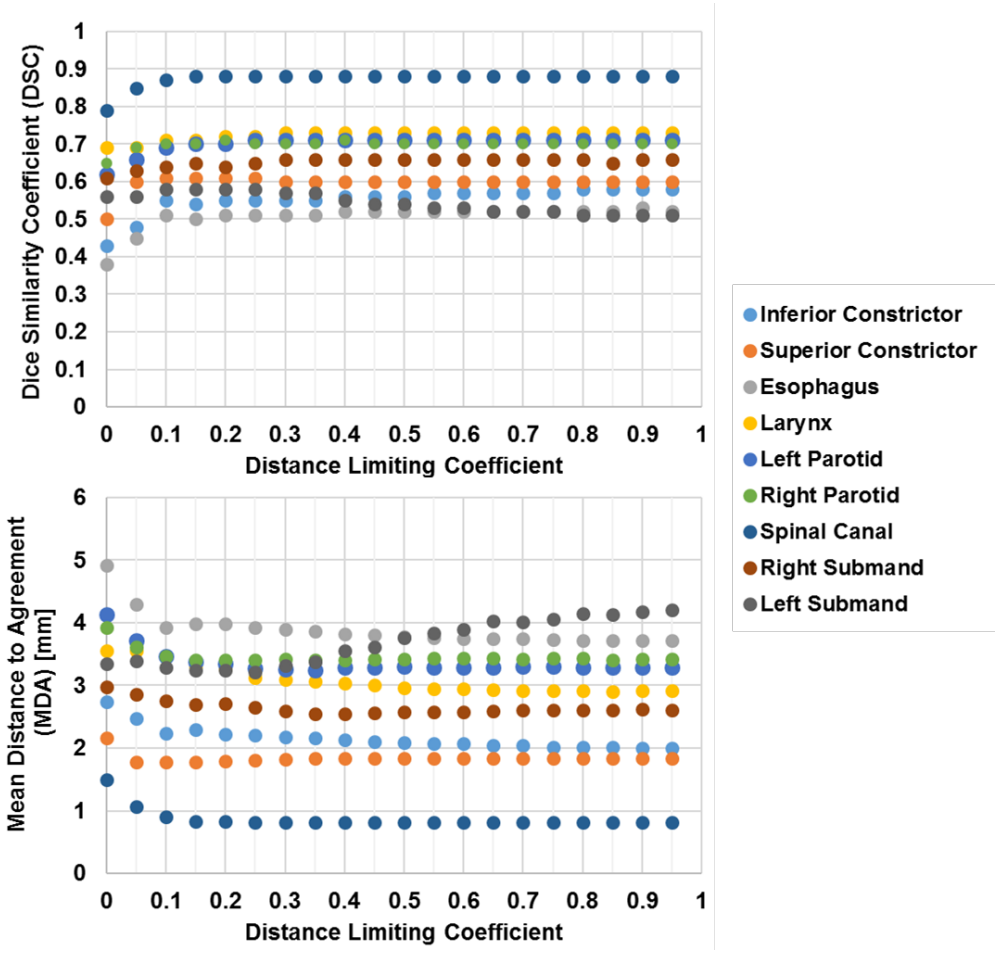


Figure 2.22. Mean population DSC and MDA of head and neck CT to CBCT with the topological regularization penalty applied using varying distance limiting coefficients. Improved registration metrics are noted for most structures when using a distance limiting coefficient between 0.25 and 0.4.

2.3.11 Optimization Process Parameters

A limited analysis of the optimization parameters was performed using the head and neck CT to CBCT cases. Generally, these parameters affect criteria for initializing and terminating the deformation process. This analysis included independently varying the maximum number of optimizer iterations, maximum number of consecutive optimizer attempts, minimum step length, and maximum step length using the parameters from the deformable multipass preset as a basis. The parameters, metric value percentage difference, relaxation factor, and gradient magnitude tolerance, were not evaluated in this study.

The maximum number of optimizer iterations affects the algorithm's ability achieve convergence at the expense of computation time. For the deformable multipass preset option, this is set at a value of 30 iterations for each resolution. In this study, the maximum number of optimizer iterations was varied from 10 to 100 in increments of 10 and set at the same value for each resolution level within a registration. Figure 2.23 shows the results for varying this parameter, demonstrating that an increased number of allowable iterations does not necessarily correlate with improved accuracy. Overall, no noticeable trend was found when varying this parameter below or above the default of 30.

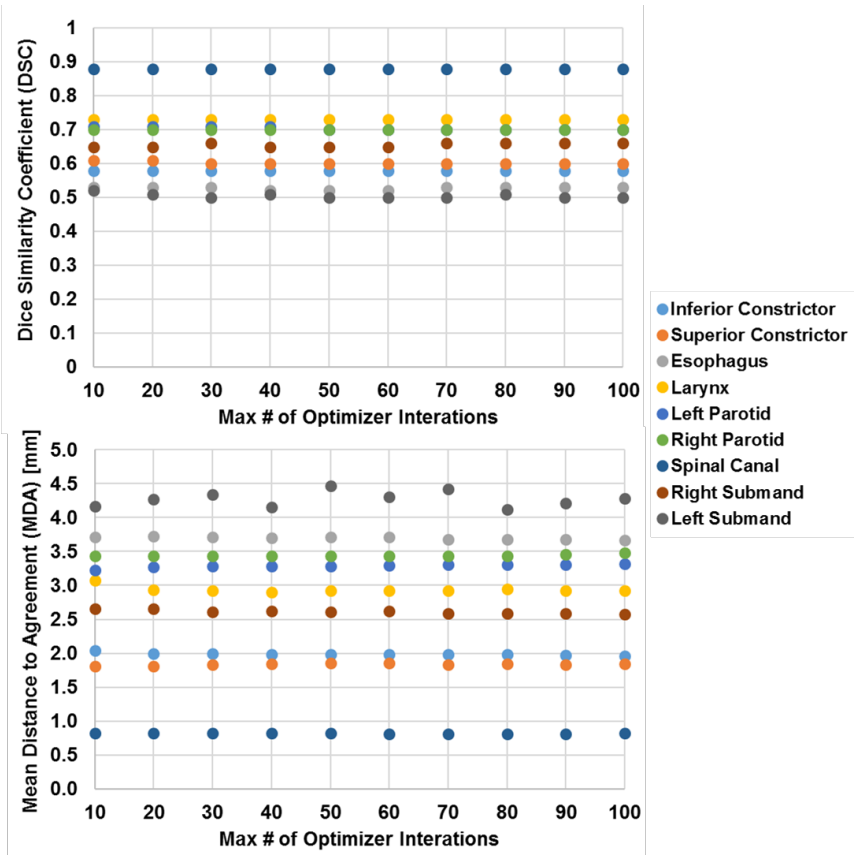


Figure 2.23. Mean population DSC and MDA of head and neck CT to CBCT with varying the maximum number of optimizer iterations. Minor variation in registration accuracy was found but it did not demonstrate a consistent trend.

Early termination criteria for the algorithm is controlled through a parameter defining the maximum number of consecutive optimizer attempts. This early termination occurs if the mutual information metric does not change more than the metric value percentage difference across a defined number of optimizer attempts. Varying this parameter from the default value of 10 did not change the registration results, demonstrating that the optimization process was not terminating early. In an automated registration setting, this observed functionality is likely preferred, as early termination of the registration algorithm benefits computation time at the potential cost of accuracy.

The maximum step length parameter sets the size of the initial step that the regular step gradient descent optimizer takes at each resolution within a registration. This step size is then decreased using the relaxation factor during the optimization process until the maximum number of iterations is reached or the step size falls below the minimum step length. Together, these parameters are mainly related to stable convergence of the deformation solution and affect robustness of the algorithm to noise in the mutual information metric. In this study, both the maximum and minimum step size were varied. Figure 2.24 shows the result of varying the maximum step length from 10 to 150. Slight changes in accuracy, particularly in the MDA metric, were found at values below 100, with the best overall registrations occurring with values around 60 to 70. When using a value above 100, accuracy begins to degrade for structures having larger known deformations. This demonstrates that decreasing the maximum step size from the default value could slightly improve the registration results. Additionally, analysis of the minimum step length parameter demonstrated that the algorithm was not sensitive to changes in the minimum step length when the value was set below 1. This suggests that the default minimum value of 0.000001 is sufficiently small not to allow the registration to terminate early based on this parameter.

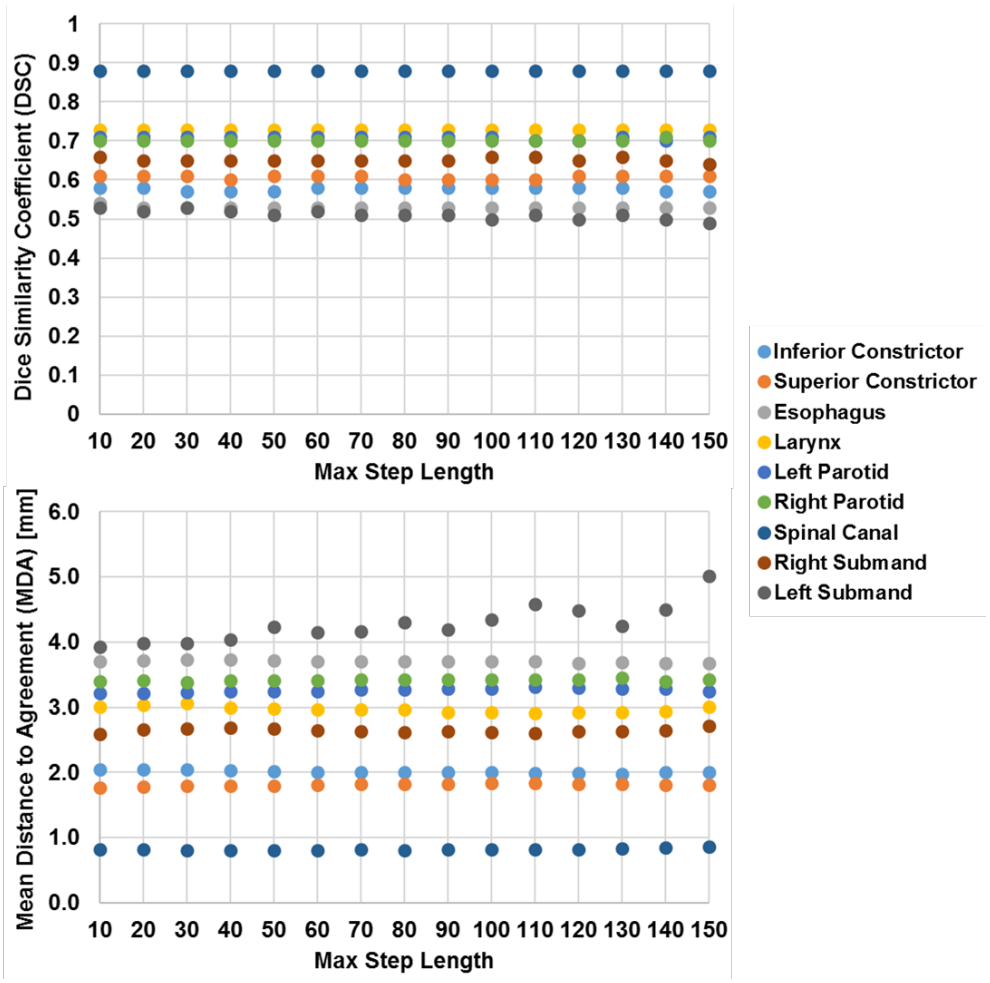


Figure 2.24. Mean population DSC and MDA of head and neck CT to CBCT with varying maximum step length. Accuracy is relatively invariant, although some structures had degraded accuracy when using a maximum step length above 100.

2.5 Conclusions

The Velocity mutual-information B-Spline DIR algorithm was quantitatively evaluated for clinical accuracy across a number of anatomical locations including head and neck, lung, liver, and pelvis for clinically relevant imaging modalities. This comprehensive clinical evaluation highlights the importance of body site-specific DIR commissioning and quantitative quality assurance testing. The current Velocity DIR algorithm fails to provide clinically acceptable registration accuracy across some of the evaluated anatomical locations. Clinically acceptable

accuracy was achieved for lung CT to CBCT. Near clinically acceptable accuracy was achieved for head and neck (CT to CT and CT to CBCT), lung 4DCT, and female pelvis MR to MR. Clinically acceptable accuracy was not achieved in liver MR to planning CT and prostate CT to CBCT. As expected, large differences in bladder filling challenge the algorithm in male and female pelvis images.

We found that the automated rigid registration, which can be used as the initial image transformation prior to the DIR, does not always provide a reasonable starting point for DIR. In some cases, this can lead to large errors in the DIR since the algorithm cannot coverage to a reasonable DVF solution. Therefore, in an unsupervised DIR environment, such as that required for broad clinical use of dose accumulation, errors in both the automated rigid registration and DIR must be considered. For manual DIR, the user should verify that the rigid initialization is reasonable prior to starting the DIR process. Further studies are required to determine the acceptable level of error in the rigid registration for the DIR algorithm to perform well.

To test DIR registration parameters, a user-developed C# program was created to programmatically determine registration parameter sensitivity for the Velocity DIR algorithm using the C# API. Through this analysis of registration parameter sensitivity, it was determined that a large improvement in registration accuracy could be achieved in lung cases, but only a marginal and variable improvement could be achieved for head and neck cases. For unsupervised dose accumulation applications of this DIR algorithm, current registration parameter tuning will likely not result in clinically acceptable registration accuracy across all anatomical locations. However, anatomically-specific tuning does offer slight improvements over default registration settings in terms of both accuracy and robustness and should be considered regardless of future algorithm development and implementation.

2.6 Extensions

2.6.1 Deformable Image Registration for MRI-based Cervical Cancer Brachytherapy

In many use cases, DIR functions as a pre-processing step to a downstream in the RT workflow. These downstream clinical processes can include treatment planning, dose accumulation, and longitudinal image analysis. Each of these potential uses of DIR likely comes with different accuracy needs and requirements. As an example, for treatment planning relying on DIR-propagated contours, only the boundaries of the propagated structures need to be accurately mapped. However, for delivered dose reconstruction and accumulation, voxel-wise accuracy may be required. Therefore, a more insightful way to evaluate DIR accuracy is to analyze its impact directly in the desired use case. As previously mentioned, brachytherapy is an integral component of locally advanced cervical cancer treatment. The field of brachytherapy has recently transitioned away from point- to volume-based planning, allowing for improved target and normal tissue visualization and more conformal dose delivery with improved OAR sparing and target coverage. Although volume-based planning has advantages, daily manual contouring of structures and plan adaptation require additional time leading to the risk of intra-fractional anatomic changes and increased patient discomfort. To address this challenge, we studied the dosimetric impact of DIR-based contour propagation on MRI-based cervical cancer brachytherapy planning.[60] Through this study, we demonstrated that DIR contour mapping could serve as a clinically acceptable replacement for manual contouring, providing significant time savings in cases that did not have dramatic changes in bladder filling state between fractions.

2.6.2 Dose Accumulation Process and Dashboard Development

Previous studies have exemplified the need for developing an automated workflow that could make widespread clinical use of adaptive radiotherapy, including dose accumulation, feasible. In part, automation is necessary due to the complexity present in the typical offline adaptive workflow, which often requires integration of DIR, dose calculation, and dose summation functionality to support clinical decision-making and if indicated, plan adaptation and re-optimization.[9] This process is referred to as *offline* since the 'dose-of-the-day' calculation, in addition to dose warping and summation into the same anatomical patient representation, occurs after an RT treatment is delivered to the patient. After this process, clinicians can review the accumulated dose metrics (i.e., dose delivered to date), which factors in anatomical changes and setup errors, and compare it to the planned dose metrics to determine if an intervention in the patient's treatment plan is required due to dose deviations or tissue response. In this method, dose could be recalculated using weekly imaging, assuming that anatomical changes occur slowly over treatment, or daily imaging, which better accounts for rapid anatomical and physiological changes in addition to daily set up errors.[153,154] In particular, implementing a daily offline adaptive workflow in a clinical setting requires automated processing since steps throughout include time-consuming computations (e.g., DIR, dose calculation) and data transfers. Additionally, the amount of new data generated through the process (e.g., daily dose distribution, daily dose metrics, structure volume changes and deformations) would be difficult for clinicians to review for each patient on a daily basis and therefore would greatly benefit from an automated analysis step to flag and report deviations that may be clinically meaningful.

Through a co-development project with Varian Medical Systems, we developed an automated offline adaptive workflow, incorporating daily dose accumulation, that utilizes features from both Velocity™ and Eclipse™. An overview of the workflow and subcomponents is outlined in Figure 2.25. Briefly, Velocity handled the deformable registration and resampling process, while Eclipse was used for dose calculation and data storage. These components were contented through the newly developed Velocity C# API and existing Eclipse API allowing for data transfer between the systems.

This workflow followed the following process for calculating the daily and accumulated dose. First the system is automatically triggered to begin the workflow after a new treatment session is completed (i.e., a patient completes a fraction of their treatment course). The planning CT and daily volumetric imaging (i.e., CBCT) is then transferred from Eclipse to Velocity, wherein an intensity-based B-spline DIR is performed to spatially map the planning CT to the CBCT. The DIR transform is then applied to the planning CT to warp the image towards the anatomy present on the CBCT which allows for dose calculation on the daily anatomy. Although differing methods exist to more efficiently calculate daily dose, we chose to calculate dose on the warped planning CT because direct dose calculation on the CBCT suffers from inaccuracies caused by the limited field-of-view and image acquisition dependent calibration. Additionally, the calculated DIR transform for warping the planning CT to the CBCT facilitates convenient dose accumulation from multiple daily dose calculations back to the original planning CT for direct comparison with the original, planned dose. The warped planning CT is then transferred back to Eclipse where the original treatment plan is copied and the dose is recalculated on the warped CT. Then the new dose distribution, closer representing the actual delivered dose to the patient, is

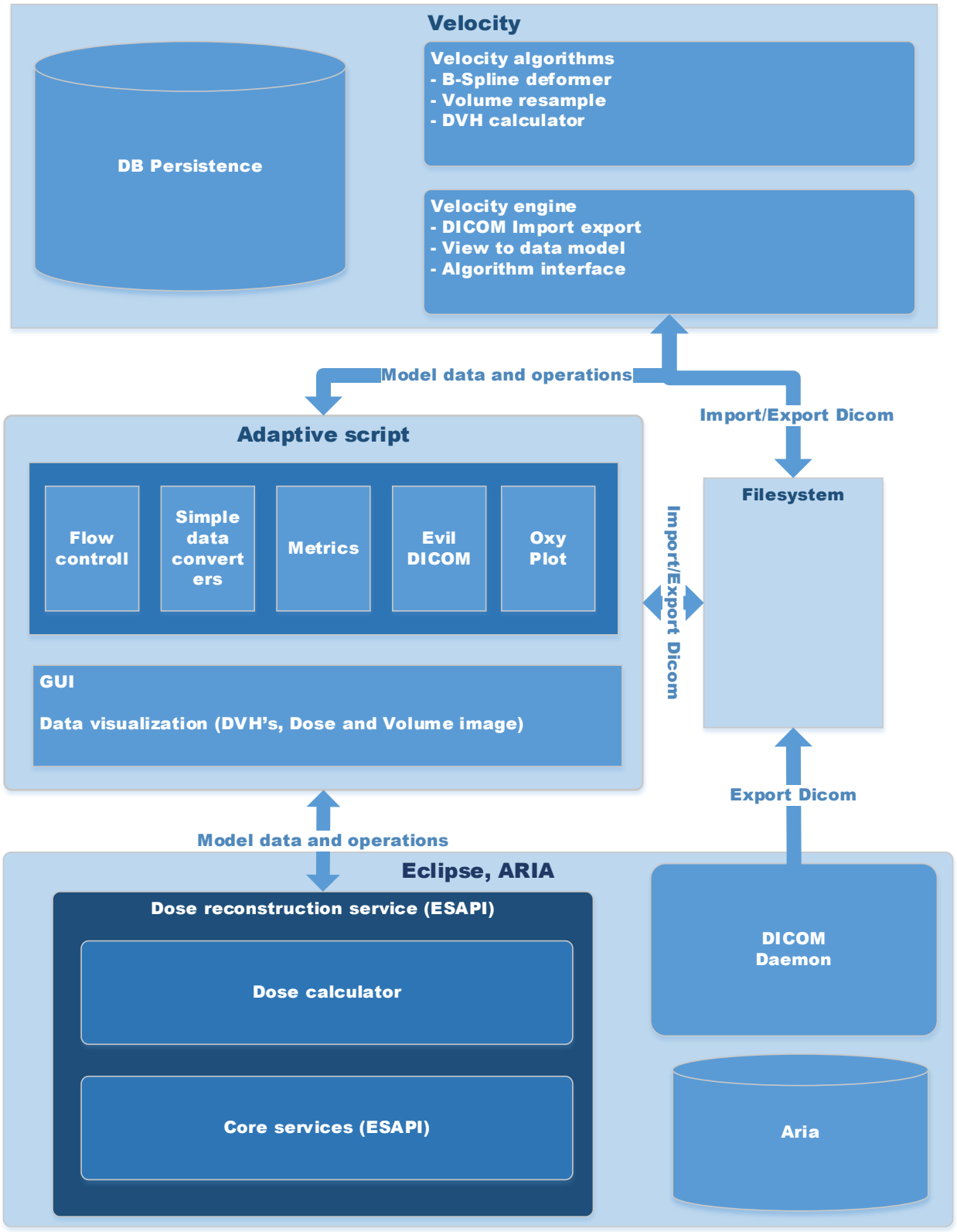


Figure 2.25. Workflow for the daily dose accumulation tool that was developed by linking Velocity™ and Eclipse™ through scripting interfaces

transferred back to Velocity where it is transformed back to the anatomy of the original planning CT (through inversion of the DIR) and is added to the accumulated dose total. Both the daily (fraction) dose distribution and accumulated dose to date is then sent back to Eclipse to allow clinicians to review the data if needed. The automated workflow was accompanied with a newly developed dashboard, shown in Appendix A, that presented users with the ability to set automated dose deviation warning/alert levels in addition to review daily and accumulated dose metrics, structure volumes, and trends. Example data from this system, calculated retrospectively for a HN cancer patient, is shown in Figure 2.26 and includes the esophagus mean fraction dose (red line, scaled to 35 fractions) and the accumulated dose (blue line, also scaled to 35 fractions) tracked across all 35 fractions. Although the final accumulated dose is similar to the planned dose, large variations in the fraction dose are noted, particularly for fraction 21. After additional analysis, this large difference was tracked back to improper patient set up for that treatment fraction. This process was successfully used retrospectively across multiple patients without issue; however, clinical validation of the process and results was not performed and clinical implementation was never achieved. Future implementation of this workflow will require clinical validation of the process and API functionality, analysis of error and failure pathways, and determination of clinically relevant warning/alert levels.

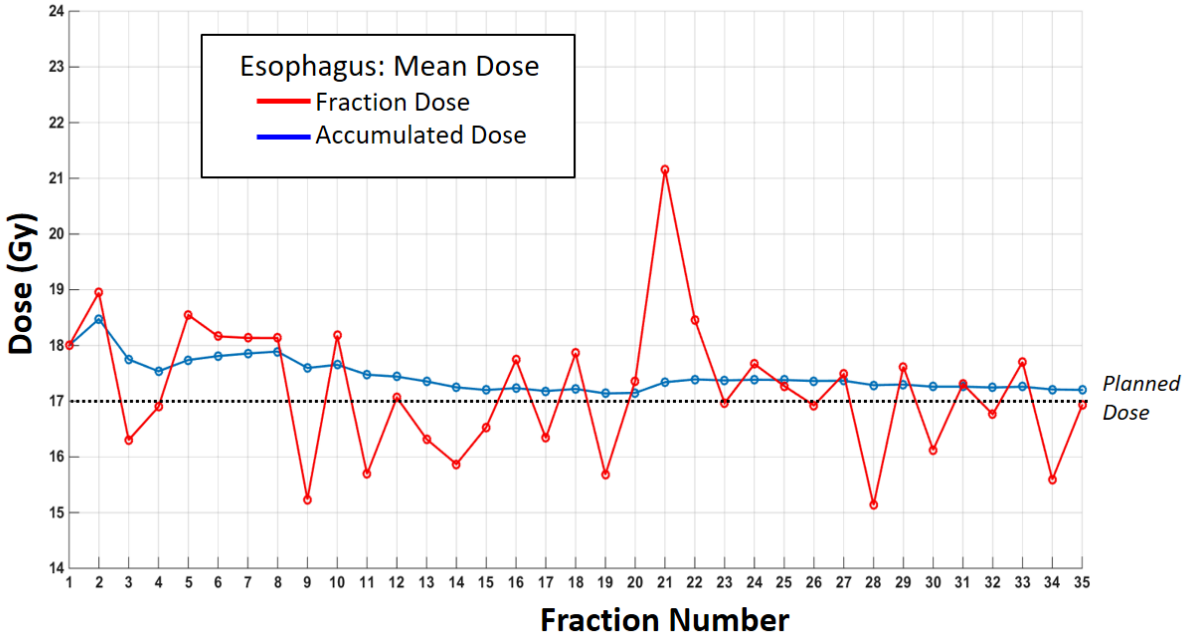


Figure 2.26. Daily dose accumulation for a HN cancer patient across a 35 fraction treatment course tracking esophagus mean dose.

Chapter 3
Characterization and Implementation of Radiation-Induced Dose-Volume Liver Response for Biomechanical Deformable Image Registration

Portions of this chapter are substantially equivalent to work the author has already published in *International Journal of Radiation Oncology Biology Physics*.[61]

3.1 Abstract

Understanding anatomical and functional changes in the liver resulting from radiotherapy is fundamental to the improvement of normal tissue complication models needed to advance personalized medicine. The ability to link pre-treatment and post-treatment imaging is often compromised by significant dose-dependent volumetric changes within the liver that are currently not accounted for in deformable image registration (DIR) techniques. This study investigates using delivered dose, in combination with other patient factors, to biomechanically model longitudinal changes in liver anatomy for follow-up care and re-treatment planning. Population models describing the relationship between dose and hepatic volume response were produced using retrospective data from 33 patients treated with focal radiation therapy. A DIR technique was improved by implementing additional boundary conditions associated with the dose-volume response in series with a previously developed biomechanical DIR algorithm. Evaluation of this DIR technique was performed on computed tomography imaging from seven patients by comparing the model-predicted volumetric change within the liver to the observed change, tracking vessel bifurcations (branching points) within the liver through the deformation process, then determining target registration error (TRE) between the predicted and identified post-

treatment bifurcation points. Evaluation of the proposed DIR technique showed that all lobes were volumetrically deformed to within the respective contour variability of each lobe. The average TRE achieved was 7.3 mm (2.8 mm LR and AP, 5.1 mm SI), with the SI component within the average limiting slice thickness (6.0 mm). This represented a significant improvement (Wilcoxon, $p < 0.01$) over the application of the previously published biomechanical DIR algorithm (10.9 mm). This study demonstrates the feasibility of implementing dose-driven volumetric response in deformable registration, enabling improved accuracy of modeling liver anatomy changes, which could allow for improved dose accumulation, particularly for patients who require additional liver radiotherapy.

3.2 Introduction

Liver cancer, including both hepatocellular carcinoma (HCC) and intrahepatic biliary duct cancer (cholangiocarcinoma), is the fourth most common cause of cancer-related mortality in the United States with an estimated 31,780 cancer deaths in 2019.[62] This is one of the few cancer types that has undergone a sustained increase in incidence over the past two decades with an average annual percentage increase of 2.5% between 2012 and 2016.[62-64] Additionally, colorectal cancer (CRC) remains one of the most common cancer types globally for both men and women, accounting for nearly 1 in 10 cancer cases and deaths.[62,65,66] For these patients, the most common site to develop metastatic disease is the liver.[65] At the time of CRC diagnosis, 20% of patients will present with synchronous liver metastases, and an additional 35–50% of patients diagnosed with CRC will develop liver metastases as the most prevalent mode of failure within 5 years of initial treatment.[67,68]

For HCC, biliary, and CRC patients presenting with unresectable liver tumors, radiation therapy (RT) provides a nonsurgical focal treatment option and has also been investigated as a

bridge to liver transplant for HCC patients.[69] However, historically the use of RT for treatment of liver tumors has been limited by the tolerance of normal liver tissue to radiation. Prior to recent advancements in RT delivery techniques, RT in the liver generally resulted in large volumes of normal liver tissue receiving damaging doses of radiation, leading to a high incidence of potentially devastating radiation induced liver disease.[70] With modern delivery techniques, such as intensity-modulated RT (IMRT) and daily image guidance, it is now possible to spare a larger percentage of the liver from radiation damage while still delivering therapeutic dose levels to the tumor. These improvements have led to increased usage of RT for hepatic tumors, particularly for small tumors < 5cm, but liver toxicity remains a concern and delivering radiation to larger tumors continues to be strongly limited by toxicity. As an example, recent clinical trials involving the use of stereotactic body RT (SBRT) for locally advanced HCC were able to achieve high rates of loco-regional control (87% at 1 year).[71] However, the rate of grade 3 or higher toxicities was found to be 30%, with 7 patient deaths possibly related to treatment.[71] Therefore, there is strong motivation to better understand radiation liver toxicity in the era of modern RT delivery techniques in an effort to reduce the prevalence of toxicity while maintaining excellent local control.

Advancements in functional imaging-based assessments, such as positron emission tomography (PET) and dynamic contrast enhanced (DCE) MRI, have improved the ability to analyze liver function beyond global measures of liver function. Beyond baseline functional analysis, the spatially-varying functional information provided by these imaging techniques can be incorporated into RT to better tailor the treatment to a specific patient. For example, liver function maps can be used in RT plan optimization to guide preferential sparing of certain liver regions from radiation in order to minimize a patient's overall risk of liver toxicity or to maximize a patient's remaining liver function after RT.[72,73] Implementing these types of treatment planning

methods depends, in part, on the ability to predictively model functional changes to subunits of the liver based on the radiation dose received by the subunit.[74-77] However, assessment and analysis of localized changes in liver function can be hindered by the inability to spatially correlate the subvolumes of tissue within the liver.[75,76] In addition, significant non-uniform changes in liver volume can compromise the correlation of the dose delivered with the longitudinal changes in the liver function observed in functional imaging. Figure 3.1 demonstrates the potential magnitude of liver changes over the course of RT treatment and the potential anatomical misalignment between the pre-treatment (RT-planning) CT image, post-treatment CT image, and post-treatment DCE-MRI corresponding to liver function. This misalignment, which is driven both by anatomical motion and volumetric changes, impairs spatial correlations between the planned-RT dose distribution (spatially linked to the planning CT) and functional changes identified in post-treatment functional imaging. Improved accuracy in the correlation of functional imaging with delivered dose could lead to modifications of currently used liver normal tissue complication probability models to better account for the likelihood of spatially localized changes in liver function.[75,76]

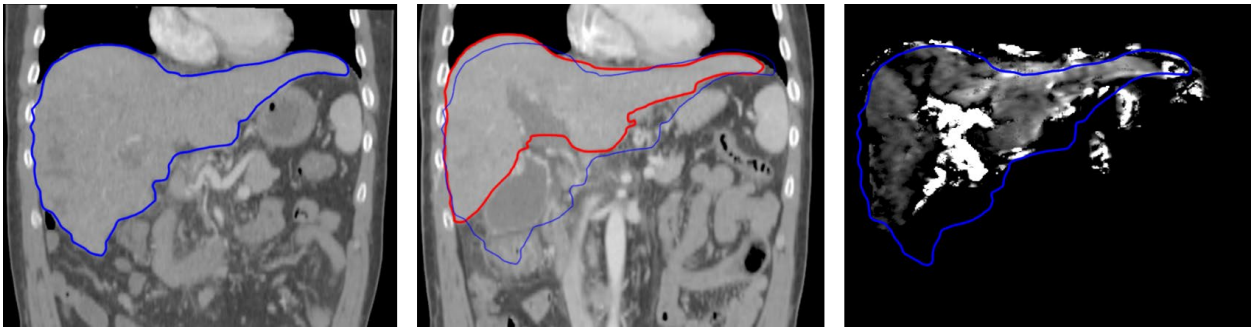


Figure 3.1. Pre-treatment (RT-planning) CT image with liver contoured in blue (left), Post-treatment CT image of the same patient with the current liver anatomy contoured in red and the blue liver contour copied from the pre-treatment CT (middle), Post-treatment DCE-MRI with the blue liver contour copied from the pre-treatment CT (right).

Furthermore, SBRT is becoming an increasingly favored treatment method for patients with unresectable oligometastases in the liver.[78-86] Development of additional liver metastases can require multiple courses of SBRT, presenting a need to accurately map previously delivered dose to the subsequent treatment plan(s) for safe treatment planning. Although current dose mapping techniques can accurately account for respiratory and other anatomical motion between treatment courses, no current technique directly accounts for potentially significant hypertrophy/atrophy volume changes observed following radiation therapy (RT).[87]

This complex dose-volume response of normal hepatic tissue has been observed in many studies.[82-86] Although precise causation remains unknown, it is hypothesized that this atrophy-hypertrophy radiation response is similar to the response commonly noted following surgical resection and other forms of liver treatment or injury.[88,89] Despite these observations, currently available deformable image registration (DIR) techniques often cannot accurately account for large volumetric changes with localized mass loss or gain.[90-92] Previously, this has been demonstrated on head and neck cases with weight loss or disease progression and prostate cases with large deviations in bladder filling.[91-93] Studies have also shown that intensity-based DIR algorithms can fail in the presence of homogenous contrast (often observed in liver computed tomography (CT) images) and large deformations.[93-95] Furthermore, many intensity-based DIR algorithms include regularization terms to ensure that the deformation field is relatively smooth across an image. These regularization methods function to decrease roughness and irregularities within a deformation field, but may consequently restrict a deformation from accurately modeling localized volume changes.[96] Although biomechanical DIR algorithms often provide more physically plausible deformation fields in the presence of large volume changes, deformation is primarily driven by external organ boundaries (generated from contours) rather than intra-organ anatomy.

In the application of liver registration, this can result in inadequate registration of liver subvolumes and intrahepatic anatomy even when the external boundary of the liver appears to be accurately registered.

In order to improve the correlation of subvolumes within the liver, this study proposes a modified biomechanical modeling method for hepatic tissue including independent dose-volume response deformation forces in addition to a previously developed spatially-constrained biomechanical DIR algorithm. Briefly, a relationship between dose and volume response, mimicking the thermal expansion relationship for non-biological materials, was developed for a population of patients using a subset of the total population. A two-step DIR algorithm was developed to include this dose-volume response in addition to spatial constraints. Using the patients excluded from the development of the population dose-volume response model, the newly developed algorithm including dose-volume response in addition to existing spatial constraints (Morfeus with Dose Boundary Conditions) was evaluated and compared to a rigid registration method and the previously developed spatially constrained DIR algorithm (Morfeus).

In summary, the purpose of this study is to improve longitudinal liver registration by developing a population-based dose-volume response model for normal liver tissue and implementing this dose-volume response as additional boundary conditions in series with previously developed biomechanical DIR algorithm. Volume- and point-based metrics were used to evaluate the performance of the newly developed DIR technique.

3.3 Materials and Methods

3.3.1 Data collection

Forty patients previously treated on an Institutional Review Board approved phase II trial of dose-escalated liver RT from 1998 to 2005 were retrospectively investigated .[97] In this clinical trial, previously described in detail,[97] patients were treated with conformal high-dose RT (median dose was 60.75 Gy) in 1.5 Gy fractions twice daily (BID). Criteria for patient selection in this DIR development and evaluation study included having a single unresectable intrahepatic primary malignancy, HCC or biliary, or liver metastasis from CRC, with a pre-RT planning CT and follow-up CT scans at least 49 days post-treatment. Potential effects of respiratory motion were mitigated by use of active breathing control and breath-hold CT scans when possible.

Liver images were previously contoured using stable anatomical landmarks to delineate the left lateral and medial segments, right anterior and posterior segments, and caudate lobe.[98] The gross tumor volume (GTV) was previously defined for clinical treatment planning purposes. Figure 3.2 shows pre-treatment liver and tumor segmentation for one of the patients included in the study. The pre- and post-RT volumes as well as mean radiation doses delivered to each contoured region were obtained from the treatment planning software (UMPLAN, University of Michigan, Ann Arbor, MI). Of the 40 patients selected, seven biliary and CRC patients were randomly chosen for accuracy evaluation while the remaining 33 were used to create a population model of normal tissue radiation response. HCC patients were not used for accuracy evaluation since the limited HCC patient cohort size did not allow enough data for both generation and evaluation of the HCC-specific response model. For the seven patients used in accuracy evaluation, a radiation oncologist selected vessel bifurcations within the liver on both the pre- and post-treatment CT scans to allow for calculation of the target registration error (TRE) to evaluate the

accuracy of the model. An example vessel bifurcation is depicted in Figure 3.2 on a pre-treatment contrast-enhanced CT image.

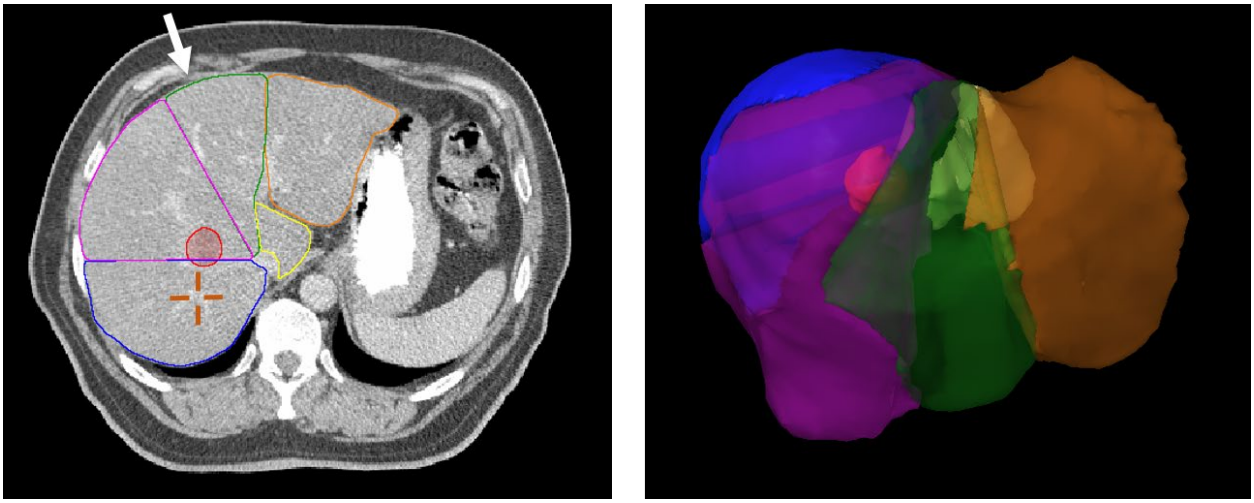


Figure 3.2. Liver segmentation shown on 2D axial CT slice (left) and 3D volume rendering (right) in the perspective indicated by the white arrow. Identified structures include the left lateral (orange), left medial (green), right anterior (purple), right posterior (blue), and caudate (yellow), GTV (red), and one example vessel bifurcation (orange crosshair).

3.3.2 Liver Response Model

To generate explicit dose-response deformation forces for hepatic biomechanical modeling, termed Dose Boundary Conditions (Dose BCs) within the deformation process, a population model was created to relate liver volume response to dose in terms of a linear expansion coefficient (α_L). This methodology was chosen so that the volumetric response could be directly modeled using existing thermal expansion tools within commercial finite element modeling (FEM) software packages. By substituting dose for temperature change in a standard isotropic thermal linear expansion equation, we can utilize standard FEM thermal boundary conditions to achieve a direct relationship and mechanical modeling method for volume change as a function of dose. To further clarify, this method is used to model the long-term biological-volumetric response (tissue atrophy/hypertrophy) and is not using dose as a surrogate for short-term temperature change associated with radiation dose deposition. For the 33 patients used in the population model, the

linear expansion coefficient was computed for each of the previously defined segments of the liver using Equation 3.1 (modified isotropic thermal linear expansion):

$$\alpha_L = \frac{V_f - V_i}{3V_i D} \tag{3.1}$$

where V_f is the component's final volume, as measured on the post-treatment imaging study, V_i is the component's initial volume, as measured on the pre-treatment imaging study, and D is the mean absorbed dose of the component (dose replaces temperature change in the standard linear expansion equation). A log-logistic sigmoid function (Equation 3.2) was selected to represent the data relationship between α_L and D since it captures the mean structure of the data well and is similar to formulas used in normal tissue complication probability and tumor control probability models. Using a gridded-search, least-squares approach, the sigmoid function was fit to this data to get α_L as a function of D. Fitting parameters included α_C and α_F representing the linear expansion ceiling and floor, respectively, D_{50} representing the dose at 50% of the linear expansion range, and γ representing the slope at D_{50}. For statistical comparison purposes, a corresponding linear least-squares fit was also generated.

$$\alpha_L = (\alpha_C - \alpha_F)\left(\frac{(D_{50}/D)^{4\gamma}}{1 + (D_{50}/D)^{4\gamma}}\right) + \alpha_F \tag{3.2}$$

Stratifications of the response model were investigated based on the hypothesis, formulated from previous studies, that dose response is correlated to tumor type (as a surrogate for the presence of underlying liver disease, e.g., cirrhosis and hepatitis) and spatial location of the tumor.[89,99-102] To maintain statistical confidence within the stratifications, stratifications were only performed in groups that maintained at least 30 samples. The resulting seven stratifications included: all patients (HCC/Biliary/CRC), HCC, Biliary, CRC, Biliary/CRC, right lobe

Biliary/CRC, left lobe Biliary/CRC. Significance of these stratifications was determined using a statistical comparison of the stratified models to the all-inclusive response model based on Spearman's rank correlation coefficients, an assessment of how well the relationship of dose and linear expansion coefficient can be described using a monotonic function, and non-linear regression analysis from the sigmoid fit compared to a linear regression.

3.3.3 Liver Deformation

This study used Morfeus, an in-house-developed biomechanical model-based deformable image algorithm, previously described in detail.[28] Briefly, two finite element models (FEM), representing the pre and post-RT livers, are generated using the CT contours of the complete livers. The pre-RT model is then deformed to the post-RT model via guided surface projections (HyperMorph, Altair Engineering, Troy MI). This biomechanical modeling technique incorporates linear-elastic material properties of liver tissue and GTV, and allows for the incorporation of additional boundary conditions within the finite element analysis (FEA).

In this study, pre-RT liver FEMs were deformed in a two-step process, shown in Figure 3.3. In Step 1, Dose BCs are applied using thermal expansion to describe the volumetric response of the liver to external beam RT (EBRT). In Step 2, surface constraints are applied, as previously published, to resolve anatomical deformation due to patient pose and physiological state.[28,93]

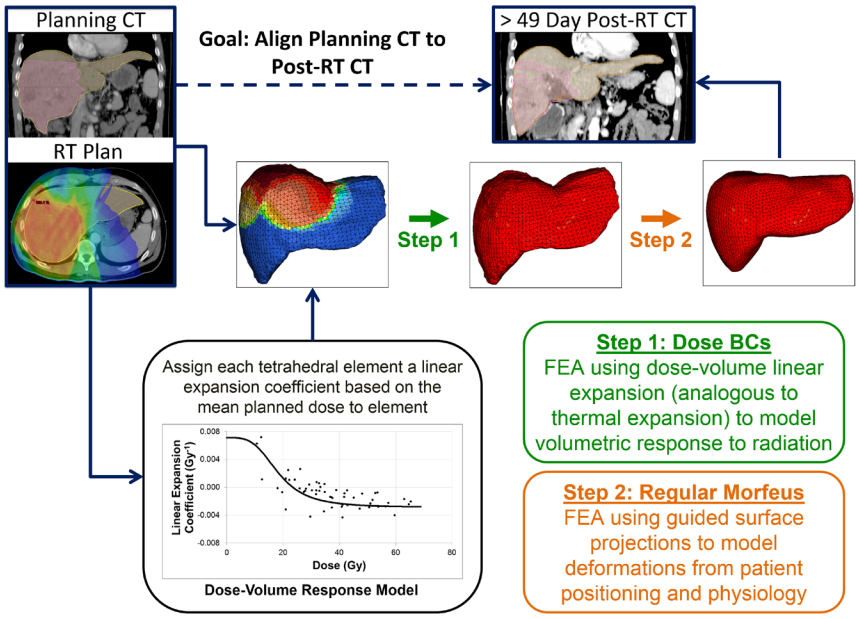


Figure 3.3. Flow chart of the proposed biomechanical deformable image registration process with the inclusion of dose-dependent volumetric response. We first create a biomechanical liver model using the planning computed tomography (CT) image. Using the linear expansion dose-volume population model and the patient's radiation therapy (RT) dose distribution, we assign each tetrahedral element in the model with a linear expansion coefficient based on the mean dose to each element (shown prior to Step 1). We then preform a finite element analysis (FEA) run, Step 1, to account for dose volume changes (result shown after Step 1). Afterwards, we apply Step 2, another FEA run, using surface correlation boundary conditions to spatially align the Step 1 deformed liver with the liver contour on the follow up CT (result shown after Step 2).

In Step 1, Dose BCs were generated by grouping tetrahedral elements within the FEM into 1 Gy dose bins by calculating the mean dose received by that tetrahedron from the 3D dose grid produced from the commercial treatment planning software (Eclipse (v11), Varian Medical Systems). Each 1 Gy dose region was then assigned a single linear expansion coefficient calculated from a population-based liver response model. These coefficients were loaded into the FEM pre-processor as linear thermal expansion coefficients, and the dose delivered to the liver was applied as the boundary condition to model the atrophy and hypertrophy observed in the liver. Since the population dose-response model is not representative of GTV response, which varies significantly between patients due to many factors, tetrahedral elements representing the GTV were given a single patient-specific thermal expansion coefficient based on measured volumetric tumor

response from the pre-RT to post-RT CT. Therefore, while the assignment of linear expansion coefficients to normal liver tissue is non-uniform and based on dose, the assignment of linear expansion coefficients to the GTV is uniform (representing an assumption of uniform volumetric response of the tumor).

Surface constraints, for Step 2 of the deformation process, were defined from guided surface projections generated after applying the FEA results from Step 1 (Dose BCs) to the FEM. The deformation results of this second FEA were applied in addition to the dose constraint analysis to spatially align the liver to the anatomical and physiological position of the liver in the follow-up image. Therefore, the final deformation vector field for complete DIR process is a summation of the tetrahedral displacement results of the Step 1 and Step 2 FEAs.

3.3.4 Characterization of the FEA volumetric response

When considering both normal liver tissue response and the tumor response, the total liver volume may change by more than 50%. Volume changes of this magnitude exceed the typical use case for thermal expansion and contraction modeling in FEA. For example, linear expansion of coefficients of many solid metals are on the order of 10^{-6}/°C. Therefore, a 1000°C increase or decrease in the temperature results in fractional length changes on the order of 0.1%. For biomechanical modeling purposes, the linear expansion coefficients for the liver and tumor are 3 magnitudes larger. Therefore, at a radiation dose of 50 Gy, the fractional length changes could be in the range of 5–10%. These relatively large volumetric changes mean that the liver FEM is prone to volumetric locking, particularly when using first-order tetrahedral elements. Volumetric locking is related to the unrealistic over-stiffening of elements when the modeled material is nearly incompressible (Poisson's ratio nearing 0.5). Since our linear-elastic FEM represents human tissue as nearly incompressible with a uniformly assigned Poisson’s ratio of 0.45, our model may be

subject to this effect. Additionally, given that the tissue elements are assigned linear expansion coefficients based on a spatially varying dose matrix, heterogeneity present in the dose matrix could cause neighboring elements to be assigned different coefficients and thermal loads. Therefore, we may be attempting to model different levels of volume change between neighboring elements. This is particularly true at the interface between elements representing normal tissue and elements representing the tumor because the tumor elements are assigned a tumor-specific, uniform linear expansion coefficient. This varying response between bordering elements may cause artificial response locking as a result of the shared sides of the neighboring elements attempting to respond differently. Given these considerations, we analyzed the volumetric response of individual elements within the FEM to the newly added thermal loads and compared it to the expected fractional (relative) element volume change. This analysis was performed for both a simple 12 tetrahedral element box FEM with a uniformly applied linear expansion coefficient and thermal load, and for the full liver FEM with varying coefficients and thermal loads. Figure 3.4 shows the expected volumetric response of the individual elements relative to dose based on a sigmoid model of the linear expansion coefficient.

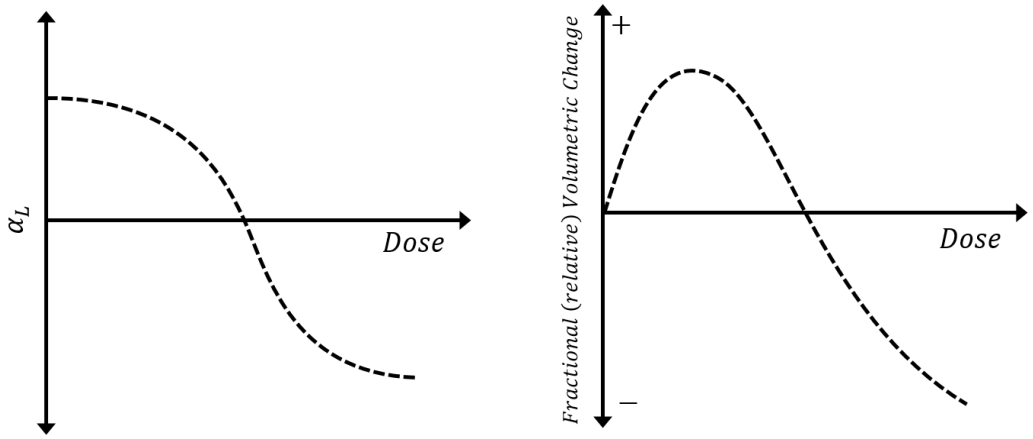


Figure 3.4. Expected fractional volumetric change of elements within the FEM (right) based on a sigmoid model of the linear expansion coefficient (left).

3.3.5 Accuracy Analysis and Comparison

Both volumetric response comparison and TRE were used to evaluate the accuracy of the model for the seven patients not included in the generation of the dose-response model. First, following the initial deformation from application of only the Dose BCs (Step 1), the volume of each deformed liver segment was compared to its actual post-treatment volume calculated from the contoured post-treatment imaging study. This comparison served as an analysis tool to determine whether the expansion and contraction generated from the population model accurately predicted the known volume change of each segment. Individual segment volume change was compared to previously reported intra-observer contouring reproducibility study for a similar liver segmentation methodology which gave the following 95% confidence intervals (CIs): left lateral segment 45 cc, left medial segment 56 cc, right lobe (superior and posterior) 84 cc, and caudate lobe 10 cc.[103] The magnitude and relative accuracy of the volumetric modeling was determined across three deformation methods: Morfeus, only Dose BCs (Step 1), and Morfeus with Dose BCs. Statistical significance of the volumetric modeling techniques was determined using a Wilcoxon signed-rank test.

TRE was calculated using the selected liver bifurcations for three registration methods: rigid, Morfeus, and Morfeus with Dose BCs. The bifurcations for the seven patients selected for accuracy analysis were selected a second time by the same observer on the follow-up image in order to quantify intra-observer variability. Statistical significance for the comparison of TRE between registration methods was determined using a Wilcoxon signed-rank test between individual corresponding point pairs.

3.4 Results

Characteristics for the seven patients excluded from the population response modeling and selected for accuracy analysis are shown in Table 3.1. The tumor type, location, and population response model used are provided. Due to varying Post-RT CT slice thickness and number of selected bifurcation points, the TRE analysis was performed on an individual patient level in addition to a population average.

Table 3.1. Characteristics of the seven patients used for evaluation of the population dose-volume response models and proposed deformable image registration method.

Patient	Tumor		Population Response Model Used	CT Slice Thickness [mm]		# Bifurcation Points
	Type	Location (Lobe)		Pre-Tx	Post-Tx	
1	Biliary	Left	Left Lobe (Biliary/CRC)	3	2	16
2	CRC	Left	Left Lobe (Biliary/CRC)	3	5	10
3	CRC	Left	Left Lobe (Biliary/CRC)	3	7	14
4	Biliary	Left	Left Lobe (Biliary/CRC)	3	5	4
5	CRC	Right	Right Lobe (Biliary/CRC)	3	7	10
6	CRC	Left and Right	Biliary/CRC	3	10	6
7	Biliary	Left and Right	Biliary/CRC	3	5	7

3.4.1 Liver Response Model and Stratifications

Table 3.2 gives statistical and curve fitting analysis for the proposed stratifications. All stratifications resulted in significant Spearman correlation coefficients suggesting good negative correlation between linear expansion coefficient and dose. Stratifications of HCC, CRC, Right Lobe Biliary/CRC, and Left Lobe Biliary/CRC provided increased correlation compared to the all-inclusive model. In all stratifications, the log-logistic sigmoid fit gave an improved coefficient of

determination when compared to a linear fit. Figure 3.5 shows the log-logistic sigmoid best-fit curves for each of the proposed stratifications. The seven patient accuracy analysis utilized the Biliary/CRC models. As noted in Table 3.1, for Patients 1 through 5, the location-specific stratification models were used since the tumors were confined to a single lobe, and for Patients 6 and 7, the general Biliary/CRC model was used since the tumors extended over both the right and left lobes. Location-specific models were used instead of tumor-specific models since the average Spearman correlation coefficients were greater.

Table 3.2. Correlation and curve fitting results for the proposed dose-volume response population stratifications.

Stratification	Sample Size	Spearman Correlation			Coefficient of Determination, r^2	
		ρ	[95% CI]	p	Linear-Fit	Sigmoid-Fit
All	117	-0.60	[-0.70, -0.47]	$p < 0.01$	0.37	0.45
HCC	32	-0.67	[-0.83, -0.42]	$p < 0.01$	0.55	0.64
Biliary	39	-0.57	[-0.75, -0.31]	$p < 0.01$	0.34	0.58
CRC	46	-0.66	[-0.80, -0.46]	$p < 0.01$	0.39	0.51
Biliary/CRC	85	-0.59	[-0.71, -0.43]	$p < 0.01$	0.33	0.43
Right Lobe (Biliary/CRC)	30	-0.64	[-0.81, -0.36]	$p < 0.01$	0.37	0.47
Left Lobe (Biliary/CRC)	55	-0.67	[-0.79, -0.49]	$p < 0.01$	0.33	0.41

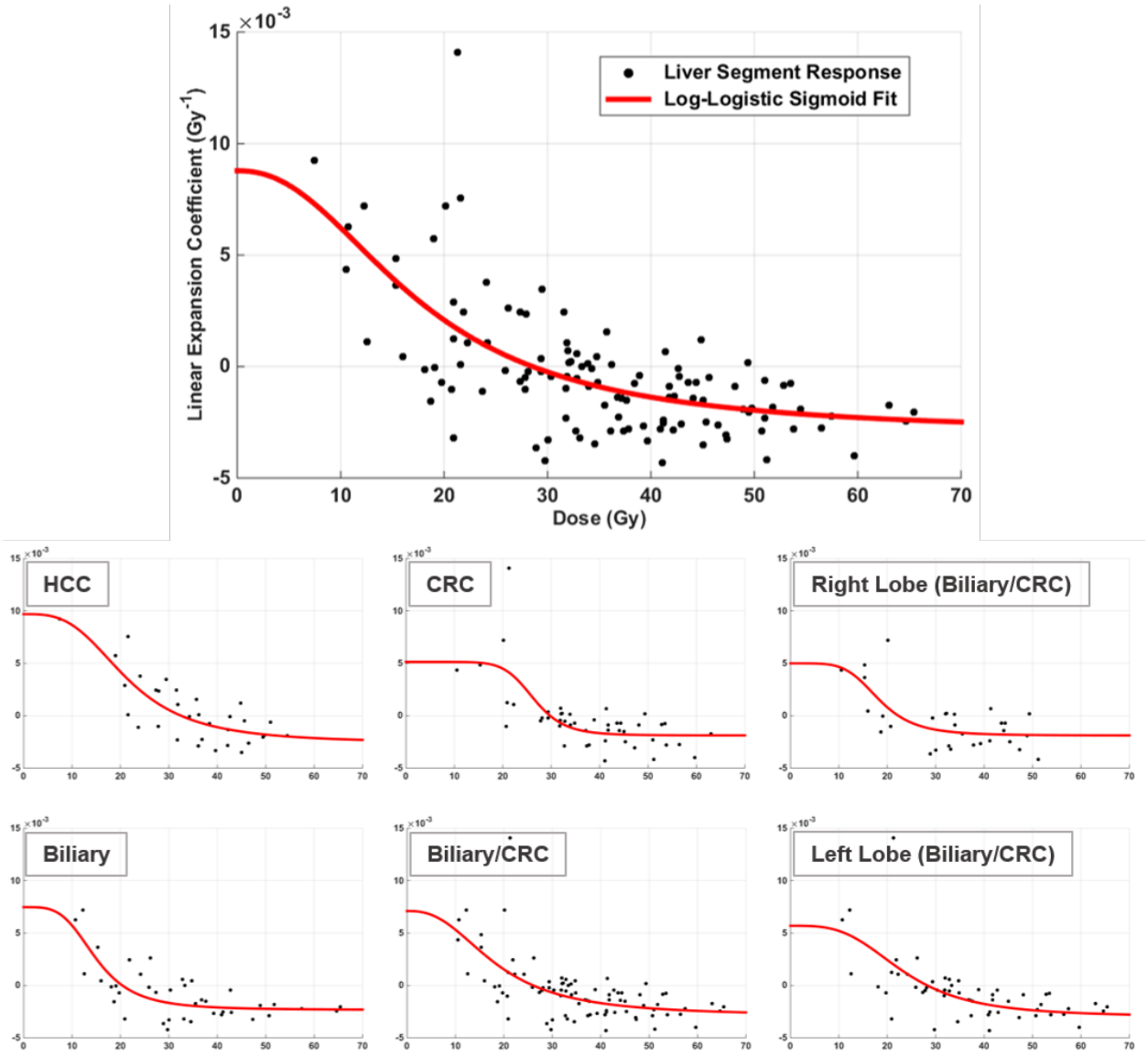


Figure 3.5. Log-logistic sigmoid function fits for each of the proposed patient population stratifications. See Table 3.2 for more information on each stratification and statistical curve fitting results.

3.4.2 Characterization of the FEA volumetric response

Volumetric response resulting from the application of thermal boundary conditions and corresponding linear expansions coefficients was analyzed on a small-scale 12 element cubic model and full-scale liver model. Figure 3.6 shows the relative error in the volumetric response compared to the applied, or planned, volume change for the cubic model. In this model, uniform

material properties and thermal loads were applied to the tetrahedral elements and the volume change was calculated over the entire cube under varying expansion and contraction levels. For applied volume changes within 15%, error in the modeling remained under 1%. However, at applied volume changes greater than 30%, modeling error could be larger than 5% for expansion and 10% for contraction. At 20% volume change, the error in contraction modeling begins to exceed that of the expansion model, which suggests that volume locking of the elements may be occurring beyond that response level. At 50% volume contraction, nearly the maximum expected volumetric response provided by the liver dose-response models shown in Figure 3.5, the relative volume error after FEA was 19%. Independent testing of second-order tetrahedral elements and decreasing the assigned Poisson's ratio did not demonstrate large volume-response modeling improvements, with second-order elements causing increased computational complexity and FEA solver time.

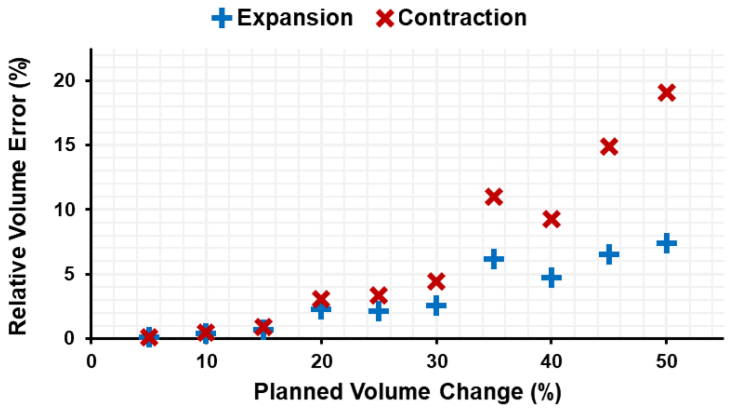


Figure 3.6. Relative error in the FEA volumetric response for a simple 12 element box model with various planned volume changes in the range of the liver-response models.

Figure 3.7 shows analysis of the FEA volume response for Step 1 of the proposed DIR method. Points on the plots represent the individual volumetric response of each tetrahedral element included in the model relative to the radiation dose at the centroid of the element. The

color of the points varies from blue to yellow based on the distance to the tumor. For both of the FEM models shown, these plots match the general expected shape of sigmoid volume response originally shown in Figure 3.4. Due to the usage of discrete dose bins for assigning the thermal expansion coefficients, the expected response of the elements, shown as black lines in Figure 3.7, is not continuous. Within each of these dose bins, a linear fit of the volumetric responses was calculated and is shown as a blue line. As expected from the results presented in Figure 3.6, the measured volumetric response begins to deviate from the expected response when the relative volume change exceeds −0.2. The response of elements representing the tumors shown as pink points on the plots and was compared to the expected volume shrinkage with uniformly applied linear expansion coefficients based on measured tumor volume change and mean dose. For Patient 1, the FEA tumor response, shown as a linear fit in gray, closely matches the expected tumor response shown as a purple line. However, for Patient 6, the FEA tumor response underrepresents the intended volume change by nearly 15%. This provides further evidence of volumetric locking in the FEM when large volume changes are applied. For Patient 1, it is also evident that the response of normal tissue elements nearby the tumor are impacted by the response assigned to the tumor elements. These bordering elements, shown as yellow points given the proximity to the tumor, demonstrate a volumetric response that is in-between the expected normal tissue tumor element response and tumor element response.

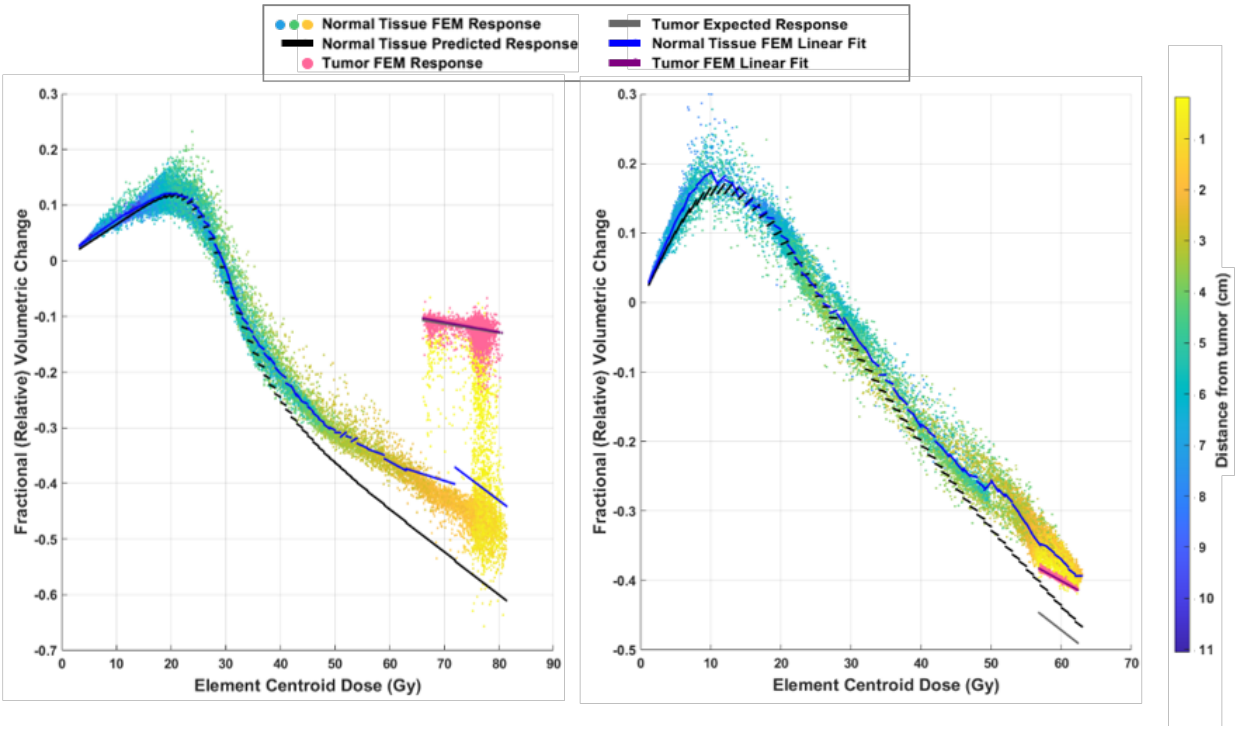


Figure 3.7. FEM analysis of tetrahedral element volume changes for Patient 1 (left) and Patient 6 (right) as a function of centroid dose compared to the predicted or expected volumetric response used as inputs in the FEA.

3.4.3 Liver Deformation Accuracy Analysis and Comparison

Volume Analysis and Comparison

Of the 29 segments modeled, 16 segments underwent volume changes exceeding the 95% CI for manual contouring of each segment location. For these 16 segments, the application of only Dose BCs resulted in 81% (13/16) of the segments modeled within the 95% CIs, whereas Morfeus (e.g., surface boundary conditions only) resulted in 63% (10/16). Application of the complete deformation process, Morfeus with Dose BCs, resulted in all (16/16) of the segments modeled to within the 95% CIs. This demonstrates that the proposed DIR technique performs volumetric modeling to within the variability of manual contouring.

Across all segments, the application of only Dose BCs resulted in final segment volumes with an average signed error of 2.3% (SD = 17.4%) and average absolute error of 14.5% (SD = 9.6%).

The average signed and absolute volumetric errors were 20.0% and 32.3%, respectively, for rigid registration, 9.8% and 16% for Morfeus, and 0.5% and 9.5% for Morfeus with Dose BCs. The application of only Dose BCs showed significant improvement over rigid registration and regular Morfeus (Wilcoxon, $p < 0.01$ and $p = 0.03$ respectively). The volumetric difference between application of only the Dose BCs and Morfeus with Dose BCs was not significant (Wilcoxon, $p = 0.50$), demonstrating that the majority of the volumetric change is occurring during the application of the Dose BCs, as intended in Step 1 of the deformation process. Tumor response, using a uniform linear expansion coefficient calculated for each patient's specific tumor response, was modeled to an average absolute error of 7.3% (6 cc, SD = 9.3%) using the Dose BCs.

Target Registration Analysis and Comparison

Average intra-observer variability for bifurcation selection on this seven patient data set, which is a limit for achievable TRE, was found to be LR: 0.9±0.2 mm, AP: 0.9±0.3 mm, and SI: 1.0±1.0 mm, giving a total vector uncertainty of 2.1±1.1 mm. Additionally, since a majority of follow-up scans had relatively large slice thicknesses (greater than 3 mm), precise localization of bifurcation points was limited axially and in-plane due to partial volume averaging affects. Figure 3.8 shows the average TRE vector for individual patients across the three registration methods: rigid, Morfeus, and Morfeus with Dose BCs. For all patients, the TRE improved with implementation of the Dose BCs.

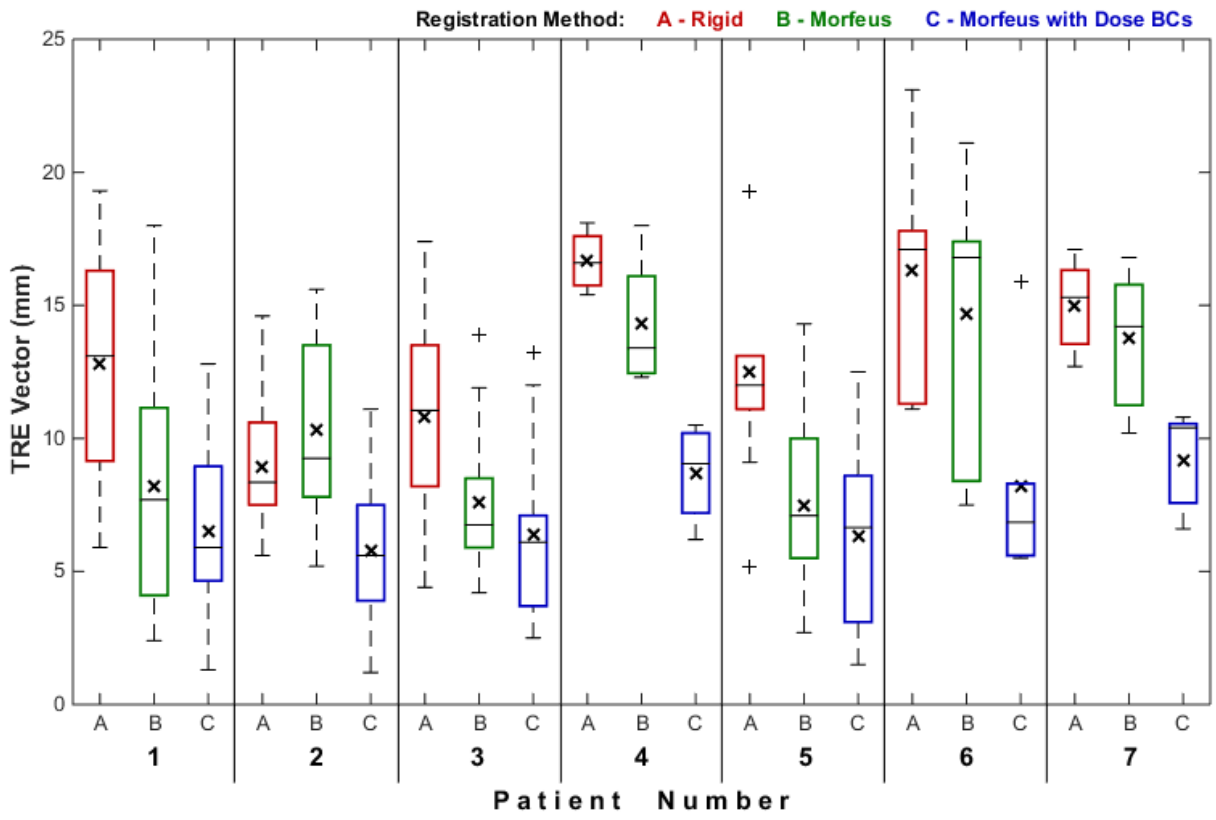


Figure 3.8. Boxplot of the mean patient Target Registration Error (TRE) for each of the three registration approaches: Rigid, Morfeus, and Morfeus with Dose Boundary Conditions (Dose BCs). The newly developed method, Morfeus with Dose BCs, showed an improvement in overall vector TRE for each of the seven patients when compared to rigid and standard Morfeus registrations.

Figure 3.9 shows the overall mean TREs for the three registration methods across all directions in addition to the overall TRE vector. In each direction, TREs were improved using Morfeus with Dose BCs, resulting in overall mean TREs of LR: 2.8±0.4 mm, AP: 2.8±1.0 mm, and SI: 5.1±2.4 mm, and giving an overall TRE vector of 7.3±1.3 mm. This represents a significant 44% and 30% improvement in the overall TRE vector when using Morfeus with Dose BCs as compared to rigid registration (Wilcoxon, $p < 0.01$) and Morfeus (Wilcoxon, $p < 0.01$), respectively.

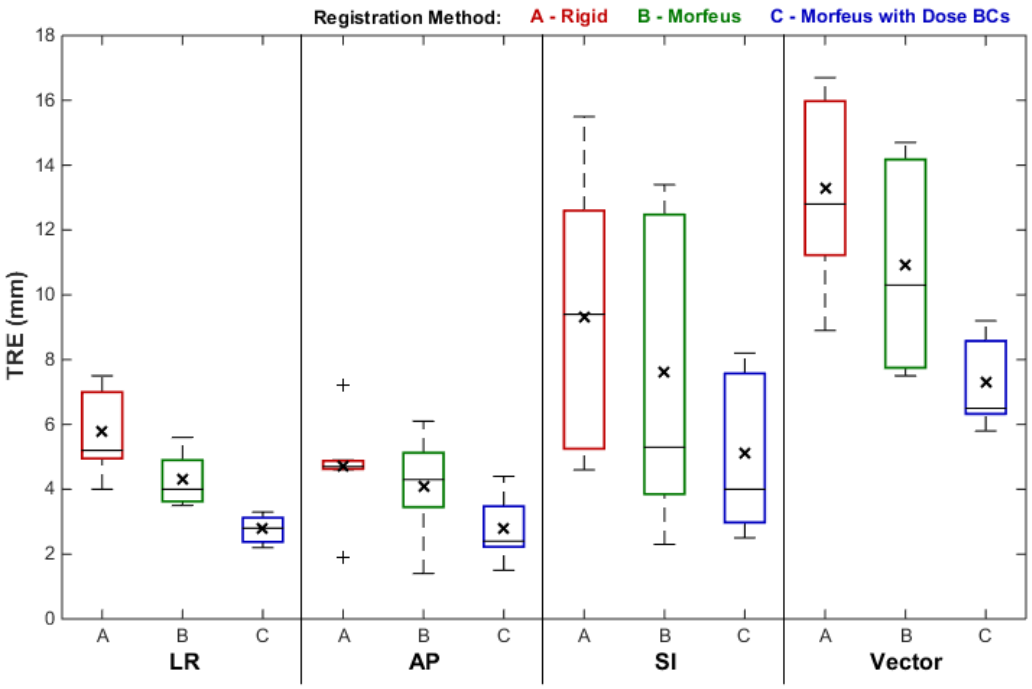


Figure 3.9. Boxplot of the mean overall Target Registration Error (TRE) for each of the three registration methods: Rigid, Morfeus, and Morfeus with Dose Boundary Conditions (Dose BCs). The newly developed method, Morfeus with Dose BCs, improvements in TRE for each direction and the overall vector when compared to rigid and standard Morfeus registrations.

3.5 Discussion and Conclusion

A biomechanical registration method has been proposed to more accurately model the complex liver deformations resulting from EBRT. The proposed DIR algorithm includes a newly developed dose-dependent volume response utilizing population-based normal tissue response models and modified FEA thermal expansion modeling to represent dose-based volume changes. The dose-volume response is applied in series with a previously developed spatial alignment biomechanical registration algorithm. This modified two-step deformation process has been evaluated for seven patients who were treated on the same treatment protocol as the patients used to generate the population dose-volume response model. In the future, this DIR technique could be applied to differing patient cohorts, such as SBRT patients, by modifying or regenerating the dose-volume response models presented in this study. For more modern RT delivery techniques, such as IMRT,

new models would likely need to be created to properly represent the presence of increased dose heterogeneity and steeper dose gradients across the liver. Additionally, modeling such patient cohorts may require additional refinement of the FEA process. Under these conditions, elements in the model may be even more prone to locking during the application of volumetric response component. To address this, the use of additional FEA steps, modification of the element types, or reworking of the material properties may be required.

Generation of the population dose-dependent volume response model showed a significant negative correlation between planned dose and volumetric response of individual liver lobes. Of the six proposed population stratifications, four stratifications demonstrated improved negative correlation when compared to the all-inclusive population model. Additionally, four of the six stratifications demonstrated improved log-logistic curved fitting metrics when compared to the all-inclusive model. When a stratification did not improve these metrics, the results were similar to the baseline model. Future work involving new and refined stratifications, including biological factors beyond tumor type and location, will require a larger sample population.

The proposed registration method, Morfeus with Dose BCs, demonstrated significant improvements in both volumetric modeling and TRE. Volumetric analysis of individual lobes showed that this deformation method successfully modeled all lobes within the respective contour variability of each lobe. The average TRE achieved with the proposed method was 7.3 mm (2.8 mm LR and AP, 5.1 mm SI), representing a significant 30% improvement over the previously published biomechanical registration algorithm (Morfeus).

Remaining error in the overall TRE vector is largely due to the residual error in the SI direction, which should be taken in the context of relatively large slice thicknesses on post-treatment scans, up to 10 mm (shown in Table 3.1). This error measurement could be improved when applying this

registration method to scans with improved slice thickness, in part because the reproducibility of manually selecting bifurcation points is dependent on slice thickness. Additionally, the modeling of the complex deformations observed in the liver might be improved by using corresponding vessel positions on the two CT scans as additional boundary conditions in the biomechanical model, as previously demonstrated in the lung.[104] However, this would rely on high-quality CT image contrast enabling visualization of vessels and the assumption of vessel preservation, which may not be a good assumption in the presence of large tumor response or cases of tumor induced thrombosis.

In this study, the development and evaluation of a population-based normal liver tissue dose-response model and application of the dose-volume response within a biomechanical DIR algorithm is reported. The use of the proposed liver registration algorithm is feasible and may aid in future studies investigating improvement in the accuracy of biomechanical deformable registration algorithms. This work has the potential for clinical impact in improving the correlation of functional imaging with delivered dose and enabling accurate longitudinal mapping of previously delivered doses to planning images for subsequent treatments. Furthermore, the methodology of the proposed DIR algorithm could be used to improve registration in other treatment locations in which structures undergo dramatic volume changes as a result of radiation, such as the parotid gland in head and neck RT.

3.5 Extensions

Normal tissue toxicities remain a major concern when treating patients with RT for head and neck cancer. Of these toxicities, xerostomia (dry mouth resulting from a reduction in saliva flow) remains prevalent and can lead to diminished quality of life after treatment.[105] Current clinical practice relies on a correlation of mean parotid gland dose to xerostomia for treatment planning; however, small animal studies have demonstrated that doses to sub-volumes of the parotid gland may be more predictive of this toxicity.[106] In the presence of patient weight loss, tumor shrinkage, and parotid gland volume changes during RT, improved DIR is required to more accurately correlate patient toxicity to the delivered dose to sub-volumes of the parotid. To address this, we applied similar dose-volume response modeling to the parotid gland for head and neck deformable registration. McCulloch et al. demonstrated that this approach significantly decreased the spatial registration error of biomechanical DIR of the parotid gland and that the improved registration accuracy could lead to meaningful differences in the estimated delivered dose when compared to a standard biomechanical method.[107] However, in order to determine the potential impact of these differences on current xerostomia toxicity models, this modeling method would need be applied across a larger cohort of patients.

Chapter 4
Incorporation of Patient-Specific Efficacy and Toxicity Estimates in Radiation Therapy Plan Optimization

Portions of this chapter are substantially equivalent to work the author has submitted for publication to *Medical Physics*.

4.1 Abstract

Current radiation therapy (RT) treatment planning relies mainly on pre-defined dose-based objectives and constraints to develop plans that aim to control disease while limiting damage to normal tissues during treatment. These objectives and constraints are generally population-based, in that they are developed from the aggregate response of a broad patient population to radiation. However, correlations of new biologic markers and patient-specific factors to treatment efficacy and toxicity provide the opportunity to further stratify patient populations and develop a more individualized approach to RT planning. We introduce a novel intensity-modulated radiation therapy (IMRT) optimization strategy that directly incorporates patient-specific dose response models into the planning process. In this strategy, we integrate the concept of utility-based planning where the optimization objective is to maximize the predicted value of overall treatment utility, defined by the probability of efficacy (e.g., local control) minus the weighted sum of toxicity probabilities. To demonstrate the feasibility of the approach, we applied the strategy to treatment planning for non-small cell lung cancer (NSCLC) patients. We developed a prioritized approach to patient-specific IMRT planning. Using a commercial treatment planning system (TPS), we calculate dose based on an influence matrix of beamlet-dose contributions to regions-

of-interest. Then, outside of the TPS, we hierarchically solve two optimization problems to generate optimal beamlet weights that can then be imported back to the TPS. The first optimization problem maximizes a patient's overall plan utility subject to typical clinical dose constraints. In this process, we facilitate direct optimization of efficacy and toxicity trade-off based on individualized dose-response models. After optimal utility is determined, we solve a secondary optimization problem that minimizes a conventional dose-based objective subject to the same clinical dose constraints as the first stage but with the addition of a constraint to maintain the optimal utility from the first optimization solution. We tested this method by retrospectively generating plans for five previously treated NSCLC patients and comparing the prioritized utility plans to conventional plans optimized with only dose metric objectives. To define a plan utility function for each patient, we utilized previously published correlations of dose to local control and grade 3–5 toxicities that include patient age, stage, microRNA levels, and cytokine levels, among other clinical factors. The proposed optimization approach successfully generated RT plans for five NSCLC patients that improve overall plan utility based on personalized efficacy and toxicity models while accounting for clinical dose constraints. Prioritized utility plans demonstrated the largest average improvement in local control (16.6%) when compared to plans generated with conventional planning objectives. However, for some patients the utility-based plans resulted in similar local control estimates with decreased estimated toxicity. The proposed optimization approach, where the maximization of a patient's RT plan utility is prioritized over the minimization of standardized dose metrics, has the potential to improve treatment outcomes by directly accounting for variability within a patient population. The implementation of the utility-based objective function offers an intuitive approach to biological optimization in which planning trade-offs are explicitly optimized.

4.2 Introduction

Advances in radiation therapy (RT) planning and delivery techniques have led to improved outcomes for cancer patients, both in terms of increased control rates and decreased treatment toxicity.[108-110] However, balancing disease control with adverse therapy effects continues to be a challenge in RT and across many other cancer treatment modalities. In external beam RT, this trade-off is inherent because delivery of radiation to the tumor often requires the use of beam trajectories that traverse adjacent normal tissue structures prior to and after converging on the target. Consequently, providing a sufficient therapeutic dose to the tumor may result in radiation-induced normal tissue toxicities, which can negatively impact a patient's quality-of-life during and after treatment. Modern RT techniques, such as intensity-modulated radiation therapy (IMRT), have made it possible to deliver highly non-uniform dose distributions with steep dose gradients between target locations and organs-at-risk (OARs). While this has enabled improved OAR sparing, optimizing an RT plan to best balance the individualized potential for disease progression and treatment toxicities remains hindered by conventional dose-metric-based planning techniques.

Conventional RT treatment planning attempts to address these trade-offs by optimizing radiation delivery based on predefined dose-based objectives that have either been previously correlated with measures of control and toxicity or determined from expert consensus and institutional standards. In inverse IMRT treatment planning, these surrogates of biological response are often incorporated into a single objective function through individually weighted linear or quadratic penalties, and the fluence map is optimized to minimize the sum of these penalties.[35] During this process, treatment guidelines and acceptable clinical trade-offs are implicitly translated and incorporated into the optimization objective by adjusting the relative importance of each dose-based penalty, relying heavily on the planner's experience and expertise

to guide the process.[111] Although a substantial amount of research has focused on weight selection to improve plan quality,[112-120] this optimization approach is limited by the inability to directly include individualized, quantitative estimates of the biological outcomes in the optimization process.

Recently, advancements in optimization approaches and availability of commercial products for biological optimization have facilitated the move beyond conventional dosimetric planning, allowing for the inclusion of tumor control probability (TCP) and normal tissue toxicity probability (NTCP) models in the plan optimization process.[37] Although implementations of available biological optimization methods differ, they are generally restricted in the ability to use both estimated biological response objectives and conventional dose-based objectives to drive the optimization process.[37,121,122] Instead, dose-based objectives may only be regarded as hard constraints in the optimization process, potentially limiting the ability to decrease dose to OARs with undefined biological response objectives.[37,121,122] Additionally, these systems continue to rely on aggregated response models and do not account for patient-to-patient variability in tissue radiosensitivity. Recent studies have focused on addressing this through the inclusion of personalized models in the planning process.[123-126] While these studies have demonstrated the ability to optimize treatments based on patient-specific outcome predictions, they remain limited to three-dimensional conformal radiotherapy (3DCRT) optimizing the beam angles and monitoring units[123-125] or prescription.[126] Additionally, these methods do not explicitly incorporate currently accepted clinical dose constraints which may impede clinical acceptance or diminish the realized benefit of the planning strategies.

To overcome these limitations, we propose an IMRT optimization method, termed Prioritized Utility Optimization (PUO), that augments the traditional dosimetric inverse treatment

planning process by directly incorporating quantitative estimates of personalized biological response while maintaining current dose-based planning objectives. This builds upon previous work that introduced the concept of utility-based planning, in which individualized optimal dose was selected by maximizing plan utility defined by a weighted combination of predicted efficacy and toxicity probabilities.[127,128] These studies demonstrated the ability of utility-based planning to increase efficacy while maintaining or reducing toxicity levels in a population, but were limited to scaling of fixed dose distributions. In the proposed IMRT optimization approach, we translate this methodology into a constrained hierarchical optimization problem that prioritizes maximization of plan utility over minimization of typical dose metrics and is subject to clinical hard constraints throughout. Through this approach, we add the ability to redistribute dose and aim to further exploit predictive models of efficacy and toxicity based on biological markers, clinical factors, and patient demographics, to improve the personalization of planning trade-offs. In this study, we introduce and evaluate the feasibility of the proposed method by applying it to a cohort of previously treated non-small cell lung cancer (NSCLC) patients.

4.3 Materials and Methods

4.3.1 Optimization Overview

Our proposed approach utilizes a two-stage optimization process to first maximize plan utility and then minimize dose-based metrics. Figure 4.1 provides the mathematical formulation of the optimization problems and Table 4.1 provides a full description of the notation used in defining the problems. In Stage 1, the objective is to maximize plan utility defined by the predicted efficacy probability based on target dose, minus the weighted sum of predicted toxicity probabilities across multiple OARs based on respective OAR doses. The efficacy and toxicity models are assumed to be monotonic functions of a dose metric, with increasing dose to the

structure never decreasing the predicted probability of the event occurring to that structure. The weighting parameter for each toxicity, θ, represents the undesirability of toxicity relative to the efficacy measure and can be adjusted based upon physician and/or patient preference, or tuned to result in an acceptable rate of predicted toxicity across a patient population.[128]

Stage-1 Optimization ***Objective: maximize overall plan utility***

$$\max.\ f_1(x) = \mathrm{P}_E(A^s x) - \sum_{\substack{s\in \mathrm{OARs} \\ m\in \mathrm{M}}} \theta^{s,m}\, \mathrm{P}_{\mathrm{T}}^{s,m}(A^s x)$$

Subject to:
(1a) $x \geq 0$
(1b) $\max(A^s x) \leq H^s \min(A^s x) : s \in$ target structure(s)
(1c) $\min(A^s x) \geq \mathrm{d}^s_{\min} : s \in$ target structure(s)
(1d) $\max(A^s x) \leq \mathrm{d}^s_{\max} : s \in$ target structure(s) and OARs with max hard constraints
(1e) $\mathrm{mean}(A^s x) \leq \mathrm{d}^s_{\mathrm{mean}} : s \in$ OARs with mean hard constraints
(1f) $\mathrm{CVaR}^{+}(A^s x, \Delta^k) \leq \mathrm{d}^{s,k}_{\mathrm{CVaR}^{+}} : s \in$ OARs with upper CVaR hard constraints

$$\text{where } \mathrm{CVaR}^{+}(A^s x, \Delta^k) = \min_{\zeta^{s,k}} \left\{ \zeta^{s,k} + \frac{1}{(1-\Delta^k)|V^s|} \sum_{j \in V^s} \max(0, A^{s,j} x - \zeta^{s,k}) \right\}$$

Stage-2 Optimization ***Objective: minimize dose to all OARs***

$$\min.\ f_2(x) = \sum_{s\in \mathrm{OARs}} \ell\mathrm{EUD}(A^s x, \alpha^s) + \mathrm{CVaR}^{+}(A^s x, \Delta^k)$$

$$\text{where } \ell\mathrm{EUD}(A^s x, \alpha^s) = \alpha^s\, \mathrm{mean}(A^s x) + (1-\alpha^s) \max(A^s x)$$

Subject to:
(2a) $x \geq 0$
(2b) $\max(A^s x) \leq H^s \min(A^s x) : s \in$ target structure(s)
(2c) $\min(A^s x) \geq \mathrm{d}^s_{\min} : s \in$ target structure(s)
(2d) $\max(A^s x) \leq \mathrm{d}^s_{\max} : s \in$ target structure(s) and OARs with max hard constraints
(2e) $\mathrm{mean}(A^s x) \leq \mathrm{d}^s_{\mathrm{mean}} : s \in$ OARs with mean hard constraints
(2f) $\mathrm{CVaR}^{+}(A^s x, \Delta^k) \leq \mathrm{d}^{s,k}_{\mathrm{CVaR}^{+}} : s \in$ OARs with upper CVaR hard constraints
(2g) $f_1 \geq \eta\, f_1^*$

Figure 4.1. Mathematical formulation of the optimization problems.

Table 4.1. Optimization notation glossary.

Symbol	Description	Symbol	Description
s	Structure index	$\theta^{s,m}$	Weight for structure s, metric m
V^s	Voxel set for structure s	P_E	Estimated probability of efficacy
x	Fluence map	$P_T^{s,m}$	Estimated probability of toxicity metric m for structure s
A	Influence matrix	H^s	Heterogeneity coefficient for structure s
A^s	Influence matrix corresponding to structure s	$CVaR^+$	Upper conditional value-at-risk
$A^{s,j}$	Influence matrix corresponding to voxel j in structure s	$d^{s,k}_{CVaR^+}$	Upper CVaR dose limit for structure s at level k
d^s_{min}	Minimum dose limit for structure s	Δ^k	Volume fraction for $CVaR^+$ metric at level k
d^s_{max}	Maximum dose limit for structure s	$\zeta^{s,k}$	Free variable for structure s, metric k
d^s_{mean}	Mean dose limit for structure s	f_1^*	Stage 1 optimal objective value
ℓEUD	Linear equivalent uniform dose	η	Relaxation parameter between utility metrics for Stages 1 and 2
α^s	ℓEUD weighting factor for structure s		

The maximization of plan utility in Stage 1 is subject to inviolable clinical criteria for the plan, guaranteeing that any solution meets dose-based hard constraints (1b–1f). These constraints are based on the following three rationales. First, for any individual patient, increasing predicted efficacy can only occur by increasing dose to the target. Since a single target dose metric is used to predict efficacy, maximization of utility through increasing efficacy may result in unacceptable target dose heterogeneity. This has previously been noted as a potential issue in biological optimization methods and is addressed through constraint (1b).[37] Second, it is possible that optimal solutions exist at dose values above or below levels that physicians would be comfortable prescribing. Hard constraints (1c) and (1d) are used to avoid these solutions and avoid

extrapolating model predictions to dose values beyond those in the dataset used to fit the models. Lastly, it is unlikely that every dose-limiting OAR will have a toxicity model corresponding to every dose metric typically used to limit OAR doses during planning. Therefore, constraints (1d–1f) are included to ensure that any solution meets or exceeds traditional OAR dose limits. To avoid non-convex functions of dose, dose-volume constraints (DVC), which are typically represented by value-at-risk (VaR) metrics (e.g., V_{20Gy}, $D_{0.1cc}$), are instead represented by upper conditional value-at risk ($CVaR^{+}$) metrics in constraint (1f). $CVaR^{+}$ represents a convex DVC that captures the *mean* upper-tail dose of a structure's dose-volume histogram and has previously been used to formulate linear RT optimization problems.[129-131] Figure 4.2 shows a visualization of both VaR and $CVaR^{+}$ on an example OAR dose-volume histogram (DVH). For a given VaR metric, calculated over the volume V_{metric}, the corresponding $CVaR^{+}$ metric is expected to be higher because VaR captures the minimum dose and $CVaR^{+}$ captures the mean dose.

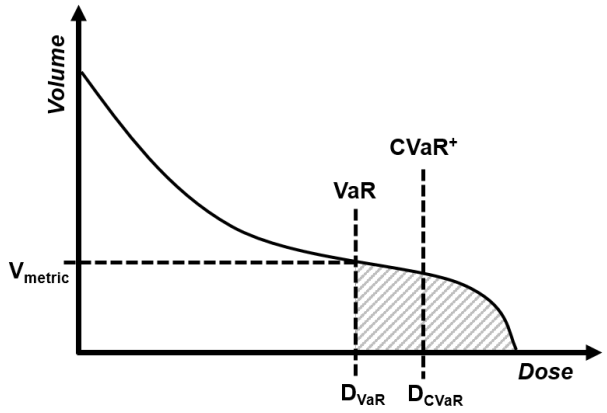


Figure 4.2. Visualization of VaR and CVaR+ on an example DVH with the relevant OAR region for calculation of the metrics shaded.

In Stage 2, the objective is to minimize dose-based metrics similar to traditional plan optimization approaches. Since Stage 1 effectively only optimizes the OAR dose metrics used in toxicity models for the utility function, OAR dose metrics not used in the toxicity models can still

potentially be reduced. Therefore, Stage 2 can be considered a solution "clean-up" or refinement step, in which all OAR doses are reduced while attempting to maintain the optimal plan utility from Stage 1. For computational efficiency, the objective function is a sum of linearized equivalent uniform dose (ℓEUD). ℓEUD is a convex piecewise-linear (PWL) approximation of generalized equivalent uniform dose and is calculated as a linear weighted combination of the mean and maximum doses for a structure.[132] In this formulation, the structure-specific parameter α defines the importance of max dose and mean dose, ranging from 0 to 1. Constraints (2a–2f) are simply carried over from constraints (1a–1f) in Stage 1. However, a new constraint, (2g), is added to preserve the optimal plan utility from Stage 1. This constraint allows for relaxation of the Stage-1 plan utility value to allow sufficient search space for the Stage-2 optimization to reduce OAR doses. The relaxation parameter, η, determines the allowable plan utility degradation between Stage 1 and Stage 2. Although a value of 1 is ideal (maintaining the optimal utility found in Stage 1), in practice a value slightly less than 1 may be required to provide the optimization algorithm sufficient search space to improve dose metrics in Stage 2.

4.3.2 Technical Implementation

Optimization first requires calculation of an influence matrix, A, which represents the dose contribution of a discretized beam to a set of voxels or points within a patient volume. To facilitate this process, we built a plugin using a research version of the Eclipse Scripting Application Program Interface (ESAPI v15.5) (Varian Medical System Inc., Palo Alto, CA) to allow integration of our process with our commercial Eclipse treatment planning system (TPS). Integration with the TPS provided access to patient images, structures, beam configuration, and dose calculation as needed throughout the planning process. Point clouds were generated using pseudorandom point sampling of pre-defined structures within a patient's body with adjustable

structure-specific point spacing. Multi-field IMRT plans were configured using pre-defined beam angle templates that were initially fit to the target structure(s) and then expanded to create a beam that could be divided into uniform beamlet sizes. To calculate the influence matrix, we used an ESAPI method (*CalculateInfluenceMatrixToMemory()*) that separates a beam into beamlets based on a user-defined size parameter, typically 0.5 cm or 1 cm based on the target size, and calculates a full-scatter influence matrix using the Analytical Anisotropic Algorithm (AAA v15.5.11) (Varian Medical System Inc., Palo Alto, CA). Given the size of the influence matrix, the data is then serialized using Google's Protocol Buffers for efficient transfer of the data to the optimization system. To reduce the overall size and complexity (i.e., number of variables and constraints) of the optimization problems, this clinically accurate influence matrix calculation was separated into primary and secondary dose contributions. Secondary dose contributions, consisting of low-dose scatter components, were removed from the influence matrix and summed for each point separately across each beam. The limit for these low-dose components was empirically determined by testing the impact of different limits on the optimization process. A limit of 0.001 Gy/100 MU was found to have marginal impact on the optimization results and resulting dose distributions. These contributions were then incorporated into the optimization dose estimation by multiplying the total low-dose component for each beam by the corresponding average beam fluence. Additionally, if the summed secondary contribution for a beam-point pair was found to be less than 0.0001 Gy/100 MU, the component was ignored since it is negligible in the context of primary dose contributions. The decision variables in the optimization problem were beamlet fluence intensities, x, and were forced to be non-negative with constraints (1a) and (2a).

Maximization of the utility function (f_1) requires that the efficacy model is concave relative to dose, and the toxicity models are convex relative to dose, within the relevant prediction range.

Although predictive models of efficacy and toxicity could take on many non-linear forms, they are typically represented by sigmoidal functions of dose (e.g., NTCP and TCP). Methods exist for transforming these types of partially convex and concave functions into purely convex or concave functions for the purposes of optimization and are useful in determining Pareto efficient solutions.[133-135] However, these transforms, including log transformations, obscure the relative importance of absolute changes in efficacy and toxicity in a weighted-sum approach and do not preserve the interpretability of the utility function and weights (θ's). Therefore, PWL relaxations of these models were implemented for the Stage-1 objective function and for constraint (2g) in Stage 2, reducing the problems to linear programs which can be efficiently solved with off-the-shelf commercial optimization solvers. These PWL approximations were uniformly spaced throughout the convex or concave envelope of the dose response function with extensions beyond the envelope acceptable within the maximum allowable error of 0.5% (absolute). Additionally, since many predictive models are based on biological dose, rather than physical dose, we implemented the ability for models to be based on equivalent dose in 2 Gy fractions (EQD_2) using PWL approximations for OARs and linear scaling for target volumes with approximation errors less than 1 Gy (EQD_2) over the relevant dose range. To eliminate approximation errors during plan evaluation, plan utility and individual components of the utility function were recalculated after optimization using the original functions. The resulting linear optimization problems were solved using a third-party commercial optimization solver Gurobi v9.0.3 (Gurobi Optimization, LLC, Beaverton, OR) through the .NET interface on a workstation with dual Intel Xeon E5-2620 v4 8-core processors and 64 GB of memory. We tested both the simplex and barrier (interior point) algorithms implemented in Gurobi to determine which method was more suitable for this set of optimization problems.

Following optimization, the optimal fluence map can be directly imported into the TPS using ESAPI, allowing for leaf sequencing, final dose calculation, and plan visualization. Since our optimization process does not include a fluence smoothing stage or intermediate full dose calculation, dose discrepancies after leaf sequencing and final dose calculation are expected but are not the focus of this initial feasibility study. Therefore, quantitative plan comparisons for this study are limited to optimization solutions prior to these steps when possible.

4.3.3 Retrospective planning study

For this initial demonstration of the proposed optimization strategy, five previously treated NSCLC patients were selected for retrospective treatment planning and plan comparison. These patients were selected to be representative of the variation in patient geometry, target size, and model inputs from a cohort of 125 stage II-III NSCLC patients previously treated on institution review board-approved prospective studies. Table 4.2 provides details for the five patients. All patients were treated with definitive standard or dose-escalated 3DCRT with or without sequential or concurrent chemotherapy. These studies included data collection previously used to generate predictive models of local regional progression-free survival at 2 years ($LRPFS_{2y}$),[136] grade ≥ 3 cardiac event within 2 years ($CE_{3+,2y}$),[137] grade ≥ 3 radiation esophagitis (RE_{3+}),[138] grade ≥ 3 radiation-induced lung toxicity ($RILT_{3+}$)[139] based on patient demographics, clinical factors, and biomarkers. We used these previously published models, further described in Table 4.3, to generate personalized predictive dose response curves, shown in Figure 4.3. These patient-specific predictions are utilized in Stage 1 of the PUO process for optimization and calculation of overall plan utility, as well as constraint (2g) of Stage 2. For this initial study, θ values were uniformly set to 1, representing equal pair-wise trade-off between efficacy and each toxicity in the utility function. To avoid use of a non-convex DVC in the optimization process, max dose was used in

place of D_{2cc} for the esophagus model, leading to slight over-estimation of RE_{3+} during optimization (typically < 0.5%, absolute). However, utility values for plan evaluation were calculated using D_{2cc} for the RE_{3+} prediction.

Table 4.2. Patient characteristics and overall plan utility metrics.

Target Details				Plan Utility ($D_{95\%}$ (Gy))				
Patient Number	PTV Volume (cc)	Laterality	Description	3DCRT	VMAT	DOO	$PUO_{S1,mod}$	PUO
1	124	Right	Near heart with separate posterior node	0.474 (64.3)	0.423 (60.1)	0.449 (60.1)	0.726 (85.4)	0.720 (83.0)
2	119	Right	Middle lobe, single target	0.298 (61.9)	0.414 (60.2)	0.427 (60.2)	0.729 (85.1)	0.728 (85.3)
3	333	Right	Superior lobe with mediastinal involvement	0.355 (62.1)	0.314 (60.0)	0.341 (60.0)	0.578 (84.8)	0.442 (65.3)
4	282	Left	Mid-superior lobe near esophagus with separate node	0.271 (71.3)	0.360 (60.0)	0.339 (60.0)	0.421 (60.6)	0.409 (58.6)
5	929	Left	Encompassing the superior lobe	0.206 (65.1)	0.294 (60.0)	0.295 (60.0)	0.382 (60.7)	0.371 (60.0)

Table 4.3. Models used in the calculation of plan utility for NSCLC. KPS = Karnofsky performance status.

Model	Reference	Structure	Dose Metric	α/β	Covariate(s)
$LRPFS_{2y}$	136	PTV	Mean EQD_2	10	Serum microRNA, age, concurrent chemotherapy, stage, KPS
$CE_{3+,2y}$	137	Heart	Mean EQD_2	2.5	Pre-existing cardiac disease
RE_{3+}	138	Esophagus	D_{2cc} EQD_2	10	Age, chemotherapy utilization, smoking status, KPS
$RILT_{3+}$	139	Lungs (-GTV)	Mean EQD2	3	Baseline cytokine levels, age, former smoking status, concurrent chemotherapy

IMRT plans using the PUO approach were generated for the five patients. These 30 fraction plans consisted of 8 non-opposing treatment fields placed at standardized beam angles based on target laterality. For the influence matrix calculation, beamlets were set to a size of 0.5 x 0.5 cm except for patient 5 which required 1.0 x 1.0 cm beamlets due to the large planning target volume (PTV) volume and computational limitations.

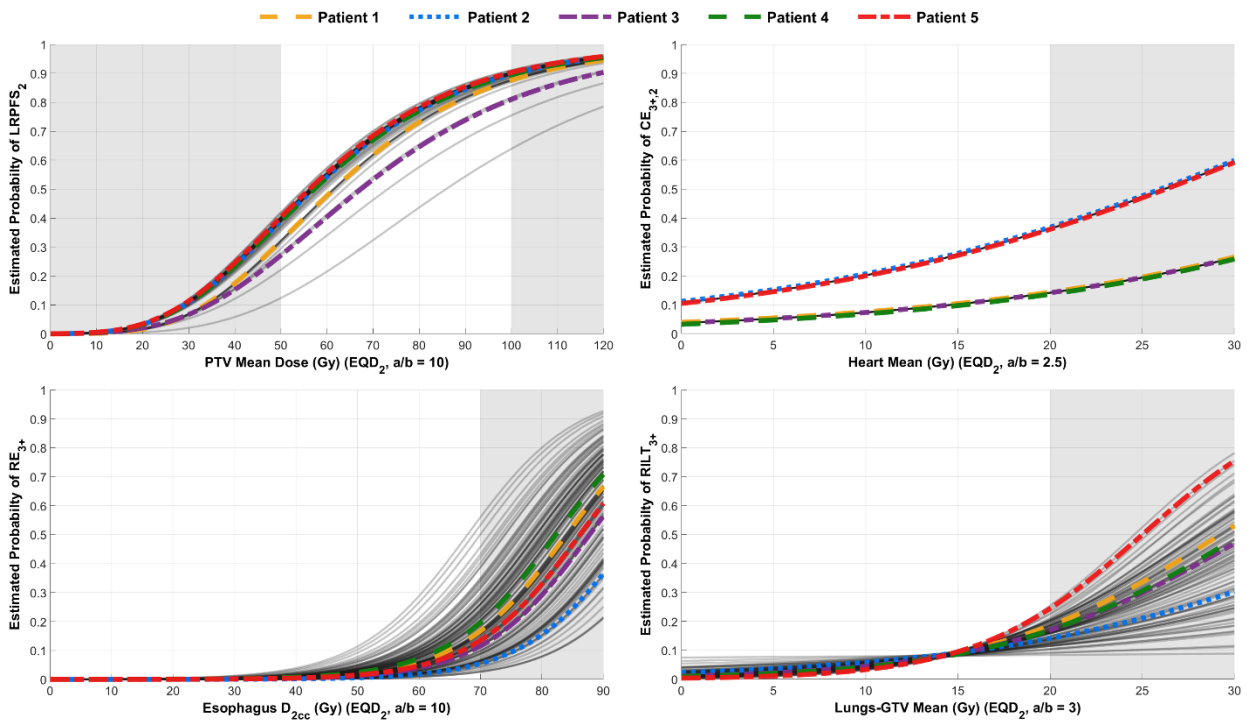


Figure 4.3. Predictive efficacy and toxicity models used to estimate overall plan utility (top left: LRPFS2y, top right: CE3+, bottom left: RE3+, bottom right: RILT3+). All cases included in the generation of the models are plotted with gray solid lines while the five cases used for plan comparison are plotted with colored dashed lines. White background areas represent the model ranges used in study, whereas the gray background represents ranges excluded by hard constraints during optimization. The CE3+ model only stratifies into two distinct groups since it is based on one binary covariate, baseline cardiac disease, with heart dose.

Structure constraints used in the optimization are shown in Table 4.4. $CVaR^{+}$ constraints were added as convex representations of the VaR-based DVCs used clinically at our institution, including lung-GTV $V_{20Gy} \leq 35\%$, heart $V_{30Gy} \leq 50\%$, and heart $V_{50Gy} \leq 25\%$. These constraints were tuned based on a fit of $CVaR^{+}$ to the corresponding VaR metric based on a retrospective analysis of IMRT plans from a separate cohort of 30 previously treated NSCLC patients. Figure 4.4 shows the correlation of $CVaR^{+}$ metrics to the corresponding VaR metrics. Given that the clinical VaR limits are stated as the percentage structure volume receiving at least a certain dose, the limit for $CVaR^{+}$ constraints was based on the relationship of VaR to $CVaR^{+}$ at the percentage structure volume limits. For the heart, VaR limits exceeded range of data available, so $CVaR^{+}$ were set conservatively based on the maximum VaR percentage structure volumes in the dataset. This

relationship is indicated as a red dashed line in Figure 4.4 and the new $CVaR^{+}$ constraints are listed in Table 4.4. Additionally, a $CVaR^{+}$ constraint for the esophagus was added as a conservative representation of $D_{2cc} < 68$ Gy. VaR dose metrics were checked after plan optimization to ensure that plans remained within the original VaR clinical limits. ℓEUD weighting factors were set to 0.5 for all structures with equal weighting between structures for the Stage-2 objective. Based on empirical testing, η was set to 0.999 (0.1% overall utility loss acceptable), constraining Stage 2 to retain nearly the same overall plan utility as Stage 1 while still providing the optimization algorithm sufficient search space.

Table 4.4. Structure hard constraints used for optimization.

Structure	Metric	Constraint
PTV	Mean	≥ 50 Gy EQD_2 ($\alpha/\beta = 10$)
	Mean	≤ 100 Gy EQD_2 ($\alpha/\beta = 10$)
	H [a]	≤ 1.15
Heart	Mean	≤ 20 Gy
	Max	≤ 107% PTV_{min}
	$CVaR^{+}_{50\%}$	≤ 45 Gy
	$CVaR^{+}_{25\%}$	≤ 55 Gy
Esophagus	Mean	≤ 34 Gy
	$CVaR^{+}_{2cc}$	≤ 68 Gy
Lungs - GTV	Mean	≤ 20 Gy
	$CVaR^{+}_{35\%}$	≤ 35 Gy
Cord	Max	≤ 45 Gy
Brachial Plexus	Max	≤ 66 Gy

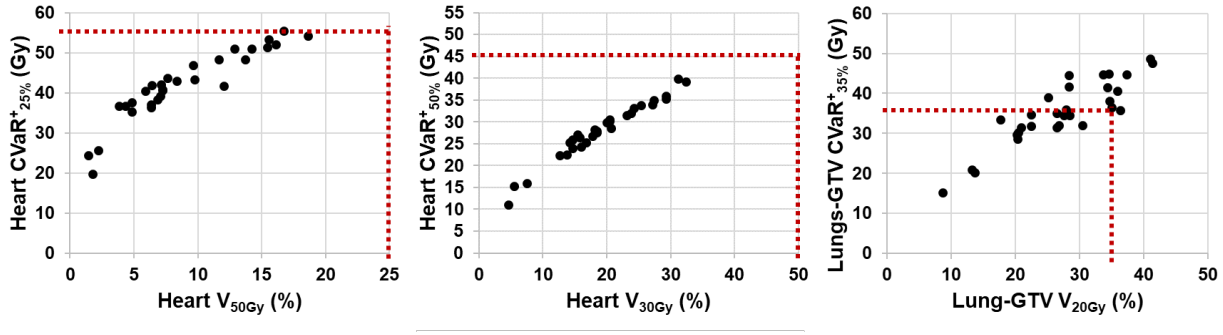


Figure 4.4. Clinical VaR (DVH) metrics and the corresponding CVaR+ metrics for 30 NSCLC patients previously treated with IMRT.

4.3.4 Planning Comparison

The proposed PUO method was compared to the clinically delivered 3DCRT plans and retrospectively generated dose-only optimization (DOO) IMRT, and volumetric modulated arc therapy (VMAT) plans. DOO plans were generated to represent plans with the same beam angles and hard constraints as the PUO plans but optimized with only a typical dose metric based objective. These plans were created using a similar workflow as the PUO plans, but only used Stage 2 of the PUO process and excluded the utility constraint (2g). An additional constraint for minimum target dose of 60 Gy (30 fractions) was added to produce plans that followed current clinical planning prescription guidelines. VMAT plans with a 60 Gy (30 fractions) target dose goal were generated in Eclipse by a dosimetrist with prior experience in clinical lung treatment planning. These VMAT plans were used as a modern, clinically deliverable benchmark for the DOO IMRT plans generated with our external optimization process and fixed beam angles. To provide a uniform comparison between planning methods, VMAT plans were normalized to the $D_{95\%}$ of the corresponding DOO plans (normalization values < 102%). 3DCRT, VMAT, DOO, and PUO plans were compared by calculating the final plan utility. DVHs of DOO and PUO plans directly from the optimal solutions were also compared. Additionally, results from Stage 1 of the optimization process without clinical OAR hard constraints(1c–1f), labeled $PUO_{S1,mod}$, were compared to the final PUO utility results to determine what effect, if any, the inclusion of OAR hard constraints had on the maximum achievable plan utility. Utility results are compared in terms of absolute utility difference. Therefore, given θ values of 1, any change in plan utility directly corresponds to a combination of changes in the predicted absolute probabilities of efficacy and toxicity. For example, if the predicted probability of efficacy increases by 5% (absolute, e.g., changing from 55% to 60% efficacy) with predicted toxicity remaining the same, then the overall

utility improvement would be 0.05. Alternatively, if the same increase in efficacy occurs, but one of the predicted toxicity probabilities increases by 7% (absolute), then the overall utility change would be 0.05−0.07 = −0.02.

4.4 Results

For all five NSCLC cases, our approach successfully generated optimal beamlet weights that maximize utility while remaining within dose-based constraints. Total (Stage 1 and Stage 2) solution times ranged from 16 to 71 minutes, with the number of beamlets and points in the optimization problems ranging from 1144 to 2610 and 41888 to 96523, respectively. On average, 78% of solver time was spent on Stage 2, of which 67% (or 52% of the total solve time) was spent specifically in crossover to produce a basic solution. For our linear problems, Gurobi's barrier method was found to be faster than primal and dual simplex methods but required modification of the numerical focus to the maximum setting (*NumericFocus = 3*) and decreasing the barrier convergence tolerance to 1e−5 to reduce crossover solver time.

Plan utility values and target $D_{95\%}$ values (representative of the prescription dose) for the various planning methods are shown in Table 4.2 with a further breakdown of the utility metrics shown in Figure 4.5. The PUO method resulted in an average absolute utility improvement of 0.21 (range: 0.09–0.43), 0.17 (0.05–0.31), 0.16 (0.07–0.30) when compared to 3DCRT, VMAT, and DOO plans, respectively. DOO IMRT plans were shown to have similar utility to VMAT plans, demonstrating that the DOO plans provide a clinically reasonable comparison for PUO plans. Analysis of individual components of the utility function showed that when compared to DOO, PUO improved absolute predicted $LRPFS_{2y}$ by, on average, 16.6% (range: 2.7–32.6%). Improvements in predicted efficacy were met with smaller changes in predicted grade 3–5 toxicities, with average changes of 0.1% (−0.5–0.5%), −0.9% (−4.0–1.7%), and 1.0% (−0.3–3.5%)

for predicted $CE_{3+,2y}$, RE_{3+}, $RILT_{3+}$, respectively. PUO resulted in increased target dose, based on analysis of PTV $D_{95\%}$, for three of the five patients, with one patient's PUO plan retaining a 60 Gy $D_{95\%}$ and another's deceasing target dose slightly. Comparison of the $PUO_{S1,mod}$ utility metrics to the final PUO plan metrics showed that the incorporation of clinically relevant hard constraints reduced the potential maximum plan utility by, on average, 0.03 (0.00–0.14). Patient 3 had the largest decrease, resulting mainly from the constraints on maximum esophagus and cord doses.

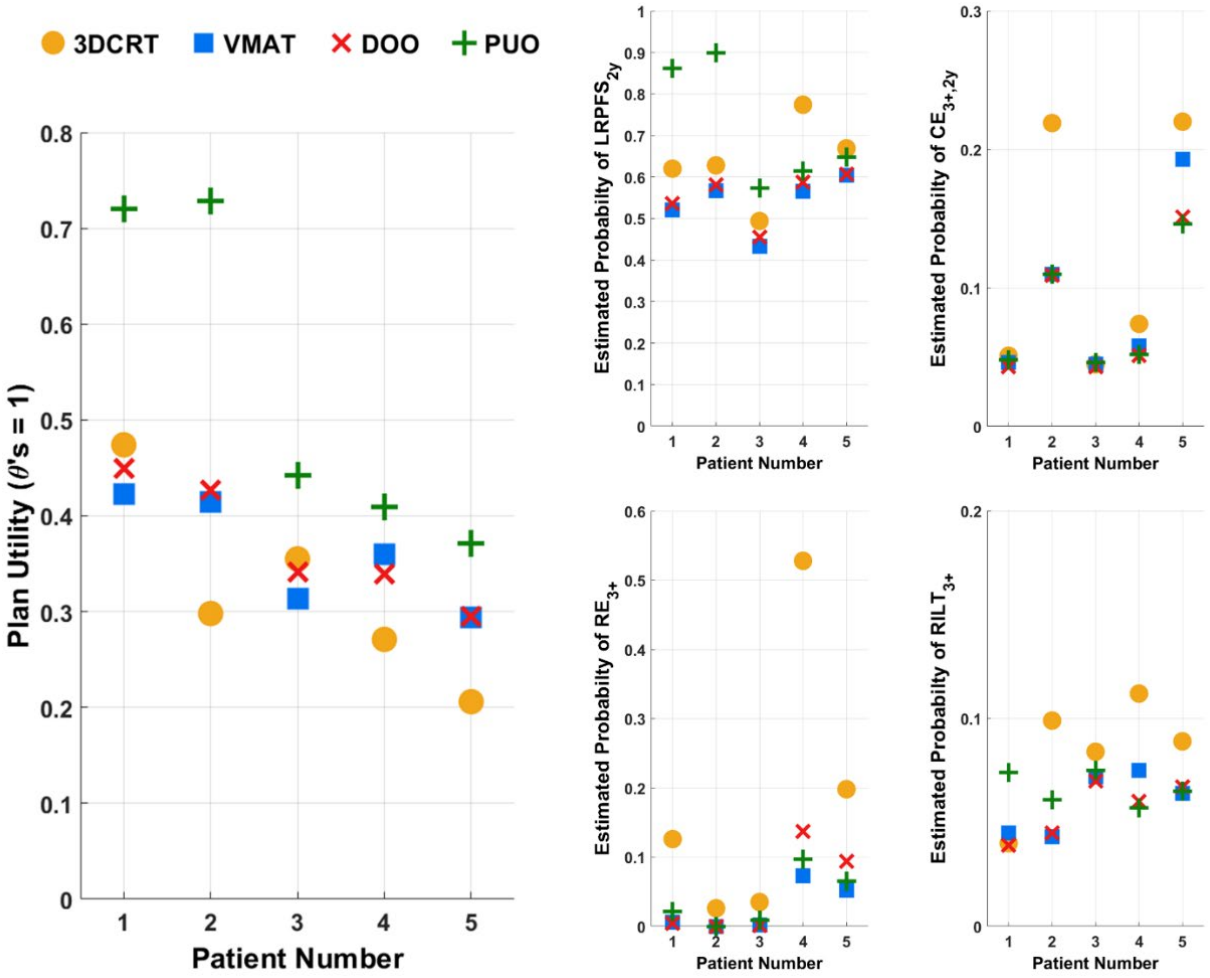


Figure 4.5. Utility results for the original clinically delivered 3DCRT plans and retrospective VMAT, DOO, and PUO replans for the five patients used for plan comparison. Results include overall plan utility (left), LRPFS2y (middle-top), CE3+ (right-top), RE3+ (middle-bottom), RILT3+ (right-bottom). As an example, Patient 1 estimated probabilities for LRPFS2y, CE3+,2y, RE3+, and RILT3+ were 53.6%, 4.3%, 0.4%, and 3.9%, respectively for Dose-Only Optimization (DOO) and 86.2%, 4.8%, 2.1%, 7.4%, respectively for Prioritized Utility Optimization (PUO). Therefore, the overall plan utility improvement using PUO is calculated as [0.862 − (0.048 + 0.021 + 0.074)] − [0.536 − (0.043 + 0.004 + 0.039)] = 0.719 − 0.450 = 0.269.

For patients 1 and 2, PUO led to large utility improvements over DOO. These improvements were mainly driven by increases in predicted $LRPFS_{2y}$, as visualized in Fig. 4.5, with the predicted probability of $RILT_{3+}$ moderately increasing and probabilities of RE_{3+} and RP_{3+} remaining similar. The PUO plan for patient 3 had a moderate increase in utility compared to the DOO plan caused by increasing predicted $LRPFS_{2y}$ with similar overall predicted toxicity risk. For patients 4 and 5, slight utility improvements in PUO plans were driven by improved predicted $LRPFS_{2y}$ and decreased probabilities of RE_{3+} and $RILT_{3+}$. For Patient 5, PUO planning also slightly decreased the predicted probably of $CE_{3+,2y}$.

Absolute and relative DVHs in Figs. 4.6 and 4.7 show comparisons of the DOO and PUO plans for patients 2 and 4, respectively. For patient 2, a large increase in target dose was noted with non-proportional increases in esophagus and cord doses. For patient 4, the PUO method slightly improved target coverage and decreased mean lung dose and maximum esophagus dose. Mean cardiac dose remained similar between the two methods, but the PUO plan increased maximum cord dose by 24.8 Gy. Variable increases in maximum cord dose were present in all PUO plans, since a cord toxicity risk model was not included in the Stage-1optimization model. However, all maximum cord doses remained below 45 Gy as enforced by the hard constraint.

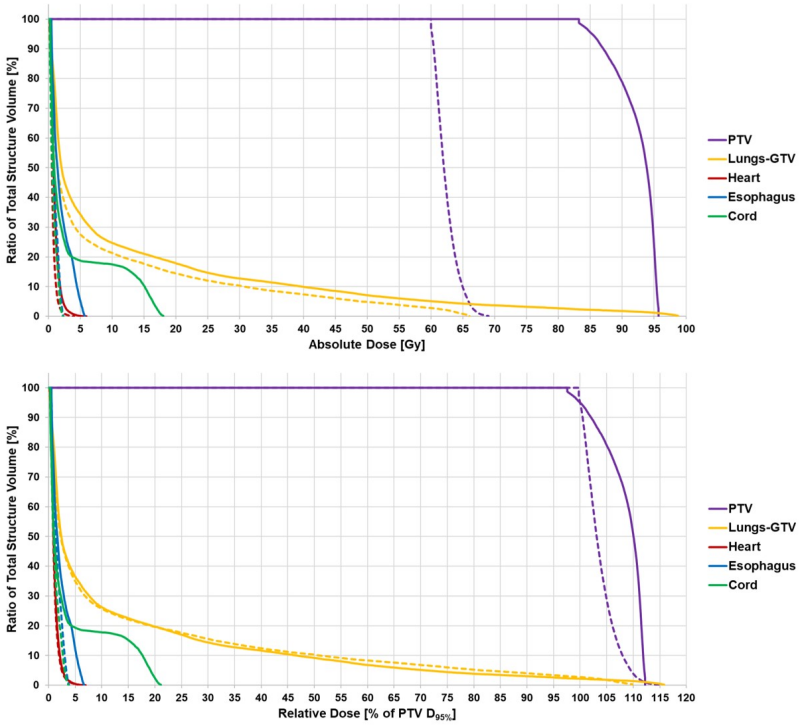


Figure 4.6. Absolute (top) and relative (bottom) DVH comparisons between the DOO plan (dashed) and the PUO plan (solid) for Patient 2.

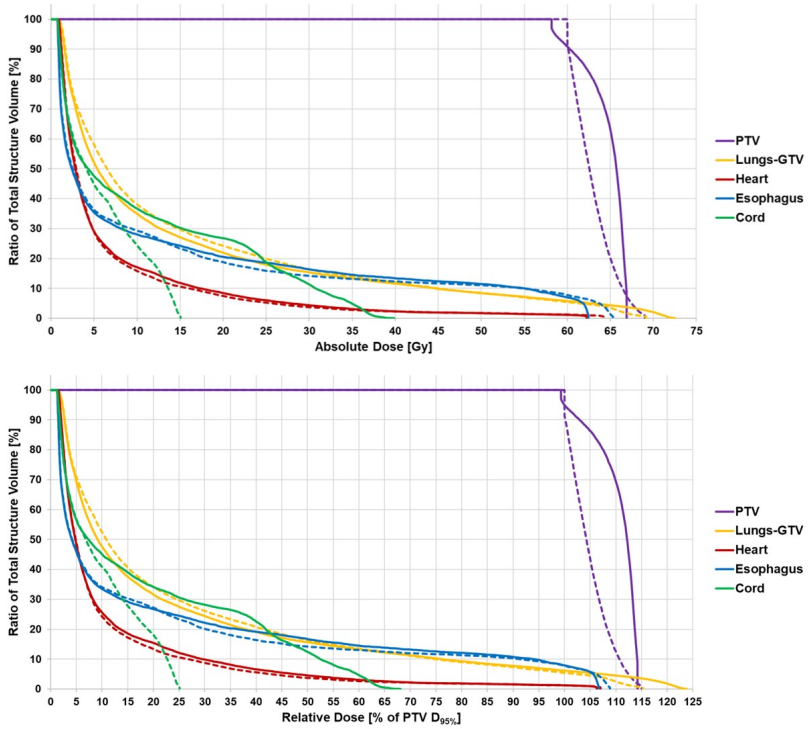


Figure 4.7. Absolute (top) and relative (bottom) DVH comparisons between the DOO plan (dashed) and the PUO plan (solid) for Patient 4.

4.5 Discussion

The current clinical standard for RT planning, employing objective functions based upon standardized dose metrics, does not adequately account for patient variability within the optimization process. In many cases, these objectives are based on historical rates of control or toxicity in a patient population without incorporating personalized factors related to the radiosensitivity of tumors and OARs. Additionally, most cost functions require user-assigned weights for the dose penalties, relying on the expertise of planners and physicians to subjectively determine the appropriate weights for a patient. Through this method, allowable planning trade-offs, based on treatment outcomes rather than dose metrics, could be inconsistently decided upon since they are not directly prioritized or optimized. The proposed method presented in this study aims to augment this standard approach by facilitating a direct exploration of planning trade-offs based on patient-specific predictive models of efficacy and toxicity. Through the implementation of a utility-based objective, we believe our strategy offers a more intuitive, tunable approach to balance a patient's potential therapeutic benefit and risk. The proposed method integrates the ability for non-uniform weighting of predicted toxicities in an interpretable manner with direct relation to the predicted probabilities of treatment outcomes. Therefore, toxicity weights (θ's) can be elicited directly by determining the relative harm associated with a given toxicity compared to a given efficacy metric. These values could be assigned independently by a clinician prior to plan creation, potentially incorporating patient input, or based on expert consensus opinion determined through a method similar to Hobbs et al.[140]

We demonstrated the feasibility of this method in a small cohort of NSCLC patients, which is a population of interest given the evident trade-offs between local control and lung, heart, and esophageal toxicities in RT planning. Additionally, failure of the RTOG 0617 randomized trial to

demonstrate improved outcomes at higher doses suggests that utilization of dose escalation in the treatment of NSCLC may only be beneficial for a subset of patients.[136,141,142] Our method provides a quantitative approach for determining which patients may benefit from dose escalation or redistribution based on patient-specific clinical factors and biomarkers while also accounting for patient geometry and OAR dose limits.

While this study illustrates the potential for the proposed planning strategy, it does have several limitations. Although we compared plan utility metrics between multiple planning methods, DVH comparison was limited to the DOO and PUO plans prior to leaf sequencing and final volumetric dose calculation due to dose discrepancies between optimal plans from the optimizer and plans imported, sequenced, and calculated inside of our TPS. This discrepancy has been previously noted in other optimization methods and is normally addressed through updating the optimization process with intermediate full dose calculations.[143] However, in our hierarchical approach, this update process could invalidate hard constraints. Future work will look at addressing this discrepancy through the implementation of a fluence smoothing optimization stage, heterogeneity constraints, or a correction step similar to that presented in Zarepisheh et al.[144]

Our study did not evaluate the effect of beam angle selection on the optimization process or achievable plan utility. Further investigation will be required to determine this impact of this and if utility could be improved through beam angle optimization. Alternatively, arc-based IMRT delivery methods, such as VMAT, could be used to alleviate the need for beam angle selection. However, constrained hierarchical optimization techniques for VMAT are currently limited and will require additional development.

The PUO approach facilitates the use of personalized dose-response models in the optimization process but aggregated response models could also be utilized. In testing this method

for NSCLC patients, we used previously developed models of response based upon patients treated with 3DCRT. Ideally, modeling would be based on a cohort that was treated with the same delivery technique as that being optimized since dose response could be, in part, correlated to the underlying dose distribution. This distribution, particularly within OARs, could systematically differ between varying treatment delivery techniques. While our approach optimizes a patient's tumor dose, rather than assigning a fixed dose goal prior to optimization, modeling for this aspect is challenging because it requires previously treated patients to receive a range of prescription doses to determine the relationship between treatment efficacy and dose. In the era of increasingly uniform prescription doses, such as the 60 Gy standard-of-care for NSCLC patients, modeling dose-based efficacy may not be fully possible. In the absence of an efficacy model, the PUO strategy could still be utilized to reduce total toxicity burden at a fixed prescription dose.

Our initial results from this method demonstrate that plan utility for NSCLC patients has the potential to be improved, although drawing additional conclusions from the results is limited by the small sample size. Additionally, we did not investigate the effects of potential error and uncertainty in the dose-response models or the selection of θ values on the optimization process and resulting plans. Future studies will aim to address this through application of the PUO strategy in additional disease sites and on larger cohorts of patients.

4.6 Conclusions

We developed and studied an inverse IMRT planning strategy where patient-specific radiosensitivities of tumors and normal tissues are directly factored into the optimization objective. Through this approach, we aim to improve a patient's overall RT outcome by balancing potential therapeutic benefit with the associated risk in an interpretable and tunable manner. First, a patient's overall plan utility, based upon personalized models of biological response, is maximized subject

to relevant clinical constraints. Then, through a hierarchical optimization technique, a typical dose-based objective function is minimized while attempting to retain the maximal plan utility.

We demonstrated the feasibility of the approach using a cohort of NSCLC patients with previously developed predictive models of treatment efficacy and toxicity based on demographics, clinical factors, and biomarkers. The proposed planning method generated plans conforming to clinical constraints with improved overall utility when compared to plans generated using typical dose-based objectives.

4.7 Extensions and Future Work

4.7.1 Visualization of planning trade-offs

Development of the previously-described, customized RT optimization framework has allowed for additional testing of optimization concepts and ideas to inform future RT planning strategies. One important consideration for future treatment planning strategies, including functional avoidance planning, is that the ability to redistribute dose between normal tissues, while still achieving acceptable disease control, may be inherently limited by current treatment planning guidelines. Depending on patient geometry and treatment goals, a treatment planner may only be able to reduce the dose to a given OAR by a limited amount before another OAR dose exceeds a guideline. Additionally, trade-offs between reducing dose to multiple OARs may not be uniform, with the reduction of dose to one OAR potentially causing a disproportional increase in dose to another OAR. In many clinical RT optimization systems, it may be challenging to understand and visualize these trade-offs within the range of acceptable clinical dose limits because systems do not allow for the inclusion of hard constraints that cannot be violated during plan optimization. Since our optimization facilitates the inclusion of hard constraints, we tested a method to map the clinically acceptable trade-off space between competing OARs. In part, this relates to the concept

of multi-criteria optimization (MCO), also known as multi-objective optimization, which has been an ongoing area of development in RT optimization for the last decade.[145-147] In MCO IMRT optimization, competing planning objectives (e.g., dose metrics, tumor coverage metrics) are assigned varying weights during plan optimization to determine the Pareto optimal front for treatment planning.[145] A given RT plan is considered to be on the Pareto optimal, or non-dominated, front if none of the planning objectives can be improved without degrading a competing planning objective. We are working to extend this concept to lung RT planning in the context of our optimization system.

To explore the clinically available trade-off space in lung RT planning, we optimized four IMRT plans for each patient with clinical dose guidelines used as hard constraints. For three of the plans, we generated non-dominated RT plans by focusing the optimization on only one relevant OAR dose metric as a planning objective, either lung-GTV mean dose, esophagus max dose, or heart mean dose. Hard constraints were used to maintain a minimum PTV dose of 60 Gy with less than 115% hot spots. We also generated a representative plan in which all three planning objectives were included in an equally-weighted objective function. Figure 4.8 shows a visual comparison between the dose metrics for the four plans in the form of triangular radar plots for two of the patients. Each vertex on the solid colored lines within the plots represent dose metrics for one of the three non-dominated plans and the dashed-line vertices represent the equally weighted plan. Although linear trade-offs are not necessarily clinically feasible, plotted linear connections between dose metrics for a given plan are used to help visualize the tradeoff space. Within each triangle, the remaining white area represents a range of doses that are feasible and clinically acceptable. The interior gray-shaded region represents plans that are clinically infeasible, the exterior gray-shaded region represents plans that are suboptimal. It is important to note that the

linear relationship between the competing dose metrics is shown for visualization purposes and may not be directly representative of solutions balancing multiple objectives.

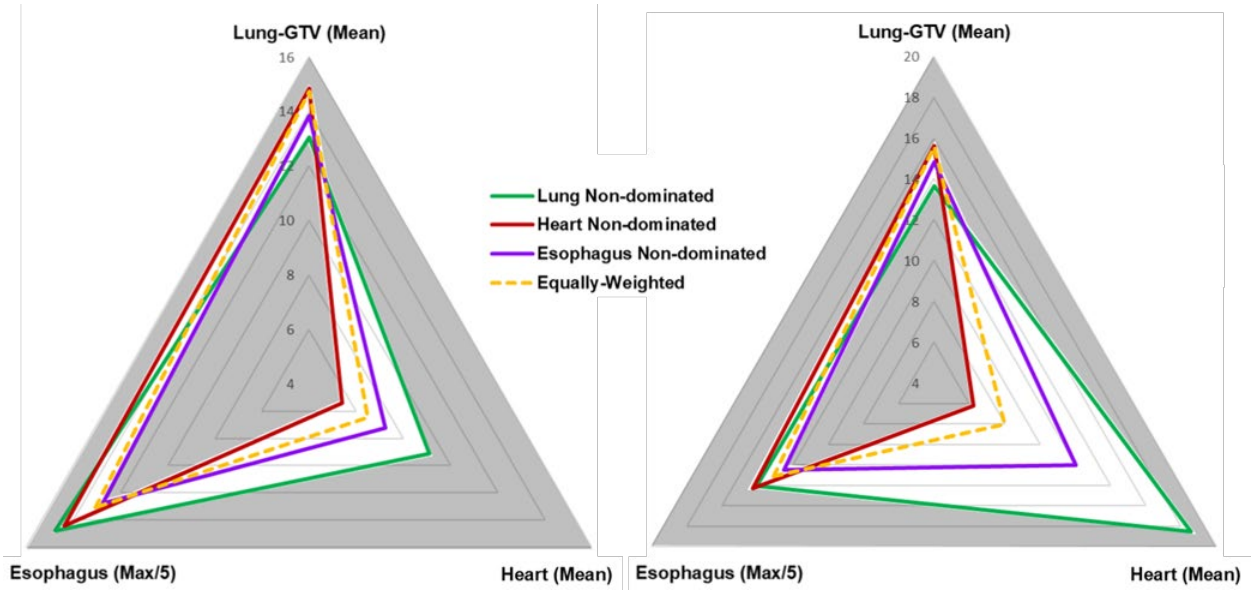


Figure 4.8. Example visualization of dose trades-off for lung IMRT plans for two patients, Patient A (left) and Patient B (right).

For both example patients, it is evident that reductions in mean dose to the lung disproportionally increase the mean dose to the heart. For Patient B, a 2 Gy decrease in mean lung dose could result in a 14 Gy increase in mean heart dose. Additionally, reducing lung dose may disproportionally increase esophagus max dose, with a 1.5 Gy decrease in mean lung dose for Patient A increasing the esophagus max dose by nearly 10 Gy. In both cases, focusing the optimization only on reducing mean lung dose provides only up to a 2 Gy decrease in dose, possibly demonstrating that reductions in mean lung dose are inhibited by the included hard constraints. While this initial work demonstrates that certain trade-offs between OAR doses may be inherently limited and that decreasing dose to certain OARs may disproportionally increase dose to other OARs, determining the underlying trends between trade-offs for lung RT will require application of this method over a larger cohort of patients.

4.7.2 Physician Survey of utility weights

One potential benefit of the utility optimization approach is that it directly incorporates the ability for physicians, and potentially patients in the future, to prioritize meaningful components of the objective function for RT plan optimization. Historically and in many current clinical applications, the priorities for treatment planning objectives are defined by setting weights for various dose metrics grouped into a weighted-sum objective function. Although these dose metrics are typically related to a treatment endpoint or toxicity, their relationship to an outcome is generally an implicit translation of a toxicity model. Given the typical sigmoid-shaped outcome models (e.g., TCP, NTCP) used in radiation therapy, the combined weight and dose metric is unlikely to accurately represent the changing probability of a treatment outcome across a range of doses. Therefore, it is difficult to derive the clinical meaning from changes in dose while comparing plans or optimizing a plan. Although previous studies have attempted to address this through the incorporation of biological models in the optimization process, current clinical biological optimization tools either do not allow weighting of the biological objective function or the clinical meaning of assigned weights are obscured due to the use of convex transforms of the models for optimization.[133-135] Unlike previously developed biological optimization tools, our implementation of utility-based planning retains the clinical meaning behind assigned weights because the models are not transformed but rather converted into highly representative PWL approximations. Since this weighting approach deviates from the current standard-of-practice in RT planning, we have started to gather physician input on how to weight the probability of toxicity outcomes for lung RT.

An online survey, shown in Appendix B, was designed and distributed using Qualtrics (Provo, UT) to gather preliminary feedback from physicians for weighting grades 2–5 radiation-induced lung toxicity, radiation esophagitis, and cardiac events within 2 years of treatment. In the survey, physicians are asked to assign relative weights to toxicity grades independently based on local-regional progression having a fixed weight of 100. In this context, an assigned relative weight of 0 means that the toxicity is not harmful, a weight of 100 means that the toxicity is equally as harmful as local-regional progression, and a weight of 200 means that the toxicity is twice as harmful as local-regional progression. These assigned weights can then be directly included in the utility function by dividing by 100 to normalize the relative weights to the efficacy metric having an implicit weight of 1. Table 4.5 shows the preliminary results from the survey completed by 7 physicians with varying experience practicing in radiation oncology and varying knowledge of the utility-based optimization project. This early feedback suggests that radiation esophagitis is considered less harmful than other toxicities across grade 2–4 toxicities. However, all grade 5 toxicities, toxicities resulting in patient death, averaged similar weighting in the range of 220. This suggests that grade 5 toxicities are consistently considered at least twice as harmful as local-regional progression. Physician variability in the assigned weights also appears to be large, with grade 3 assigned toxicity weights ranging from 5 to 80, and grade 4 assigned toxicity weights ranging from 20 to 154. It is yet to be determined whether this large range is a result of variable understanding of the relative weight concept, or if it represents true variability in how physicians currently consider treatment toxicities. Additional survey responses and feedback will be required to help make these determinations. Furthermore, additional consideration is required to convert the individually weighted toxicity grades into a single weight for toxicities grouped across a range of grades. For example, in our lung utility planning study, toxicity outcomes include grades 3–5

in a single model due to overall low incidence of the individual toxicity events within each grade. Although we could ask physicians to assign weights based on grouped grades of toxicity, initial feedback showed that it is challenging to assign a weight without a clear understanding of the underlying distribution of individual toxicity grades within a group.

Table 4.5. Preliminary results from an online survey asking physicians to quantify relative weights for lung utility-based planning.

	Average relative weight (range)			
Toxicity	Grade 2	Grade 3	Grade 4	Grade 5
Radiation-induced lung toxicity	15 (0−50)	25 (5−75)	80 (20−135)	229 (150−500)
Radiation esophagitis	10 (0−25)	28 (5−50)	71 (25−125)	221 (150−500)
Cardiac events within 2 years of treatment	17 (0−50)	42 (5−80)	83 (25−154)	229 (150−500)

4.7.3 Head and neck total toxicity burden optimization

Head and neck cancer (HNC) RT typically results in high doses of radiation to OARs, contributing to decreased quality-of-life (QOL) resulting from of xerostomia and dysphagia.[148-150] Current RT planning methods attempt to control for these types of toxicities by minimizing dose-volume parameters during optimization. However, dosimetric trade-offs made during treatment planning may be inconsistent with improving or balancing a patient's overall predicted total toxicity burden (TTB) because the relationship between physical dose and the response of OARs is typically considered non-linear. Prior studies have incorporated toxicity models in the treatment planning process but are limited by the use of a combined, singular dose and biological objective function, or by the implementation of convexified toxicity models that obscure the relative importance of each toxicity metric in the objective function.[135] By modifying our utility

optimization approach, we can directly optimize and potentially reduce the predicted TTB of a patient's RT plan while remaining within current clinical dosimetric guidelines.

We implemented a prioritized approach to TTB optimization using the in-house optimization plugin developed for prioritized utility-based planning. In this case, Stage 1 of the optimization process was reconfigured to minimize TTB, defined as the weighted-sum of toxicities. This change is effectively equivalent to using a fixed efficacy metric in the utility equation. Prioritized TTB (P-TTB) plans were optimized using a similar two stage hierarchal method for inverse IMRT planning. In the first stage, TTB is minimized subject to high-priority OAR dose limits, target prescription, and target coverage requirements. Weights for the toxicities directly represent the relative undesirability of toxicities. In the second stage, a weighted-sum dose-metric objective is minimized, similar to current clinical practice. This stage is subject to the same constraints as the first stage, but an additional constraint for TTB is incorporated based on the optimal solution of the first stage.

Due to the complexity of HNC plans, with multiple small OARs typically in close proximity to the target volumes, an improved point sampling method was implemented. Instead of generating sets of pseudorandom points for the influence matrix calculation, we implemented point sampling based on a low discrepancy (quasirandom) sequence with the ability to sample a portion of points directly on the surface of the structure. The additive recurrence sequence, when combined with pseudorandom localized jittering, improved mean distance between points and minimum distance between sampling points by approximately 20%. With the addition of surface point sampling, this allowed for decreased point sampling density, reducing optimization model complexity, while still maintaining an accurate representation of the structure dose. Additionally, to reduce optimization solver time, the usage of crossover to produce a basic solution from the

barrier solver solution was eliminated. While this means that small infeasibilities may still exists in the final solution, these infeasibilities were found to have negligible impact on the final dose distribution. Initial tests suggest that these modifications reduce model complexity and solver time without degrading plan quality, and may also improve agreement between the optimized plan and deliverable plan (generated after leaf sequencing and final dose calculation). In the future, these modifications will be applied to lung prioritized utility-based planning to determine if these changes improve the optimization process and impact optimization results.

Using this approach, we retrospectively generated IMRT plans for 5 HNC patients and compared the resulting plans to IMRT plans generated without the added TTB constraint. Both plans consisted of 9 equally-spaced, non-opposing beams with the same OAR dose limits and dose-metric objective. Target dose requirements were based on simultaneous integrated boost prescriptions of 70 and 56 Gy delivered in 35 fractions. TTB was calculated using previously published NTCP models for dysphagia and xerostomia.[151,152] Dysphagia was based on increase in aspiration or HNQOL score > 81 as a function of pharyngeal constrictor (PC) mean dose.[152] Xerostomia was based on parotid flow ratio <25% of pretreatment (grade 4) or xerostomia questionnaire > 73 as a function of parotid gland (PG) mean dose.[151]

Our optimization engine successfully generated P-TTB plans for all five patients while maintaining similar target coverage to the standard dosimetrically optimized plans. P-TTB plans resulted in average absolute NTCP reductions of 15.5% (range: 0.2–53.4%) for dysphagia and 20.5% (1.9–68.5%) for xerostomia. These changes correlate to average mean dose decreases of 6.9 Gy (0.2–22.7 Gy) for PC and 10.2 Gy (1.2–33.5 Gy) for PGs. P-TTB planning resulted in greater fluence modulation and higher doses to OARs, such as the larynx, not included in the calculation of TTB.

The P-TTB planning, similar to the utility method, offers an intuitive optimization approach to balancing trade-offs between various treatment outcomes and has the ability to directly reduce and redistribute predicted toxicity which may aid in improving QOL for HNC patients receiving RT. Future studies will focus on the incorporation of personalized toxicity models for HNC patients in addition to addressing concerns of increased plan modulation and dose to non-modeled structures.

4.7.4 Virtual Radiotherapy Trial Suite

Retrospective testing of new treatment planning strategies is critical to inform design and determine the potential clinical benefit from new interventions. With increasing complexity and personalization in modern treatment planning, the evaluation of new treatment planning techniques prior to clinical implementation would benefit from more robust testing across larger cohorts of patients. However, testing and evaluation has historically required a large amount of manual effort dedicated to generating treatment plans, manipulating parameters, and extracting data for analysis. To address this, we are currently developing a software framework, called the virtual radiotherapy trial suite, that will allow larger scale testing and analysis of new treatment planning techniques, algorithms, and optimization strategies.

This software suite interacts with our TPS through ESAPI to access clinical images and commercially available planning algorithms (e.g., dose calculation, knowledge-based planning). Data from the TPS, along with data from external sources, can then be used to test various treatment strategies under consideration for clinical implementation, such as utility or TTB optimization. In addition to testing treatment planning techniques, the framework can also be used as a data mining and extraction tool for rapidly gathering and exporting data from the Eclipse/Aria environment with limited user-interface interaction. As shown in Figure 4.9, the system is set up

as a flexible framework that manages batch tasks that compose a 'trial' for a set of patients. The tasks are developed independently of the system and compiled into function code blocks that interact with the suite through a common pathway for data transfer and progress reporting. While this system is currently still under development, we have recently demonstrated the ability to rapidly test different dose calculation algorithms on a cohort of lung patients (example user workflow shown in Appendix C). In the future, we will utilize this system to test prioritized utility and TTB optimization across larger cohorts of patients to determine the potential clinical impact of the methods. Additionally, this will allow us to more rapidly test the effects of differing biological models, component weights, patient factors, and optimization parameters on the treatment planning process with the goal of developing clinically relevant implementations of the proposed treatment planning strategies.

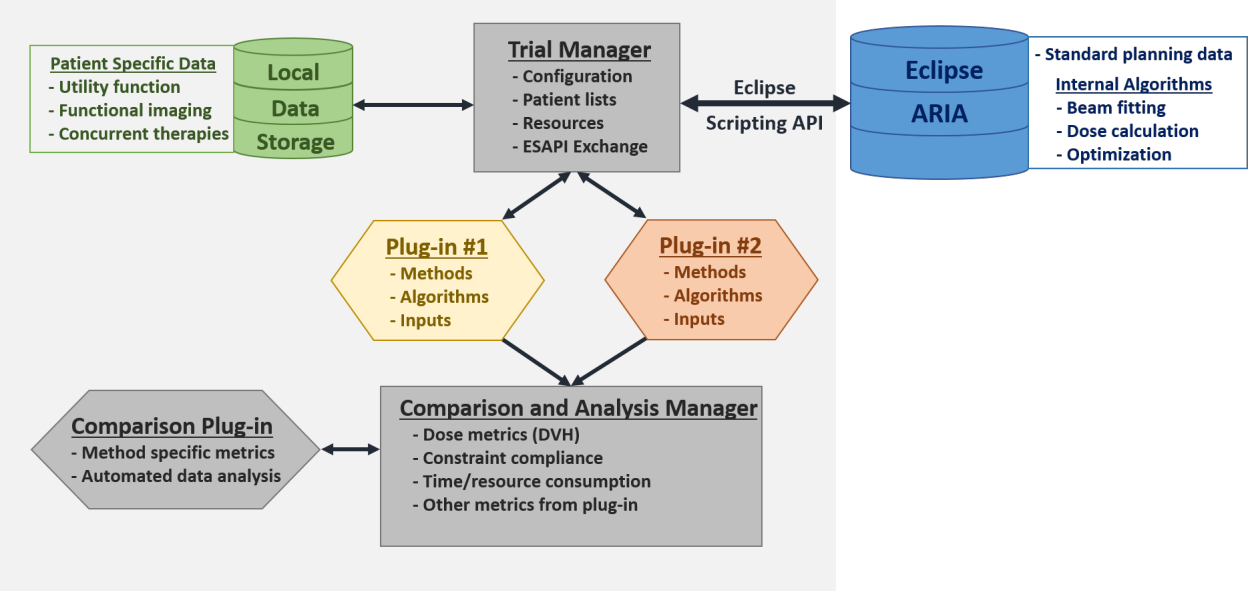


Figure 4.9. Conceptual diagram for the virtual radiotherapy trial suite (gray shaded background) connected to the TPS, Eclipse, through ESAPI.

Chapter 5
Conclusion, Summary, and Future Directions

This dissertation focuses on multiple aspects related to increasing personalization in RT with the goal of improving disease control and lowering toxicity rates. Broadly across RT, additional improvements and technological innovations are needed to move the field into an era of precision medicine, in which a patient's RT treatment is guided by the patient's individual characteristics rather than historical population level models and simple patient stratifications. This will require advancements in imaging, spatial correlation, treatment modalities, delivery methods, outcomes collection, predictive modeling, and treatment optimization in addition to advancements in many other sub-fields of RT.

Advancements in imaging and deformable registration will allow better monitoring of patients for anatomical and physiological changes throughout the RT treatment course and provide clinicians with the ability to determine when a patient may benefit from modification of the treatment plan in a timely manner. Additionally, advancements in optimization and outcome modeling will drive our ability to tailor the treatment specifically to the patient through personalized dosing, normal tissue sparing, and plan adaptation based on patient-specific predicted response, accounting for the individualized potential for disease control and risk of treatment associated toxicity. Overall these advancements can not only aid in improving the therapeutic ratio for individual patients, but can help capture more accurate treatment data which, when paired with robust patient characteristic and outcome reporting, can be used to develop modern, highly-predictive patient response models that go beyond a 'one size fits all' approach currently employed

in RT. In this dissertation, specifically, improvements in DIR and plan optimization are explored as contributions towards improving personalization in RT.

In Chapter 2, we characterized the clinical accuracy of an intensity-based DIR algorithm and explored tuning of algorithm parameters to improve accuracy and robustness. Although intensity-based DIR has the benefit of not requiring contouring or landmark identification prior to registration, which makes it an attractive option for widespread clinical use and use in unsupervised offline adaptive RT, we found that the accuracy of the B-spline, mutual information based algorithm tested only achieved clinically acceptable accuracy in certain clinical imaging scenarios. Algorithm parameter tuning, which is not typically analyzed prior to clinical use of these types of DIR algorithms, was found to improve registration accuracy, but only for lung imaging scenarios where bronchial and vessel bifurcations can drive the registration algorithm to deform locally based on clear tissue differentiation. For HN imaging, registration parameters deviating from the default values were shown to have little clinical impact on registration accuracy. This suggests that different approaches, beyond algorithm tuning, may be required to improve accuracy for imaging scenarios without large contrast differences between deforming tissues or organs. These approaches could include modifications to the B-spline method, such as irregular spacing of B-spline control points to better model deformations at the boundaries between tissues. Additionally, intensity-based DIR accuracy is highly dependent on the quality of the images being registered. Therefore, improving imaging quality prior to registration through tuning acquisition parameters or improving reconstruction techniques should be considered, particularly for CBCTs in the use case of daily offline adaptive RT using DIR.[155]

In Chapter 3, we demonstrated that incorporating local volumetric response into a previously developed biomechanical model-based DIR algorithm could improve longitudinal liver

registration accuracy. In this study, we first modeled liver normal-tissue volumetric response based on radiation dose and applied modified thermal linear expansion coefficients to drive volume change at the element level rather than relying on external modeling forces to model response. We then applied the external, surface forces to resolve the spatial differences between the two models of the liver resulting from typical anatomical and physiological motion. This novel addition to the biomechanical DIR process has the ability to improve our ability to correlate functional changes that occur in the liver during and after RT and spatially relate observed changes to delivered dose.[75,76] Beyond the liver, this method has the ability to be applied to other anatomical locations (e.g., parotid gland, stomach, bladder, rectum, etc.) where local mass loss or gain impacts the accuracy of current DIR technologies and can potentially aid in resolving these volumetric changes without generating unrealistic deformation patterns. This DIR method, along with other geometric DIR methods, has the downside that corresponding contours and landmarks need to be identified in the images prior to registration.[25] Although historically this has relied on manual contouring, a time-consuming process, advancements in machine learning, particularly those in 3D convolutional neural networks, have improved the ability to accurately segment structures automatically with limited human intervention.[156-159] The continued improvement of these auto-segmentation technologies for the liver and other structures will lead to improved accuracy and efficiency in geometric-based DIR, potentially leading to more widespread clinical adoption of these methods.

In Chapter 4, we demonstrated a potential clinical application for modern, personalized outcome models in RT. We developed and implemented a novel, but also generalizable, optimization strategy, based on the concept of plan utility, that facilitates direct incorporation of patient-specific efficacy and toxicity models into the treatment planning process without disrupting

current clinical standards of practice. By merging personalized, biological optimization with traditional dose-metric based optimization in a prioritized approach, we offer clinicians an opportunity to start exploring individualized treatment planning tradeoffs in an intuitive manner without requiring a complete overhaul of the current treatment planning process. Using previously developed models for efficacy and toxicity that factor in patient-specific clinical factors and predictive biomarkers for NSCLC patients, we demonstrated that this approach is feasible and has the potential to improve treatment outcomes through non-uniform dose escalation and OAR sparing in a limited patient population. The results from this feasibility study are encouraging and support the need for further investigation of this method across a larger cohort of patients. Additionally, a reformulation of the prioritized utility method, accounting only for toxicity burden, was introduced for HN cancer treatment and was demonstrated to be a feasible optimization strategy when applying current NTCP models. Future work will include developing new patient-specific toxicity models for HN cancer patients, similar to those available for NSCLC patients, to improve personalization in this method.

The work encompassed in this dissertation highlights two important aspects to help shift RT towards an era of precision medicine. First, improvements in DIR offer the increased ability to accurately determine spatial relationships and patient-specific tissue response between imaging acquired at different time points throughout a patient's treatment course. Accurate spatial mapping of functional imaging could facilitate the generation of new dose response models that more directly correlate localized changes, computed as longitudinal functional changes during treatment, with delivered dose. Additionally, through offline adaptive RT including dose accumulation, improved DIR accuracy and automation should translate into better estimations of a patient's delivered dose, which in turn could increase the accuracy of efficacy and toxicity

prediction models which have historically been based on planned dose.[160,161] Once refined, patient-specific predictive models are generated, the models need to be incorporated back into the treatment planning process for future patients to realize benefits of this work. Our proposed optimization strategy, which directly incorporates these types of patient-specific models, offers a potential solution for real clinical application, wherein newly developed predictive models can be progressively integrated into the planning process without explicitly requiring such models for every clinically relevant structure. In summary, the ability to generate patient-specific response models and intervene based on such models offers exciting opportunities in RT that aim to address patient variability and improve each patient's quality of life.

Appendices

Appendix A
Deformable Dose Accumulation Dashboard

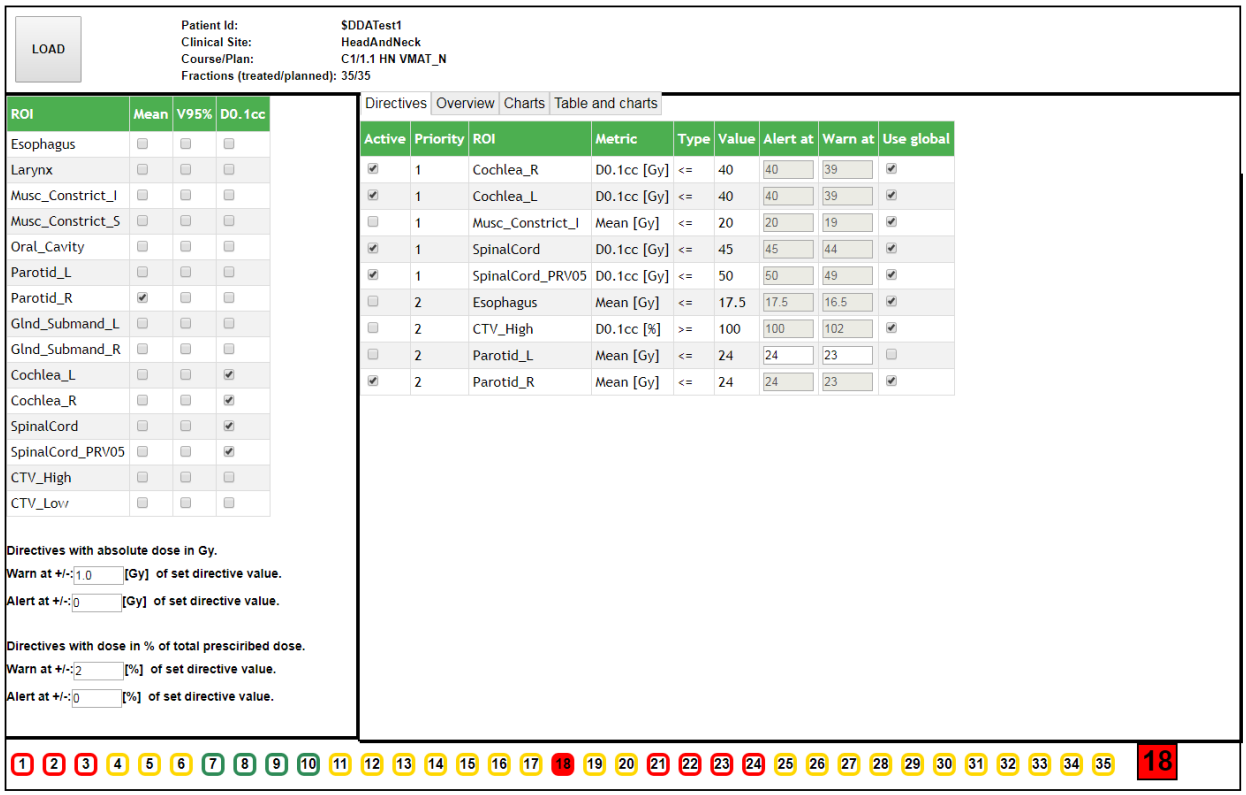
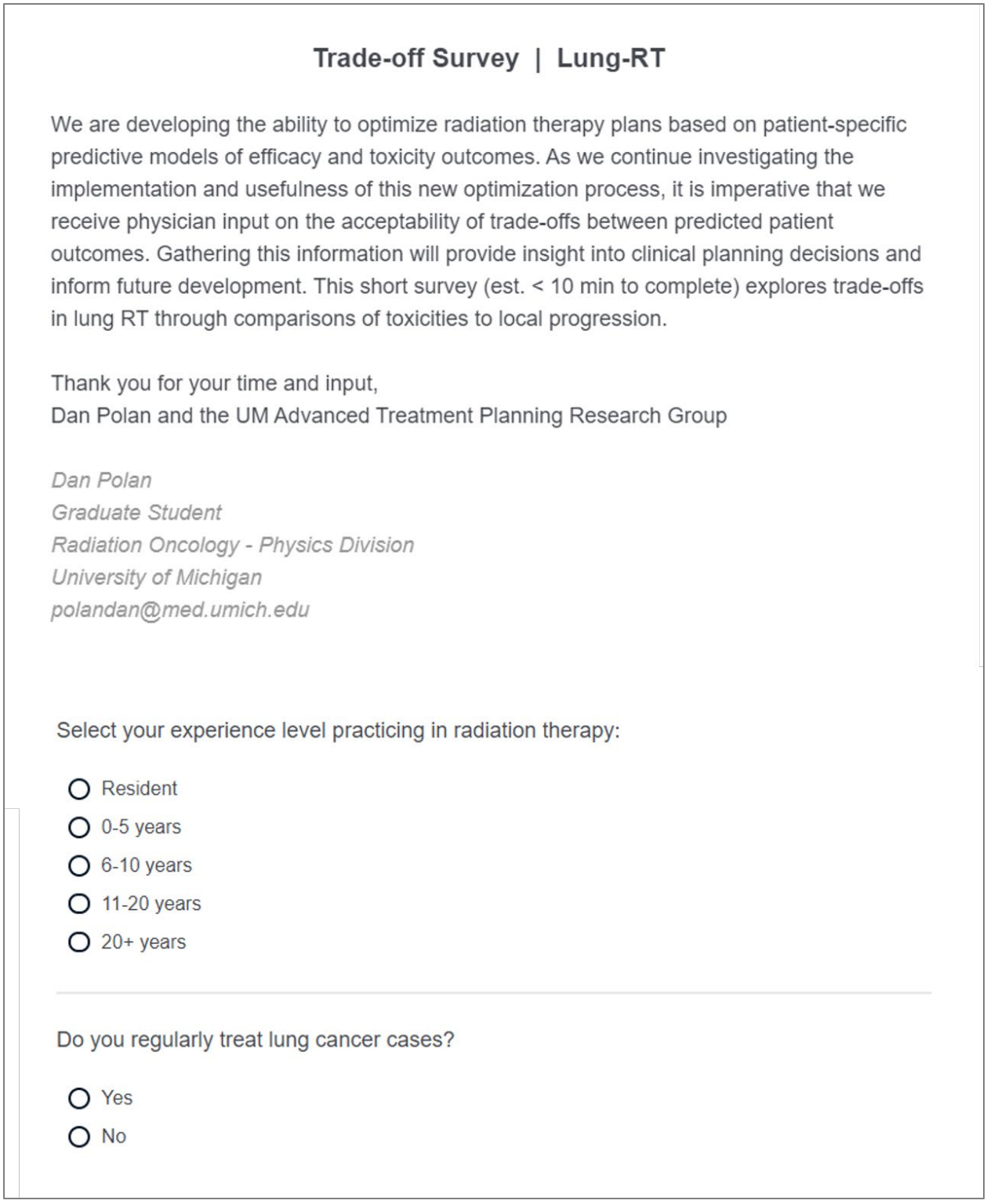


Figure A.1. Fully functional dose accumulation prototype dashboard showing the customizable treatment directives view with programmable dose warning and alert levels.

Figure A.2. Fully functional dose accumulation prototype dashboard showing dose and motion tracking of a patient's right parotid gland.

Appendix B
Physician Survey of Lung Radiotherapy Utility Weights Virtual Radiotherapy Trial Suite



Trade-off Survey | Lung-RT

We are developing the ability to optimize radiation therapy plans based on patient-specific predictive models of efficacy and toxicity outcomes. As we continue investigating the implementation and usefulness of this new optimization process, it is imperative that we receive physician input on the acceptability of trade-offs between predicted patient outcomes. Gathering this information will provide insight into clinical planning decisions and inform future development. This short survey (est. < 10 min to complete) explores trade-offs in lung RT through comparisons of toxicities to local progression.

Thank you for your time and input,
Dan Polan and the UM Advanced Treatment Planning Research Group

Dan Polan
Graduate Student
Radiation Oncology - Physics Division
University of Michigan
polandan@med.umich.edu

Select your experience level practicing in radiation therapy:

- ○ Resident
- ○ 0-5 years
- ○ 6-10 years
- ○ 11-20 years
- ○ 20+ years

Do you regularly treat lung cancer cases?

- ○ Yes
- ○ No

Figure B.1. Pages 1 and 2 of the online survey to gather utility weights from physicians.

On the next page, we ask that you quantitatively estimate the relative harm associated with a patient experiencing toxicity during and after definitive RT treatment for NSCLC. Using the slider bars (example below), assign a weight for various grades of radiation-induced lung toxicity, radiation esophagitis, and cardiac events within 2 years of treatment. The weight, ***w***, of each toxicity represents how harmful the toxicity is ***relative*** to local-regional progression within 2 years (***w*** = 100 for local-regional progression).

Weight (*w*)	**Interpretation of toxicity**
0	not harmful
1-99	less harmful than local-regional progression within 2 years
100	equally as harmful as local-regional progression within 2 years
101-150	more harmful than local-regional progression within 2 years

Example slider bar: Drag the button on the slider bar to the relative weight (*w*) for local-regional progression within 2 years. Reminder: ***w*** = 100 for local-regional progression within 2 years (so the slider should be set to ***w*** = 100 below).

Example slider bar for local-regional progression within 2 years

Please note the following:

1. The scale is ***linear***. Therefore, if you assign ***w*** = 50 for a toxicity, then the toxicity is considered half as harmful as local-regional progression within 2 years. Similarly, between toxicities, if you assign ***w*** = 30 for toxicity 'A' and ***w*** = 60 for toxicity 'B', then toxicity 'A' is considered half as harmful as toxicity 'B'.

2. Severity grades are based on the Common Terminology Criteria for Adverse Events (CTCAE version 4.0) with the following general guidelines: Grade 2 = moderate, Grade 3 = severe or medically significant but not immediately life-threatening, Grade 4 = a life-threatening condition requiring urgent intervention, Grade 5 = death related to the adverse event.

Figure B.2. Page 3 of the online survey to gather utility weights from physicians.

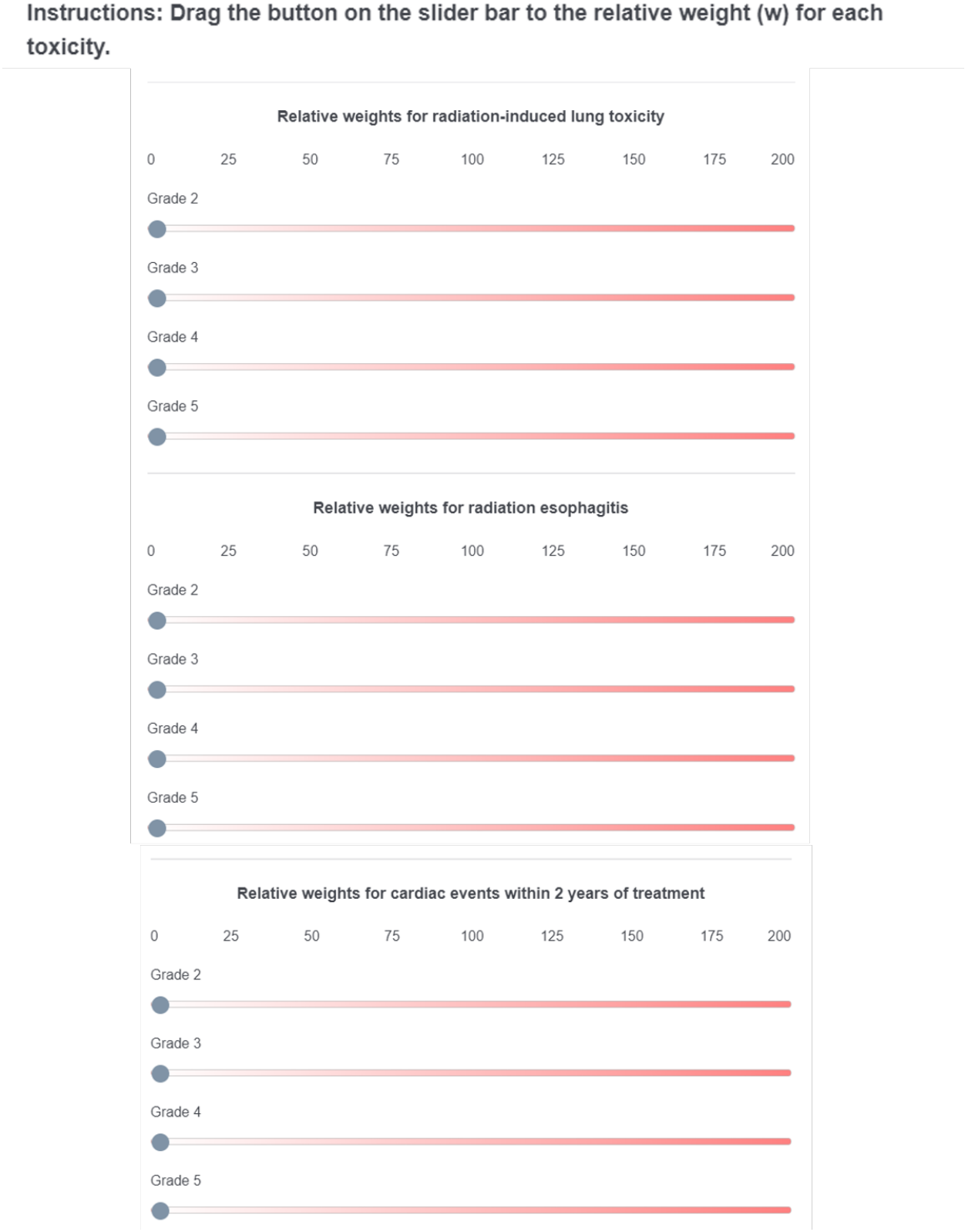


Figure B.3. Page 4 of the online survey to gather utility weights from physicians.

Appendix C
Virtual Radiotherapy Trial Suite

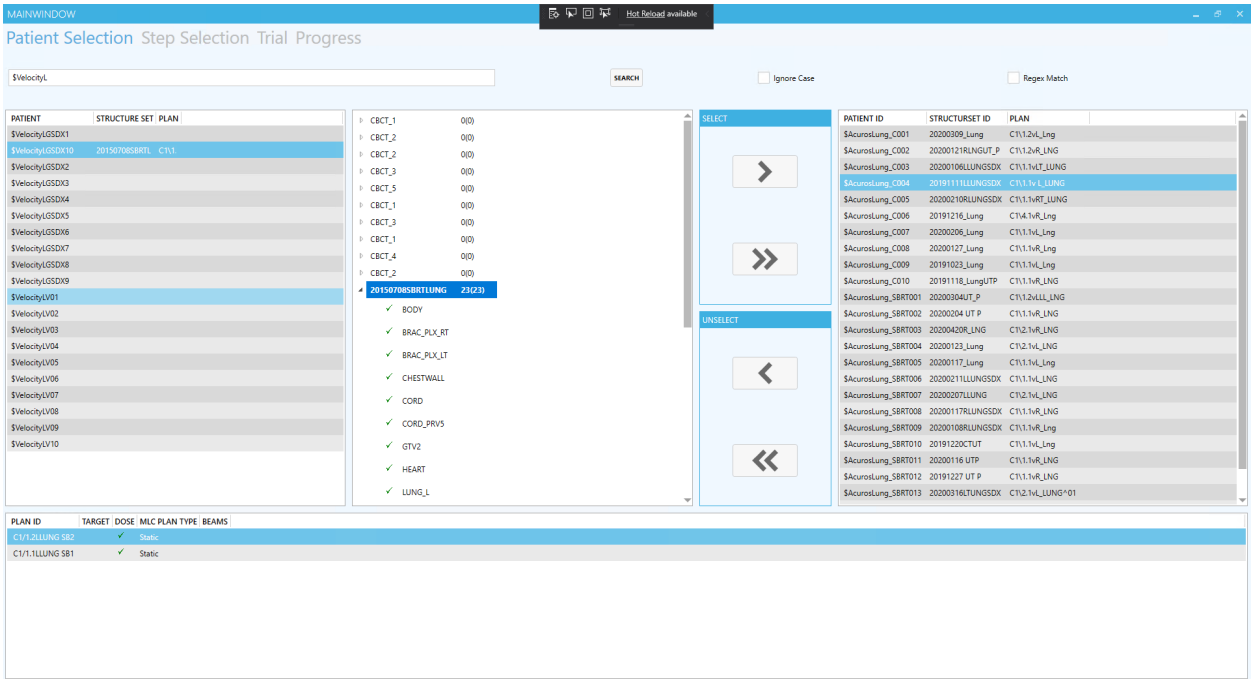


Figure C.1. Virtual radiotherapy trial suite workflow Step 1: Searching for and selecting patients from Aria database.

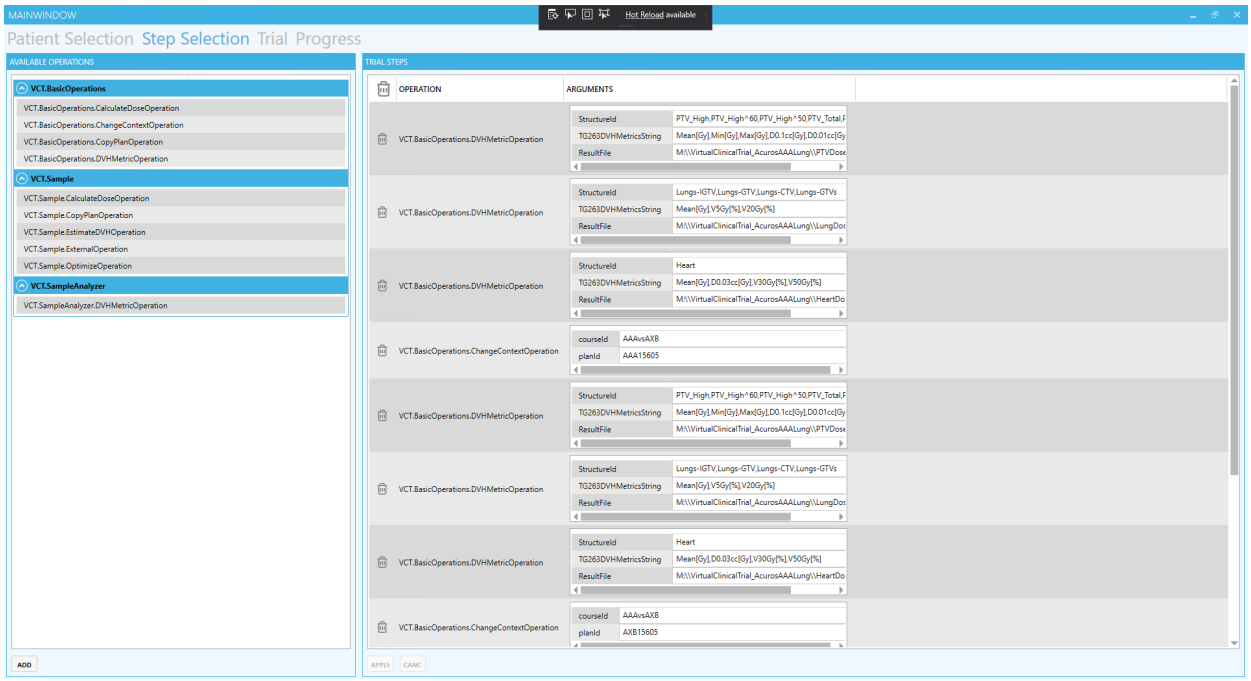


Figure C.2. Virtual radiotherapy trial suite workflow Step 2: Loading in functional .dll blocks and selecting/configuring trial steps.

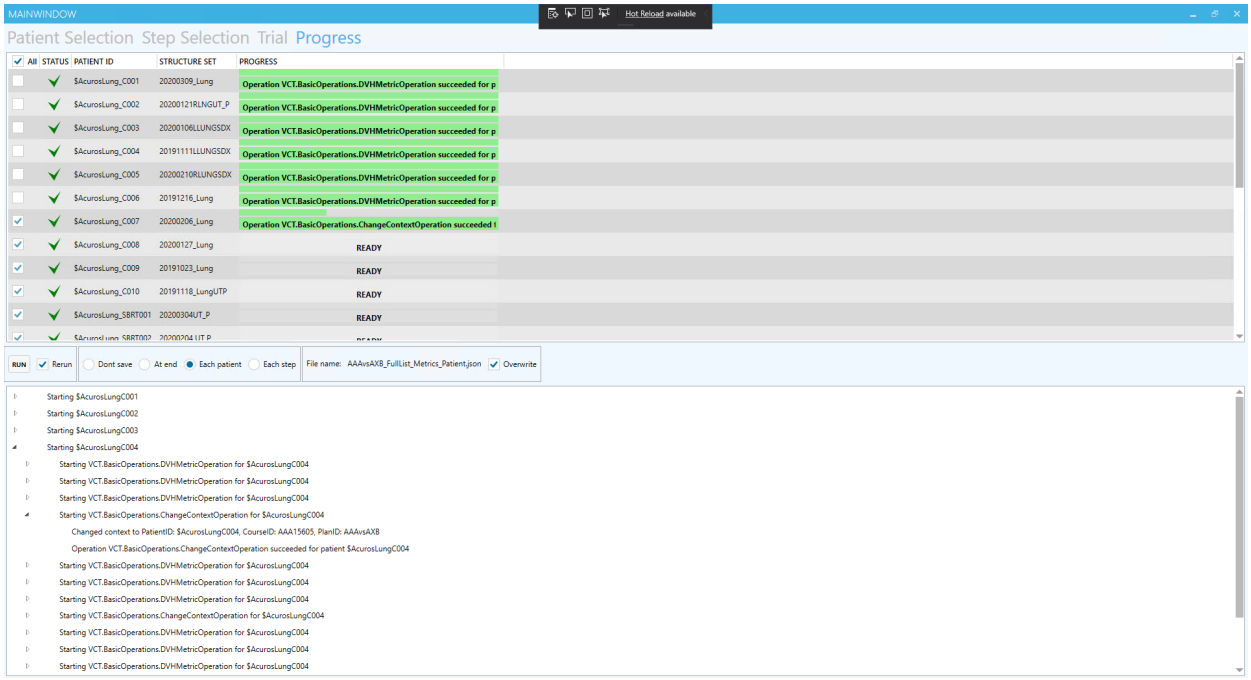


Figure C.3. Virtual radiotherapy trial suite workflow Step 3: Running the trial and checking the progress of individual trial steps per patient.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660

2. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*. 2019;69(5):363-385. doi:10.3322/caac.21565

3. Baskar R, Itahana K. Radiation therapy and cancer control in developing countries: Can we save more lives?. *Int J Med Sci*. 2017;14(1):13-17. Published 2017 Jan 1. doi:10.7150/ijms.17288

4. Metcalfe P, Liney GP, Holloway L, et al. The potential for an enhanced role for MRI in radiation-therapy treatment planning. *Technol Cancer Res Treat*. 2013;12(5):429-446. doi:10.7785/tcrt.2012.500342

5. Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. *Br J Cancer*. 2019;120(8):779-790. doi:10.1038/s41416-019-0412-y

6. Hoover DA, Capaldi DP, Sheikh K, et al. Functional lung avoidance for individualized radiotherapy (FLAIR): study protocol for a randomized, double-blind clinical trial. *BMC Cancer*. 2014;14:934. Published 2014 Dec 11. doi:10.1186/1471-2407-14-934

7. Moteabbed M, Sharp GC, Wang Y, Trofimov A, Efstathiou JA, Lu HM. Validation of a deformable image registration technique for cone beam CT-based dose verification. *Med Phys*. 2015 Jan;42(1):196-205. doi: 10.1118/1.4903292.

8. Taylor A, Powell ME. Intensity-modulated radiotherapy--what is it?. *Cancer Imaging*. 2004;4(2):68-73. Published 2004 Mar 26. doi:10.1102/1470-7330.2004.0003

9. Brock KK. Adaptive Radiotherapy: Moving Into the Future. *Semin Radiat Oncol*. 2019;29(3):181-184. doi:10.1016/j.semradonc.2019.02.011

10. Fass L. Imaging and cancer: a review. *Mol Oncol*. 2008;2(2):115-152. doi:10.1016/j.molonc.2008.04.001

11. West CM, Huddart RA. Biomarkers and Imaging for Precision Radiotherapy. *Clin Oncol (R Coll Radiol)*. 2015 Oct;27(10):545-6. doi: 10.1016/j.clon.2015.06.021. Epub 2015 Jul 11. PMID: 26173954.

12. Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. *Br J Cancer*. 2019;120(8):779-790. doi:10.1038/s41416-019-0412-y

13. Kessler ML, Pitluck S, Petti P, Castro JR. Integration of multimodality imaging data for radiotherapy treatment planning. *Int J Radiat Oncol Biol Phys*. 1991;21(6):1653-1667. doi:10.1016/0360-3016(91)90345-5

14. Mera Iglesias M, Aramburu Núñez D, Del Olmo Claudio JL, et al. Multimodality functional imaging in radiation therapy planning: relationships between dynamic contrast-enhanced MRI, diffusion-weighted MRI, and 18F-FDG PET. *Comput Math Methods Med*. 2015;2015:103843. doi:10.1155/2015/103843

15. Dawson LA, Sharpe MB. Image-guided radiotherapy: rationale, benefits, and limitations. *Lancet Oncol*. 2006;7:848–858.

16. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. *Nat Rev Clin Oncol*. 2012;9(12):688-699. doi:10.1038/nrclinonc.2012.194

17. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol*. 1997;42(1):123-132. doi:10.1088/0031-9155/42/1/008

18. Sonke JJ, Aznar M, Rasch C. Adaptive Radiotherapy for Anatomical Changes. *Semin Radiat Oncol*. 2019;29(3):245-257. doi:10.1016/j.semradonc.2019.02.007

19. Keall P, Poulsen P, Booth JT. See, Think, and Act: Real-Time Adaptive Radiotherapy. *Semin Radiat Oncol*. 2019;29(3):228-235. doi:10.1016/j.semradonc.2019.02.005

20. Matuszak MM, Kashani R, Green M, et al. Functional Adaptation in Radiation Therapy [published correction appears in Semin Radiat Oncol. 2021 Jan;31(1):93]. *Semin Radiat Oncol*. 2019;29(3):236-244. doi:10.1016/j.semradonc.2019.02.006

21. Chetty IJ, Rosu-Bubulac M. Deformable Registration for Dose Accumulation. *Semin Radiat Oncol*. 2019;29(3):198-208. doi:10.1016/j.semradonc.2019.02.002

22. Shelley LEA, Scaife JE, Romanchikova M, et al. Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy. *Radiother Oncol*. 2017;123(3):466-471. doi:10.1016/j.radonc.2017.04.008

23. Swaminath A, Massey C, Brierley JD, et al. Accumulated Delivered Dose Response of Stereotactic Body Radiation Therapy for Liver Metastases. *Int J Radiat Oncol Biol Phys*. 2015;93(3):639-648. doi:10.1016/j.ijrobp.2015.07.2273

24. McCulloch MM, Muenz DG, Schipper MJ, Velec M, Dawson LA, Brock KK. A simulation study to assess the potential impact of developing normal tissue complication probability models with accumulated dose. *Adv Radiat Oncol*. 2018;3(4):662-672. Published 2018 May 16. doi:10.1016/j.adro.2018.05.003

25. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44(7):e43-e76. doi:10.1002/mp.12256

26. Broit C. Optimal registration of deformed images [*dissertation*] Pennsylvania, PA: University of Pennsylvania; 1981.

27. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging*. 2003 Jan;22(1):120-8. doi: 10.1109/TMI.2003.809072. PMID: 12703765.

28. Brock KK, Sharpe MB, Dawson LA, Kim SM, Jaffray DA. Accuracy of finite element model-based multi-organ deformable image registration. *Med Phys*. 2005;32(6):1647-1659. doi:10.1118/1.1915012

29. Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol*. 2005;50(12):2887-2905. doi:10.1088/0031-9155/50/12/011

30. Coselmon MM, Balter JM, McShan DL, Kessler ML. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med Phys*. 2004;31(11):2942-2948. doi:10.1118/1.1803671

31. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Digit Imaging*. 2019;32(4):582-596. doi:10.1007/s10278-019-00227-x

32. Fraass BA. The development of conformal radiation therapy. *Med Phys*. 1995;22(11 Pt 2):1911-1921. doi:10.1118/1.597446

33. Brewster L, Mohan R, Mageras G, Burman C, Leibel S, Fuks Z. Three dimensional conformal treatment planning with multileaf collimators. *Int J Radiat Oncol Biol Phys*. 1995;33(5):1081-1089. doi:10.1016/0360-3016(95)02061-6

34. Wu VW, Sham JS, Kwong DL. Inverse planning in three-dimensional conformal and intensity-modulated radiotherapy of mid-thoracic oesophageal cancer. *Br J Radiol*. 2004;77(919):568-572. doi:10.1259/bjr/19972578

35. Kessler, M.L., Mcshan, D.L., Epelman, M.A. et al. Costlets: A Generalized Approach to Cost Functions for Automated Optimization of IMRT Treatment Plans. *Optim Eng 6*, 421–448 (2005). https://doi.org/10.1007/s11081-005-2066-2

36. Cho B. Intensity-modulated radiation therapy: a review with a physics perspective [published correction appears in Radiat Oncol J. 2018 Jun;36(2):171]. *Radiat Oncol J*. 2018;36(1):1-10. doi:10.3857/roj.2018.00122

37. Allen Li X, Alber M, Deasy JO, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM. *Med Phys*. 2012;39(3):1386-1409. doi:10.1118/1.3685447

38. Jaffray DA, Das S, Jacobs PM, Jeraj R, Lambin P. How Advances in Imaging Will Affect Precision Radiation Oncology. *Int J Radiat Oncol Biol Phys*. 2018;101(2):292-298. doi:10.1016/j.ijrobp.2018.01.047

39. Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. *J Appl Clin Med Phys*. 2019;20(3):105-114. doi:10.1002/acm2.12554

40. Mera Iglesias M, Aramburu Núñez D, Del Olmo Claudio JL, et al. Multimodality functional imaging in radiation therapy planning: relationships between dynamic contrast-enhanced MRI, diffusion-weighted MRI, and 18F-FDG PET. *Comput Math Methods Med*. 2015;2015:103843. doi:10.1155/2015/103843

41. Den RB, Doemer A, Kubicek G, et al. Daily image guidance with cone-beam computed tomography for head-and-neck cancer intensity-modulated radiotherapy: a prospective study. *Int J Radiat Oncol Biol Phys*. 2010;76(5):1353-1359. doi:10.1016/j.ijrobp.2009.03.059

42. van de Lindt T, Sonke JJ, Nowee M, et al. A Self-Sorting Coronal 4D-MRI Method for Daily Image Guidance of Liver Lesions on an MR-LINAC. *Int J Radiat Oncol Biol Phys*. 2018;102(4):875-884. doi:10.1016/j.ijrobp.2018.05.029

43. Cavalieri, R., Gay, H.A., Liu, J. et al. Total error shift patterns for daily CT on rails image-guided radiotherapy to the prostate bed. *Radiat Oncol* 6, 142 (2011). https://doi.org/10.1186/1748-717X-6-142

44. Liu JX, Chen YS, Chen LF. Fast and accurate registration techniques for affine and nonrigid alignment of MR brain images. *Ann Biomed Eng*. 2010;38(1):138-157. doi:10.1007/s10439-009-9840-9

45. Csapo I, Davis B, Shi Y, Sanchez M, Styner M, Niethammer M. Longitudinal image registration with non-uniform appearance change. *Med Image Comput Comput Assist Interv*. 2012;15(Pt 3):280-288. doi:10.1007/978-3-642-33454-2_35

46. Brock KK, Dawson LA, Sharpe MB, Moseley DJ, Jaffray DA. Feasibility of a novel deformable image registration technique to facilitate classification, targeting, and monitoring of tumor and normal tissue. *Int J Radiat Oncol Biol Phys*. 2006;64(4):1245-1254. doi:10.1016/j.ijrobp.2005.10.027

47. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol*. 2004;49(14):3067-3087. doi:10.1088/0031-9155/49/14/003

48. Heath E, Collins DL, Keall PJ, Dong L, Seuntjens J. Quantification of accuracy of the automated nonlinear image matching and anatomical labeling (ANIMAL) nonlinear registration algorithm for 4D CT images of lung. *Med Phys*. 2007;34(11):4409-4421. doi:10.1118/1.2795824

49. Crum WR, Tanner C, Hawkes DJ. Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Phys Med Biol*. 2005;50(21):5153-5174. doi:10.1088/0031-9155/50/21/014

50. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal*. 1998;2(3):243-260. doi:10.1016/s1361-8415(98)80022-4

51. Noe KØ, Tanderup K, Lindegaard JC, Grau C, Sørensen TS. GPU accelerated viscous-fluid deformable registration for radiotherapy. *Stud Health Technol Inform*. 2008;132:327-332.

52. Brock KK; Deformable Registration Accuracy Consortium. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys*. 2010;76(2):583-596. doi:10.1016/j.ijrobp.2009.06.031

53. Kadoya N, Nakajima Y, Saito M, et al. Multi-institutional Validation Study of Commercially Available Deformable Image Registration Software for Thoracic Images. *Int J Radiat Oncol Biol Phys*. 2016;96(2):422-431. doi:10.1016/j.ijrobp.2016.05.012

54. Kubli A, Pukala J, Shah AP, et al. Variability in commercially available deformable image registration: A multi-institution analysis using virtual head and neck phantoms. *J Appl Clin Med Phys*. 2021;22(5):89-96. doi:10.1002/acm2.13242

55. Jacobson TJ, Murphy MJ. Optimized knot placement for B-splines in deformable image registration. *Med Phys*. 2011;38(8):4579-4582. doi:10.1118/1.3609416

56. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging*. 2003;22(1):120-128. doi:10.1109/TMI.2003.809072

57. Castillo R, Castillo E, Guerra R, Johnson VE, McPhail T, Garg AK, Guerrero T. 2009. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol* 54 1849-1870.

58. Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. 2009. Four-dimensional deformable image registration using trajectory modeling. *Phys Med Biol* 55 305-327.

59. Se Young Chun and J. A. Fessler, "Regularized methods for topology-preserving smooth nonrigid image registration using B-spline basis," *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Paris, 2008, pp. 1099-1102.

60. Chapman CH, Polan D, Vineberg K, et al. Deformable image registration-based contour propagation yields clinically acceptable plans for MRI-based cervical cancer brachytherapy planning. *Brachytherapy*. 2018;17(2):360-367. doi:10.1016/j.brachy.2017.11.019

61. Polan DF, Feng M, Lawrence TS, Ten Haken RK, Brock KK. Implementing Radiation Dose-Volume Liver Response in Biomechanical Deformable Image Registration. *Int J Radiat Oncol Biol Phys*. 2017;99(4):1004-1012. doi:10.1016/j.ijrobp.2017.06.2455

62. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA Cancer J Clin* 2019; 69(1):7-34. https://doi.org/10.3322/caac.21551

63. Center MM, Jemal A. International trends in liver cancer incidence rates. *Cancer Epidemiol Biomarkers Prev* 2011;20(11):2362-2368. http://dx.doi.org/10.1158/1055-9965

64. Henley SJ, Ward EM, Scott S, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer*. 2020;126(10):2225-2249. doi:10.1002/cncr.32802

65. Kow AWC. Hepatic metastasis from colorectal cancer. *J Gastrointest Oncol*. 2019;10(6):1274-1298. doi:10.21037/jgo.2019.08.06

66. Bray F, Ferlay J, Soerjomataram, et al. Global cancer statistics 2018: GLOBOSCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424. 10.3322/caac.21492

67. Chung KY, Kemeny N. Regional and systemic chemotherapy for primary hepatobiliary cancers and for colorectal cancer metastatic to the liver. *Semin Radiat Oncol* 2005;15(4):284–298. http://dx.doi.org/10.1016/j.semradonc.2005.04.007

68. Tong D, Russell AH, Dawson LE, Wisbeck W. Second laparotomy for proximal colon cancer. Sites of recurrence and implications for adjuvant therapy. *Am J Surg* 1983;145(3):382-386. http://dx.doi.org/10.1016/0002-9610(83)90207-6

69. Gresswell S, Tobillo R, Hasan S, et al. Stereotactic body radiotherapy used as a bridge to liver transplant in patients with hepatocellular carcinoma and Child-Pugh score ≥8 cirrhosis. *J Radiosurg SBRT*. 2018;5(4):261-267.

70. Koay EJ, Owen D, Das P. Radiation-Induced Liver Disease and Modern Radiotherapy. *Semin Radiat Oncol*. 2018;28(4):321-331. doi:10.1016/j.semradonc.2018.06.007

71. Bujold A, Massey CA, Kim JJ, et al. Sequential phase I and II trials of stereotactic body radiotherapy for locally advanced hepatocellular carcinoma. *J Clin Oncol*. 2013;31(13):1631-1639. doi:10.1200/JCO.2012.44.1659

72. Bowen SR, Saini J, Chapman TR, et al. Differential hepatic avoidance radiation therapy: Proof of concept in hepatocellular carcinoma patients. *Radiother Oncol*. 2015;115(2):203-210. doi:10.1016/j.radonc.2015.04.011

73. Wu VW, Epelman MA, Wang H, et al. Optimizing global liver function in radiation therapy treatment planning. *Phys Med Biol*. 2016;61(17):6465-6484. doi:10.1088/0031-9155/61/17/6465

74. Wang H, Feng M, Jackson A, Ten Haken RK, Lawrence TS, Cao Y. Local and Global Function Model of the Liver. *Int J Radiat Oncol Biol Phys*. 2016;94(1):181-188. doi:10.1016/j.ijrobp.2015.09.044

75. Cao Y, et al. Liver function after irradiation based on computed tomographic portal vein perfusion imaging. *Int J Radiat Oncol Biol Phys* 2008;70(1):154-160. http://dx.doi.org/10.1016/j.ijrobp.2007.05.078

76. Cao Y, et al. Prediction of Liver Function by Using Magnetic Resonance-based Portal Venous Perfusion Imaging. *Int J Radiat Oncol Biol Phys* 2013;85(1):258-263. http://dx.doi.org/10.1016/j.ijrobp.2012.02.037.

77. Wang H, Feng M, Jackson A, Ten Haken RK, Lawrence TS, Cao Y. A Local and Global Function Model of the Liver. *Int J Radiat Oncol Biol Phys*. 2016;94(1):181-188. http://dx.doi.org/10.1016/j.ijrobp.2015.09.044.

78. Goodman KA, et al. Dose-escalation study of single-fraction stereotactic body radiotherapy for liver malignancies. *Int J Radiat Oncol Biol Phys* 2010;78(2):486–449. http://dx.doi.org/10.1016/j.ijrobp.2009.08.020

79. Chang DT, et al. Stereotactic body radiotherapy for colorectal liver metastases: a pooled analysis. *Cancer* 2011;117(17):4060-4069. http://dx.doi.org/10.1002/cncr.25997

80. Schefter TE, Kavanagh BD. Radiation therapy for liver metastases. *Semin Radiat Oncol* 2011;21(4):264-270. http://dx.doi.org/10.1016/j.semradonc.2011.05.006

81. Scorsetti M, et al. Is stereotactic body radiation therapy an attractive option for unresectable liver metastases? A preliminary report from a phase 2 trial. *Int J Radiat Oncol Biol Phys* 2013;86(2):336-342. http://dx.doi.org/10.1016/j.ijrobp.2012.12.021

82. Ben-Josef E, et al. Phase II trial of high-dose conformal radiation therapy with concurrent hepatic artery floxuridine for unresectable intrahepatic malignancies. *J Clin Oncol* 2005;23(34):8739-8747. http://dx.doi.org/10.1200/JCO.2005.01.5354

83. Lee MT, et al. Phase I study of individualized stereotactic body radiotherapy of liver metastases. *J Clin Oncol* 2009;27(10):1585-1591. http://dx.doi.org/10.1200/JCO.2008.20.0600

84. Rusthoven KE, et al. Multi-institutional phase I/II trial of stereotactic body radiation therapy for liver metastases. *J Clin Oncol* 2009;27(10):1572-1578. http://dx.doi.org/10.1200/JCO.2008.19.6329

85. Schefter TE, et al. A phase I trial of stereotactic body radiation therapy (SBRT) for liver metastases. *Int J Radiat Oncol Biol Phys* 2005;62(5):1371-1378. http://dx.doi.org/10.1016/j.ijrobp.2005.01.002

86. Tse RV, et al. Phase I study of individualized stereotactic body radiotherapy for hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Clin Oncol* 2008;26(4):657-664. http://dx.doi.org/10.1200/JCO.2007.14.3529

87. Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tomé WA. Accurate Accumulation of dose for Improved Understanding of Radiation Effects in Normal Tissue. *Int J Radiat Oncol Biol Phys* 2010;76(3 Suppl):S135-139. http://dx.doi.org/10.1016/j.ijrobp.2009.06.093

88. Kim RD, Kim JS, Watanabe G, Mohuczy D, Behrns KE. Liver Regeneration and the Atrophy-Hypertrophy Complex. *Semin Intervent Radiol* 2008;25(2):92-103. http://dx.doi.org/10.1055/s-2008-1076679.

89. Gaba RC, et al. Radiation lobectomy: preliminary findings of hepatic volumetric response to lobar yttrium-90 radioembolization. *Ann Surg Oncol* 2009;16(6):1587-1596. http://dx.doi.org/10.1245/s10434-009-0454-0

90. Thörnqvist S, Petersen JB, Høyer M, Bentzen LN, Muren LP. Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration. *Acta Oncol* 2010;49(7):1023-1032. http://dx.doi.org/10.3109/0284186X.2010.503662

91. Nithiananthan S, et al. Demons deformable registration of CT and cone-beam CT using an iterative intensity matching approach. *Med Phys* 2011;38(4):1785-1798. http://dx.doi.org/10.1118/1.3555037

92. Varadhan R, Magome T, and Hui S. Characterization of deformation and physical force in uniform low contrast anatomy and its impact on accuracy of deformable image registration. *Med Phys* 2016;43(1):52-61. http://dx.doi.org/10.1118/1.4937935

93. Cazoulat G, et al. Surface-Constrained Nonrigid Registration for Dose Monitoring in Prostate Cancer Radiotherapy. *IEEE Trans Med Imaging* 2014;33(7):1464-1474. http://dx.doi.org/10.1109/TMI.2014.2314574

94. Velec M, Juang T, Moseley JL, Oldham M, Brock KK. Utility and validation of biomechanical deformable image registration in low-contrast images. *Pract Radiat Oncol* 2015;5(4):e401-e408. http://dx.doi.org/10.1016/j.prro.2015.01.011

95. Juang T, Das S, Adamovics J, Benning R, Oldham M. On the need for comprehensive validation of deformable image registration, investigated with a novel 3D deformable dosimeter. *Int J Radiat Oncol Biol Phys* 2013;87(2):414-421. http://dx.doi.org/10.1016/j.ijrobp.2013.05.045

96. Kirby N, Chuang C, Ueda U, and Pouliot J. The need for application-based adaptation of deformable image registration. *Med Phys* 2013;40(1):011702. http://dx.doi.org/10.1118/1.4769114

97. Ben-Josef, et al. Phase II trial of high-dose conformal radiation therapy with concurrent hepatic artery floxuridine for unresectable intrahepatic malignancies. *J Clin Oncol* 2005;23(34):8739-8747. http://dx.doi.org/10.1200/JCO.2005.01.5354

98. Pan CC, Krishnan U, Normolle D, Lawrence TS, Ten Haken RK. Liver Regeneration in Patients with Intrahepatic Malignancies Treated with Focal Liver Radiation Therapy. *Int J Radiat Oncol Biol Phys* 2007;69(3):S81. http://dx.doi.org/10.1016/j.ijrobp.2007.07.145

99. Gaba RC, et al. Chemoembolic lobectomy: imaging findings of hepatic lobar volume reduction after transcatheter arterial chemoembolization. *Diagn Interv Radiol* 2011;17(2):177-180. http://dx.doi.org/10.4261/1305-3825.DIR.3166-09.1

100. Jakobs TFet al. Fibrosis, portal hypertension, and hepatic volume changes induced by intra-arterial radiotherapy with 90yttrium microspheres. *Dig Dis Sci* 2008;53(9):2556-2563. http://dx.doi.org/10.1007/s10620-007-0148-z

101. Imada H, et al. Compensatory enlargement of the liver after treatment of hepatocellular carcinoma with carbon ion radiotherapy - relation to prognosis and liver function. *Radiother Oncol* 2010;96(2):236-242. http://dx.doi.org/10.1016/j.radonc.2010.03.025

102. Robertson JM, et al. Long-term results of hepatic artery fluorodeoxyuridine and conformal radiation therapy for primary hepatobiliary cancers. International journal of radiation oncology, biology, physics. *Int J Radiat Oncol Biol Phys* 1997;37(2):325-330. http://dx.doi.org/.1118/1.1915012

103. Hermoye et al. Liver Segmentation in Living Liver Transplant Donors: Comparison of Semiautomatic and Manual Methods. *Radiology* 2005;234(1):171-178. http://dx.doi.org/10.1148/radiol.2341031801

104. Cazoulat G, Owen D, Matuszak MM, Balter JM, Brock KK. Biomechanical deformable image registration of longitudinal lung CT images using vessel information. *Phys Med Biol* 2016;61(13):4826-4839. http://dx.doi.org/10.1088/0031-9155/61/13/4826

105. Vainshtein JM, Moon DH, Feng FY, Chepeha DB, Eisbruch A, Stenmark MH. Long-term quality of life after swallowing and salivary-sparing chemo–intensity modulated

radiation therapy in survivors of human papillomavirus–related oropharyngeal cancer. *International Journal of Radiation Oncology* Biology* Physics* 2015;91:925-933.

106. Konings AW, Cotteleer F, Faber H, van Luijk P, Meertens H, Coppes RP. Volume effects and region-dependent radiosensitivity of the parotid gland. *Int J Radiat Oncol Biol Phys* 2005;62:1090-1095.

107. McCulloch MM, Cazoulat G, Ford AC, et al. Biomechanical modeling of radiation dose-induced volumetric changes of the parotid glands for deformable image registration. *Phys Med Biol*. 2020;65(16):165017. Published 2020 Aug 31. doi:10.1088/1361-6560/ab8bf1

108. Kachnic LA, Tsai HK, Coen JJ, et al. Dose-painted intensity-modulated radiation therapy for anal cancer: a multi-institutional report of acute toxicity and response to therapy. *Int J Radiat Oncol Biol Phys*. 2012;82(1):153-158. doi:10.1016/j.ijrobp.2010.09.030

109. Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol*. 2011;12(2):127-136. doi:10.1016/S1470-2045(10)70290-4

110. Klopp AH, Yeung AR, Deshmukh S, et al. Patient-Reported Toxicity During Pelvic Intensity-Modulated Radiation Therapy: NRG Oncology-RTOG 1203 [published correction appears in J Clin Oncol. 2019 Mar 20;37(9):761] [published correction appears in J Clin Oncol. 2020 Apr 1;38(10):1118]. *J Clin Oncol*. 2018;36(24):2538-2544. doi:10.1200/JCO.2017.77.4273

111. Deasy JO, Alaly JR, Zakaryan K. Obstacles and advances in intensity-modulated radiation therapy treatment planning. *Front Radiat Ther Oncol*. 2007;40:42-58. doi:10.1159/000106027

112. Xing L, Li JG, Donaldson S, Le QT, Boyer AL. Optimization of importance factors in inverse planning. *Phys Med Biol*. 1999;44(10):2525-2536. doi:10.1088/0031-9155/44/10/311

113. Wu X, Zhu Y. An optimization method for importance factors and beam weights based on genetic algorithms for radiotherapy treatment planning. *Phys Med Biol*. 2001;46(4):1085-1099. doi:10.1088/0031-9155/46/4/313

114. Bedford JL, Webb S. Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy. *Med Phys*. 2003;30(7):1788-1804. doi:10.1118/1.1582471

115. Lee T, Hammad M, Chan TC, Craig T, Sharpe MB. Predicting objective function weights from patient anatomy in prostate IMRT treatment planning. *Med Phys*. 2013;40(12):121706. doi:10.1118/1.4828841.

116. Chan T, Craig T, Lee T, Sharpe M. Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy. *Operations Research*. 2014;62(3), 680-695. doi:10.1287/opre.2014.1267

117. Lee T, Hammad M, Chan TC, Craig T, Sharpe MB. Predicting objective function weights from patient anatomy in prostate IMRT treatment planning. *Med Phys*. 2013;40(12):121706. doi:10.1118/1.4828841

118. Boutilier JJ, Lee T, Craig T, Sharpe MB, Chan TC. Models for predicting objective function weights in prostate cancer IMRT. *Med Phys*. 2015;42(4):1586-1595. doi:10.1118/1.4914140

119. Babier A, Boutilier JJ, Sharpe MB, McNiven AL, Chan TCY. Inverse optimization of objective function weights for treatment planning using clinical dose-volume histograms. *Phys Med Biol*. 2018;63(10):105004. Published 2018 May 10. doi:10.1088/1361-6560/aabd14

120. Shen C, Nguyen D, Chen L, et al. Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning. *Med Phys*. 2020;47(6):2329-2336. doi:10.1002/mp.14114

121. Feng Z, Tao C, Zhu J, et al. An integrated strategy of biological and physical constraints in biological optimization for cervical carcinoma. *Radiat Oncol*. 2017;12(1):64. Published 2017 Apr 4. doi:10.1186/s13014-017-0784-1

122. Kan MW, Leung LH, Yu PK. The use of biologically related model (Eclipse) for the intensity-modulated radiation therapy planning of nasopharyngeal carcinomas. *PLoS One*. 2014;9(11):e112229. Published 2014 Nov 5. doi:10.1371/journal.pone.0112229

123. Modiri A, Stick LB, Rice SR, et al. Individualized estimates of overall survival in radiation therapy plan optimization - A concept study. *Med Phys*. 2018;45(11):5332-5342. doi:10.1002/mp.13211

124. Stick LB, Vogelius IR, Modiri A, et al. Inverse radiotherapy planning based on bioeffect modelling for locally advanced left-sided breast cancer. *Radiother Oncol*. 2019;136:9-14. doi:10.1016/j.radonc.2019.03.018

125. Rechner LA, Modiri A, Stick LB, et al. Biological optimization for mediastinal lymphoma radiotherapy - a preliminary study. *Acta Oncol*. 2020;59(8):879-887. doi:10.1080/0284186X.2020.1733654

126. Scott JG, Sedor G, Scarborough JA, et al. Personalizing Radiotherapy Prescription Dose Using Genomic Markers of Radiosensitivity and Normal Tissue Toxicity in NSCLC. *J Thorac Oncol*. 2021;16(3):428-438. doi:10.1016/j.jtho.2020.11.008

127. Schipper MJ, Taylor JM, TenHaken R, Matuzak MM, Kong FM, Lawrence TS. Personalized dose selection in radiation therapy using statistical models for toxicity and efficacy with dose and biomarkers as covariates. *Stat Med*. 2014;33(30):5330-5339. doi:10.1002/sim.6285

128. Li P, Taylor JMG, Kong S, Jolly S, Schipper MJ. A utility approach to individualized optimal dose selection using biomarkers. *Biom J*. 2020;62(2):386-397. doi:10.1002/bimj.201900030

129. Romeijn HE, Ahuja RK, Dempsey JF, Kumar A, Li JG. A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Phys Med Biol*. 2003;48(21):3521-3542. doi:10.1088/0031-9155/48/21/005

130. Kishimoto S, Yamashita M. A successive LP approach with C-VAR type constraints for IMRT optimization. *Operations Research for Health Care*. 2018;17, 55-64. doi:10.1016/j.orhc.2017.09.007

131. Wu VW, Epelman MA, Pasupathy KS, Sir MY, Deufel CL. A new optimization algorithm for HDR brachytherapy that improves DVH-based planning: Truncated Conditional Value-at-Risk (TCVaR). *Biomed. Phys. Eng. Express*. 2020;6. doi:10.1088/2057-1976/abb4bc

132. Thieke C, Bortfeld T, Küfer KH. Characterization of dose distributions through the max and mean dose concept. *Acta Oncol*. 2002;41(2):158-161. doi:10.1080/028418602753669535

133. Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models. *Phys Med Biol*. 2004;49(10):1991-2013. doi:10.1088/0031-9155/49/10/011

134. Hoffmann AL, den Hertog D, Siem AY, Kaanders JH, Huizenga H. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. *Phys Med Biol*. 2008;53(22):6345-6362. doi:10.1088/0031-9155/53/22/006

135. Kierkels RG, Korevaar EW, Steenbakkers RJ, et al. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. *Radiother Oncol*. 2014;112(3):430-436. doi:10.1016/j.radonc.2014.08.020

136. Sun Y, Hawkins PG, Bi N, et al. Serum MicroRNA Signature Predicts Response to High-Dose Radiation Therapy in Locally Advanced Non-Small Cell Lung Cancer. *Int J Radiat Oncol Biol Phys*. 2018;100(1):107-114. doi:10.1016/j.ijrobp.2017.08.039

137. Dess RT, Sun Y, Matuszak MM, et al. Cardiac Events After Radiation Therapy: Combined Analysis of Prospective Multicenter Trials for Locally Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol*. 2017;35(13):1395-1402. doi:10.1200/JCO.2016.71.6142

138. Soni PD, Boonstra PS, Schipper MJ, et al. Lower Incidence of Esophagitis in the Elderly Undergoing Definitive Radiation Therapy for Lung Cancer. *J Thorac Oncol*. 2017;12(3):539-546. doi:10.1016/j.jtho.2016.11.2227

139. Hawkins PG, Boonstra PS, Hobson ST, et al. Radiation-induced lung toxicity in non-small-cell lung cancer: Understanding the interactions of clinical factors and cytokines with the dose-toxicity relationship. *Radiother Oncol*. 2017;125(1):66-72. doi:10.1016/j.radonc.2017.09.005

140. Hobbs BP, Thall PF, Lin SH. Bayesian Group Sequential Clinical Trial Design using Total Toxicity Burden and Progression-Free Survival. *J R Stat Soc Ser C Appl Stat*. 2016;65(2):273-297. doi:10.1111/rssc.12117

141. Bradley JD, Paulus R, Komaki R, et al. 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 factorial phase 3 study. *Lancet Oncol*. 2015;16(2):187-199. doi:10.1016/S1470-2045(14)71207-0

142. Bradley JD, Hu C, Komaki RR, et al. Long-Term Results of NRG Oncology RTOG 0617: Standard- Versus High-Dose Chemoradiotherapy With or Without Cetuximab for Unresectable Stage III Non-Small-Cell Lung Cancer. *J Clin Oncol*. 2020;38(7):706-714. doi:10.1200/JCO.19.01162

143. Siebers JV, Lauterbach M, Tong S, Wu Q, Mohan R. Reducing dose calculation time for accurate iterative IMRT planning. *Med Phys*. 2002;29(2):231-237. doi:10.1118/1.1446112

144. Zarepisheh M, Hong L, Zhou Y, et al. Automated intensity modulated treatment planning: The expedited constrained hierarchical optimization (ECHO) system. *Med Phys*. 2019;46(7):2944-2954. doi:10.1002/mp.13572

145. Hong TS, Craft DL, Carlsson F, Bortfeld TR. Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head. *Int J Radiat Oncol Biol Phys*. 2008;72(4):1208-1214. doi:10.1016/j.ijrobp.2008.07.015

146. Pardo-Montero J, Fenwick JD. An approach to multiobjective optimization of rotational therapy. II. Pareto optimal surfaces and linear combinations of modulated blocked arcs for a prostate geometry. *Med Phys*. 2010;37(6):2606-2616. doi:10.1118/1.3427410

147. Müller BS, Shih HA, Efstathiou JA, Bortfeld T, Craft D. Multicriteria plan optimization in the hands of physicians: a pilot study in prostate cancer and brain tumors. *Radiat Oncol*. 2017;12(1):168. Published 2017 Nov 6. doi:10.1186/s13014-017-0903-z

148. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J Clin Oncol* 2008;26:3770–6.

149. Jensen AB, Hansen O, Jorgensen K, Bastholt L. Influence of late side-effects upon daily life after radiotherapy for laryngeal and pharyngeal cancer. *Acta Oncol* 1994;33:487–91.

150. Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *Lancet Oncol*. 2010;11(1):85-91. doi:10.1016/S1470-2045(09)70231-1

151. Dijkema T, Raaijmakers CP, Ten Haken RK, et al. Parotid gland function after radiotherapy: the combined michigan and utrecht experience. *Int J Radiat Oncol Biol Phys*. 2010;78(2):449-453. doi:10.1016/j.ijrobp.2009.07.1708

152. Eisbruch A, Kim HM, Feng FY, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. *Int J Radiat Oncol Biol Phys*. 2011;81(3):e93-e99. doi:10.1016/j.ijrobp.2010.12.067

153. Xu Y, Diwanji T, Brovold N, et al. Assessment of daily dose accumulation for robustly optimized intensity modulated proton therapy treatment of prostate cancer. *Phys Med*. 2021;81:77-85. doi:10.1016/j.ejmp.2020.11.035

154. Giacometti V, King RB, Agnew CE, et al. An evaluation of techniques for dose calculation on cone beam computed tomography. *Br J Radiol*. 2019;92(1096):20180383. doi:10.1259/bjr.20180383

155. Gardner SJ, Mao W, Liu C, et al. Improvements in CBCT Image Quality Using a Novel Iterative Reconstruction Algorithm: A Clinical Evaluation. *Adv Radiat Oncol*. 2019;4(2):390-400. Published 2019 Jan 10. doi:10.1016/j.adro.2018.12.003

156. Anderson BM, Lin EY, Cardenas CE, et al. Automated Contouring of Contrast and Noncontrast Computed Tomography Liver Images With Fully Convolutional Networks. *Adv Radiat Oncol*. 2020;6(1):100464. Published 2020 May 16. doi:10.1016/j.adro.2020.04.023

157. Rigaud B, Anderson BM, Yu ZH, et al. Automatic Segmentation Using Deep Learning to Enable Online Dose Optimization During Adaptive Radiation Therapy of Cervical

Cancer. *Int J Radiat Oncol Biol Phys*. 2021;109(4):1096-1110. doi:10.1016/j.ijrobp.2020.10.038

158. Rhee DJ, Jhingran A, Rigaud B, Netherton T, Cardenas CE, Zhang L, Vedam S, Kry S, Brock KK, Shaw W, O'Reilly F, Parkes J, Burger H, Fakie N, Trauernicht C, Simonds H, Court LE. Automatic contouring system for cervical cancer using convolutional neural networks. *Med Phys*. 2020 Nov;47(11):5648-5658. doi: 10.1002/mp.14467. Epub 2020 Oct 9. PMID: 32964477; PMCID: PMC7756586.

159. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol*. 2019;29(3):185-197. doi:10.1016/j.semradonc.2019.02.001

160. McCulloch MM, Lee C, Rosen BS, et al. Predictive Models to Determine Clinically Relevant Deviations in Delivered Dose for Head and Neck Cancer. *Pract Radiat Oncol*. 2019;9(4):e422-e431. doi:10.1016/j.prro.2019.02.014

161. Heukelom J, Kantor ME, Mohamed ASR, et al. Differences between planned and delivered dose for head and neck cancer, and their consequences for normal tissue complication probability and treatment adaptation. *Radiother Oncol*. 2020;142:100-106. doi:10.1016/j.radonc.2019.07.034