12-2016

INVESTIGATION OF RADIATION INJURY IN THE ESOPHAGUS FROM DEFINITIVE CHEMORADIATION THERAPY USING NOVEL IMAGING BIOMARKERS

Joshua S. Niedzielski

INVESTIGATION OF RADIATION INJURY IN THE ESOPHAGUS FROM DEFINITIVE CHEMORADIATION THERAPY USING NOVEL IMAGING BIOMARKERS

A

DISSERTATION

Presented to the Faculty of

The University of Texas
Health Science Center at Houston

and

The University of Texas
MD Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by

Joshua Scott Niedzielski, B.S.

Houston, Texas
September, 2016

Dedicated to my canine companion, Alexander Ovechkin (Ovie). You taught me true friendship, woof.

Acknowledgements

I must begin by thanking the individual that had the most positive impact on my work as well as my development as a researcher, my advisor, Dr. Laurence Court. Our countless discussions and brainstorming sessions about the work presented in this dissertation was the catalyst for my growth in the scientific process.

In addition, I would like to thank Jinzhong Yang, for answering every tedious question I could produce about scientific computing. I also would like to thank my advisory committee members: Mary Martel, Tina Briere, Francisco Stingo, and Daniel Gomez. Their guidance and suggestions helped steered my research into the most productive of directions.

I would like to thank my lab members that helped in numerous ways: Adam Yock, Joey Cheung, Henry Yu, Luke Hunter, Scott Ingram, Ashley Rubinstein, Xenia Fave, David Fried, Kelly Kisling, Rachel McCarroll.

I would also like to thank my classmate Shane Krafft. Our endless discussions about machine learning and statistical analysis helped push this work into a fruitful direction, instead of spinning into dead ends.

Finally, I would like to thank my friends and family that helped and encouraged me: Robert Niedzielski, Judy Niedzielski, Rachael Copsey, Joseph Niedzielski, and everyone else, which are too numerous to name.

INVESTIGATION OF RADIATION INJURY IN THE ESOPHAGUS FROM DEFINITIVE CHEMORADIATION THERAPY USING NOVEL IMAGING BIOMARKERS

Joshua Scott Niedzielski, B.S.

Advisory Professor: Laurence Court, Ph.D.

ABSTRACT

Radiation injury in the esophagus occurs with high frequency from the treatment of non-small cell lung cancer (NSCLC). Radiation esophagitis is an acute normal tissue toxicity that negatively affects treatment efficacy by limiting dose and potentially interrupting radiation therapy. Clinical quantification of this toxicity is typically achieved by utilizing physician grading scales, assigning complication severity on an ordinal scale of symptom presentation and/or physician chosen interventions. These criteria are subjective in nature, both from the physician assigning the grade and the patient reporting the symptom. Furthermore, radiation therapy planning guidelines for the esophagus are derived from toxicity prediction models utilizing these subjective grading scores as complication endpoints. Not only does this schema of toxicity analysis leads to a lack of consistency between models from different study populations, and thereby radiation therapy planning recommendations for the esophagus, but inherent patient radiosensitivity is not considered, leading to suboptimal treatment regimens.

The purpose of this work was to investigate radiation injury in the esophagus by first developing *in-vivo* imaging biomarkers of radiation-response in the esophagus using 4-dimensional computed tomography (4DCT) and ^{18}fluorodeoxyglucose positron emission tomography (FDG-PET), separately. These imaging biomarkers were then compare with radiation esophagitis grade, using traditional and machine learning techniques, and shown to objectively quantify esophageal radiation toxicity. Metrics describing the esophageal radiation response from either imaging modality were strong classifiers of radiation esophagitis grade. Multivariate models to predict maximum esophagitis treatment grade (4DCT), and esophagitis symptom progression (FDG-PET) were developed and had strong performance for both scenarios.

These imaging biomarkers were then used to comprehensively investigate the influence of dose-geometry and radiation type (photon or proton) on esophageal response. Using these radiation-response biomarkers in esophageal dose-response analysis, dose metrics with spatial information of esophageal dose coverage, (e.g. dose to a subregion of the esophagus with specific percent cross-sectional area coverage), as well as without spatial information, (traditional dose-volume histogram), was analyzed separately using machine learning methods. No detectable difference in response was observed when comparing dose metrics with and

without spatial information. Statistical analysis showed no significant difference ($p<0.05$) in biomarker value when comparing patient populations of different radiation type (intensity-modulated photon radiation therapy versus passive-scatter proton therapy).

Inherent patient radiation sensitivity was investigated using esophageal expansion and delivered dose to the corresponding esophageal subregion. Cluster analysis was used to group patient patients based on their maximum expansion and delivered dose to the analyzed subregion of the esophagus. Patients clustered with proportionally higher expansion per delivered dose were considered radiosensitive. These results were then applied to NTCP toxicity modelling by using patient radiosensitivity cluster membership as a predictor variable. Models with the radiosensitive predictor outperformed models not including the cluster membership variable for prediction of grade 3 esophagitis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

NSCLC	Non-Small-Cell Lung Cancer
PEG	Percutaneous Endoscopic Gastrostomy
RTOG	Radiation Therapy Oncology Group
MED	Mean Esophageal Dose
CTCAE	Common Terminology Criteria for Adverse Events
NTCP	Normal Tissue Complication Probability
DVH	Dose Volume Histogram
FDG-PET	Fluorodeoxyglucose-Positron Emission Tomography
CT	Computed Tomography
4DCT	Four Dimensional Computed Tomography
IMRT	Intensity Modulated Radiation Therapy
PSPT	Passive Scatter Proton Therapy
HIPAA	Health Insurance Portability and Accountability Act
BHCT	Breath Hold Computed Tomography
DIR	Deformable Image Registration
JM	Jacobian Map
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
FWHM	Full Width Half Maximum
BIC	Bayesian Information Criterion
LASSO	Least Absolute Shrinkage and Selection Operator
DICOM	Digital Imaging and Communications in Medicine
SUV	Standard Uptake Value
GMM	Gaussian Mixture Model
EM	Expectation Maximization

Chapter 1

Introduction

Radiation Esophagitis is a prevalent normal tissue toxicity with tremendous negative impact on quality of life for patients with non-small-cell lung cancer (NSCLC) treated with radiation therapy.[1] Typically, radiation esophagitis presents as an acute toxicity during radiotherapy for NSCLC, with occurrence rates of approximately 25% for concurrent chemoradiation therapy.[2-6] Mild esophagitis symptoms can be treated or managed with proton pump inhibitors, local anesthetics, oral analgesics, narcotics, and alteration of diet.[7,8] If symptoms become severe enough, intravenous fluids, total parenteral nutrition, or percutaneous endoscopic gastrostomy (PEG) tube, with or without possible hospitalization may be prescribed. These interventions represent large financial burden for the patient and time-consuming resources for the hospital or supporting care facility.

In addition to the negative impact on patient quality of life, acute radiation esophagitis can affect treatment outcomes. Normal esophageal tissue dose, out of concern for radiation esophagitis complications, can limit the amount of prescribed dose to the malignant tumor volume. It is common that the tumor volume is located near the esophagus, and even with advances in treatment planning and delivery (conformation) of dose, the esophagus routinely receives sufficient radiation to induce toxicity. Furthermore, if radiation esophagitis becomes severe enough it can result in treatment interruption. On-time completion of radiation therapy has been identified as the most crucial factor in treatment outcome for unresectable lung cancer.[4]

Several strategies to prevent radiation esophagitis have been studied. One such strategy is the use of radioprotectors such as amifostine. While tumor control is increased by adding concurrent chemotherapy to radiation therapy, this also increases the risk of esophagitis. The concept behind utilizing radioprotectors is to ameliorate the esophagitis symptoms while reaping the tumor control benefit of concurrent chemoradiation therapy, thereby increasing the therapeutic ratio. However, the use of radioprotectors with chemoradiation therapy has had mixed results. Several small trials of radiation therapy combined with the use of radioprotectors, most commonly amifostine, have showed a reduction in radiation esophagitis incidence.[9-13] Conversely, a large randomized trial of amifostine did not show a significant reduction in the occurrence of severe radiation esophagitis.[14]

The most directly controllable clinical factors associated with radiation esophagitis are the dose-constraints to the esophagus during the radiation treatment planning process. For example, the radiation therapy oncology group (RTOG) 0617 trial was designed with a recommended mean esophageal dose (MED) to be less than 34 Gy, as well as the volume of esophagus receiving at least 60 Gy (V60) to be recorded.[2,15] These dose-constraints leave a wide margin of possible dose-volume configurations available for radiation treatment planning. Furthermore, several studies have investigated increased restriction of esophageal dose in the treatment planning process, while maintaining tumor coverage, showing potential optimization of radiotherapy plans for minimizing esophagitis incidence. However, sparing of the esophagus can cause increase in radiation dose to other normal tissues, such as the lung, heart and spinal cord, potentially exceeding the dose-constraints for these organs.[16-18]

Radiation esophagitis is commonly quantified on an ordinal scale with increasing score or value, corresponding to increasing esophagitis symptom severity. One frequently used esophagitis grading system is the Common Terminology Criteria for Adverse Events (CTCAE).[19] The esophagitis severity quantification of CTCAE 3.0 is given in Table 1.1. The most recent version of CTCAE, version 4.0, removed the intravenous fluid criterion for grade determination.[20]

Table 1.1: *CTCAE 3.0 esophagitis grading system.*

Grade	Criteria
1	Asymptomatic pathologic, radiographic, or endoscopic findings only
2	Symptomatic; altered eating or swallowing (altered dietary habits, oral supplements); intravenous fluids indicated for less than 24 hours
3	Symptomatic and severely altered eating or swallowing (inadequate oral caloric or fluid intake); intravenous fluids, tube feedings, or total parenteral nutrition indicated for more than 24 hours
4	Life-threatening consequences
5	Death

There are numerous drawbacks to this methodology of quantifying esophagitis severity. The subjective nature of grades themselves creates uncertainty in the scoring process; the grades are determined from physician chosen interventions based on patient reported symptoms, with variation between both the patient reporting the symptoms and the physician prescribing the intervention. Another drawback to grading of esophagitis severity is the lack of

localization of the toxic region of the esophagus. Under grading schema, the whole organ must be treated as toxic (e.g. grade 3) or asymptomatic (e.g. grade 0). This ignores the possibility that toxicity is located in a sub volume of the esophagus, which should be the primary region of study for esophagitis amelioration and prevention. One important drawback specifically with grading criteria and radiation therapy is the desire to understand cause and effect associated with the radiation dose. Esophagitis symptoms can be caused by conditions unrelated to radiation therapy such as gastroesophageal reflux or esophageal infection. Therefore, it would be ideal to isolate toxicity solely from radiation when investigating esophageal toxicity.

The primary objective of investigating esophageal toxicity is to prevent complications from occurring, rather than minimizing symptom severity after presentation. The development of toxicity prediction models, commonly referred to as normal tissue complication probability (NTCP) models, which are based on previous clinical experience, and then applying these models to future patients with the goal of reducing toxicity incidence. Common practice is to create multivariate logistic regression models with many clinical and dosimetric parameters as model predictors, and esophagitis grade as the dichotomized endpoint. In statistical learning, this is deemed a classification problem, with esophagitis grade being the item sought to be classified.

The literature contains copious amounts of NTCP studies for the esophagus.[2,3,5,6] Comparison of many NTCP studies shows a lack of agreement between specific dosimetric predictors and esophagitis complication. Furthermore, a commonality exists where NTCP models fit the training dataset well, but lack generalizability to external datasets.[2,3,21-24] This is a common pitfall in prediction modelling, where over fitting the training data leads to poor

predictive model performance on external datasets. As previously stated, the esophagitis endpoint itself has associated uncertainty, which will affect any toxicity model's predictive performance.

Another source of uncertainty in the NTCP modelling process includes variability of esophagus positioning on the planning CT, which can affect dosimetric accuracy. In one study, variability of esophagus position in Cartesian space was measured from different phases of 4DCT for 29 patients. Motion of the central axis of the esophagus was as extreme as 4mm in the medial/lateral direction.[25] Another study found esophageal motion as high as 13.8mm in the cranio-caudal direction.[26] A study of 236 lung cancer patients showed correcting the esophagus dose-volume histogram (DVH) for anatomical uncertainties improved correlation of DVH dose metrics to esophagitis grade.[27]

Another explanation for the underwhelming performance of esophagitis prediction models is the lack of consideration for inherent patient radiosensitivity. If we consider two patients receiving similar radiation therapy dose to the esophagus, it is quite possible one patient may develop severe radiation esophagitis, while the other patient may be asymptomatic. If the NTCP model does not have some means to quantify this difference in radiosensitivity, then model uncertainty has been increased as both patients have similar model predictors, but different classification endpoints. Genomic data has recently been introduced into the NTCP modelling process in an attempt to account for radiosensitivity.[28-31] However, validating genetic predisposition to radiation toxicity using esophagitis grade is still beholden to the uncertainty of the subjective complication endpoint.

The inclusion of spatial dose terms, or dose-geometry, can potentially determine if partial sparing of the esophagus is achievable. Most studies use dose-volume metrics such as volume of esophagus receiving at least a particular dose, but some studies have included dose metrics that yield spatial information of dose conformity instead of simply using a gross average or threshold of dose-volume.[32-35] One example of spatial dose metrics is the length of the esophagus receiving a particular dose with a specific percent of axial coverage across the cross-sectional area of the imaged slice of the esophagus.

Previously, optimization of treatment planning for reduced esophagus dose showed feasibility for external beam radiation therapy of lung malignancies.[17] In this study, retrospective analysis of patients that exhibited grade 3 esophagitis during treatment showed sizeable reductions in dose-volume above 50 Gy were achievable on re-planning with heavy constraint on the esophagus, with preservation of tumor dose coverage. Therefore, there exists clinical utility and viability in optimizing dose-geometry of normal esophagus. While investigation of partial sparing has had some success in other normal tissues, such as the salivary glands or bone marrow, up to this point there is no clear influence of dose-geometry and esophageal toxicity.[32-39] It is possible that any dose-geometry effect of toxicity in the esophagus, if it exists, could be undetectable with an endpoint such as esophagitis grade, and may require a more sensitive measure of toxicity. Moreover, if partial sparing is achievable in the esophagus, lack of utilizing dose-geometry could be one explanation for the variability in esophagitis toxicity prediction models.

Esophageal expansion, or swelling, is an inflammatory response that is discernible on CT imaging.[40] Another study has shown radiation esophagitis can be visualized on 4DCT as an in-

vivo technique throughout the course of radiotherapy.[41] Furthermore, this study showed that the change in the esophagus contour over the course of radiation therapy can be used as a surrogate for esophageal expansion. In addition, the average relative expansion of the esophagus was significantly different between patients with grade 0 and grade 3 esophagitis. Therefore, the esophageal expansion response may be a reliable quantification of radiation response in the esophagus.

^{18}Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) is a functional imaging technique that uses glucose metabolism as a surrogate for tissue inflammation. Primarily used in the radiation oncology setting for tumor detection, delineation, and treatment assessment, FDG-PET has been shown to quantify radiation-induced lung toxicity.[42-44] The slope of FDG uptake and lung dose has been shown to correlate to radiation pneumonitis grade for both lung and esophageal cancers in external beam radiation therapy.[45-49] Another study examined FDG uptake correlation to esophagitis grade using FDG-PET imaging at follow up from radiation therapy.[50] This study showed the addition of FDG uptake into a NTCP model increased prediction of grade 2 and higher esophagitis.

The purpose of this work is to comprehensively enhance understanding of radiation injury in the esophagus, which can thereby optimize radiation therapy treatment outcomes and improve patient quality of life. Esophageal expansion and inflammation, from 4DCT and FDG-PET respectively, will be investigated as radiation response and toxicity measures. This is to address the uncertainty inherent in current toxicity grading systems. Any improvement in toxicity prediction models will be explored and then used to investigate dose-response in the esophagus, with the previously mentioned prediction model sources of uncertainty addressed.

This will be applied in both broad dose-response models, and a spatially localized response analysis to investigate any influence in esophageal dose-geometry and toxicity, to determine if partial sparing can be achieved in the esophagus. Any influence of radiation type and esophageal response will also be examined. Potentially, expansion and inflammation can yield information about patient-specific response to radiation, and therefore inherent patient radiosensitivity. The efficacy of the expansion mechanism to identify radiosensitive patients, and this information's impact on dose-response and radiation injury in the esophagus will be investigated.

Chapter 2

Central Hypothesis and Specific Aims

2.1 Central Hypothesis

The overall objective of this project is to identify objective, *in-vivo* biomarkers of esophageal radiation-response as quantifications of toxicity and to utilize these metrics to improve prediction modelling of radiation injury in the esophagus, as well as to determine if dose-response in the esophagus is dependent on dose-geometry or radiation therapy modality, and to investigate the use of these radiation-response measures as patient-specific biomarkers of radiosensitivity.

The **central hypothesis** is that objective metrics quantified from 4DCT and FDG-PET imaging, during radiation therapy, are biomarkers of radiation-response in the esophagus, which can be used to improve understanding of radiation injury in the esophagus.

2.2 Specific Aims

To test the central hypothesis, the following specific aims were developed and investigated:

Specific Aim 1: Analysis of CT-based esophageal expansion to quantify radiation-response in the esophagus during radiation therapy.

Hypothesis: CT-based esophageal expansion is a biomarker of radiation-response that can quantify radiation injury in the esophagus and can be used to improve outcome modelling of radiation-induced esophagitis.

To test the hypothesis, the following projects were conducted:

Project 1.1: Analyze whether CT-based expansion is a quantification of esophgeal radiation toxicity

Project 1.2: Analyze the utility of CT-based expansion metrics in outcome modelling

Specific Aim 2: Analysis of FDG-PET to quantify esophageal radiation-response in the esophagus during radiation therapy.

Hypothesis: FDG-PET uptake is a biomarker of radiation-response that can quantify radiation injury in the esophagus and can be used to predict esophagitis symptom progression during radiation therapy.

To test the hypothesis, the following projects were conducted:

Project 2.1: Analyze whether FDG uptake in the esophagus is a quantification of esophageal radiation toxicity

Project 2.2: Analyze whether FDG uptake in the esophagus can predict esophagitis symptom progression during radiation therapy

Specific Aim 3: Analysis of esophageal dose-response using radiation-response biomarkers.

Hypothesis: Esophageal expansion will identify if dose-geometry or radiation type contributes to radiation injury in the esophagus, and that expansion can be used to quantify patient-specific radiosensitivity.

To test the hypothesis, the following projects were conducted:

Project 3.1: Analyze whether dose-geometry influences esophageal radiation response.

Project 3.2: Analyze whether there exists any difference in esophageal radiation response between photon-based intensity modulated radiation therapy (IMRT) and passive-scatter proton therapy (PSPT).

Project 3.3: Analyze whether esophageal expansion can quantify patient-specific radiosensitivity and determine if this knowledge can improve understanding of esophageal dose-response.

The outcomes of this work show esophageal expansion and normalized FDG uptake are objective, *in-vivo* radiation response measures that can be used as surrogates for radiation injury in the esophagus. Moreover, when utilized as endpoints, these biomarkers improve toxicity prediction modelling in the esophagus. Finally, the expansion biomarker and corresponding dose show inherent patient-specific radiosensitivity. This work will have positive impact by developing and applying *in-vivo* toxicity biomarkers that can potentially enhance personalization of radiation therapy, leading to improved treatment outcomes and increase patient quality of life.

2.3 Dissertation Structure

The remaining chapters of the dissertation are organized to address each specific aim in chronological order. Chapters 3 and 4 together address specific aim 1, by first developing a computational method to calculate esophageal expansion and correlating this measure of radiation-response to esophageal toxicity (project 1.1), and then examining predictive model performance with expansion as an endpoint (project 1.2). Chapter 5 addresses specific aim 2 by calculating normalized FDG uptake and correlating this measure of response to esophageal toxicity (project 2.1), and then using this biomarker to predict symptom progression (project 2.2).

Chapter 6, 7, and 8 address specific aim 3. Chapter 6 presents the investigation of dose-geometry and its relationship with esophageal expansion (project 3.1). The work in Chapter 7 analyzes esophageal response from different therapeutic radiation types (proton or photon radiation, project 3.2). Chapter 8 describes the use of the objective metrics to identify patient-specific radiosensitivity, and then apply these radiosensitivity assays to improve prediction modelling of radiation-induced esophagitis (project 3.3).

Chapter 9 presents a summary and the conclusions of this work. This chapter also presents directions for future work based on the results of this dissertation.

Chapter 3

Esophageal Expansion Quantified from 4-Dimensional Computed Tomography as a Measure of Esophageal Radiation Injury

---

The work in this chapter investigates esophageal expansion, as quantified from 4DCT, as a radiation-response and toxicity measure for radiation injury in the esophagus. Robust metrics of esophageal expansion are derived for classification of toxicity. These expansion metrics are then shown to be robust biomarkers of esophageal radiation-response and esophageal toxicity, by statistical analysis with physician scored radiation esophagitis grade. This chapter comprises project 1.1, of specific aim 1.

A substantial portion of this chapter is based on the following publication:

Niedzielski JS, Yang J, Stingo F, Martel MK, Mohan R, Gomez DR, Briere TM, Liao Z, Court LE. Objectively Quantifying Radiation Esophagitis with Novel Computed Tomography-based Metrics. *Int J Radiat Oncol Biol Phys* **2016** 94(2):385-393.

doi: 10.1016/j.ijrobp.2015.10.010

3.1 Methods and Materials

3.1.1 Patient Population

For the work in this chapter, 94 patients were selected from a prospective clinical trial of radiation therapy using either IMRT or passive-scatter proton therapy (PSPT) for stage III NSCLC at University of Texas-MD Anderson Cancer Center, treated with intensity-modulated radiation therapy and concurrent chemotherapy (paclitaxel and carboplatin), with tumor prescription doses of 60 (n=4), 66 (n=28), or 74 (n=53) Gy in 2-Gy fractions over 6-8 weeks. Due to sample size we utilized the patients from the IMRT arm of this trial. Of the 94 study patients, 85 patients had weekly 4DCT imaging, while a total of 9 patients with weekly breath hold CT (BHCT) imaging. Three patients were excluded from the original 97 IMRT patient cohort due to poor image quality.

These patients had prospective weekly esophagitis scoring by the radiation oncologist according to Common Terminology Criteria for Adverse Events version (CTCAE) 3.0.[19] The grading scale can be summarized as: grade 0, no esophagitis; grade 1, asymptomatic with only clinical or diagnostic observations; grade 2, symptomatic with altered eating/swallowing and oral supplements; grade 3, severely altered eating/swallowing with tube feeding, total peritoneal nutrition, or hospitalization; grade 4, life-threatening consequences; and grade 5 is death.[20] The clinical esophagitis symptom management was: liquid narcotic medication, topical anesthetics, and antacid medication for grade 2, and IV fluids with possible feeding tube for grade 3. The distribution of maximum esophagitis grades during treatment was: 24 were grade 0, 45 were grade 2, and 16 were grade 3. There were no grade 1 patients in this study, as asymptomatic diagnostic assessment of esophagitis was not conducted. This study was

approved by the University of Texas-MD Anderson Cancer Center Institutional Review Board and was compliant with HIPAA regulations.

CT scans were acquired on General Electric Lightspeed Discovery ST or Lightspeed RT16 (GE Healthcare, Waukesha, WI) or Philips Brilliance 64 (Philips Healthcare, Bothell, WA) CT scanners operated at 120 kV. Voxel dimensions were 0.98x0.98x2.50 mm in the right-left direction, anterior-posterior, and superior-inferior direction, respectively, with a 512×512 pixel area. Treatment planning for all patients whose data were used in our study was conducted using the Pinnacle treatment planning system (Phillips Healthcare), including segmentation. Esophageal contours were segmented in Pinnacle version 9.8 in the axial plane, from the cricoid cartilage to the gastroesophageal junction by an experienced treatment planner and verified by the radiation oncologist.

3.1.2 Computational Framework and Jacobian Map Algorithm

The computational framework used to calculate esophageal expansion on the treatment plan for any treatment week can be summarized in 3 steps: segmentation, deformable image registration, and the Jacobian map algorithm. A general overview of this process is shown in Figure 3.1. Esophagus contours were delineated on the planning image. Next, deformable registration was performed between the plan and weekly CT images to obtain deformation vector fields, and propagate segmentation to weekly CT images. Finally, the algorithm uses the deformation vector fields, planning, and weekly contours to calculate the local esophageal volume change with correction due to anatomical variability as an average of slices of the esophagus along the cranio-caudal axis.

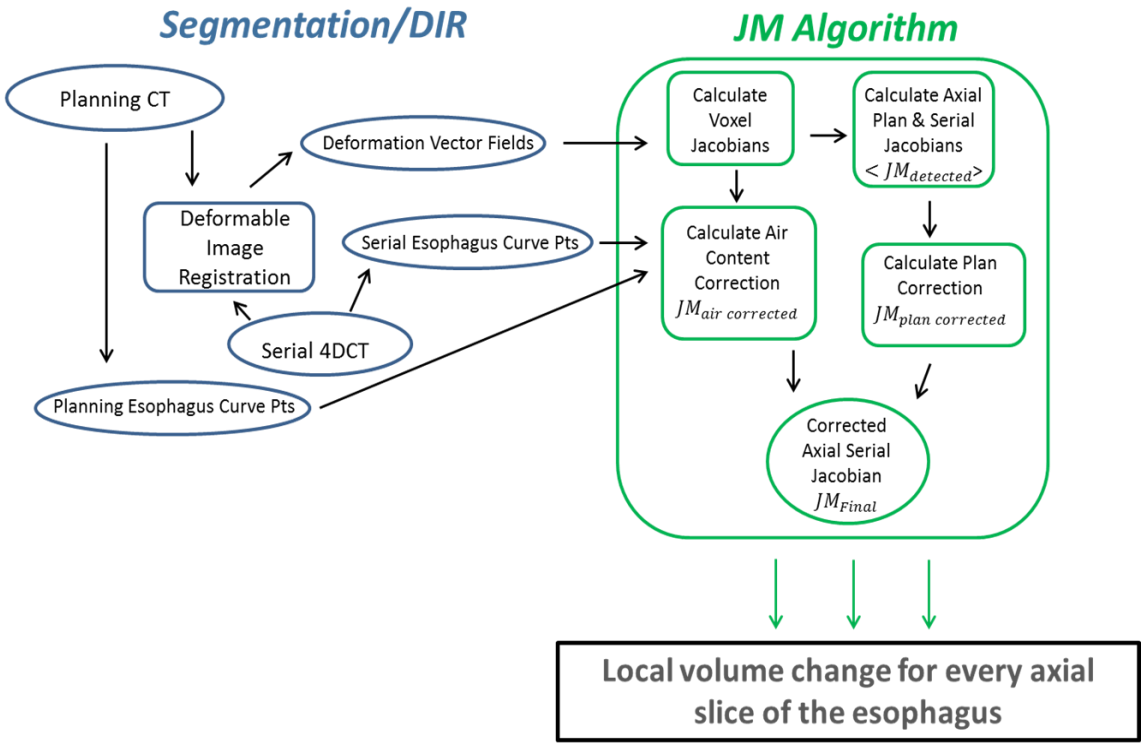

Figure 3.1 *General overview of the computational framework to calculate esophageal expansion.*

A demons algorithm was used to perform deformable image registration from the planning CT to the weekly CT image set; this algorithm was validated for thoracic patients.[7,8] Let $\boldsymbol{du} = (dx, dy, dz)$ denote the deformation vector pointing from voxel $\boldsymbol{u} = (x, y, z)$ in the planning image to the weekly CT image voxel $\boldsymbol{u}' = (x', y', z')$. The voxel mapping from the plan to the weekly image becomes:

$$\boldsymbol{u}'(x', y', z') = \boldsymbol{u}(x, y, z) + \boldsymbol{du}(dx, dy, dz) \tag{3.1}$$

The Jacobian Map is calculated by taking the determinant of the Jacobian of the transformation defined in equation 3.1, using $\boldsymbol{du}$ for every voxel in the esophagus. The Jacobian represents the local voxel-volume change, and thus voxel esophageal expansion, from the deformable registration of the plan to the weekly CT imageset.[52-55] A Jacobian map of 1.0, >1.0,

and <1.0, represents no volume change, relative expansion, and relative shrinkage, respectively, for the voxel anatomy from the planning to the corresponding weekly CT.

For each of the 85 patients, after deforming the planning exhale phase all the other plan phases, and all weekly 4DCT phases, the deformation vector fields were used to calculate the corresponding Jacobians. The Jacobians were then averaged at each axial slice of the esophagus and esophageal expansion metrics were calculated.

To validate this computational framework, a digital phantom representing esophageal expansion was developed consisting of a pair of CT image sets containing: (1) a uniform cylindrical volume of tissue equivalent CT numbers, and (2) another cylindrical volume of tissue equivalent CT numbers, but regions of differing volume in the axial plane to represent expansion in the esophagus. Figure 3.2 illustrates this phantom, in which the expansion image has a region of the same area as the uniform image and regions of twice and thrice the original uniform area ('2A' and '3A' in Figure 3.2, respectively), as well as regions of changing cross-sectional area (white region in 'target image' of Figure 3.2). The observed axial expansion is tested against the ground truth of relative volume change, which is known for a given slice.

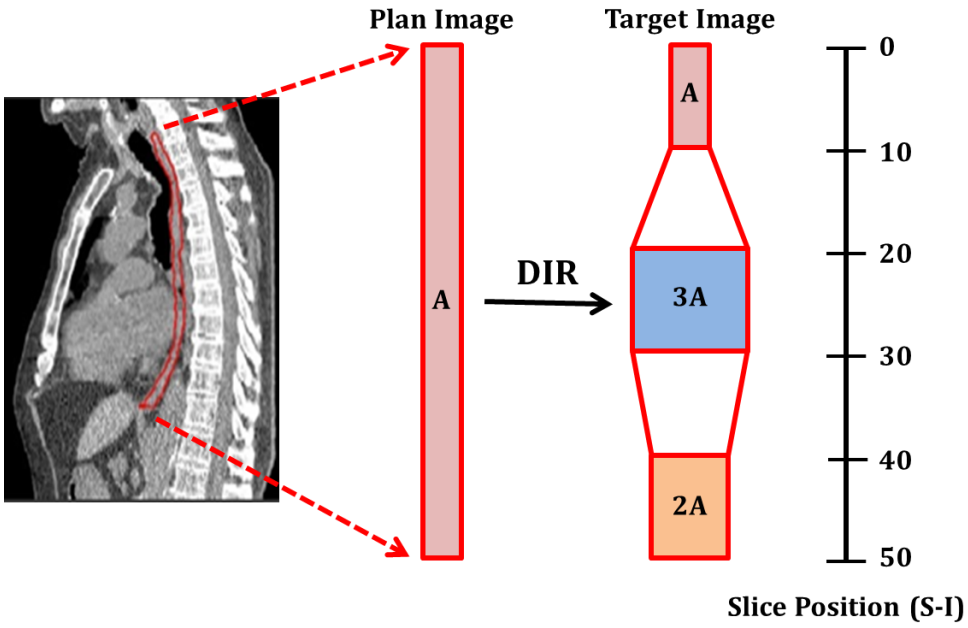

Figure 3.2: *Representation of the digital phantom image pair that is used to validate the jacobian map algorithm used to quantify esophageal expansion. The plan image of uniform cross-sectional area is deformed to a target image that has the axial area known for each slice in the esophagus, along the superior-inferior direction. The deformation vector field from the deformable image registration is entered into the Jacobian map algorithm and relative volume change for each axial slice of the phantom esophagus is calculated. Uncertainty in the Jacobian map calculation is then assessed using the known and measured axial volume values.*

3.1.3 Anatomic Volume Variability and Correction Methods

Physiological effects that skew esophageal volume calculations are numerous, including motility of the esophagus, dilation, swallowing during CT acquisition, and esophageal air. The presence of air in the esophagus is particularly challenging. Because deformable image registration uses CT number as a measure of image similarity, any 2 images in which esophageal air is present in one image but not the other have a high chance to cause deformable registration error in this region, propagating into the Jacobian calculation. The uncertainty caused by esophageal air makes expansion quantification on the voxel-level unreliable. However, quantifying expansion for an axial slice is possible and a method to correct uncertainty associated with esophageal air can be accomplished by utilizing the axial anatomy.

To reduce errors caused by air content, an "air content correction factor" can be calculated for each axial slice as the relative ratio of tissue in the esophagus on the planning image to that in the weekly image:

$$JM_{air\ corrected} = JM_{detected} * \psi, \qquad where\ \psi = \frac{relative\ weekly\ treatment\ esophageal\ tissue}{relative\ plan\ esophageal\ tissue} \tag{3.2}$$

The quantity $JM_{air\ corrected}$ is the axial-averaged Jacobians corrected for air content in both planning and weekly images, and $JM_{detected}$ is the originally calculated Jacobian. The quantity ψ is the air content correction factor, computed at each axial slice of the esophagus.

Miscalculations of local volume change caused by intra-scan transient effects are minimized in 2 ways. The first method corrects transient effects on the plan image by deforming the exhale phase to all phases of the planning image and then computing an axial-Jacobian-averaged from all the phase deformations, which yields the plan normalization correction factor:

$$JM_{plan\ norm\ corrected} = \frac{<JM_{detected}>}{\Phi}, \qquad where\ \Phi = < JM_{exhale\ phase\ to\ plan\ phases} > \tag{3.3}$$

The mean axial-averaged Jacobian from the exhale phase to the planning phase deformations is the plan normalization correction factor, denoted Φ. Dividing the Jacobian of the planning exhale phase to all weekly phases, denoted $< JM_{detected} >$, by Φ yields the Jacobian corrected for transient effects $JM_{plan\ norm\ corrected}$. Random esophageal motion and swallowing during CT acquisition are transient anatomical effects that may reduce the accuracy of the Jacobian calculation.

The second method reduces contributions to volume change from non-treatment-related sources on the weekly CT. Similar to the first correction, we utilize all phases of the

weekly images to create a mean axial-averaged Jacobian. The mean of the plan exhale phase Jacobian to all weekly phases averaged at each axial segment of the esophagus, applying both anatomic variability correction factors, becomes:

$$JM_{Final} = < JM_{Corrected} > = \frac{<JM_{detected}>}{\Phi} * \psi \quad (3.4)$$

where $< JM_{Corrected} >$ is the weekly 4DCT phased-averaged axial Jacobian with all correction factors applied, resulting in JM_{Final}, which is the weekly Jacobian used to derive all metrics of esophageal expansion for the given treatment week.

To test the correction methodology and quantify uncertainties, the Jacobian from planning to first treatment week is used to measure inter-CT scan variability without radiation-response, because insufficient time or dose has been delivered to induce esophageal expansion. The effect of anatomic variability correction is quantified by 2 metrics, with and without anatomic corrections. The first metric is the absolute difference between the axial Jacobians and a value of 1.0. Jacobian calculation without any radiation-induced expansion should show values close to 1.0. The second metric is the full-width half-maximum (FWHM) of the distribution of axial Jacobian values for all phases of the first week; accurate deformations have normal distributions of Jacobians centered near 1.0, with small FWHM values.

3.1.4 Esophageal Expansion Metrics

Various metrics were created to quantify esophageal swelling, including the following:

- Mean axial esophageal expansion (MeanExp)
- Maximum axial esophageal expansion (MaxExp)
- Esophageal length ≥ specified percent of axial expansion (LenExp), ranging from 20% to 100% in 10% increments (LenExp20% for 20% expansion)
- Peak esophageal expansion of 3, 5, and 7 axial slices of the esophagus (PeakExp3 for 3-slice peak expansion)
- Percentile of esophageal expansion (PerExp), ranging from sixtieth to ninetieth in increments of 10, as well as ninety-eighth (PerExp60% for sixtieth percentile of expansion)

These quantifications allowed us to examine expansion using the average (MeanExp), maximum (MaxExp), spatial-length dependence (LenExp), and volume dependence of the expansion (PerExp). The MaxExp3-MaxExp7 metrics are meant to overcome uncertainties in quantifying MaxExp1 (which measures only a single axial response point). LenExp represents the physical esophageal length that increases in volume at least a given percentage. PerExp represents the expansion value for which the given percentiles of all other expansion values are below. An axial comparison of esophagi before and during treatment is shown in Figure 3.3. The temporal relationship between expansion and esophagitis grade was also investigated by comparing the timing of maximal expansion and esophagitis grade.

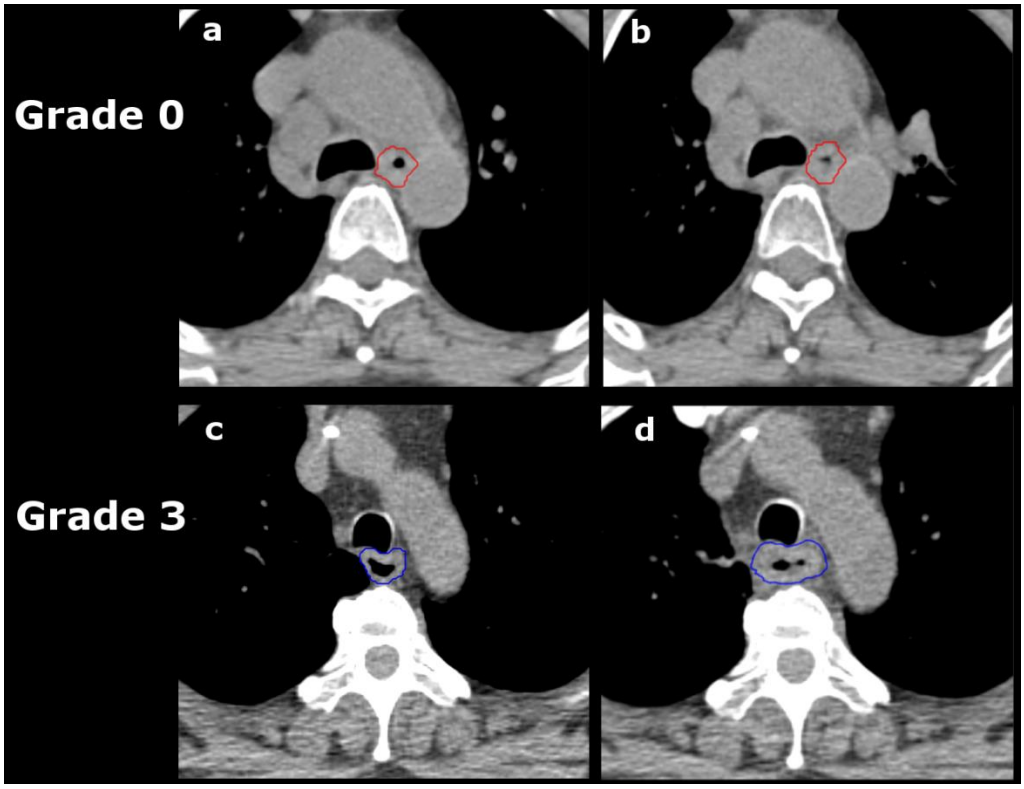

Figure 3.3: *Example of axial expansion of esophagus as relative change from planning (a,c) to treatment week 6 (b,d). Top patient (a,b) has maximum grade 0 esophagitis and no change in esophagus volume (esophagus is outlines in red). Bottom patient (c,d) has maximum grade 3 esophagitis and considerable expansion (esophagus is outlines in blue).*

3.1.5 Expansion & Toxicity Analysis

An analysis was conducted to determine any relationship between the expansion metrics and radiation esophagitis severity. The treatment week with maximal expansion was compared to the patient's maximum esophagitis grade during treatment. Normal tissue complication probability (NTCP) from univariate logistic regression and Spearman rank correlation coefficients were calculated for expansion metrics with grade 2 and grade 3 esophagitis endpoints, with the metric value with 50% probability of grade 2 and grade 3 esophagitis calculated for each expansion metric. P-values were calculated using the likelihood ratio chi-square test. The Benjamini-Hochberg false discovery rate procedure was applied to all

p-values. Receiver operating characteristic (ROC) analysis was used to quantify the performance of each expansion metric in classifying esophagitis. For all analyses, $p<0.05$ after application of the Benjamini-Hochberg procedure was considered statistically significant. All statistical analyses were carried out in Matlab version 7.9 or version 8.2 (Mathworks, Natick, MA).

3.1.6 Esophageal Expansion & Breath Hold CT

To determine the suitability of this expansion calculation methodology with breath hold CTs (BHCT), all 9 patients from the clinical trial patient pool with weekly BHCT imaging during treatment were analyzed. Of these nine patients, 5 had both weekly 4DCT and BHCT imaging for a single treatment week, in addition to a plan 4DCT image set. This allowed a direct comparison of esophageal expansion calculated from either 4DCT or BHCT imagining. The expansion computational framework was tested on these BHCT patients in the same manner as the weekly 4DCT patient cohort.

The planned BHCT was deformed to all 3 BHCTs taken during the planning process, as well as 3 BHCTs acquired for every treatment week. The expansion metrics were then calculated for every treatment week with an available serial CT. In addition, the variance of the esophageal volume was calculated for the plan BHCT to the other BHCTs taken on the planning date and the serial BHCTs. The same methodology of correcting esophageal anatomic variability used for the patients with primary 4DCT imaging was applied to BHCT patients, with multiple BH scans used in place of the phases of the 4DCT image set.

3.2 Results

3.2.1 Validation of Jacobian Map Algorithm and Anatomic Volume Variability

The result from the digital phantom test of the Jacobian map algorithm is shown in Figure 3.4. The top pane of Figure 3.4 shows good agreement between calculated and known anatomical volume change with the pattern of volume change along the superior-inferior direction of the esophagus, as originally shown in Figure 3.2. Furthermore, this test shows that uncertainty in the Jacobian Map algorithm correctly calculating the relative volume change of the esophagus is below 5.0% for any given esophageal slice.

Next, the anatomical variability correction was tested on 75 of the 85 study patients. Since treatment week 1 images were not available for 10 of these 85 patients, these patients were excluded from the anatomic correction analysis. The top panel in Figure 3.5 shows the axial Jacobian profile along the esophagus for the planning exhale phase deformed to the planning image for one patient. The volume variability of each phase-Jacobian can be observed on the planning image, and variability was reduced by taking the mean Jacobian of all phases from each axial location (dashed black line). The mean axial Jacobian was close to 1.0 for all points along the esophageal-length, representing an accurate calculation. A similar trend was observed with the Jacobian profiles for the first treatment.

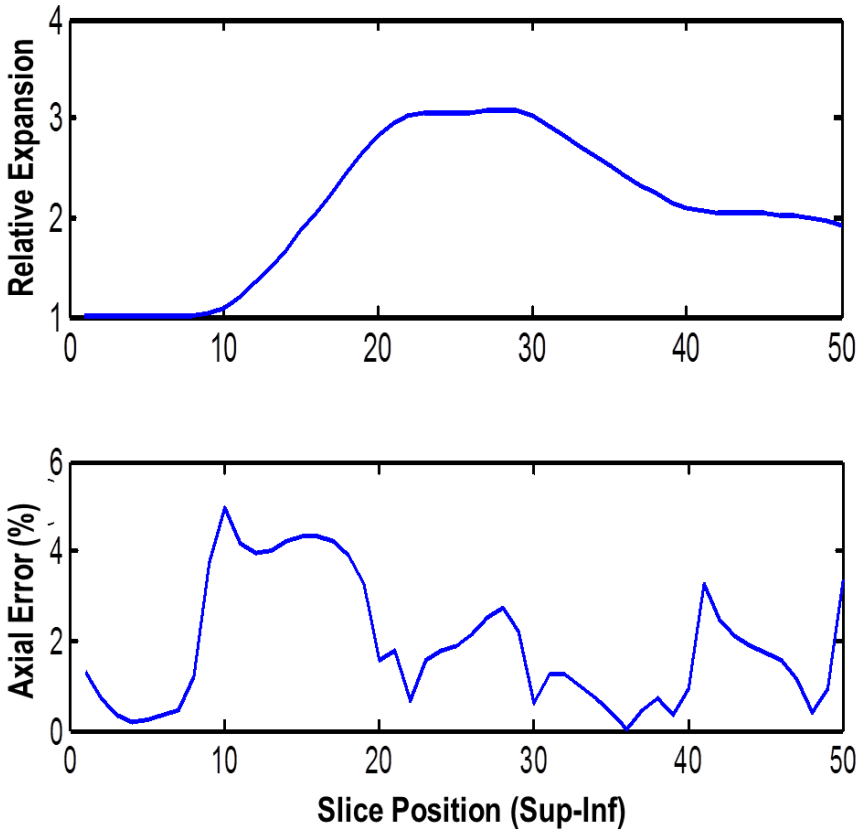

Figure 3.4: *Results of the digital phantom test to validate relative volume change clouted using the Jacobian map algorithm along the superior-inferior length of the phantom. The relative expansion calculated from the Jacobian map algorithm is shown at the top. The percent error in calculated versus known axial volume is shown at the bottom.*

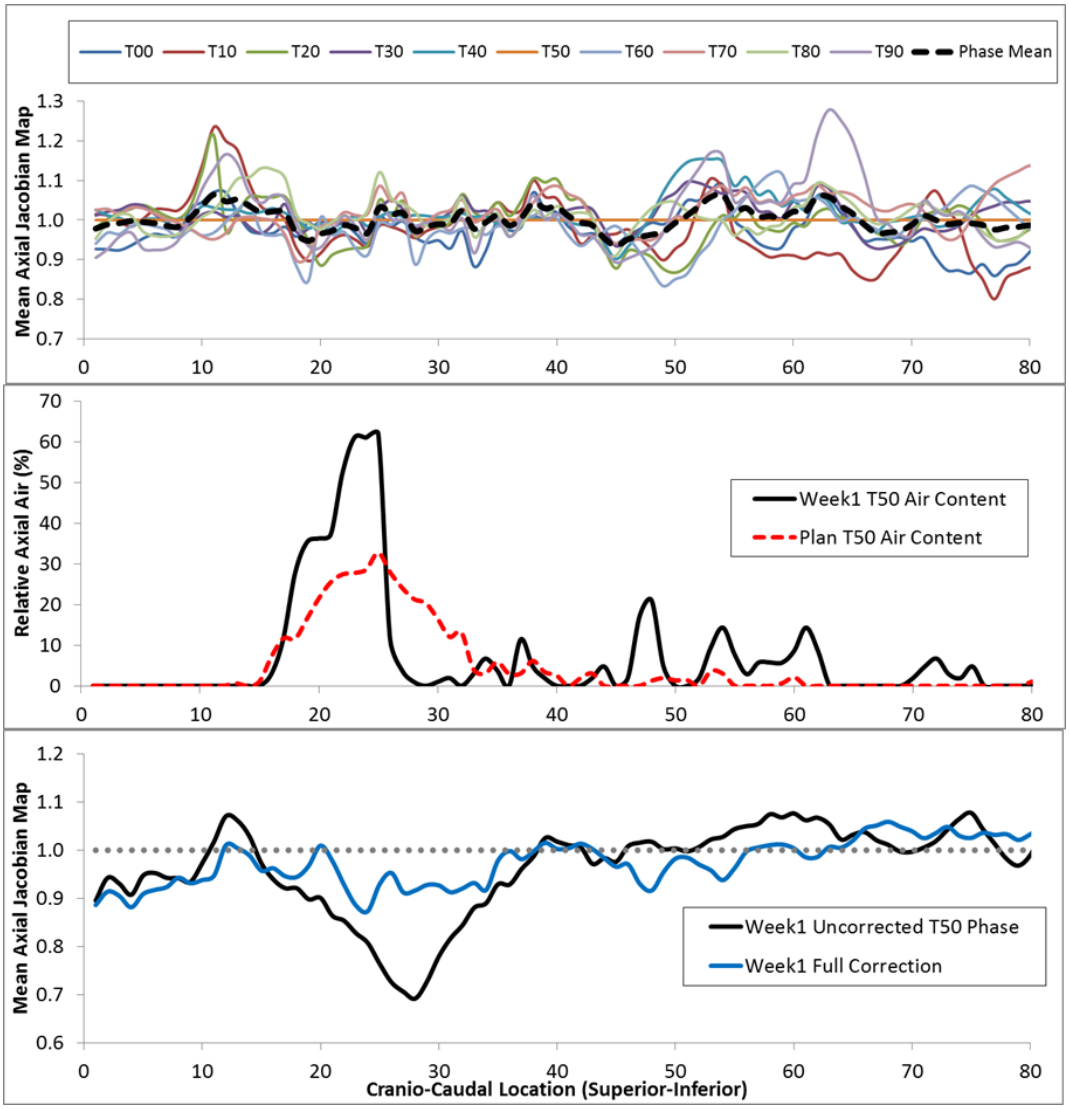

Figure 3.5: *(Top) Axial averaged Jacobian map for all phases and mean Jacobian (black dashed line) of all phases for the planning 4DCT image set for one patient. T00 represents the inspiration phase and T50 the exhale phase. (Middle) Relative axial cross-sectional area of air content for the planning (solid black line) and week 1 (dashed red line) T50 phases for one patient. (Bottom) Axial averaged Jacobian map of the week 1 T50 phase, uncorrected (black line) and with full anatomic correction (blue line), for one patient.*

The middle panel in Figure 3.5 shows the relative air content in plan and week 1 exhale images for the same patient. A large difference in air content occurred around slice 25, causing Jacobian miscalculation at the corresponding slice, shown in the bottom panel of Figure 3.5 (solid black line). By applying the anatomic corrections, we obtained a more accurate Jacobian value (solid blue line, bottom panel).

Figure 3.6a,b shows the plan to week1 Jacobian distributions with and without anatomic variability correction for one patient. The distribution had a long asymmetric tail to small Jacobians before correction and a more symmetric distribution with smaller FWHM values and a peak centered closer to 1.0 after correction.

Figure 3.6c is a boxplot showing the distribution of FWHM values for the plan to week1 Jacobian distributions before and after anatomic correction was applied. Applying the anatomic correction reduced the FWHM values of the Jacobian distributions by 10.3% (±5.6%), average percent-difference with standard deviation in brackets. For all axial slices, the mean absolute percent-differences between the Jacobians and a value of 1.0 ranged from 13.3% (±5.4%) to 9.2% (±5.6%) after the anatomic correction were applied.

3.2.2 Expansion Metrics and Esophagitis Severity

Expansion metric distributions were grouped according to esophagitis grade (Figure 3.7a-c). This analysis illustrated a strong relationship between increased expansion values and increased esophagitis grades; the relationship was most pronounced for the MaxExp-based metrics. For most metrics, a gap was evident between the highest value for grade 0 and the lowest value for grade 3 esophagitis.

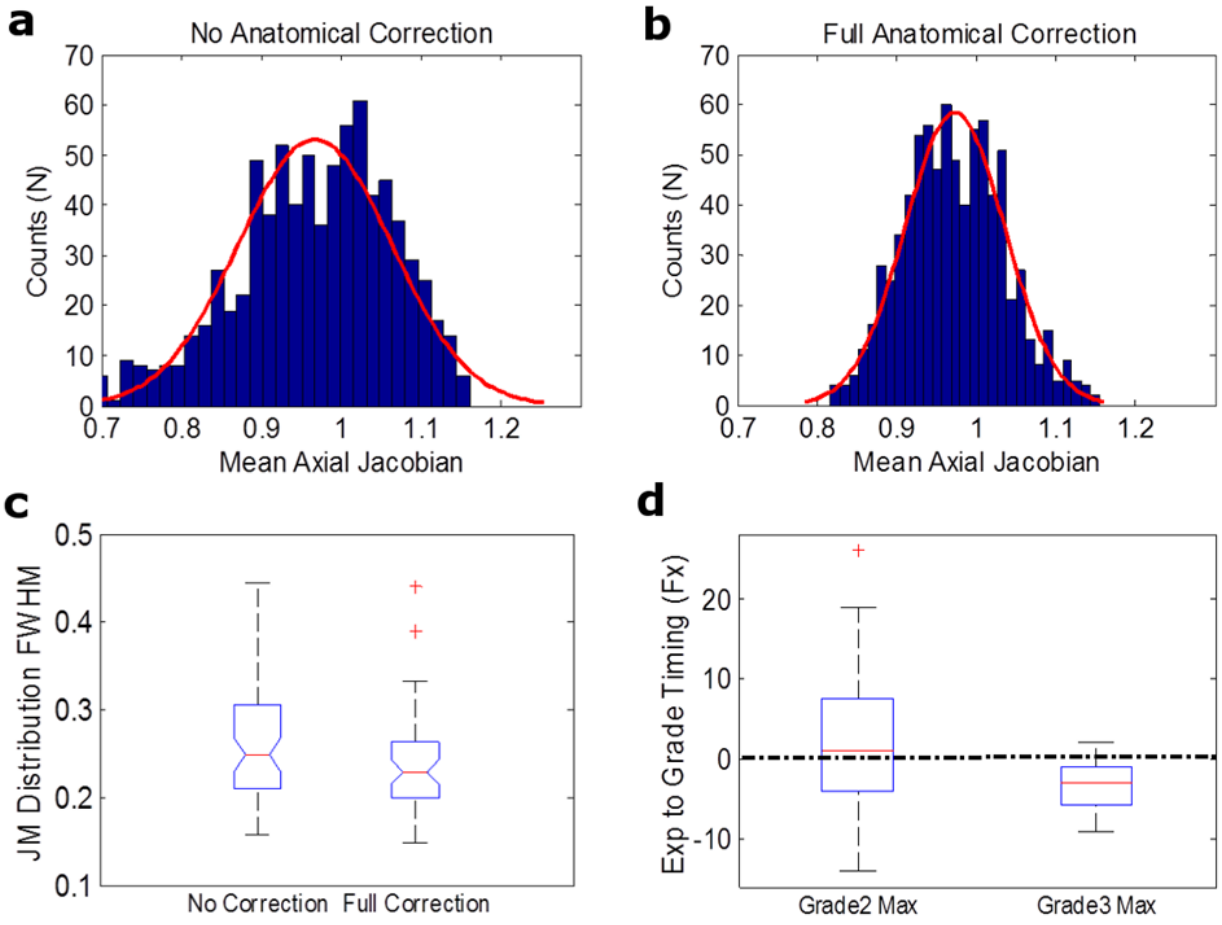

Figure 3.6: *(a) Histogram showing the distribution of planning to week 1 Jacobian values for one patient, without anatomic corrections. Red line represents a normal fit. (b) Histogram showing the distribution of planning to week 1 Jacobian values for one patient, with all anatomic corrections applied. Red line represents a normal fit. (c) Boxplot of the planning to week 1 Jacobian full width half maximum (FWHM) values without and with anatomic corrections. (d) Boxplot of fraction of maximum expansion (MaxExp1) minus the fraction of maximum esophagitis grade, with a dotted line at zero.*

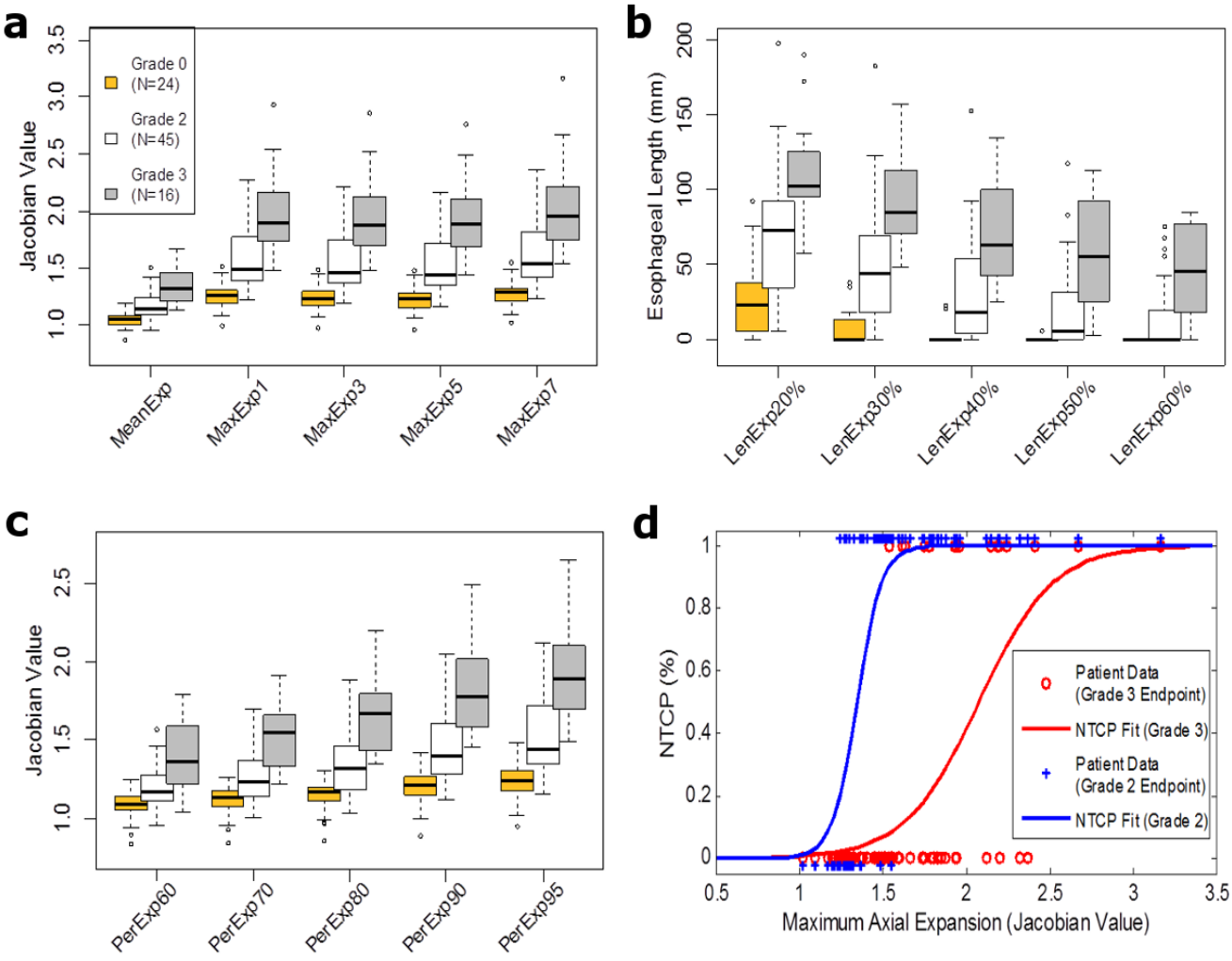

Figure 3.7: *(a-c) Boxplots of the esophageal expansion metrics grouped according to esophagitis grade (yellow is grade 0, white is grade 2, and gray is grade 3). The box edges represent the 75% (top edge) and 25% (bottom edge) quartile values, the middle line represents the median value, the whiskers represent the range of values, and the circles represent outliers. (d) Plot of the NTCP function for grade 2 (blue), and grade 3 (red) complication thresholds with individual patient result above (1.0) and below (0.0) the given threshold (blue +, grade 2; red o, grade 3). Expansion metrics are defined in the Methods and Materials section.*

Table 3.1 summarizes the statistical relationships between expansion metrics and esophagitis grade. All expansion metrics were very highly significantly ($p < 0.001$) associated with esophagitis grade2 and 3 according to logistic regression. Spearman rank analysis showed most metrics to have correlation coefficients in the range of 0.50-0.67. While area under the curve (AUC) values from ROC analysis indicated metric performance varied slightly around the binary cutoff for grade 2 or grade 3, all MaxExp-based, PerExp95, LenExp30%, and LenExp40% metrics performed strong with AUC>0.88 for both esophagitis endpoints, indicating these metrics' ability to classify esophagitis (Table 3.1).

Table 3.1: *Logistic regression analysis of the relationship between expansion metrics and esophagitis grade (n = 85).*

	Grade 2				Grade 3			
Expansion metric*	**p**	**AUC†**	**Spearman rank**	**50% Complication Value**	**p**	**AUC†**	**Spearman rank**	**50% Complication Value**
MeanExp	5.39E-08	0.855	0.553	1.191	3.87E-07	0.880	0.515	1.511
MaxExp1‡	*1.51E-11*	*0.928*	*0.668*	*1.445*	*2.70E-07*	*0.899*	*0.540*	*2.123*
MaxExp3	1.72E-11	0.921	0.657	1.208	2.70E-07	0.893	0.532	1.880
MaxExp5	2.44E-11	0.915	0.648	1.139	2.70E-07	0.899	0.540	2.145
MaxExp7	5.98E-11	0.910	0.639	1.371	2.70E-07	0.900	0.542	2.039
LenExp20%	2.08E-08	0.866	0.572	26.911	4.85E-06	0.860	0.488	127.848
LenExp30%	*5.98E-11*	*0.909*	*0.642*	*12.625*	*3.87E-07*	*0.901*	*0.546*	*98.625*
LenExp40%	4.10E-10	0.900	0.639	2.536	6.41E-07	0.899	0.553	80.554
LenExp50%	1.26E-09	0.851	0.579	0.708	6.41E-07	0.894	0.564	64.979
LenExp60%	1.56E-07	0.762	0.487	0.393	3.87E-07	0.889	0.606	49.769
LenExp70%	8.41E-07	0.738	0.448	0.500	5.21E-06	0.831	0.530	40.571
LenExp80%	1.93E-05	0.689	0.382	0.000	5.38E-05	0.784	0.492	31.333
LenExp90%	1.48E-03	0.588	0.170	0.000	5.41E-05	0.783	0.475	23.571
PerExp60	6.72E-07	0.815	0.491	1.133	1.29E-06	0.834	0.453	1.456
PerExp70	4.66E-08	0.842	0.533	1.112	2.88E-07	0.894	0.534	1.764
PerExp80	6.53E-09	0.864	0.568	1.083	2.70E-07	0.907	0.551	1.774
PerExp90	4.10E-10	0.885	0.601	1.246	2.70E-07	0.902	0.545	1.910
PerExp95	1.32E-10	0.902	0.627	1.252	2.70E-07	0.908	0.552	1.915

**Expansion metrics as defined in section 3.1.4.*

†Area under the curve.

‡Italic text indicates highest performing metrics for both grade 2 and grade 3 esophagitis endpoints.

The timing of maximum expansion and esophagitis grade showed a strong temporal correlation, as both esophagitis endpoints occurred on average around the same treatment fraction as the maximum expansion (Figure 3.7d). In addition, 8 patients had breaks in treatment due to esophagitis symptoms, with an average reduction of 14.3% in maximum expansion.

3.2.3 Expansion & Breath Hold CT

Similar to the 4DCT study cohort the anatomical correction methodology reduced the variance of Jacobian values. The mean FWHM for all 9 patients with weekly BHCT imaging was reduced from 0.27 (±0.38) to 0.23 (±0.55) after anatomical correction, yielding a 17.69% (±4.84) absolute percent difference.

The comparative analysis of expansion calculated from BHCTs and 4DCTs for a single treatment week for 5 patients using the absolute difference between the expansion metric values is shown at the top of Table 3.2. For these five patients, the amount of expansion is consistent between both CT types for the given patient. The metrics values of the treatment week with largest expansion are consistent with the 4DCT cohort of the same esophagitis grade, as shown in the bottom of Table 3.2.

Table 3.2: *Comparison of the absolute difference in expansion metric values calculated from either BHCT or 4DCT for 5 patients in relative units (top), and the expansion metrics values grouped according to maximum treatment esophagitis grade for 9 patients with weekly BHCT. The values in parentheses in the lower portion of the table represent the range of expansion metric values observed.*

	MaxExp1	MaxExp3	MaxExp5	LenExp30%	LenExp40%	PerExp98
Mean Abs Difference	0.098	0.104	0.102	15.5	11.0	0.09
Max Abs Difference	0.160	0.160	0.150	70.0	55.0	0.14
Min Abs Difference	0.061	0.061	0.040	0.0	0.0	0.07

Grade	MaxExp1	MaxExp3	MaxExp5	LenExp30%	LenExp40%	PerExp98
G0 (N=4)	1.17 (1.13-1.27)	1.15 (1.10-1.24)	1.12 (1.05-1.21)	0.00 (0.00-0.00)	0.00 (0.00-0.00)	1.07 (0.94-1.17)
G2 (N=2)	1.55 (1.43-1.67)	1.51 (1.36-1.65)	1.49 (1.35-1.62)	56.25 (27.50-85.00)	28.75 (2.50-55.00)	1.44 (1.35-1.52)
G3 (N=2)	1.98 (1.93-2.03)	1.95 (1.89-2.01)	1.91 (1.83-1.98)	121.25 (85.00-157.50)	98.75 (77.50-125.00)	1.77 (1.71-1.83)

3.3 Chapter Discussion

To the best of our knowledge, this study was the first to propose an alternative measure of radiation esophagitis using objective, imaging biomarkers. In addition, we developed a method to reduce uncertainty in Jacobian calculations caused by esophageal anatomic variability. This study's findings can be summarized as: first, the localized esophageal volume change, from planning to any weekly-treatment time point, can be calculated using the Jacobian; second, this correction methodology improves Jacobian calculation accuracy; third, the expansion metrics examined were significantly correlated to radiation esophagitis, with maximum esophagitis occurring near week of maximum expansion.

Although transient effects can lead to erroneous Jacobian calculations, we found our correction methodology reduced these uncertainties. Using our correction methodology, we were able to reduce uncertainty by 10.0%, with air content producing the most error. Air in the esophagus is common, making censoring patients or sections of the esophagus with air not feasible.[56] By utilizing the air content correction, we can obtain a more accurate Jacobian calculation in air-containing esophageal regions.

We quantified esophageal expansion to measure esophagitis severity in a novel way. Previous studies observed expansion of the esophageal wall in many forms of esophagitis, including radiation esophagitis.[12,51] However, these studies were presented as clinical observations, not thorough radiation-response analyses. In addition, these studies did not assess esophagitis severity. Previously, the relationship between esophagitis grade and the ratio of esophageal cross-sectional areas of weekly to planning CT images during treatment were studied and found that this ratio increased with grade, and these increases occurred in regions receiving the highest radiation doses.[41] On the basis of this work, we improved our analysis by using a 3-dimensional measure of expansion. We developed localized measures of expansion and identified the highest correlated metrics to esophagitis grade. Furthermore, the timing relationship of maximum expansion and esophagitis grade was investigated. In addition, we corrected for anatomic variation to reduce the associated uncertainties.

Of the various esophageal expansion metrics we tested, most performed well and were highly correlated to esophagitis grade, as shown in Table 3.1. The highest performing metrics were maximum axial expansion (MaxExp1) and esophageal length ≥30% axial expansion (LenExp30%). The maximum axial expansion metrics seem intuitive as measures of high-grade

esophagitis, if the functional subunit of the esophagus is considered a cross-sectional segment, and the organ is serial. In addition, MaxExp1 and LenExp30% were combined into a multivariate logistic model for both endpoints and computed AUC. An improvement in classification was observed for the grade 3 esophagitis endpoint, with AUC of 0.93 for grade 2 and AUC of 0.91 for grade 3.

The timing of maximum expansion is correlated to esophagitis grade. On average, patients with maximum grade 2 esophagitis had maximum expansion occur at the same treatment fraction (Figure 3.7d). In addition, 15 of the 16 grade 3 max patients had expansion occur before grade 3 esophagitis. Whether expansion precedes grade 3 symptoms is not currently discernible as expansion and esophagitis scores are quantified weekly. The relationship with grade 2 esophagitis had more variance, but both this could be a product of subjectivity within grade 2 assessment.

The breath hold patient cohort of this study reflected the results of the 4DCT cohort. The anatomical corrections had a similar reduction in uncertainty. The extent of expansion metrics for patient esophagitis grade was consistent with observed values in the 4DCT cohort. The analysis of weekly 4DCT expansion metrics to those quantified on BHCT were consistent for most of the patients when weekly treatment CT time and patient grade were considered. Two of the five patients, however had discrepancies in expansion metric values.

Our study had some limitations. First, a local error in deformable registration may cause miscalculation of the Jacobian. We reduced the potential impact of this miscalculation by implementing our correction methodology. In addition, there is no direct method to validate anatomic uncertainty late in treatment, and we assumed that the plan-to-week 1 variance is

representative of variance later in treatment. Nevertheless, patients with grade 0 esophagitis did not show any appreciable radiation-response, and even in extreme cases these patients exhibited little esophageal expansion. For every patient, expansion is a localized effect, with the expanding region only existing within the irradiated esophagus. Although dose-response may be considered a paradigm of radiation therapy, the goal of the current work was to show that esophageal expansion can quantify esophagitis, and dose was not a focus in our study. How dose induces esophageal expansion will be presented soon in a future study. We also did not conduct pre-treatment esophageal contrast studies, which allows for identification of pre-existing thickening of the esophageal wall. In addition, chemotherapy does increase occurrence of high-grade esophagitis. How chemotherapy contributes to expansion is not thoroughly investigated. However, every patient in this study had the same course on concurrent chemotherapy, and no appreciable expansion was observed outside the irradiated esophagus. Due to small sample size, the results from the analysis of expansion calculated from BHCT only shows feasibility and needs to be verified with a larger data set.

Quantifying esophagitis with expansion is an attractive method of quantifying esophagitis severity. The continuous nature of the expansion metrics may allow esophagitis severity to be described in mathematical rather than qualitative terms. The spatial localization of expansion allows geometric dose-response information, allowing for a deeper understanding of radiation injury in the esophagus, which was previously unavailable. Because esophagitis is an endpoint in most thoracic radiation therapy trials, expansion may potentially provide an objective measure for comparison of treatment modalities, as well as *in-vivo* measures of the effectiveness of radioprotectors.

This provides new options for toxicity prediction modeling. The binary endpoint of logistic regression can be chosen with flexibility. As shown in Figure 3.7, most expansion metrics had a gap between the minimal metric value for grade 3 esophagitis and the maximal metric value for grade 0. This gap as well as the expansion values of 50% probability of complication (Table 3.1) represents candidates for dichotomy. NTCP modeling is common practice to predict radiation esophagitis at treatment planning.[2,3,5,57,58] In previous studies, variation in outcome reporting and differing grading scales presented challenges for obtaining effective NTCP models.[2,21,24] Review studies by Werner-Wasik et al and Rose et al showed that although many NTCP-based studies have been performed with collectively thousands of patients, no common model can predict esophagitis with high accuracy in external data sets.[2,3] The continuous nature of expansion metrics may improve prediction model performance In addition, modeling techniques other than logistic regression may be studied.

In conclusion, esophageal expansion is an imaging biomarker of radiation-response that is a suitable surrogate for toxicity. The highest performing expansion metrics were maximum axial esophageal expansion and axial length with at least 30% expansion. Expansion metrics may be useful to quantify response associated with new treatment techniques and clinical trials. The uncertainty in esophageal Jacobian calculations can be reduced by using anatomic correction methods. Use of breath-hold CT to calculate esophageal expansion is feasible. Breath hold CT would require multiple serial acquisitions to derive the anatomical variability correction factors, but these may be less impactful on the Jacobian calculation. However, more studies would be required using breath hold CT.

Chapter 4

Prediction Modelling of Toxicity Using Esophageal Expansion

In the previous chapter, the viability of using the esophageal expansion as a radiation-response biomarker to quantify toxicity was shown. A natural application of a toxicity surrogate metric is its performance in toxicity prediction modelling. The work in this chapter investigates the use of esophageal expansion as an endpoint in the toxicity prediction modelling process with the goal of increasing model predictive performance. To accomplish this, toxicity prediction models were constructed with esophageal expansion endpoints and were then compared to toxicity prediction models constructed with traditional grade-based endpoints for severe radiation esophagitis. Multiple methods of toxicity prediction model construction were implemented within a cross-validation procedure.

4.1 Methods and Materials

4.1.1 Patient Population

The patients utilized for analysis in this chapter comprise of the same 85 patients from the IMRT arm of the prospective clinical trial at University of Texas-MD Anderson Cancer Center that had weekly 4DCT imaging, which were described in chapter 3. Patients had weekly 4DCT imaging which allows for the computation of esophageal expansion at each week during treatment. Esophagitis was graded according to Common Terminology Criteria for Adverse Events version (CTCAE) 3.0.[19] The distribution of maximum esophagitis grades during treatment was: 24 were grade 0, 45 were grade 2, and 16 were grade 3. Out of a pool of 88 patients with weekly 4DCT and IMRT treatment from this clinical trial, the same 3 patients excluded in chapter 3's analysis were excluded in this work due to poor image quality. The work in this chapter was approved by the University of Texas-MD Anderson Cancer Center Institutional Review Board and was HIPAA compliant.

As previously described in chapter 3, 4DCT scans were acquired on CT scanners operated at 120 kV. Voxel dimensions were 0.98x0.98x2.50 mm in the right-left direction, anterior-posterior, and superior-inferior direction, respectively, with a 512×512 pixel area. Patient treatment planning and segmentation was conducted using the Pinnacle treatment planning system (Phillips Healthcare), with esophageal contours segmented from the cricoid cartilage to the gastroesophageal junction, in the axial plane, with Pinnacle version 9.8.

4.1.2 Logistic Regression

Since NTCP models use dichotomization for outcome quantification, with 1 representing complication occurrence, and 0 representing no complication occurrence, logistic regression with is a natural choice for statistical analysis. We define the logistic model as:

$$NTCP = \frac{1}{1+e^{-(\alpha_0+\alpha_1 * x_1+\ldots+\alpha_n * x_n)}} \quad (4.1)$$

for a model with n predictor variables, denoted x_i, where α_i are the regression coefficients.

4.1.3 Forward, Stepwise Model Selection

While there are many ways to select model predictor variables, one of the most traditional is the forward, stepwise selection method.[6,33,34] Starting with a null model, each predictor variable is added and the ratio likelihood test is used to identify which individual predictor performs best for the given model order. In addition, Bayesian information criterion (BIC) is calculated for each model order; when BIC is minimized, this is selected as the optimal model order to prevent overfitting.

4.1.4 Least Absolute Shrinkage and Selection Operator (LASSO)

While stepwise selection is common for NTCP model creation, the LASSO has been shown to have improved prediction ability.[64-66] The predictor variable selection methodology in LASSO identifies the subset of the regression coefficients, α_i, equal to zero, using the penalty term, λ, by minimizing the following equation:

$$\sum_{i=1}^{n}(Y_i - \frac{1}{1+e^{-(\alpha_0+\alpha_1 * x_1+\ldots+\alpha_n * x_n)}})^2 + \lambda\sum_{i=1}^{n}|\alpha_i| \quad (4.2)$$

for n number of predictors, where Y_i is the complication outcome. For a larger λ, less variables will be utilized in the model.[62]

4.1.5 Random Forest Classification

Random Forests classification is a wholly different modelling approach than stepwise or LASSO logistic regression, since there is no longer an attempt to fit a logit analytical function.[62] Random Forests is an ensemble decision tree approach with an implementation of bootstrap aggregation, or bagging, and a random draw of a set-sized subset of model predictors at each tree node.[69] The result of this procedure is an ensemble of de-correlated trees that should increase generalizability of the model. In our implementation of Random Forests, the number of randomly drawn predictors from the training data was 7 for each node (the square root of the number of predictors). The number of trees in the ensemble was 300, to allow for sufficient stabilization, and the minimum leaf observation size was 1 for all trees.

4.1.6 Objective Esophagitis Endpoints

As shown in chapter 3, esophageal expansion calculated from the radiation therapy treatment planning CT to weekly treatment 4DCTs is a surrogate quantification of radiation toxicity in the esophagus. In addition, the analysis in chapter 3 showed that the maximum axial expansion of one slice (MaxExp1), and the axial esophageal length with at least 30% expansion (LenExp30%), were the highest performing endpoints for quantifying grade 3 esophagitis. In this chapter's analysis, these two expansion biomarkers are examined separately as endpoint in the toxicity prediction modelling process. The scientific question of this analysis is: can these surrogate endpoints produce better performing prediction models, in terms of prediction performance, than using an esophagitis grade input.

While the objective expansion metrics are continuous, to get a direct comparison of NTCP performance with traditional grade-based endpoint models, the patients had their MaxExp1 and LenExp30% biomarker values dichotomized according to a threshold value for the given metric. This threshold value was chosen to be the intersection value of the grade 0 and grade 3 distributions for the particular metric: 50% for MaxExp1, and 45mm for LenExp30%. An example for the determination of this intersection metric value is given by the dotted red line for MaxExp1 in Figure 4.1.

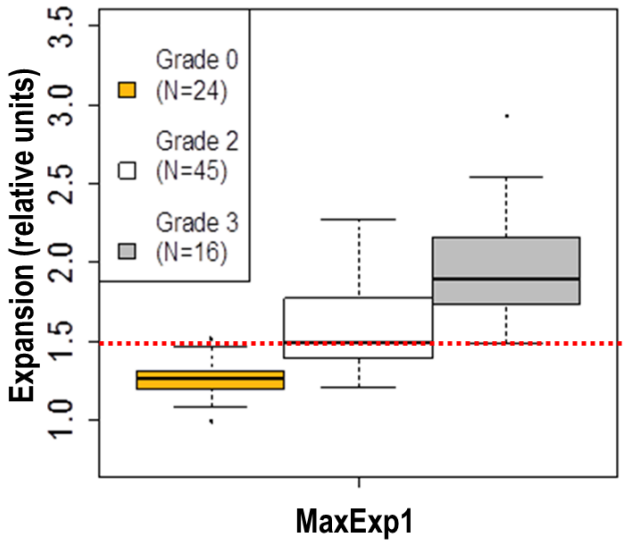

Figure 4.1: *Determination of the threshold value for endpoint dichotomization of patients as either having the complication or not having the complication for the MaxExp1 expansion biomarker. The dotted red line represents the intersection of grade 0 and grade 3 MaxExp1 metric values and is the threshold metric value (MaxExp1 of 1.5 in relative units and 50% in units of percent expansion).*

4.1.7 NTCP Model Assessment: Repeated Holdout Cross-Validation

For all model construction methods, repeated holdout cross-validation, also known as Monte Carlo cross-validation was utilized to randomly partition patient data in the model construction process.[62,70] The patient data are comprised of the clinical factors and DVH dose metrics that are used as model predictors, as well as the individual patients' endpoints, specifically esophagitis grade, MaxExp1, and LenExp30% objective values. The toxicity predictor variables were clinical factors consisting of: age, gender, tumor prescription dose, tumor location, tumor stage, tumor histology, gross tumor volume, nodal involvement, if the patient had induction chemotherapy, and smoking status (former smoker, current smoker, never smoked). Esophagus DVH dose metrics included: maximum (Dmax) and mean esophagus dose (MED), absolute volume of dose from 10Gy to 70Gy in 5Gy increments, esophageal length

receiving ≥ 10Gy to 50Gy in 10Gy increments for both 25% and 100% circumference coverages of the esophagus.

Approximately 75% of the data (64 patients) are partitioned into the training data set for model construction and fit assessment. The remaining 25% of the original patient data set is held out for assessing model prediction performance. This process starts by randomly splitting the dataset into the aforementioned training and test data sets. A toxicity prediction model is constructed on the training data set for all three model types simultaneously (stepwise, LASSO, and Random Forests). There is a unique step in the LASSO model methodology, whereby the optimal penalization parameter, λ, is calculated by 5-fold cross-validation with deviance as the optimization measure on the training data set.

The fit performance of the model is assessed by predicting the known outcomes of patients in the training set by using each patient's respective predictor variables and corresponding outcome, in the derived model. The model order of the fitted models, as well as the predictors present in the models are all cataloged. Model predictive performance is then assessed by applying the developed model to each patient in the test data set. In both the fit and predictive model assessments, area under the curve (AUC) from receiver operating characteristics analysis is used to quantify performance. This process is repeated for 100 iterations to account the randomness of partitioning the full data into training and test sets. This model construction methodology is performed for all 3 endpoint types (Grade, MaxExp1, and LenExp30%), separately. The distribution of $AUC_{Prediction}$ values are tested for significant differences between all model methods and endpoint types with a paired T-test ($p<0.05$ for significance).

Figure 4.2 illustrates the iterative process of the model construction methodology. The model order and predictor variables present in the model are recorded. Since Random Forests is an ensemble approach, recording model order and predictor variable occurrence in the model is not informative. Therefore, predictor importance is quantified with the out-of-bag permutation error, which is the average increase in prediction error is as variables are permuted on observations that are out-of-bag for the entire ensemble.

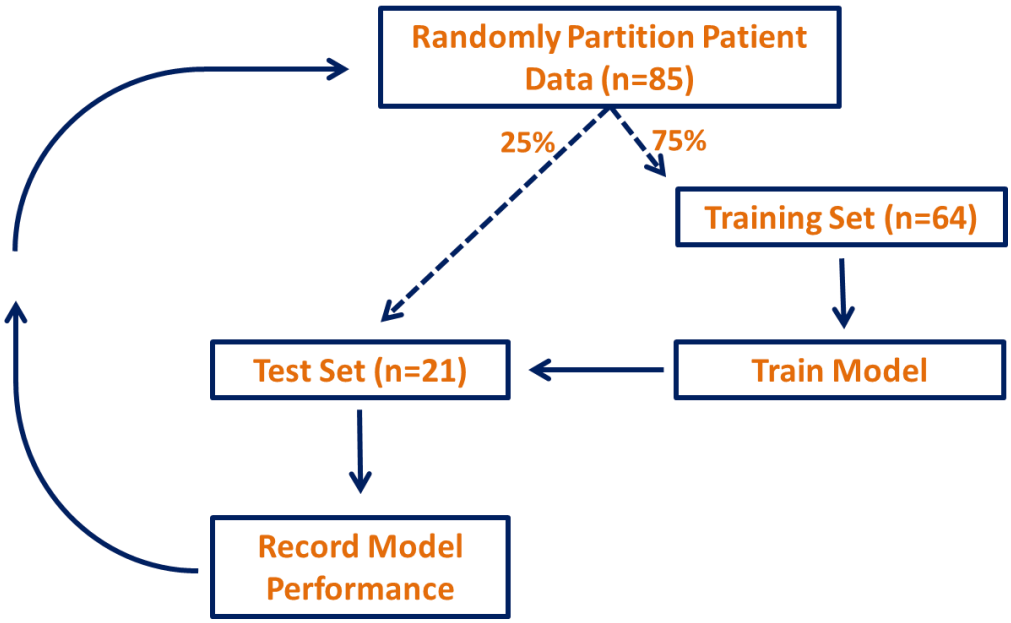

Figure 4.2: *Schematic of the model construction and assessment process. This represents a single iteration of 1000 total iterations of the repeated cross-validation process. For each iteration, a model is trained on the training set from 75% of the full patient data, based on a random partition. The trained model is assessed on the test set consisting of the remaining 25% of full patient data, to simulate external validation and prediction performance. This process is executed 1000 times with a different random partition of test and training data.*

4.1.8 Computational Implementation

All predictor variables were standardized by subtracting the mean variable value of all patients from each individual patient value, and then dividing the result by the standard deviation.[70] All computations were conducted in MATLAB (Mathworks, Natick, MA) version 8.2. Forward stepwise logistic regression and Random Forests classification models were constructed with MATLAB's machine learning and statistics toolbox. LASSO models were constructed using the open source glmnet package implemented in MATLAB.[72] It should be noted that for each model construction method using different endpoints, the same random number seed is initialized before any computations.

4.2 Results

4.2.1 Patient Population

Clinical factors, Radiotherapy Prescription dose, and maximal esophagitis grade during treatment for the 85 study patients are summarized Table 4.1. The median week of maximal esophageal expansion was week 6 (range, 4-8). A summary of DVH dose metrics for this population is shown in Table 4.2.

Table 4.1: *Demographic characteristics of the 85 study patients.*

Characteristic	*Datum*
Median age (range)	
All	65 years (43-85 years)
Male	65 years (47-80 years)
Female	66 years (43-85 years)
Sex	
No. of Males	45
No. of Females	40
Histologic findings	
Squamous cell carcinoma	29
Adenocarcinoma	50
Large cell carcinoma	3
Other	3
Smoking history	
Current smoker	18
Former smoker	58
Never smoked	9
Stage	
IIa	3
IIb	3
IIIa	39
IIIb	36
IV	4
Treatment dose, Gy	
74	53
66	28
60	4
Max Esophagitis Grade	
Grade 0	24
Grade 2	45
Grade 3	16

Table 4.2: *Summary statistics for dose-volume histogram metrics for the 85 study patients, including mean, standard deviation, maximum, and minimum metric values. The units are Gy for MED and Dmax, cm^3 for V10-V70, and esophgeal length in cm for LE metrics.*

DVH Metrics	*Mean*	*S.D.*	*Max*	*Min*
MED	71.53	11.11	83.06	20.64
Dmax	30.02	10.50	53.17	3.99
V10	20.11	8.37	51.72	3.24
V15	18.66	8.43	49.81	0.25
V20	17.31	8.39	47.86	0.01
V25	16.04	8.23	44.57	0.00
V30	14.97	8.00	43.71	0.00
V35	13.92	7.67	42.67	0.00
V40	12.85	7.41	41.19	0.00
V45	11.73	7.10	38.57	0.00
V50	10.47	6.74	35.06	0.00
V55	9.09	6.14	30.24	0.00
V60	7.10	5.02	21.21	0.00
V65	4.61	3.98	18.86	0.00
V70	2.46	3.39	16.09	0.00
$LE10_{25\%}$	13.39	3.33	20.00	4.75
$LE20_{25\%}$	12.05	3.36	19.00	2.75
$LE30_{25\%}$	10.52	3.68	18.50	0.00
$LE40_{25\%}$	9.23	3.85	18.00	0.00
$LE50_{25\%}$	8.10	3.72	17.50	0.00
$LE60_{25\%}$	6.23	3.13	14.75	0.00
$LE10_{100\%}$	12.06	3.73	19.00	0.50
$LE20_{100\%}$	9.48	3.86	18.00	0.50
$LE30_{100\%}$	7.63	3.68	17.50	0.25
$LE40_{100\%}$	6.18	3.15	15.00	0.25
$LE50_{100\%}$	4.93	2.68	12.75	0.25
$LE60_{100\%}$	3.43	1.84	7.00	0.25

Abbreviations: MED = mean esophagus dose; Dmax = maximum esophagus dose; V10 = volume of esophagus receiving at least 10 Gy; $LE10_{25\%}$ = esophageal length with at least 10 Gy to at least 25% of the cross-sectional area to axial slice of the esophagus; $LE10_{100\%}$ = esophageal length with at least 10 Gy to at least 100% of the cross-sectional area to axial slice of the esophagus.

4.2.2 Model Construction

A total of 49 predictor variables were inputs to the toxicity prediction model construction process, as summarized in Table 4.3. For model construction with the grade 3 endpoint, the most common model order was 7 terms for the forward stepwise, and 9 for the LASSO implementation. Models using the objective MaxExp1 ≥ 50% endpoint had most common model orders of 6 for stepwise, and 11 for LASSO. LenExp30% ≥ 45mm endpoint-based models had an average model order of 6 for stepwise and 7 for LASSO.

The most common predictor variables in all models are recorded in Table 4.3. For grade-based endpoint models, clinical factors were the highest recurring predictor type with $LE60_{100\%}$ being the sole DVH metric in the top-five recurring model predictors. This was a similar result in the grade-based endpoint LASSO models, where $LE60_{100\%}$ and V70 were the second and fifth most recurring predictors in the fitted models, respectively. This was not the case for the Random Forests grade endpoint models, where no clinical factors were in the top-five of most recurring model predictors. For both LenExp30% and MaxExp1 objective endpoint-based models, dosimetric predictors were the most recurring, specifically $LE40_{100\%}$ which was the first or second highest recurring predictor in all model types. Mean esophageal dose was also present in all model types.

The fit performance of each model type for each endpoint is also displayed in Table 4.3. All scenarios except one had average AUC_{Fit} >0.90, showing strong calibration of each model. Stepwise model type showed the highest average AUC_{Fit}.

Table 4.3: *Results of toxicity prediction model construction for 85 patients using stepwise logistic regression, LASSO logistic regression, and Random Forest classification. Each model type was constructed with complication endpoints of: esophagitis ≥ grade 3, maximum axial expansion of one slice (MaxExp1) ≥ 50%, and esophageal length with expansion of at least 30% (LenExp30%) ≥ 45mm, separately. The area under the curve from model fitting on training data and model prediction on test data is denoted AUC_{Fit} and $AUC_{Prediction}$, respectively. Highest recurring predictors are listed in order of most to least recurring, for all model predictors occurring in at least 500 out of 1000 iterations of the cross-validation procedure.*

Model Type	Endpoint	AUC_{Fit} (S.D.)	$AUC_{Prediction}$ (S.D.)	Highest Recurring Predictors
Stepwise	**Esophagitis≥ Grade 3**	0.99 (±0.01)	0.58 (±0.15)	*Age, Tumor Location, Induction Chemo, $LE60_{100\%}$, Smoking Status*
LASSO		0.91 (±0.10)	0.64 (±0.12)	*Tumor Location, $LE60_{100\%}$, Age, Smoking Status, V70*
Random Forest		0.97 (±0.05)	0.55 (±0.07)	*V55, $LE50_{100\%}$, $LE40_{100\%}$, MED, V40*
Stepwise	**MaxExp1 ≥ 50%**	0.96 (±0.02)	0.66 (±0.18)	*$LE40_{100\%}$, V50, Stage, MED, Dmax*
LASSO		0.90 (±0.06)	0.75 (±0.10)	*$LE40_{100\%}$, V70, Stage, MED, Dmax*
Random Forest		0.99 (±0.02)	0.76 (±0.10)	*$LE40_{100\%}$, $LE50_{100\%}$, $LE30_{100\%}$, V70, MED*
Stepwise	**LenExp30% ≥ 45mm**	0.96 (±0.02)	0.69 (±0.13)	*Stage, $LE40_{100\%}$, MED, Tumor Location, Dmax*
LASSO		0.84 (±0.08)	0.73 (±0.10)	*$LE40_{100\%}$, MED, Stage, $LE30_{25\%}$, $LE50_{100\%}$*
Random Forest		0.99 (±0.03)	0.73 (±0.10)	*$LE50_{100\%}$, $LE40_{100\%}$, V55, MED, $LE30_{100\%}$*

Abbreviations: MED = mean esophagus dose; Dmax = maximum esophagus dose; V10 = volume of esophagus receiving at least 10 Gy; $LE10_{25\%}$ = esophageal length with at least 10 Gy to at least 25% of the cross-sectional area to axial slice of the esophagus; $LE10_{100\%}$ = esophageal length with at least 10 Gy to at least 100% of the cross-sectional area to axial slice of the esophagus.

4.2.3 Model Prediction

The prediction performance of the different model types with the three endpoints is quantified with $AUC_{Prediction}$ and is displayed in Table 4.3. The mean $AUC_{Prediction}$ value for models using the grade 3 endpoint is 0.58, 0.64, and 0.55 for the stepwise, LASSO, and Random Forests, respectively. The mean $AUC_{Prediction}$ value for models using the objective MaxExp1≥50% endpoint is: 0.66, 0.75 and 0.76 for the stepwise regression, LASSO, and Random Forests, respectively. The mean $AUC_{Prediction}$ value for models using the objective LenExp30%≥45mm endpoint is: 0.69, 0.73 and 0.73 for the stepwise regression, LASSO, and Random Forests, respectively. Standard deviation of $AUC_{Prediction}$ values are generally lower for LASSO and Random Forests models using either objective endpoint than either the grade-based endpoint models, or the stepwise model type. A boxplot of $AUC_{Prediction}$ values from all iterations of the model construction and assessment methodology is shown in Figure 4.3. The $AUC_{Prediction}$ values for the Random Forests and LASSO methods using either objective endpoint were significantly higher than $AUC_{Prediction}$ values from all methods using the grade endpoint. Additionally, $AUC_{Prediction}$ values from LASSO and Random Forests methods were significantly higher than the stepwise method when using either objective endpoint.

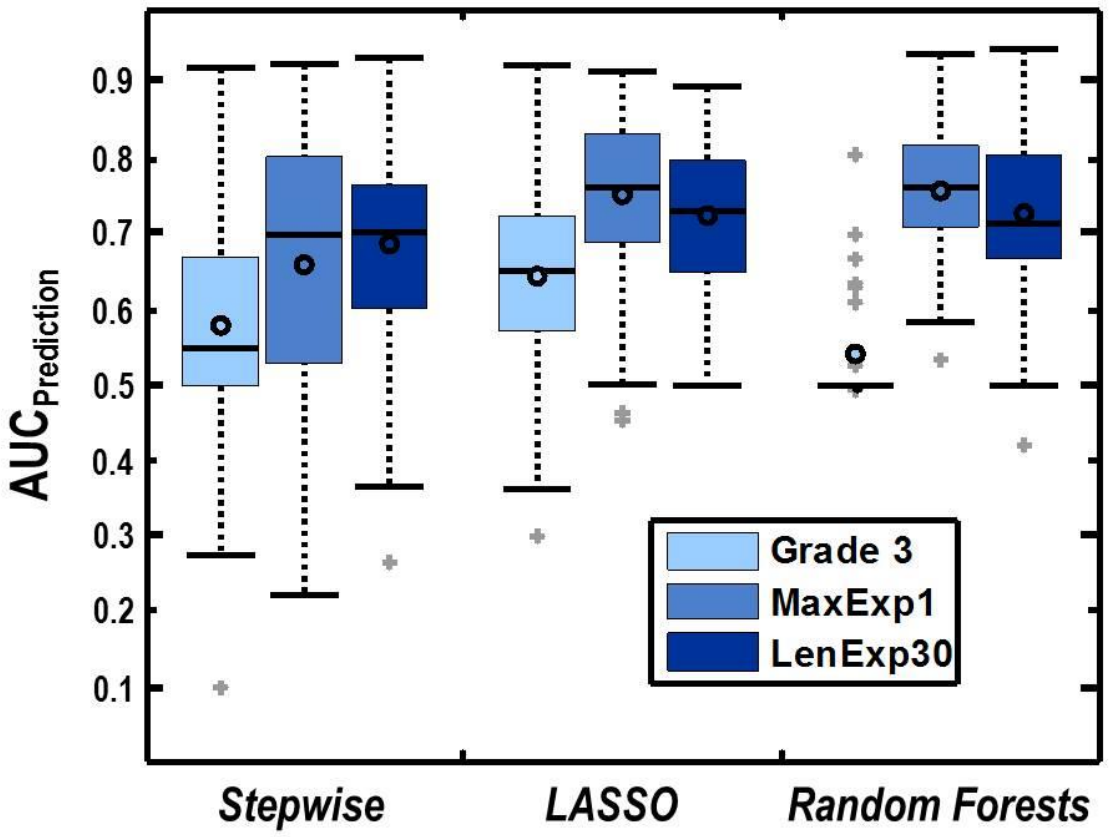

Figure 4.3: *Boxplot of AUC values of model prediction using holdout test data for all 100 iterations of the repeated cross-validation procedure. Endpoints were grade 3 esophagitis (light blue), and the objective endpoint MaxExp1≥50% (medium blue), and LenExp30%≥45mm (dark blue). The first group of three individual boxplots are models constructed with stepwise logistic regression, the second grouping is LASSO, and the final group is Random Forests. The individual boxes represent the respective quartiles, with the solid black line in the box representing the median AUC value, as well as the solid black circle representing the mean AUC value. Outliers are denoted with gray '+' symbols.*

4.3 Chapter Discussion

The work in this chapter represents a novel investigation of the use of objective measures of esophageal toxicity in the prediction modelling process. Several comprehensive review publications in the literature identified the need for objective imaging biomarkers with application in toxicity prediction modelling.[67,68] The work in this chapter specifically addresses this issue. Prediction models were constructed with two different logistic regression methods and a Random Forests classifier, all using the traditional grade 3 endpoint, and directly comparing the modelling process with objective esophagitis measures as endpoints in the form of the MaxExp1 ≥ 50% and LenExp30% ≥ 45mm.

There was consistency in predictor variables chosen for model selection, regardless of model construction method or model endpoint. Mean esophageal dose (MED) and length of esophagus receiving at least 40Gy to 100% of the esophageal circumference ($LE40_{100\%}$), were consistently selected as variables for the prediction models. This is consistent with previous esophagitis toxicity studies.[2,3] Mean esophageal dose and length of esophagus receiving at least 40Gy to the entire esophageal circumference were commonly selected in all objective endpoint-based scenarios, which showed higher prediction performance then the grade-based endpoint models. Interestingly, irradiated length of the esophagus was shown to be correlated to toxicity and may infer an ability to partially spare the esophagus. A strong fit was obtained on the training data for all model types and endpoints, as shown in Table 4.3.

It is important to note that true prediction ability must be tested on data not used in the NTCP model process. As an example, the stepwise method with the grade endpoint has a mean AUC_{Fit} of 0.99. Without cross-validation one may conclude this is a robust prediction modelling

method. However, the mean $AUC_{Prediction}$ value is only 0.58 for this model, indicating severe overfitting.

The prediction ability, as quantified by AUC on the test sets of data, is illustrated in figure 2 for all 9 modelling scenarios (3 model construction methods, 3 endpoints). From this figure, a general trend of increase $AUC_{Prediction}$ with a smaller variance can be seen with Random Forests and LASSO models using objective endpoints compared to the grade endpoint. In addition, the prediction performance is increased by using LASSO or Random Forests compared to stepwise logistic regression, with the exception of the Random Forests method using the grade 3 endpoint.

Comparing the AUC_{Fit} to $AUC_{Prediction}$ for respective model method and endpoint type shows a smaller difference between model training and testing performance for the LASSO and Random Forests methods using the objective endpoints, when compared to the stepwise method and grade endpoints. This indicates the LASSO and Random Forests models, as well as the objective biomarker endpoints are more generalizable model construction methods than stepwise regression or the grade 3 endpoint. The observed superiority of LASSO to stepwise regression for NTCP modelling is consistent with previous studies.[64-66]

There were several limitations with this analysis. Generally, a larger sample size is desired in prediction modelling, and the 85 patient sample in this study is on the lower end of common experience.[2,3] However, the need for weekly 4DCT acquisition during treatment to calculate esophageal expansion limits the amount of patients currently available for analysis. Another limitation was the inability of our study to report a single 'superior' model. Typically, NTCP modelling studies report a full logistic regression equation with predictor variables and

coefficients for comparison with existing and future models. However, our goal was not to find a single 'superior' model, but to analyze the performance of objective esophagitis endpoints in the NTCP modelling process, with comparison to traditional grade based endpoints. In addition, we wanted to compare the performance of a common predictor selection method (stepwise regression) with two promising techniques, LASSO and Random Forests.

The results from this chapter show the utility of esophageal expansion as objective biomarker endpoints for toxicity prediction modelling. In addition, we have shown the superior predictive performance of LASSO and Random Forests compared to stepwise models in the toxicity prediction modelling process for esophagitis. Further investigations into using objective esophagitis endpoints for more robust prediction models may now be explored.

In conclusion, objective, localized measures of esophageal toxicity in the form the expansion biomarkers maximum axial expansion of one slice of the esophagus (MaxExp1) and esophageal length with at least 30% expansion (LenExp30%), with respective complication thresholds of 50% and 45mm, have higher predictive performance than CTCAE grade for prediction models of severe esophagitis. All model types showed irradiated length of the esophagus as a common predictor of toxicity. LASSO and Random Forests model types outperform forward, stepwise model selection for predicting severe esophagitis.

Chapter 5

Normalized Uptake from ^{18}Fluorodeoxyglucose Positron Emission Tomography as a Measure of Esophageal Radiation Injury

---

This chapter investigates normalized FDG uptake, as quantified from mid-treatment FDG-PET/CT imaging, as a radiation-response and toxicity measure for esophageal radiation injury. Robust metrics of normalized FDG uptake are derived for classification of toxicity. A methodology to normalize FDG uptake as a radiation-response metrics is tested. These normalized FDG uptake metrics are then shown to be robust biomarkers of esophageal radiation-response and esophageal toxicity, by statistical analysis with physician scored radiation esophagitis grade. The highest performing normalized uptake metrics are then analyzed as predictors of esophagitis symptom development from patients who are asymptomatic at the time of the FDG-PET study. In addition, the relationship between FDG uptake and esophageal expansion was examined. This chapter comprises specific aim 2.

A substantial portion of this chapter is based on the following publication:

Niedzielski JS, Yang J, Liao Z, Gomez DR, Stingo F, Mohan R, Martel MK, Briere TM, Court LE. ^{18}F-fluorodeoxyglucose-positron emission tomography can quantify and predict esophageal injury during radiation therapy. *Int J Radiat Oncol Biol and Phys* **2016** 96(3):670-678.

doi: 10.1016/j.ijrobp.2016.07.012

5.1 Methods and Materials

5.1.1 Patient Population

The patient population studied in this chapter was from the same pool of patients from a prospective clinical trial for treatment of NSCLC. We identified 79 patients for this retrospective analysis that had a single FDG-PET/CT scan acquired during treatment. As mentioned in previous chapters, all patients in this clinical trial were treated with concurrent chemotherapy (paclitaxel and carboplatin) and either intensity-modulated photon radiation therapy or passive-scatter proton radiation therapy, with a tumor prescription dose of 60-74 Gy in 2-Gy fractions over 6-8 weeks. Of these 79 patients for this FDG-based study, 67 patients were included that also had expansion calculated during treatment. Of these 67 patients, 36 were included in the analyses from chapters 3 and 4.

Esophagus contours were segmented in the axial plane from the cricoid cartilage to the gastroesophageal junction. Esophagitis was prospectively scored weekly throughout treatment using the Common Terminology Criteria for Adverse Events 4.0. The distribution of esophagitis grades was as follows: 43 grade 0, 30 grade 2, and 6 grade 3 at the time of the PET studies, and progressed to maximum esophagitis of 17 grade 0, 40 grade 2, and 22 grade 3, by treatment completion. For the 43 grade 0 patients at time of the PET study, 26 would progress to have esophagitis (23 grade 2, 3 grade 3).

All PET/CT scans were acquired with a General Electric Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI) with $3.9\times3.9\times3.3mm^3$ voxels. PET/CT studies were acquired in the treatment position with standard patient immobilization used for radiation therapy planning CT scans. All PET studies were attenuation-corrected with an accompanying CT scan.

PET scans were conducted at a median of 100.1 minutes (range, 58.1-159.2 minutes) after injection of a median of 377.0 MBq (range, 241.4-507.2 MBq) of FDG. The PET/CT scans were deformed using a B-splines image registration algorithm in the planning CT frame of reference and resampled into $0.98\times0.98\times2.50mm^3$ voxels using Velocity AI 3.0.1 (Velocity Medical Solutions, Atlanta, GA). The timing of the PET studies during treatment was not uniform. On average, the PET scan was acquired at fraction 23 (±2.4 standard deviation, 18-28 range).

5.1.3 Normalized FDG Radiation-Response

We developed in-house code for data extraction, uptake calculation, and analysis using MATLAB (Mathworks, Natick, MA). This software converted the planning CT esophagus segmentations into binary masks for identification of esophageal voxels in image space. Using spatial reference points from the corresponding digital imaging and communications in medicine (DICOM) header file, spatial positioning in Cartesian coordinate space of each voxel can be identified for both the FDG-PET and radiation dose arrays. From this information, the FDG-uptake and radiation dose for each voxel in the esophagus can be identified. FDG uptake was quantified as the standard uptake value (SUV) according to the bodyweight calculated SUV equation, with voxel SUV_{Mean} and SUV_{Max} calculated.[73,74]

To control intra-patient FDG uptake variability, uptake was normalized as a patient-specific radiation-response quantification. For each patient, we calculated a normalization factor of the mean SUV value for esophageal voxels irradiated up to a X Gy low-dose threshold, for delivered dose at the time of the PET study, and then divided the remaining esophageal PET voxels above the low-dose threshold by this normalization factor.

The normalized SUV equation is:

$$nSUV(x,y,z) = \left(\frac{SUV(x,y,z)_{>X\ Gy}}{SUV_{Mean<X\ Gy}} - 1\right) * 100 \quad (1)$$

Where SUV_{Mean} < X Gy is the average FDG uptake in the low-dose region (less than X Gy) of the esophagus and uptake value. Normalization low-dose cutoff value was analyzed for threshold values of X = 1, 2, 5, 8, and 10-Gy. The multiple choices of the low-dose threshold were examined to test the sensitivity of the choice of the threshold value. The normalized SUV (nSUV) represents the percent uptake increase from the low-dose esophageal region due to the corresponding radiation dose. SUV(x,y,z)>X Gy is the given voxel FDG uptake.

Localized uptake metrics were derived by averaging nSUV at each axial segment of the esophagus in two ways: axial-averaged maximum nSUV for 1, 3, 5, and 7-slices (e.g., $nSUV_{AxMax7}$ for the 7-slice average) and axial length of esophagus with at least a given percentage of axial-averaged nSUV response (e.g., $nSUV_{Len40}$ for esophageal length with nSUV ≥ 40% axial-averaged normalized response), with an nSUV increase ranging from 20% to 60% in 10% increments. Voxel nSUV mean ($nSUV_{Mean}$), maximum ($nSUV_{Max}$), and percentile nSUV values from sixty-fifth to ninety-fifth percentile (e.g., $nSUV_{Perc65}$ for the sixty-fifth percentile) were also calculated. Standard SUV values of SUV_{Max} and SUV_{Mean} were also reported.

5.1.4 Esophagitis Timing and Progression

The relationship between esophagitis grade (both maximum grade and grade at the time of the PET study) and normalized uptake was examined. The timing of esophagitis severity and progression to maximum grade was analyzed in terms of treatment fractions between the PET scan and the escalation of grade, and then compared with nSUV.

The ability of mid-treatment nSUV to predict maximum esophagitis grade was also investigated for patients who were asymptomatic during the PET study. These patients were grouped as: grade 0 at the time of the PET study that remained grade 0 throughout treatment (G0-0), grade 0 that became grade 2 or 3 (G0-2/3), grade 2 that stayed grade 2 (G2-2), grade 2 that became grade 3 (G2-3), and grade 3 that stayed grade 3 (G3-3). Progression for G2-3 patients was not analyzed due to low sample size. No patients had esophagitis grade escalate after completion of radiation therapy. Prediction of symptom progression from asymptomatic patients during the PET study (n = 43) was created using LASSO logistic regression and nSUV metric values. Model construction is described in section 5.1.6.

5.1.5 Normalized FDG Uptake & Esophageal Expansion

A natural question arises of how FDG uptake and esophageal expansion relate for a given patient. A total of 67 out of the 79 patients in this analysis had esophageal expansion quantified calculated from 4DCTs for the same week of FDG-PET acquisitions. This allowed the FDG uptake and expansion to be compared at the same time point during treatment for a given patient.

5.1.6 Statistical Considerations

FDG uptake and dose metrics were compared with esophagitis grade, both at the time of the PET study and maximum treatment grade, for ≥ grade 2 esophagitis complication using univariate logistic regression and Spearman rank analysis. Model fit was assessed using area under the curve (AUC) from receiver operating characteristics (ROC) analysis. Reported Logistic regression p-values were calculated using the likelihood ratio chi-square test. Prediction of maximum esophagitis grade from mid-treatment nSUV was tested using the Mann-Whitney U test. All p-values were reported after applying the Benjamini-Hochberg false discovery procedure. A p-value of ≤0.05 was considered statistically significant.

To test the relationship of nSUV and esophagitis grade in a multivariate analysis, least absolute shrinkage selection operator (LASSO) penalized logistic regression was utilized. LASSO has the ability to ignore redundant features and improve predictive ability when compared to stepwise logistic regression.[62,64,65,66] Model features in the form of nSUV were used to predict ≥ grade 2 esophagitis at time of PET study, and ≥ grade 2 maximum treatment grade. A 3-fold cross-validation procedure (folds of 1 training, 1 validation, and 1 test) was repeated for 100 iterations, with AUC calculated from NTCP predictions (probability of ≥grade 2 esophagitis incidence) on the test fold for every iteration. The AUC_{Mean} and standard deviation of AUC values quantifies the robustness of the trained nSUV model to classify esophagitis on the test set. LASSO model parameters were derived from the validation set and the tested model was constructed from the training set.

To investigate the added value of nSUV in classifying esophagitis, dosimetric features were examined as separate model features and then repeated with nSUV features to predict

≥grade 2 esophagitis in the model construction method previously described. Dosimetric features were quantified using dose-volume histogram (DVH) metrics in the form of esophagus volume receiving at least X dose (VX), length of esophagus irradiated to at least X dose with complete esophageal axial coverage (LEX), and mean (MED) and max (Dmax) esophageal dose. Dose metrics were also quantified and separately tested in model construction in the form of fractional DVH metrics, for each patient's fraction at time of the PET study.

The previously described model construction method was implemented with asymptomatic patients at time of the PET study (n=43), to predict symptom progression by completion of treatment (≥ grade 2). Model construction of nSUV-only, dose-only, and nSUV and dose features were conducted.

5.2 Results

5.2.1 Patient Characteristics

Clinical factors of the 79 patients studied are summarized in Table 5.1. No clinical factors were associated with esophagitis grade, either maximum grade or grade during the PET study. Neither injected FDG dose nor PET scan times were correlated to nSUV metrics. In addition, the timing of the PET study was not associated with esophagitis or nSUV metrics' value.

5.2.2 Normalization of FDG Uptake

The normalization factor was calculated for threshold values of 1, 2, 5, 8, and 10 Gy for 50 of the 79 patients studied. The 5Gy threshold was selected. The average percent difference between using 5Gy normalization factor versus other examined threshold values was 2.36% (±2.00% standard deviation), with four patients more than 6.00% (10.48% maximum). In addition, nSUV metrics were calculated for these 50 patients and the effect of different threshold values was analyzed. None of the metrics had an average difference of more than 4.8% for an individual patient. Since effect of threshold choice was minimal, two patients had nSUV metrics calculated with different normalization factor thresholds (1Gy, 8Gy) to acquire an adequate amount of normalization voxels for analysis. The necessity of normalizing the SUV value can be illustrated by comparing Figure 5.2 (no normalization) to Figure 5.3a,b (with normalization). The stratification of esophagitis grade by SUV magnitude is stronger with normalization (Figure 5.3) than without (Figure 5.2). In addition, the meaning of the SUV metric is stronger and more appropriate for analyzing dose-response, as the metric has been normalized according to a patient's radiation response characteristics.

5.2.3 Normalized FDG Uptake and Esophagitis Severity

Normalized uptake generally increased along with esophagitis severity. This was true of both esophagitis grade measured during the PET study, and treatment maximum. For each patient, increased normalized uptake was confined to the region of highest radiation dose along the length of the esophagus (Figure 5.1).

Table 5.1: *Demographic characteristics of the 79 study patients.*

Characteristic	Datum
Median age (range)	
All	66 years (38-80 years)
Male	65 years (51-79 years)
Female	66 years (38-80 years)
Sex	
No. of Males	46
No. of Females	33
Histologic findings	
Squamous cell carcinoma	31
Adenocarcinoma	41
Large cell carcinoma	3
Other	4
Smoking history	
Current smoker	27
Former smoker	46
Never smoked	6
Stage	
IIa	3
IIb	5
IIIa	29
IIIb	38
IV	4
Treatment dose, Gy	
74	46
66	28
60	5

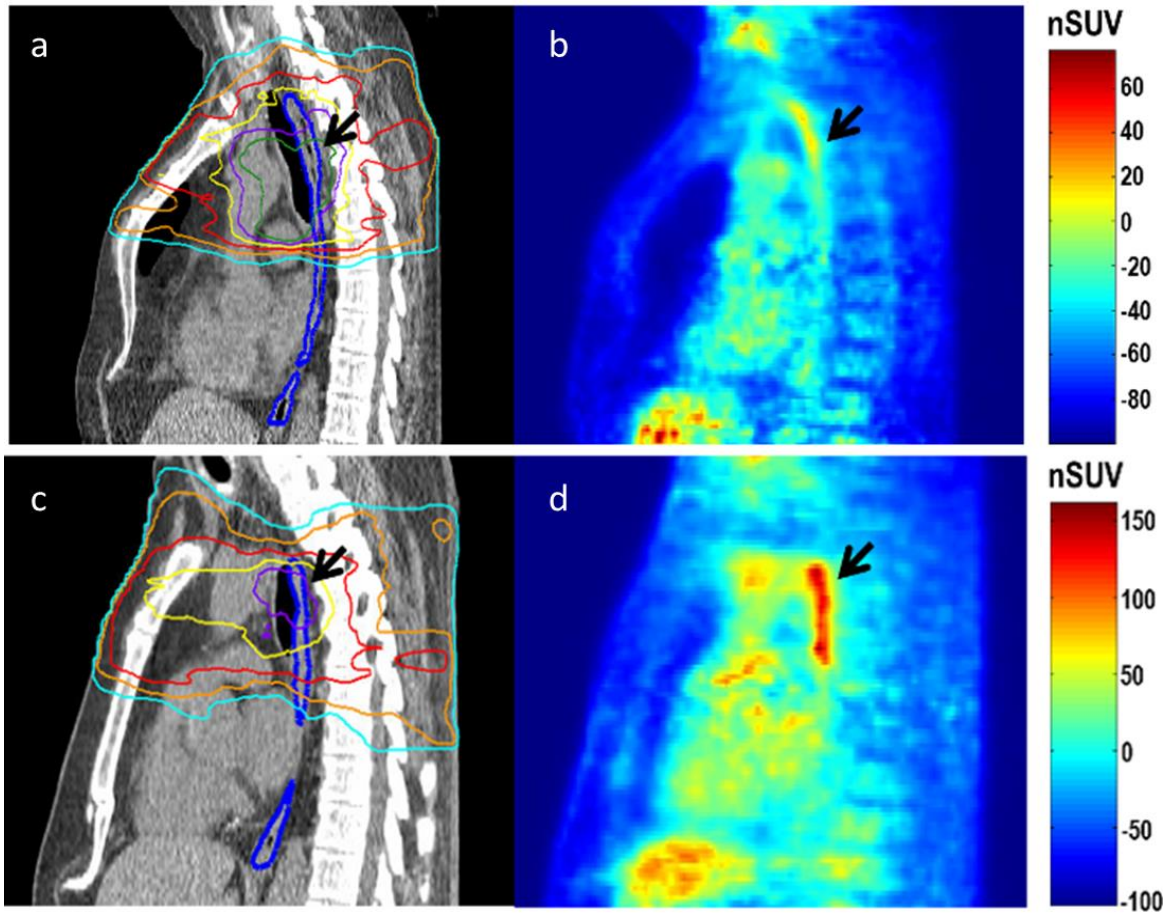

Figure 5.1: *Sagittal view of esophageal anatomy as measured by computed tomography (CT; left panels) with radiation dose and normalized uptake in the sagittal plane (right panels) for an asymptomatic patient (68 year old male), a,b, and a patient (73 year old male) with grade 3 esophagitis at the time of the positron emission tomography (PET) study, c,d. The esophagus is outlined in blue on the CT scans, with radiation planning isodose lines of 20 Gy (light blue), 30 Gy (orange), 40 Gy (red), 50 Gy (yellow), 60 Gy (purple), and 70 Gy (dark green) . The region of high esophageal dose and corresponding normalized standard uptake value (nSUV) is highlighted with the black arrows.*

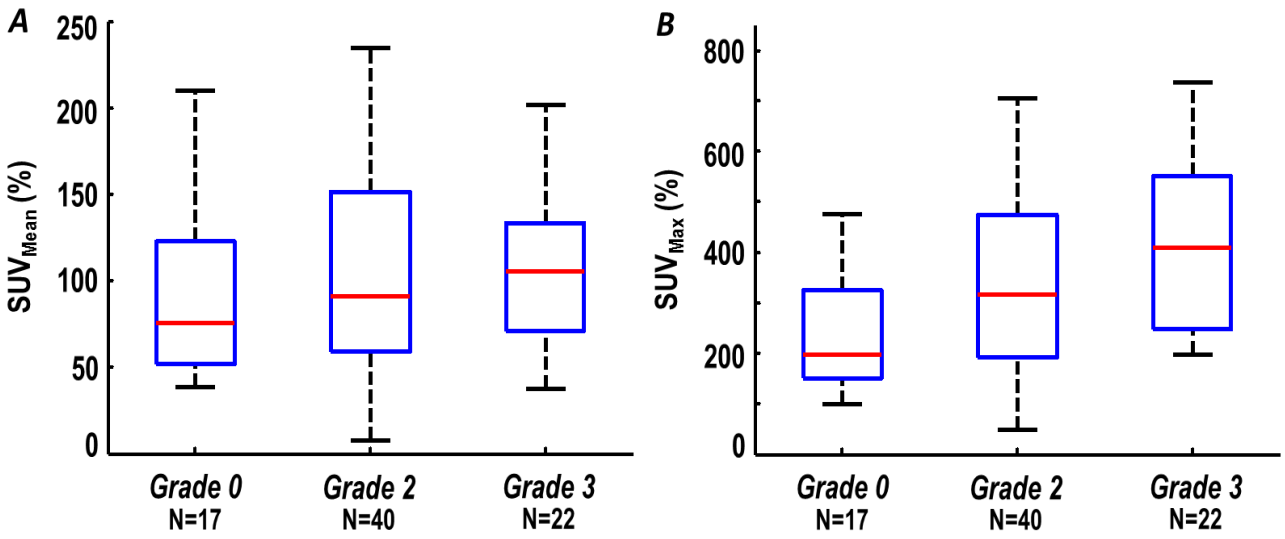

Figure 5.2: *Esophagus FDG standard uptake values without normalization for voxel mean uptake (SUV_{Mean}) in (A) and voxel max uptake (SUV_{Max}) in (B) for patients with grade 0, 2, or 3 maximum esophagitis grades,. Boxplot corners represent 25th and 75th percentiles and middle line the median value. Outliers are denoted as '+'.*

The ability of nSUV to stratify patients on the basis of esophagitis grade at the time of the PET study, as well as the maximum esophagitis grade, is shown for $nSUV_{AxMax1}$ in Figure 5.3a,b. A strong trend of increasing nSUV with increasing esophagitis severity is observed. Grouping patients' esophagitis grade during the PET study and the maximum grade showed a trend of increasing normalized uptake with increasing esophagitis severity (Figure 5.3c). Figure 5.3 shows the ability of nSUV to stratify esophagitis severity, as well as esophagitis progression, from the time of the FDG-PET study to treatment completion; for comparison, boxplots of the mean and maximum esophageal dose grouped according to esophagitis grade are presented in the supplemental materials.

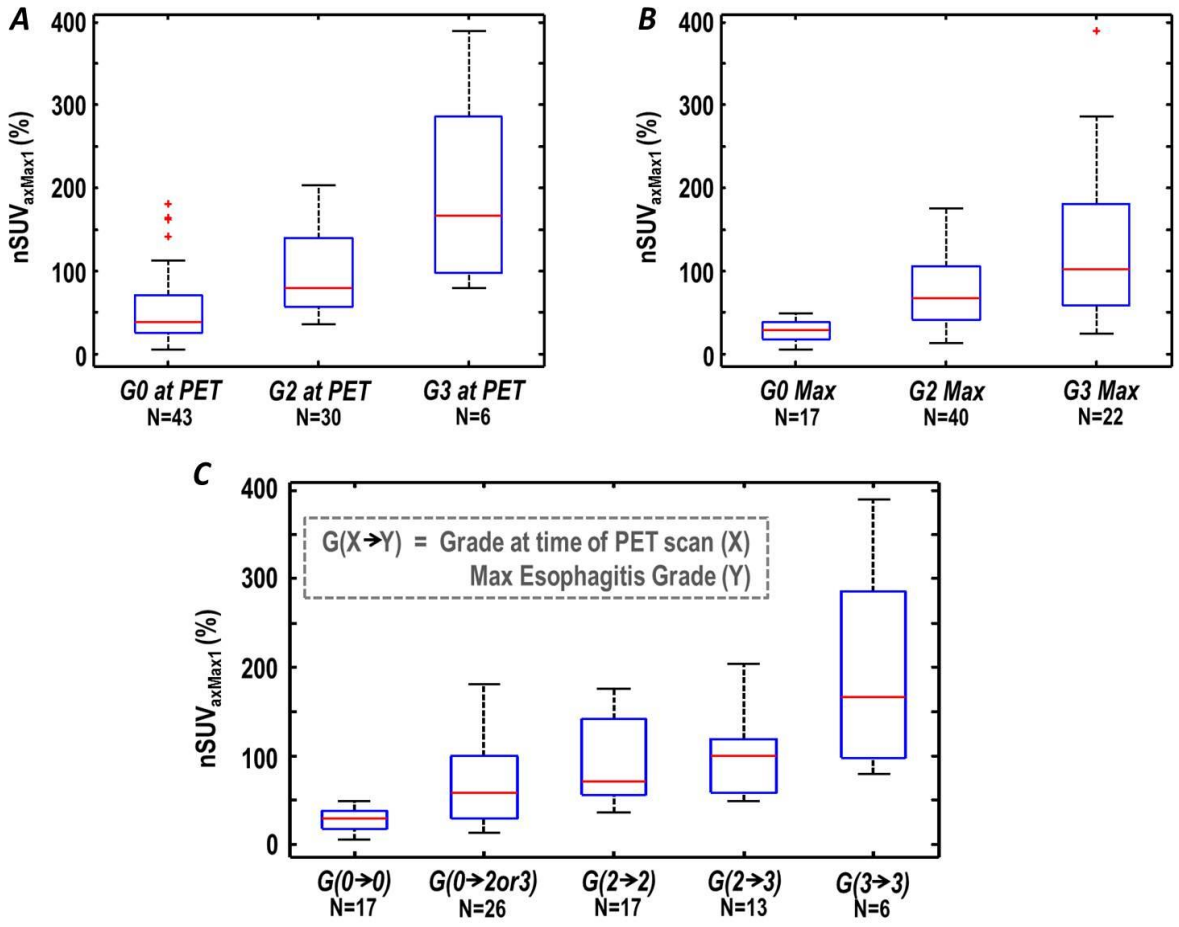

Figure 5.3: *Esophagus FDG axial-averaged maximum normalized standard uptake values for 1 slice ($nSUV_{AxMax1}$) for patients with grade 0, 2, or 3 esophagitis at the time of the PET study, A, grade 0, 2, or 3 maximum esophagitis grades, B, and $nSUV_{AxMax1}$ grouped by patient progression of esophagitis grade at the time of the PET study to maximum grade, C. Boxplot corners represent 25th and 75th percentiles and middle line the median value. Outliers are denoted as '+'.*

In the univariate analysis, most nSUV metrics were significantly correlated with esophagitis grade ($p < 0.05$) for the endpoints investigated (Table 5.2). The highest performing metrics for both endpoints were $nSUV_{AxMax1}$ and $nSUV_{Len40}$, with AUC values ≥ 0.85 for ≥ grade 2 esophagitis at the time of the PET study and AUC values ≥ 0.91 for maximum esophagitis ≥grade 2. SUV_{Mean} was associated only with ≥grade 2 esophagitis at the time of the PET study. SUV_{Max} was significantly correlated ($p < 0.05$) with both endpoints, but performance was much lower than any significant normalized uptake metric (Table 5.2).

The results of the multivariate analysis are listed in Table 5.3. Inclusion of only nSUV metrics for model features outperformed dose-only models for both ≥ grade 2 esophagitis endpoint scenarios (AUC 0.83 and 0.88 vs. 0.52 and 0.76, for grade at time of PET scan and treatment completion, respectively). The difference in AUC between model types was significant for both grade at time of PET scan ($p<0.05$) and treatment completion ($p<0.001$), respectively. Models that combined nSUV and dose metrics showed slight improvement in AUC, albeit with higher model complexity. In addition, nSUV metrics had higher model occurrence and lower penalization on average than dose metrics.

Table 5.2: *Statistical analysis of the relationship between 18F-fluorodeoxyglucose (FDG) uptake metrics and esophagitis grade at the time of the FDG-positron emission tomography (PET) scan, and maximum treatment esophagitis grade (n = 79). Univariate logistic regression p-values reported have been corrected with the Benjamini-Hochberg false discovery rate procedure.*

	≥Grade 2 during PET scan		≥Grade 2 Treatment Maximum	
FDG uptake metric	**P Value**	**AUC***	**P Value**	**AUC***
SUV_{mean}	2.04E-02	0.67	2.35E-01	0.61
SUV_{max}	5.23E-03	0.73	1.88E-03	0.75
$nSUV_{mean}$†	8.16E-05	0.84	4.09E-05	0.87
$nSUV_{max}$	1.55E-04	0.82	1.26E-06	0.91
$nSUV_{axMax1}$‡	8.16E-05	0.85	1.17E-06	0.91
$nSUV_{axMax3}$	5.32E-05	0.85	9.53E-06	0.87
$nSUV_{axMax5}$	5.75E-05	0.84	1.52E-05	0.86
$nSUV_{axMax7}$	8.16E-05	0.81	3.65E-05	0.84
$nSUV_{len20}$**	5.32E-05	0.80	7.37E-06	0.89
$nSUV_{len30}$	5.32E-05	0.84	1.17E-06	0.91
$nSUV_{len40}$	8.16E-05	0.85	6.52E-08	0.92
$nSUV_{len50}$	7.15E-04	0.81	1.26E-06	0.87
$nSUV_{len60}$	1.32E-02	0.74	4.23E-05	0.80
$nSUV_{perc65}$††	8.31E-05	0.83	1.12E-05	0.89
$nSUV_{perc75}$	8.31E-05	0.83	2.48E-06	0.90
$nSUV_{perc85}$	8.16E-05	0.84	2.48E-06	0.90
$nSUV_{perc95}$	4.69E-04	0.83	1.26E-06	0.91

**Area under the curve.*

†Voxel mean normalized standard uptake value.

‡Axial-averaged maximum nSUV for 1, 3, 5, and 7 slices (e.g., nSUVaxMax1 for the 1-slice average).

***Axial length of esophageal tissue with at least a given percentage of axial-averaged nSUV response (e.g., nSUVlen20 for a length of esophagus with nSUV ≥ 20% axial-averaged response over the baseline SUV value).*

††Voxel percentile nSUV values from the sixty-fifth to the ninety-fifth percentiles (e.g., nSUVperc65 for the sixty-fifth percentile).

Table 5.3: *Statistical analysis of the relationship between ^{18}F-fluorodeoxyglucose (FDG) uptake metrics and maximum esophagitis grade (n = 79). P values reported have been corrected with the Benjamini-Hochberg false discovery rate procedure.*

	Grade 2			Grade 3		
FDG uptake metric	**P Value**	**Spearman coefficient**	**AUC***	**P Value**	**Spearman coefficient**	**AUC***
SUV_{mean}	2.35E-01	0.17	0.62	3.13E-01	0.15	0.60
SUV_{max}	1.88E-03	0.38	0.78	3.02E-02	0.33	0.71
$nSUV_{mean}$†	4.09E-05	0.50	0.86	1.44E-02	0.39	0.75
$nSUV_{max}$	1.26E-06	0.56	0.91	1.41E-02	0.40	0.75
$nSUV_{AxMax1}$‡	1.26E-06	0.56	0.90	4.43E-03	0.40	0.75
$nSUV_{AxMax3}$	9.53E-06	0.50	0.86	4.43E-03	0.45	0.78
$nSUV_{AxMax5}$	1.52E-05	0.50	0.86	4.43E-03	0.43	0.77
$nSUV_{AxMax7}$	3.65E-05	0.48	0.85	4.43E-03	0.41	0.76
$nSUV_{Len20}$**	7.37E-06	0.53	0.89	7.15E-03	0.37	0.73
$nSUV_{Len30}$	1.17E-06	0.56	0.90	4.43E-03	0.41	0.76
$nSUV_{Len40}$	6.52E-08	0.67	0.91	4.43E-03	0.43	0.77
$nSUV_{Len50}$	1.26E-06	0.57	0.86	7.15E-03	0.42	0.76
$nSUV_{Len60}$	4.23E-05	0.46	0.79	1.98E-02	0.38	0.72
$nSUV_{perc65}$††	1.12E-05	0.53	0.89	3.57E-02	0.35	0.72
$nSUV_{perc75}$	2.48E-06	0.55	0.90	1.47E-02	0.37	0.74
$nSUV_{perc85}$	1.17E-06	0.56	0.91	4.43E-03	0.41	0.76
$nSUV_{perc95}$	1.26E-06	0.56	0.91	4.43E-03	0.45	0.79

**Area under the curve.*

†Voxel mean normalized standard uptake value.

‡Axial-averaged maximum nSUV for 1, 3, 5, and 7 slices (e.g., nSUVaxMax1 for the 1-slice average).

***Axial length of esophageal tissue with at least a given percentage of axial-averaged nSUV response (e.g., nSUVlen20 for a length of esophagus with nSUV ≥ 20% axial-averaged response over the baseline SUV value).*

††Voxel percentile nSUV values from the sixty-fifth to the ninety-fifth percentiles (e.g., nSUVperc65 for the sixty-fifth percentile).

5.2.4 Normalized Uptake and Esophagitis Timing

On average, symptoms were clinically reported after nSUV quantification. For all 26 asymptomatic patients during the PET study that would develop ≥grade 2 esophagitis by treatment end (G0-2/3), esophagitis occurred a median of 6 fractions after the PET scan. For patients who had grade 2 esophagitis during the PET scan that became grade 3 esophagitis by the end of the treatment, the median time to onset was 7 fractions.

5.2.5 Normalized Uptake to Predict Esophagitis Progression

Figure 5.1C shows the difference in nSUV for the 2 asymptomatic patient groups during the PET study. Patients who eventually developed grade ≥2 esophagitis had markedly higher nSUV values during the PET study than those who did not develop esophagitis. The Mann-Whitney U test showed differences in nSUV metric distributions between these 2 groups were significantly different ($p < 0.05$) for all nSUV metrics examined; differences in SUV_{Max} or SUV_{Mean} were not significant. The performance of symptom progression models was AUC_{Mean} = 0.67 and 0.75 for the dosimetric and nSUV-based models, respectively (Table 5.3). The combined features of dose and nSUV did not improve model AUC_{Mean} over nSUV-only models (0.72). The top recurring features were: $nSUV_{Len30}$, $nSUV_{Len40}$, and $nSUV_{AxMax1}$ in both model types.

Table 5.3: *LASSO regression multivariate analysis of nSUV metrics and ≥esophagitis grade 2 (N=79), and the symptom progression prediction model construction from grade 0 esophagitis patients at time of the PET study (N=43) Using 3-fold cross-validation repeated for 100 iterations. The median model order, mean AUC value with standard deviation, and the most prevalent recurring model features in the cross-validation process are listed.*

Feature Class	Endpoint	Model Order	AUC_{mean}* (S.D.)	Top Recurring Predictors†
nSUV**	Mid-treatment ≥Grade 2	5	0.83 (±0.07)	$nSUV_{Mean}$,$nSUV_{Len30}$, $nSUV_{AxMax3}$, $nSUV_{Len20}$, $nSUV_{AxMax1}$
	Treatment Max ≥Grade 2	4	0.88 (±0.05)	$nSUV_{Len40}$, $nSUV_{Mean}$, $nSUV_{Len30}$, $nSUV_{AxMax1}$
DVH††	Mid-treatment ≥Grade 2	3	0.52 (±0.07)	$LE10_{100\%}$, $LE50_{100\%}$, MED
	Treatment Max ≥Grade 2	3	0.76 (±0.12)	D_{max}, $LE40_{100\%}$, V50
nSUV** & DVH††	Mid-treatment ≥Grade 2	6	0.81 (±0.07)	$nSUV_{Mean}$, $nSUV_{AxMax5}$, $nSUV_{Len40}$, $LE50_{100\%}$, $nSUV_{AxMax1}$
	Treatment Max ≥Grade 2	5	0.91 (±0.06)	$nSUV_{Mean}$, $LE50_{100\%}$, $nSUV_{Len40}$, $LE30_{100\%}$, $nSUV_{Len30}$
DVH††	Symptom Progression ≥Grade 2	3	0.67 (±0.13)	D_{max}, V30, $LE30_{100\%}$
nSUV**	Symptom Progression ≥Grade 2	3	0.75 (±0.10)	$nSUV_{Len30}$, $nSUV_{Len40}$, $nSUV_{AxMax1}$
nSUV** & DVH††	Symptom Progression ≥Grade 2	4	0.72 (±0.12)	$nSUV_{Len30}$, $nSUV_{Len40}$, $nSUV_{AxMax1}$, $LE40_{100\%}$

**Mean area under the curve from repeated cross-validation.*

†Dose and normalized uptake abbreviations defined in the methods section.

***Normalized FDG uptake*

††Dose-volume histogram

5.2.6 Normalized Uptake & Esophageal Expansion

The relationship between normalized FDG uptake and esophageal expansion both quantified at the same time during radiation therapy is shown for 67 patients in Figure 5.4. Patients have are identified by esophagitis grade for both esophagitis at time of quantification (Figure 5.4a), and esophagitis maximum treatment grade (Figure 5.4b). MaxExp1 and $nSUV_{AxMax1}$ expansion and FDG metrics were statistically correlated with a Pearson coefficient of 0.66. An approximately linear correlation between MaxExp1 and $nSUV_{AxMax1}$ is observed. Additionally, for both timings of esophagitis grade a strong stratification of toxicity and both metrics values is observed.

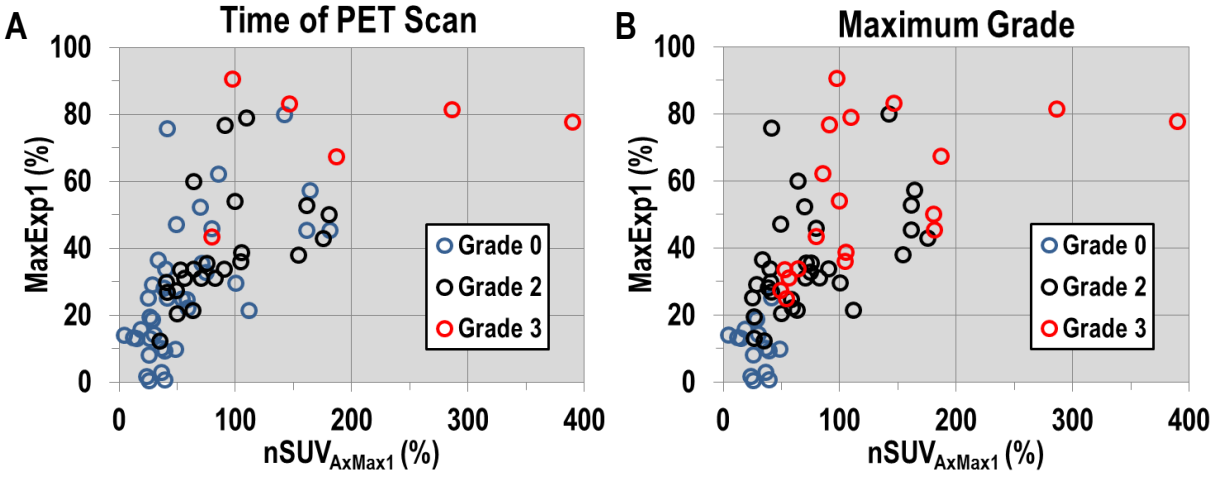

Figure 5.4: *Esophgeal expansion (MaxExp1) and FDG uptake ($nSUV_{AxMax1}$) values for 67 patients with grade 0, 2, or 3 esophagitis at the time of the PET study (A), and maximum esophagitis grades (B).*

5.3 Chapter Discussion

In this chapter, we demonstrate that FDG uptake is a radiation-response imaging biomarker that can be used as an objective quantification of esophagitis during radiation therapy for NSCLC. In addition, it is shown that FDG uptake can predict symptom progression for asymptomatic patients at the time of the PET study. The goal of this chapter was to develop an in-vivo method of quantifying esophagitis that provides spatial information about the specific location and extent of injury. We accomplished this by deriving localized metrics of normalized uptake from PET studies performed during radiation treatment and establishing the relationship between these metrics and esophagitis grade.

Previous studies have shown that FDG uptake can quantify normal tissue response in lung for patients with NSCLC and esophageal cancer.[44-49] Nijkamp et al also examined the use of post-treatment FDG-PET studies to quantify esophageal injury from radiation therapy.[50] Lyman-Kutcher-Burman models were created with and without FDG uptake to predict esophageal injury, and these models showed that prediction was improved by adding PET information to the dose-response model.[58] Another study showed an increase in FDG uptake at 6 discrete points along the length of the esophagus, with uptake normalized to the aortic arch.[75] In another study, it was shown that the change in the 95th percentile of SUV of the esophagus from pre-treatment to weeks 2,4, and 7 of chemoradiotherapy is correlated to esophagitis using rigid PET image alignment, for 27 patients.[76] However, these previous studies did not describe uptake in terms of esophagus geometry or conduct prediction modeling for symptom progression. In addition, various metrics with localization of response were not derived and tested in a multivariate analysis.

FDG uptake was normalized to the < 5Gy region of irradiated esophagus to reduce inter-patient variability of SUV. Standard SUV_{Mean} and SUV_{Max} were not able to stratify or predict the progression of esophagitis. The choice of normalization has been utilized in studies of radiation-induced lung toxicity using FDG-PET.[44-49] Another study used change in FDG-PET SUV during radiation therapy relative to pre-treatment status.[76] Since pre-treatment FDG-PET studies acquired in the treatment position were not available for our patient cohort, this normalization method was not feasible.

Normalized uptake metrics performed well in classifying not only esophagitis grade at the time of the PET scan, but also maximum esophagitis grade. Combining multiple nSUV metrics in logistic regression models to classify esophagitis did not substantially improve AUC value over using a single nSUV metric. DVH metrics alone performed poorly in classifying esophagitis grade. When DVH metrics were combined with nSUV metrics, AUC improved slightly for only the ≥grade 2 treatment maximum endpoint. In addition, nSUV metrics were chosen with higher frequency than DVH metrics in the nSUV/Dose-combined model construction. This shows the robustness of a single nSUV metric to quantify esophagitis.

In addition to classifying esophagitis, the normalized uptake has the ability to predict symptom progression from patients that have grade 0 esophagitis at the time of the PET study. The ability of the LASSO models to predict if a patient will become ≥ grade 2 esophagitis by treatment completion was greater for nSUV-only models than DVH-only models, or the combined nSUV and DVH metric models, as quantified by the cross-validated AUC_{Mean}. The combined nSUV and DVH metric models outperformed the DVH-only models, with nSUV metrics chosen with higher frequency. The cross-validated AUC_{Mean} of 0.75 could potentially be

improved with a larger sample size and increase standardization of CT/PET scan acquisition and processing. In the realm of radiation oncology, preemptively identifying patients that will become symptomatic may allow for preemptive interventions (e.g. change in diet, anti-inflammatory medicine, radioprotectors), that may help reduce adverse effects from therapy, thereby improving patient quality of life and preventing costly interventions (e.g., feeding tube, hospitalization). In addition, the comparison of nSUV and DVH models shows normalized uptake is providing unique response information we cannot obtain from the DVH.

The timing of esophagitis symptoms and normalized uptake magnitude showed that increased uptake occurs before presentation of clinical symptoms. This supports the use of FDG uptake as a preemptive diagnostic tool for esophagitis development. Although the current study was conducted in patients with radiation-induced esophagitis, it may be possible to extend this methodology to other types of esophagitis.

In chapter 3, the esophageal swelling on CT scans were shown to be a robust, objective measure of radiation esophagitis. Interestingly, the highest performing CT-based expansion biomarkers were based on axial maximum expansion (MaxExp1) and esophageal length with expansion greater than 30% (LenExp30%), which is consistent with the results of the nSUV metrics. The performance of normalized uptake and esophageal expansion metrics most correlated to the ≥grade 2 treatment maximum esophagitis endpoint were similar with AUC values above 0.90. Furthermore, this chapter showed there is a strong correlation between expansion and FDG biomarkers. One difference in the analyses of chapter 3 and the work in this chapter was the metric sampling frequency; weekly for expansion and once during treatment for normalized uptake. Normalized uptake's classification performance could potentially be

increased by sampling multiple times during treatment. Because swelling is an inflammatory response, it is feasible that this method could be applied to other forms of esophagitis for detection and quantification of injury.

Our study had limitations. Because only 1 PET scan was acquired during radiation treatment, the change of normalized uptake throughout treatment, including normalization value, could not be quantified for individual patients. Therefore, it is uncertain how early in radiation treatment quantification of esophagitis may be achievable using FDG uptake. However, several symptomatic patients did have large responses at fraction 18 (36 Gy delivered prescription dose). In addition, specification of normalized uptake magnitude and the onset of symptoms were skewed for patients who already had esophagitis symptoms at the time of the PET study. Another limitation is the frequency of esophagitis scoring. Esophagitis was scored at the weekly radiation oncology symptom clinic, which introduces uncertainty when trying to establish time differences between uptake magnitude and esophagitis severity. In addition, esophagitis scoring is a subjective process and other endpoints may be examined as well (e.g. weight loss, patient reported outcomes, etc.).

In conclusion, FDG-PET can be used to quantify esophagitis during radiation therapy. This quantification is noninvasive and objective, and also provides spatial information about the location and extent of esophageal injury which may be useful for studying esophageal dose-response, specifically with the inclusion of spatial information. Normalized uptake is a patient-specific radiation response that can predict whether patients who are asymptomatic at the time of the PET study may develop symptoms by the end of radiation treatment.

Chapter 6

Influence of Dose-Geometry on Esophageal Expansion

In previous chapters, esophageal expansion was established as a radiation-response biomarker and a surrogate endpoint of esophageal toxicity. This chapter uses the expansion biomarker to probe if dose-geometry influences esophageal-response. Esophageal regions with the highest radiation therapy plan dose were analyzed by examining the standard deviation of voxel dose across each cross-sectional slice of the sub volume, for each patient. Any correlation between the expansion biomarker and the spatial distribution of voxel dose across the esophageal slices is examined. This analysis is executed by examining the esophageal sub volume near the maximum axial esophagus dose, for any one slice, for a region of 9 esophageal slices (constant region volume) and the region with slices receiving at least 90% of the single maximum axial esophagus slice dose (region with similar axial esophagus dose).

6.1 Methods and Materials

6.1.1 Patient Population

The patient population for project 3.1 is the same cohort as those in chapters 3 and 4 (projects 1.1 and 1.2). Eighty-five patients were selected from a prospective clinical trial at University of Texas-MD Anderson Cancer Center for stage III NSCLC, treated with concurrent intensity-modulated radiation therapy and chemotherapy (paclitaxel and carboplatin), with tumor prescription doses of 60 (n=4), 66 (n=28), or 74 (n=53) Gy in 2-Gy fractions over 6-8 weeks. Patients had weekly 4DCT imaging and esophagitis scoring according to Common Terminology Criteria for Adverse Events version (CTCAE) 3.0. The distribution of maximum esophagitis grades during treatment was: 24 were grade 0, 45 were grade 2, and 16 were grade3. Since asymptomatic diagnostic assessment of esophagitis was not conducted, there were no grade1 patients in this study. We selected 85 of 97 possible patients from this clinical trial for the present study, excluding 3 patients due to image quality and 9 for not having 4DCT imaging. Our study was approved by the University of Texas-MD Anderson Cancer Center Institutional Review Board and we complied with HIPAA regulations.

CT scans were acquired on the same scanners using identical protocols as described in previous chapters. Patient treatment planning and segmentation was conducted using the Pinnacle treatment planning system (Phillips Healthcare), with esophageal contours segmented from the cricoid cartilage to the gastroesophageal junction, in the axial plane, with Pinnacle version 9.8.

6.1.2 Dose-Geometry Quantification

The scientific question investigated in this chapter is: what influence, if any, does dose-geometry have on the expansion-response mechanism. Dose-geometry in this context is considered the conformity of the dose distribution across the cross-sectional area of a given axial slice of the esophagus. A comparison of differing dose-geometries for a single axial slice of the esophagus is illustrated in Figure 6.1. In this figure, the deviation of dose across the esophgeal slice is markedly different between the two representations. Figure 6.1B shows a partial sparing of dose for the given slice. This work examines if these differences in dose conformity yields significantly different observed esophageal expansion. The expansion-response allows a precise quantification to test if differing dose-geometries can spare the esophagus.

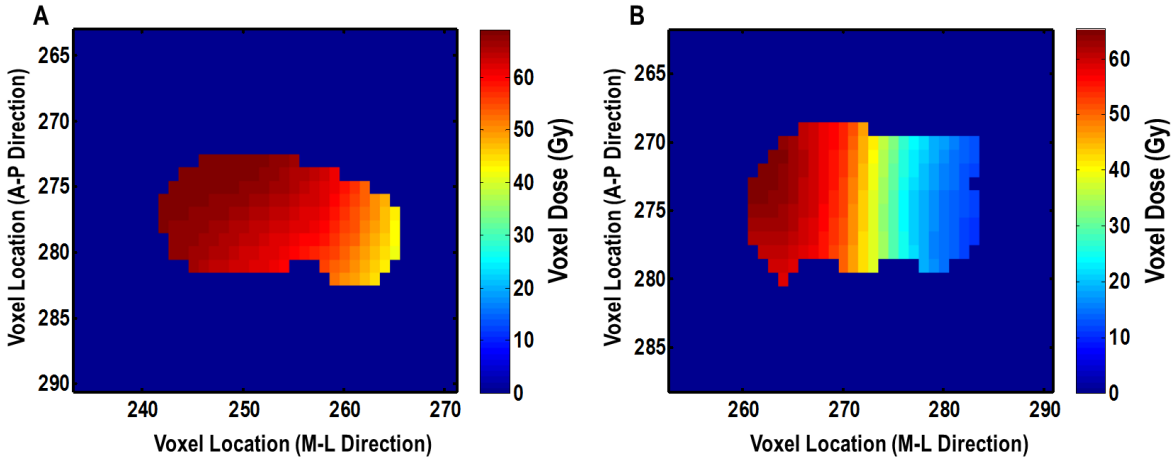

Figure 6.1: *Plot of voxel dose within the esophagus in the axial plane for two different example dose conformities. The dose-geometry in (A) is more uniform, when compared to (B), which has a partial-sparing of dose towards the lateral end of the esophagus.*

To examine any influence of dose-geometry on expansion-response, a robust measure of dose-geometry needs to be utilized. We summarize dose-geometry based on two principal quantities: the mean dose to the slice and the variation of dose across the slice. Dose-geometry was quantified by first calculating mean axial dose to each slice, for every slice, within the analyzed sub volume of the esophagus, individually for each patient. Next, the standard deviation of voxel dose across the cross-sectional area of the esophgeal slice was quantified for all slices in the analysis region. The standard deviation of dose is used as the representation of dose variation across a given slice of the esophagus. These two dose quantities are also examined in proportion to each other as the ratio of mean slice dose to standard deviation of dose across the given slice throughout this chapter. This arises from the fact that the expansion-response depends on the mean dose to the axial slice, yet the influence of dose deviation across the slice is being studied in this chapter. To put this into context, what if two slices had the same deviation of dose across each slice (dose-geometry), yet the mean doses were greatly different? It would follow that the slice with a larger mean dose would have a larger expansion-response, but this would not be a consequence of the dose-geometry. Additionally, this allows for analysis using a single metric instead of a combination of metrics.

Dose-geometry is analyzed with two different types of analysis regions for this work. The attractiveness of using expansion as a response quantification, is its ability to localize the response. Therefore, we can discard much of the study patient's esophagus from the analysis, only considering regions were response physically occurs. A question of how to choose the subregion of analysis naturally arises: should the region be dose dependent, or should the region be volume dependent, as both region types may effectively be different size and could

yield different results. Therefore, the esophageal subregion of analysis can either be volume constant, or dose constant. This work considers both subregion types in separate analyses, which are described in-depth in the following sections.

6.1.3 Analysis of Constant Volume of Esophagus

To hold the volume of the region of analysis approximately constant, the response from a given patient's dose-geometry was analyzed in a 9 slice region of the esophagus, centered on the axial slice with the highest axial dose. Dose-geometry, as quantified according to the previous section, was compared to the axially-averaged expansion in this 9 slice region of the esophagus. Each patient's axial expansion was quantified at the given treatment week with maximal expansion. Variation of voxel dose across each slice in the analysis region is compared to the corresponding slices' axial-average expansion. This expansion metric is the MaxExp9, as defined in chapter 3. While the MaxExp1 was shown to be the most robust measure of esophageal radiation response compared to toxicity, using MaxExp9 provides multiple data points of expansion and dose-geometry for a given patient, while not sacrificing the expansion and toxicity relationship, since MaxExp9 was highly correlated to esophagitis grade. In addition, using the 9 slice region allows for uncertainties associated with expansion and dose to be quantified in the superior-inferior direction along the esophagus, not just the axial plane, to be considered in the analysis.

6.1.4 Analysis of High-Dose Region of Esophagus

While the previous analysis holds the volume of analysis approximately constant, the expansion-response from a region with similar axial-averaged dose should also be examined. Potentially, a partial-sparing may be affected from regions with significantly different mean axial doses, and this must be held constant in a thorough analysis. Therefore, the high-dose analysis region of the esophagus was defined as all slices with a mean axial dose of at least 90% of the single maximum axial slice for a given patient. An example of the high-dose region of analysis is illustrated for an example patient in Figure 6.2. From this figure, the difference in response within slices receiving similar mean doses is shown in the red circled area. This analysis examines if variation in dose across the esophageal slices can explain the differences in slice expansion.

All slices considered in the analysis must form a connected region with the axial slice that has the maximum mean slice dose. The axially-averaged expansion from each of these slices from the high-dose region is then compared to the corresponding standard deviation of dose across each slice within the analysis region. Each patient's axial expansion was quantified at the given treatment week with maximal expansion.

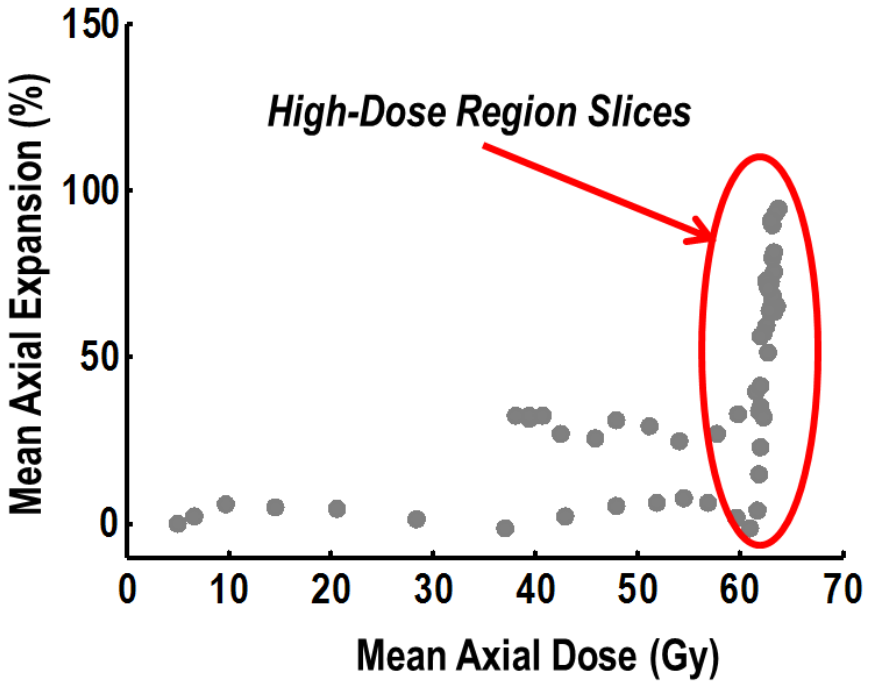

Figure 6.2: *Plot of axial expansion and corresponding axial dose for all slices of the esophagus at the week of maximal expansion for a single patient (gray circles). The red circled region represents the slices that comprise the high-dose region in the constant dose analysis of dose-geometry.*

6.1.5 Statistical Analysis

Individual patient's slice-based dose-geometry and corresponding slice expansion was examined using Spearman rank analysis. For each patient, the individual esophageal slice's dose deviation, as quantified by the slice's mean axial dose divided by the standard deviation of voxel dose, and the corresponding esophageal slice's axially-averaged expansion value were paired together for every slice in the analysis region and compared with the Spearman rank correlation coefficient. This test was carried out for both the constant volume region of the esophagus and the high-dose region analyses, independently for each patient. Any dependence of expansion on dose-geometry would therefore be analyzed on an individual patient basis.

The patient population was then analyzed collectively for correlation of the esophageal region mean expansion and the region's dose-geometry. This analysis was conducted for both the high-dose region and the constant volume region, independently. Spearman correlation coefficients were calculated between mean regional dose, standard deviation of regional dose, the ratio of regional mean dose to the regional standard deviation of dose, the region mean expansion value, and the maximum esophagitis grade. The ratio of regional mean dose to the regional standard deviation of dose is analogous to the ratio of mean slice dose to standard deviation of dose across the slice, except in this metric dose and standard deviation of dose is quantified for the analyzed volume of esophagus. Linear models of dose-geometry using mean regional dose, standard deviation of regional dose, and the ratio of mean slice dose metrics, were covariates to fit regional mean expansion value. The goodness-of-fit metric was used to assess model fit. Statistical significance of linear model variables was test with a t-test at the $p<0.05$ significance level.

6.2 Results

6.2.1 Patient-Specific Slice-based Dose-Geometry

Spearman correlation coefficients of axial-averaged expansion and corresponding dose deviation for each esophageal slice within a specific patient's analysis region is shown in Figure 6.3 for the analysis region of constant volume and analysis region of similar slice dose, respectively. The Spearman rank coefficients show a correlation between slice-based expansion and corresponding dose deviation for some patients. However, this effect is inversely proportional in many patients.

Moreover, the majority of patients had correlation coefficients in the range of (-0.5, 0.5), shown by the gray shaded region in the figure, and indicates poor correlation of dose-geometry and slice expansion. This trend is observed for both analysis region types.

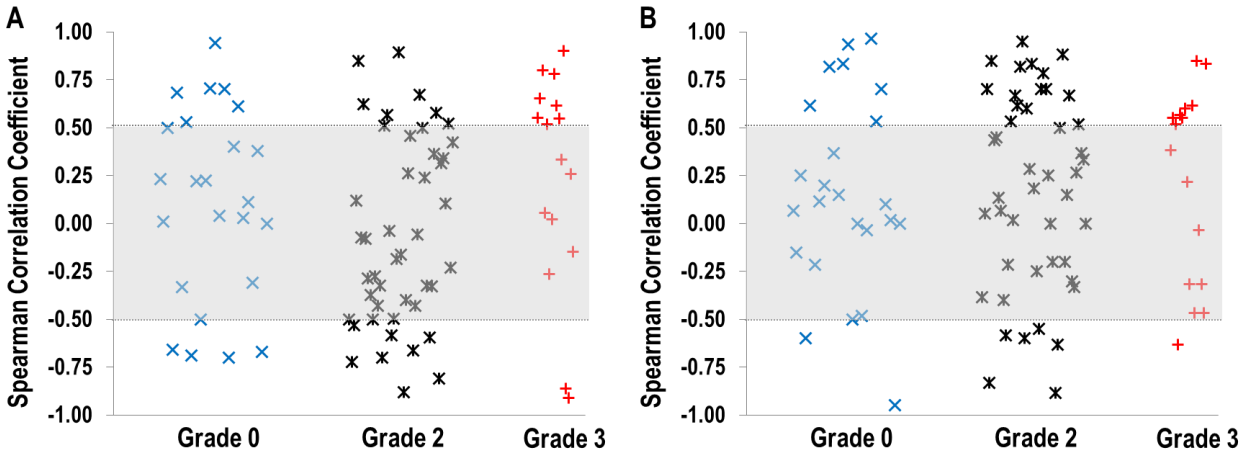

Figure 6.3: *Plot of Spearman correlation coefficient of axially-averaged expansion and the ratio of mean slice dose to standard deviation of dose across a given slice, for all slices in the high-dose region of analysis (A), and the 9 esophageal slice region of analysis (B), for every individual patient grouped according to maximum esophagitis grade. The gray shaded region represents coefficients in the range of (-0.5,0.5).*

6.2.2 Population-based Dose-Geometry

Figure 6.4 shows the relationship of dose-geometry to expansion for the constant volume region analysis, for patients that had either maximum grade 0 or grade 3 esophagitis during radiation therapy and mean region dose of at least 25 Gy. From this figure, an overlap of grade 0 and grade 3 patients with similar mean regional doses can be observed. A similar trend exists with constant dose region analysis. The Spearman correlation analysis is summarized in Figure 6.5 for both region types analyzed. The highest Spearman coefficient of any dose variable to any outcome was 0.495 for mean regional dose and ≥ grade 2 esophagitis, for the

constant-dose region. The highest Spearman coefficient of any dose variable to any outcome was 0.496 for mean regional dose and mean regional expansion, for the constant-volume region. In both analysis region types, the ratio of mean regional dose to standard deviate on of dose had Spearman coefficients of less than 0.400 for any of the three outcomes.

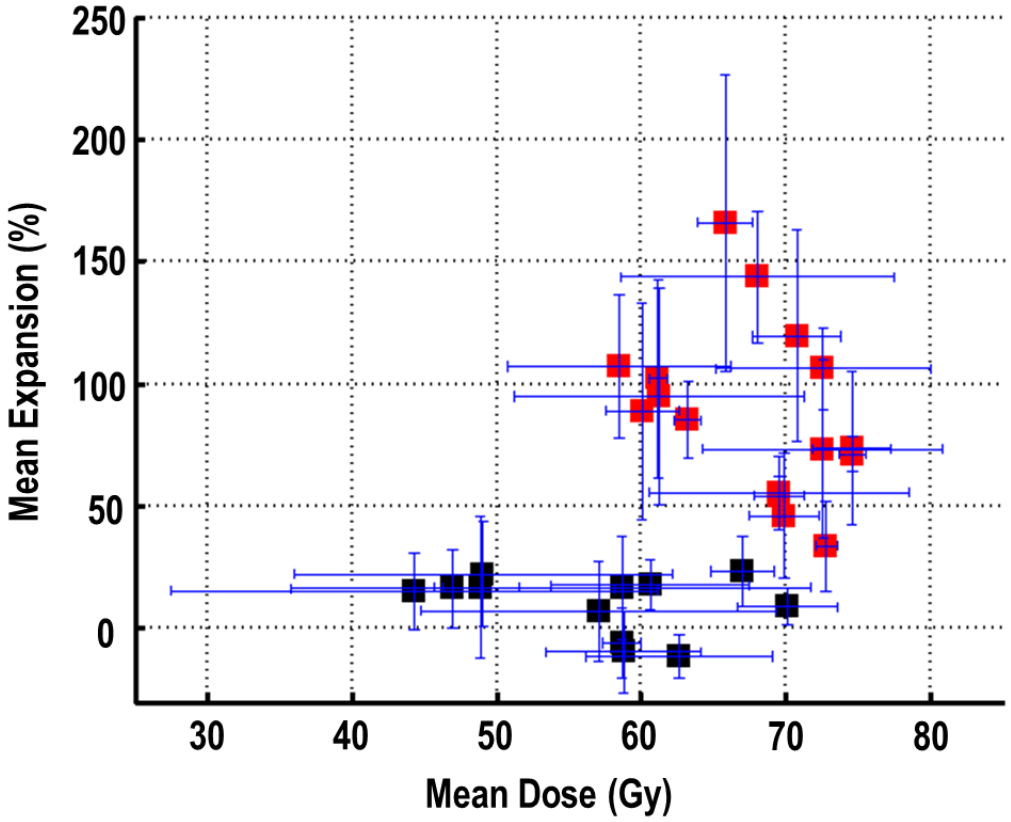

Figure 6.4: *Plot of mean regional dose and mean regional expansion for the 9 slice analysis region for 11 grade 0 patients (black squares) and 16 grade 3 patients (red squares). The vertical bars represent each patient's standard deviation of axial expansion. The horizontal bars represent the standard deviation of dose in the analysis region.*

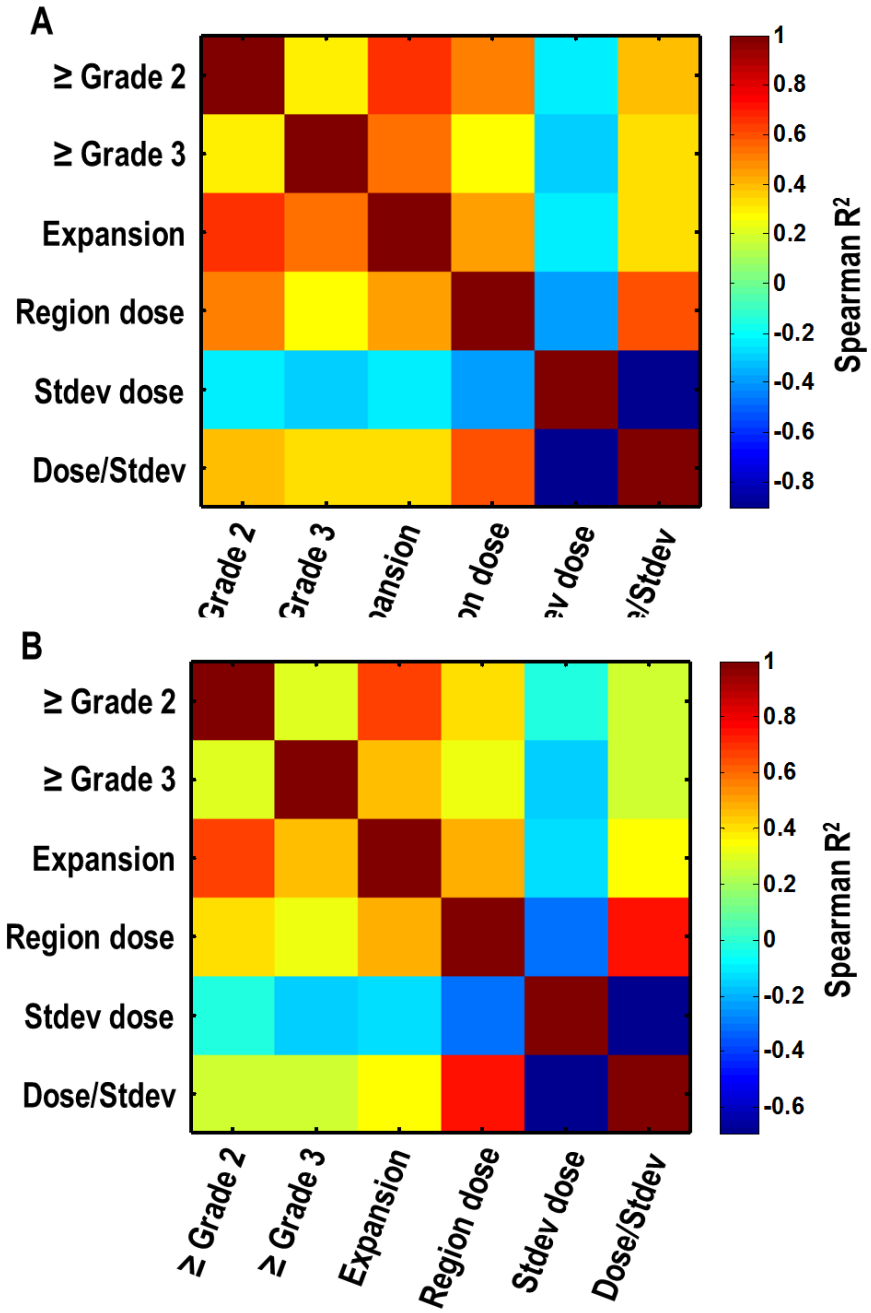

Figure 6.5: *Plot of self-correlation matrix of Spearman correlation coefficients for ≥grade 2 and grade 3 esophagitis, region mean expansion value, region mean dose value, the mean standard deviation of axial dose, and the ratio of region mean axial dose to the corresponding standard deviation of dose, for the high-dose region of analysis (A), and the 9 esophageal slice region of analysis (B). Color shade represents the Spearman correlation coefficient value, as given by the corresponding color bar, between the corresponding variables for the row/column combination.*

Results of the linear model of expansion and dose-geometry are displayed in Table 6.1. The goodness-of-fit metric was 0.107 and 0.153 for the constant dose and constant volume regions of analysis, respectively. Neither standard deviation of regional dose nor the ratio of mean regional dose to the standard deviation of dose was significant in either linear model. The only significant coefficient in either region type linear model was the mean regional dose for the constant volume region of analysis.

Table 6.1: *Linear model of expansion in the analysis region, with mean regional dose, standard deviation of regional dose, and the ratio of mean regional dose to standard deviation of dose, as the respective covariates, from all 85 patients. The upper table represents the region of near constant dose and the lower table represents the 9 slice volume region of analysis.*

High-Dose Region	Coefficient	SE	tStat	p-value
Intercept	33.05	20.28	1.63	0.107
Mean Dose	0.50	0.27	1.85	0.068
Stdev Dose	-1.07	1.29	-0.83	0.410
Dose/Stdev	0.14	0.26	0.54	0.589

Constant-Volume Region	Coefficient	SE	tStat	p-value
Intercept	9.32	12.40	0.75	0.455
Mean Dose	0.73	0.21	3.38	0.001
Stdev Dose	-0.66	0.90	-0.73	0.465
Dose/Stdev	-0.10	0.27	-0.36	0.717

6.2 Chapter Discussion

This project was designed to elucidate any dependence of expansion on dose-geometry with the hypothesis that a higher standard deviation of dose across the cross-sectional area of a given slice of the esophagus will provide a partial sparing effect, thereby reducing expansion for an increasing standard deviation of dose.

The effect of varying dose across a given axial slice of the esophagus did show a decrease in expansion for some patients, for slices examined in the volume with all slices of at least 90% of the maximum axial-averaged dose (constant dose analysis region). This was also observed for some patients in the 9 slice volume centered on the slice with the maximum axial-averaged dose (constant volume analysis region). However, most patients did not show a correlation of decreasing expansion for a corresponding increase in dose deviation. Additionally, some patients observed the opposite effect where dose deviation across esophageal slices had an accompanying increase in the slice expansion.

The population-based analysis averaged the expansion and dose, as well as dose deviation across the axial slices, from all slices in the analysis region (for both analysis region types). This summarized the expansion and dose-geometry for individual patients and then combined them into a single cohort to examine the entire population. In this analysis, Spearman rank correlation did not show a strong influence of dose-geometry on expansion. In addition to expansion, esophagitis grade was also examined as a response. Similar to expansion-response, dose-geometry did not show a strong correlation to esophagitis grade. This should be expected given the strong relationship to maximum expansion of one axial slice of the esophagus (MaxExp1) and esophagitis grade, as shown in Chapter 3. The linear models

of dose-geometry with expansion had poor goodness-of-fit metrics, with both region types having R^2 of less than 0.160 in either case. Furthermore, as illustrated in Figure 6.4, there is no clear partial sparing effect of increasing axial dose-deviation in either grade 0 or grade 3 patients. The expectation would be that if dose-geometry can reduce expansion, there should be an observed trend of grade 0 patients with large dose-deviation compared to grade 3 patients with small dose-deviations, for similar mean doses for the analysis region. However, we see an opposite effect; many grade 3 patients have larger dose-deviations than grade 0 patients of similar mean region dose.

This study was not without limitations. Dosimetric accuracy is of primary concern when examining voxel dose in an organ that has a small cross-sectional area, such as the esophagus. This is especially true in our study as axial slices of the esophagus and the corresponding deviation of dose across the slice was this study's focus. To reduce uncertainty associated with dose calculation, we accumulated dose on weekly 4DCT study's and then deformed the calculated weekly dose back into the plan 4DCT frame of reference and summed all weekly doses. Another limitation is the uncertainty in expansion. While expansion allows study of sub volumes of the esophagus to localize radiation response, it does have uncertainty associated with its calculation. It is possible that the individual patient-specific analysis of dose-geometry may require a level of precision not attainable presently. However, the population-based analysis showed no influence of dose-geometry on expansion that is clinically useful.

In conclusion, dose-geometry does not have a clinically detectable influence on radiation response in the esophagus. The amount of variability in dose-deviation, mean regional dose, and corresponding expansion shows dose-geometry is not influential for esophageal response.

Chapter 7

Influence of Modality on Esophageal Expansion and FDG Uptake: IMRT versus Proton Therapy

---

In the previous chapter, the esophageal expansion biomarker was utilized to examine if radiation dose-geometry was influential on radiation response. In this chapter, the question of radiation treatment modality having an influence on radiation-response and toxicity in the esophagus is addressed. First, the most robust expansion and normalized uptake biomarkers are identified for patients receiving passive scatter proton therapy (PSPT), using the same methodology as in chapters 3 and 5. Next, the dose was compared between IMRT and PSPT patients for the study population. Additionally, dynamics of biomarker and toxicity response were examined throughout treatment between the two modalities. Any influence of modality on expansion was investigated with a cross-validated Random Forests regression model. Finally, modality's impact on toxicity prediction modelling was examined with LASSO logistic regression.

7.1 Methods and Materials

7.1.1 Patient Population

It is important to note that two separate cohorts are being analyzed in this chapter. One cohort consists of patients analyzed using the FDG-based biomarker response (n=79), and the other cohort being patients analyzed using esophageal expansion-based biomarker response (n=134). A total of 68 patients were in both the FDG and expansion cohorts. The FDG-based cohort is identical to the cohort studied in chapter 5. The expansion-based patient cohort for this chapter (project 3.2) contains the same cohort as those in chapters 3 and 4 (projects 1.1, 1.2, 2.1 and 2.2), but 49 additional patients treated with passive-scatter proton therapy (PSPT) are also included in this project's study cohort.

All patients were selected from the prospective clinical trial at University of Texas-MD Anderson Cancer Center for stage III NSCLC, described in previous chapters. Esophagitis scoring was conducted in accordance to Common Terminology Criteria for Adverse Events version (CTCAE) 3.0. The demographics and distribution of maximum esophagitis grades during treatment for patients in the expansion and FDG-PET analyses are listed in Tables 7.1 and 7.2, respectively. Our study was approved by the University of Texas-MD Anderson Cancer Center Institutional Review Board and was compliant with HIPAA regulations.

As described in previous chapters, CT scans were acquired on General Electric Lightspeed Discovery ST or Lightspeed RT16 (GE Healthcare, Waukesha, WI) or Philips Brilliance 64 (Philips Healthcare, Bothell, WA) CT scanners operated at 120 kV. Voxel dimensions were 0.98x0.98x2.50 mm in the right-left direction, anterior-posterior, and superior-inferior direction, respectively, with a 512×512-pixel area. Patient treatment planning and segmentation was

conducted using the Pinnacle treatment planning system (Phillips Healthcare), with esophageal contours segmented from the cricoid cartilage to the gastroesophageal junction, in the axial plane, with Pinnacle version 9.8.

Table 7.1: Demographics of patients from the expansion analyses (n=134).

Characteristic	Datum	Proton	IMRT
Median age (range)			
All	66 (38-85)	67 (38-76)	65 (43-85)
Male	66 (43-85)	68 (57-76)	65 (47-80)
Female	65 (38-80)	65 (38-75)	66 (43-85)
Sex			
No. of Males	75	30	45
No. of Females	59	19	40
Histologic findings			
Squamous cell carcinoma	47	18	29
Adenocarcinoma	75	25	50
Large cell carcinoma	5	2	3
Other	7	4	3
Smoking history			
Current smoker	44	26	18
Former smoker	79	21	58
Never smoked	11	2	9
Stage			
IIa	5	2	3
IIb	9	6	3
IIIa	59	20	39
IIIb	56	20	36
IV	5	1	4
Treatment dose, Gy			
74	88	35	53
66	38	10	28
60	8	4	4
Maximum Esophagitis Grade			
Grade 0	33	9	24
Grade 2	95	49	46
Grade 3	26	11	15

Table 7.2: Demographics of patients from the FDG-PET normalized uptake analyses (n=79).

Characteristic	Datum	Proton	IMRT
Median age (range)			
All	66 years (38-80 years)	66	66
Male	65 years (51-79 years)	68	65
Female	66 years (38-80 years)	62	68
Sex			
No. of Males	46	20	26
No. of Females	33	11	22
Histologic findings			
Squamous cell carcinoma	31	10	21
Adenocarcinoma	41	15	26
Large cell carcinoma	3	1	2
Other	4	3	1
Smoking history			
Current smoker	27	15	12
Former smoker	46	15	31
Never smoked	6	2	4
Stage			
IIa	3	2	1
IIb	5	3	2
IIIa	29	11	18
IIIb	38	14	24
IV	4	1	3
Treatment dose, Gy			
74	46	19	27
66	28	9	19
60			4
Maximum Esophagitis Grade			
Grade 0	14	6	8
Grade 2	36	16	20
Grade 3	21	9	12

7.1.2 Proton Treatment Radiation Response and Toxicity

This chapter seeks to determine if response, whether quantified with expansion or FDG-uptake, differs based on the type radiation therapy modality used for treatment (IMRT or Proton therapy). Therefore, the differences in radiation dose must be analyzed between the treatment modality subgroups. This is to ensure that if there exists a difference in response, it can be elucidated whether the variation in response is caused by discrepancy in treatment dose or because there is an inherent difference in esophageal response of radiation treatment from protons or photons.

The dosimetric differences were analyzed between patients treated with PSPT or IMRT. The Kruskal-Wallis analysis of variance (ANOVA) was utilized to test any significant difference ($p<0.05$) between dose metrics in the form of: equivalent uniform dose (EUD), maximum esophagus dose (Dmax), mean esophageal dose (MED), and V20 to V70, in 10 Gy increments. This was carried out separately for both expansion and FDG analyses cohorts. To determine the appropriate scaling parameter to calculate EUD, Lyman-Kutcher-Burman (LKB) models were fit to the expansion and FDG-PET cohorts for a grade 3 complication endpoint, in separate models, to determine the slope of the sigmoid curve, the dose of 50% complication, and the volume parameter of the corresponding LKB model. These derived parameters were then used to calculate equivalent uniform dose for the corresponding study cohort.

For the expansion cohort, the correlation of radiation-response biomarkers and esophagitis grade was analyzed for the 49 proton patients. This is a repeat of the analysis conducted in chapter 3, where univariate logistic regression and AUC were calculated to identify the highest performing expansion metrics for patients receiving IMRT. Spearman rank

coefficients and p-values ($p<0.05$ for significance after applying the Benjamini-Hochberg false discovery rate procedure) from univariate logistic regression were also calculated.

The highest performing expansion and FDG-based biomarkers for both modality cohorts were compared. The Kruskal-Wallis ANOVA was used to test any significant difference in biomarker values between all patients grouped according to treatment modality. This was repeated for only patients that had ≥ grade 2 maximum esophagitis, and then again ≥ grade 3 maximum esophagitis, in separate tests. The temporal dynamics of esophagitis and expansion biomarker value throughout treatment were also examined.

7.1.3 Modality and Expansion Regression Modelling

To probe if modality influences expansion biomarker response in a multivariable analysis, Random Forests regression was utilized. Repeated cross-validation was utilized to reduce bias and obtain a more generalizable modelling process for 1000 iterations of randomly drawn data folds consisting of training and test sets in a respective 70%/30% split of the patient cohort. This model construction process is identical to that utilized in chapter 4, with the exception of solving a regression and not a classification problem. Additionally, the minimum observation per leaf was set to 5 and the number of randomly drawn predictors at each node is 16 (the number of predictors divided by 3). Model predictors consisted of clinical (tumor stage, tumor location, histology, gender, age, and smoking status) and dosimetric factors (Dmax, MED, esophagus EUD, V10-V70 in 5Gy increments, LE10-LE60, in 10 Gy increments for 25% and 100% axial esophagus coverage). These are the same model predictors described with more detail in chapter 4.

Model calibration is assessed with the goodness-of-fit metric on the training set between model-calculated and known expansion values. To assess model predictive performance, the root mean squared error (RMSE) was calculated on the test sets from model-computed and known expansion values. The model construction process was first executed without modality as a predictor and model performance was assessed. Next, modality information was included as a model predictor; the performance was assessed, and then compared to the models constructed without modality information as a model predictor. Modality predictor influence in the constructed models was computed with the out-of-bag permutation error. This allows predictors to be ranked in terms of importance in the model.

7.1.4 LASSO NTCP Modelling and Modality

In chapter 4, multiple machine learning methods were utilized to create NTCP models for esophageal toxicity. This section uses the same LASSO logistic regression implementation as found in chapter 4, with prediction models created with and without modality as a predictor (dichotomous 1 or 0 indicating PSPT or IMRT modality, respectively). The same repeated cross-validation procedure is utilized for 1000 iterations of randomly drawn training and test data folds, with stratification. The models are constructed to classify the ≥ grade 3 esophagitis endpoint in both model types. Model calibration and predictive performance is assessed with AUC on the training and test sets, respectively. To quantify predictor importance, the most common recurring predictors in individual models were cataloged for each iteration of model construction.

7.2 Results

7.2.1 Modality, Radiation Response, and Toxicity

Overall, DVH dose metrics were similar between modality for both expansion and normalized uptake cohorts. Figure 7.1 illustrates the distribution of patient DVH metrics between modality for both biomarker type. For both analyses, the distributions of DVH metrics for patients that develop grade 3 esophagitis are displayed in Figure 7.2.

The equivalent uniform dose calculation parameters derived from the LKB models were consistent for both the expansion and FDG-PET cohorts. The volume parameter was determined to be 0.33 for both cohorts, which is consistent with other LKB models of esophagitis in the literature.[32, 50, 77,78]

For patients treated with PSPT, we observe higher dose-volume in the expansion cohort, with similar mean and max doses. The difference between dose-volume was greatest among patients that develop grade 3 esophagitis (Figure 7.2). This difference was shown to be statistically significant for V20 to V60, for the grade 3 patient group. Neither EUD, Dmax, nor MED showed significant differences in any patient complication group from the expansion analysis. Results of the statistical tests are shown in Table 7.3.

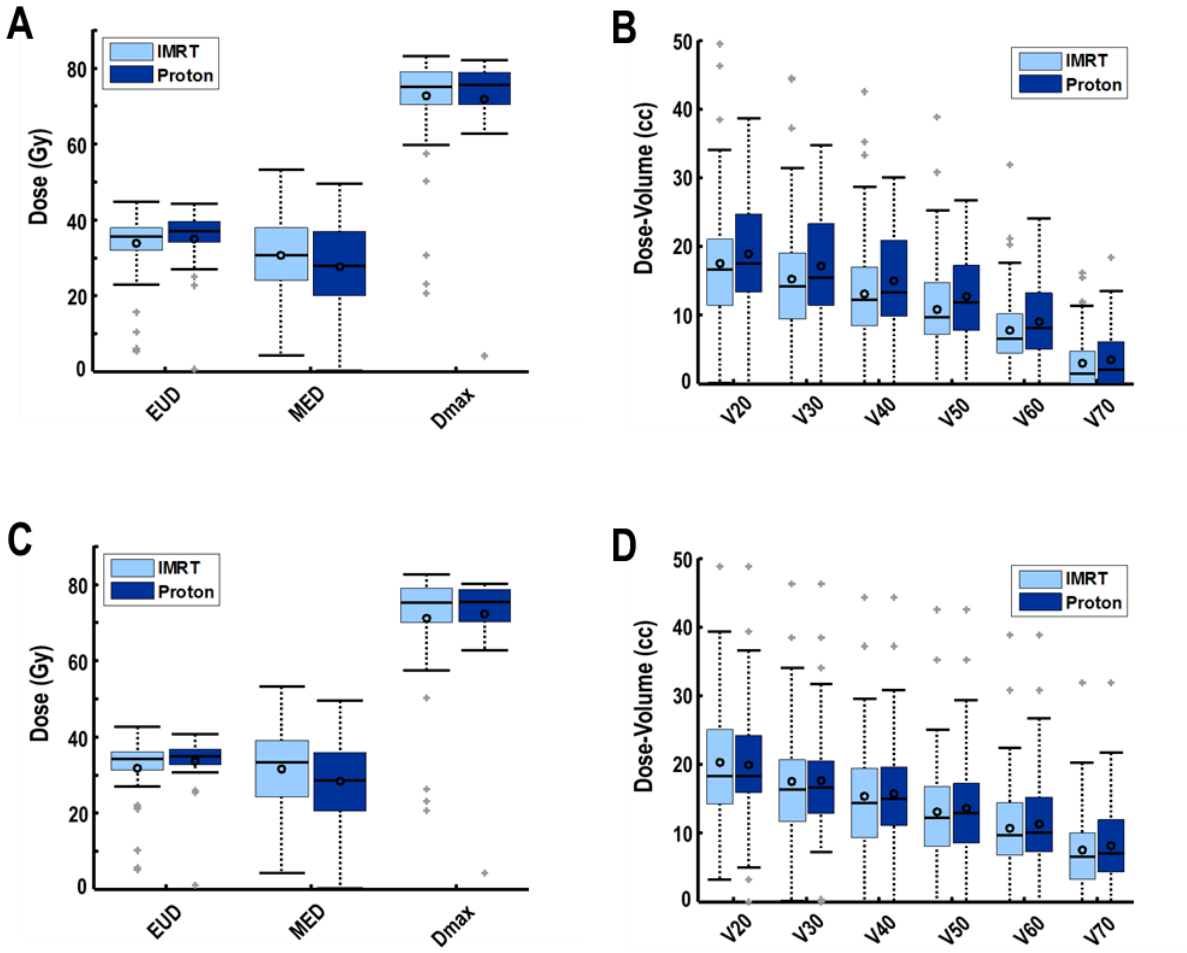

Figure 7.1: *Boxplots of the distribution of esophageal equivalent uniform dose and DVH metrics for the expansion analyses (A,B) and the normalized uptake analyses (C,D). The gray '+' represent outliers, the box vertical edges represent the 25th and 75th quartiles, the solid line in each box represents the corresponding median value, the black circles within each box represents the corresponding mean metric value, and the whiskers represent the metric value ranges.*

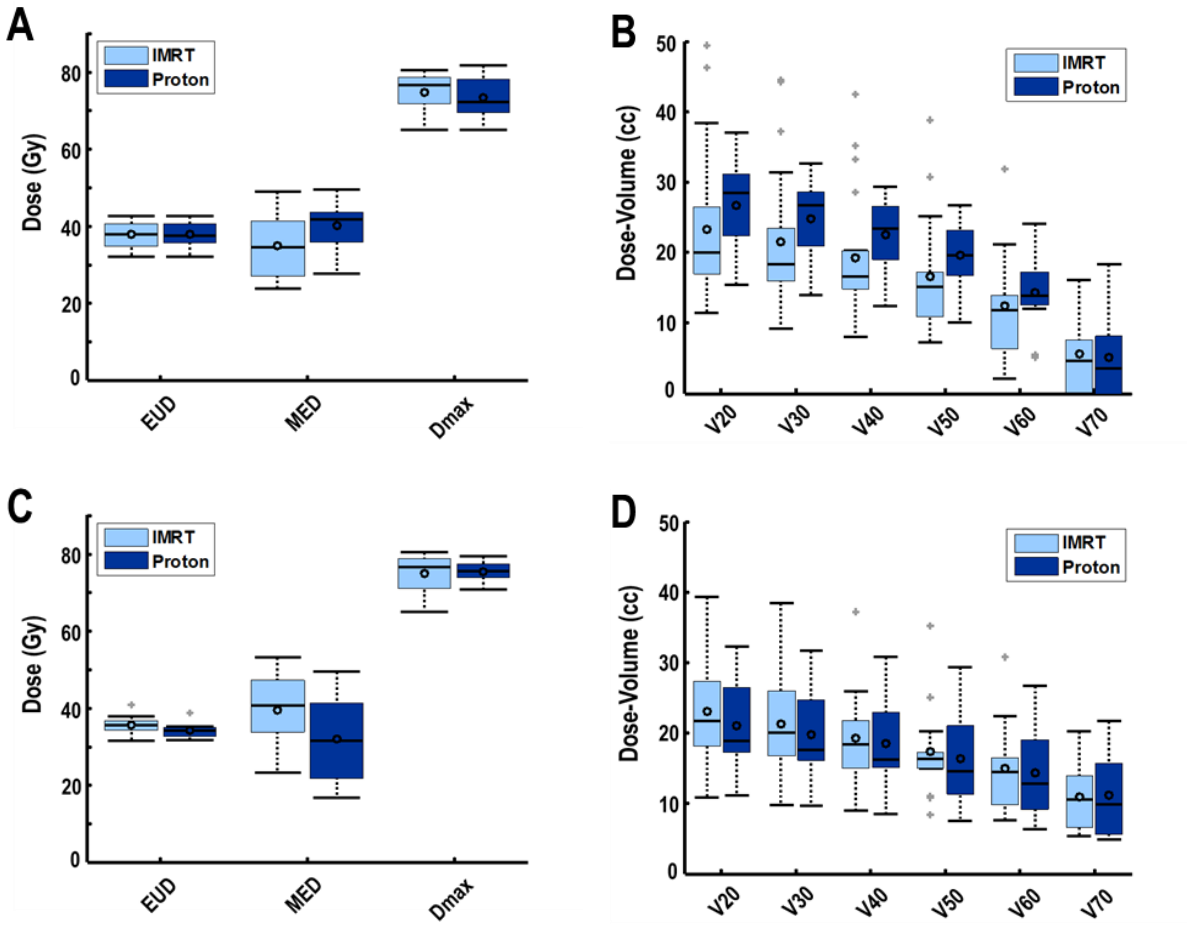

Figure 7.2: *Boxplots of the distribution of esophageal equivalent uniform dose and DVH metrics for the expansion analyses (A,B) and the normalized uptake analyses (C,D) for patients that develop grade 3 esophagitis. The gray '+' represent outliers, the box vertical edges represent the 25th and 75th quartiles, the solid line in each box represents the corresponding median value, the black circles within each box represents the corresponding mean metric value, and the whiskers represent the metric value ranges.*

Table 7.3: *p-values of Kruskal-Wallis ANOVA tests for DVH metrics between patients treated with PSPT and IMRT for the expansion biomarker cohort. Significance level is p<0.05, n=134.*

Kruskal-Wallis ANOVA (Dose-Expansion Patients)									
Endpoint	***EUD (Gy)***	***Dmax (Gy)***	***MED (Gy)***	***V20 (cc)***	***V30***	***V40***	***V50***	***V60***	***V70***
All	0.085	0.928	0.224	0.315	0.125	0.130	0.180	0.197	0.639
≥Grade 2	0.378	0.899	0.092	0.338	0.147	0.136	0.468	0.401	0.917
≥Grade 3	0.844	0.693	0.076	0.026	0.030	0.016	0.002	0.030	0.960

For the normalized uptake cohort analysis, DVH metrics were not significantly different for either the entire study group of patients or patients with grade 2 or grade 3 esophagitis complication. The results of the statistical tests are shown in Table 7.4.

Table 7.4: *p-values of Kruskal-Wallis ANOVA tests for DVH metrics between patients treated with PSPT and IMRT for the FDG-PET biomarker cohort. Significance level is p<0.05, n=79.*

Kruskal-Wallis ANOVA (Dose-FDG Patients)									
Endpoint	***EUD (Gy)***	***Dmax (Gy)***	***MED (Gy)***	***V20 (cc)***	***V30***	***V40***	***V50***	***V60***	***V70***
All	0.240	0.783	0.633	0.900	0.928	0.759	0.222	0.716	0.175
≥Grade 2	0.5491	0.983	0.526	0.780	0.672	0.983	0.175	0.865	0.119
≥Grade 3	0.1166	0.764	0.570	0.483	0.243	0.483	0.241	0.780	0.151

The dynamics of patient toxicity throughout treatment, separated by treatment modality, is presented for patients in the expansion biomarker analysis in Figure 7.3. The onset and proportionality of toxicity throughout treatment is similar for both treatment types.

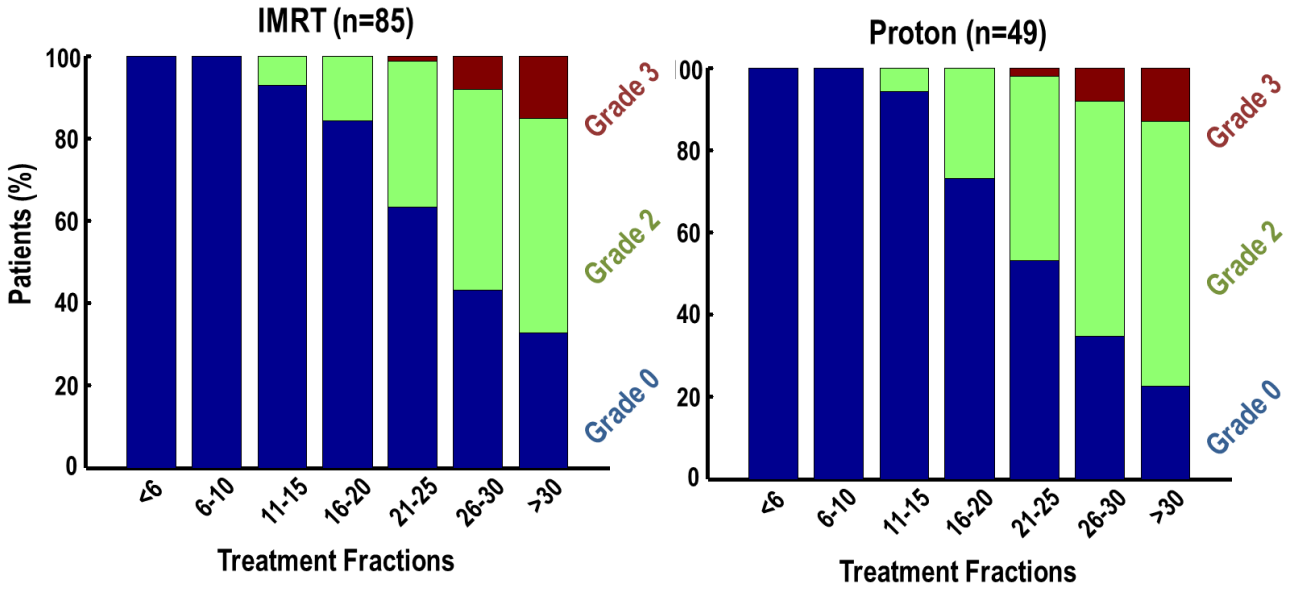

Figure 7.3: *Stacked bar chart of distribution of esophagitis grade in 5 fraction time intervals throughout treatment for patients receiving IMRT (left) and PSPT (right), in the expansion biomarker analysis. Blue represents grade 0, green grade 2, and red represents grade 3 maximum esophagitis grades.*

Similar to toxicity onset, expansion dynamics peak at the end of treatment and in a similar fashion. The patient MaxExp1 in 5 fraction time intervals is illustrated in Figure 7.4.

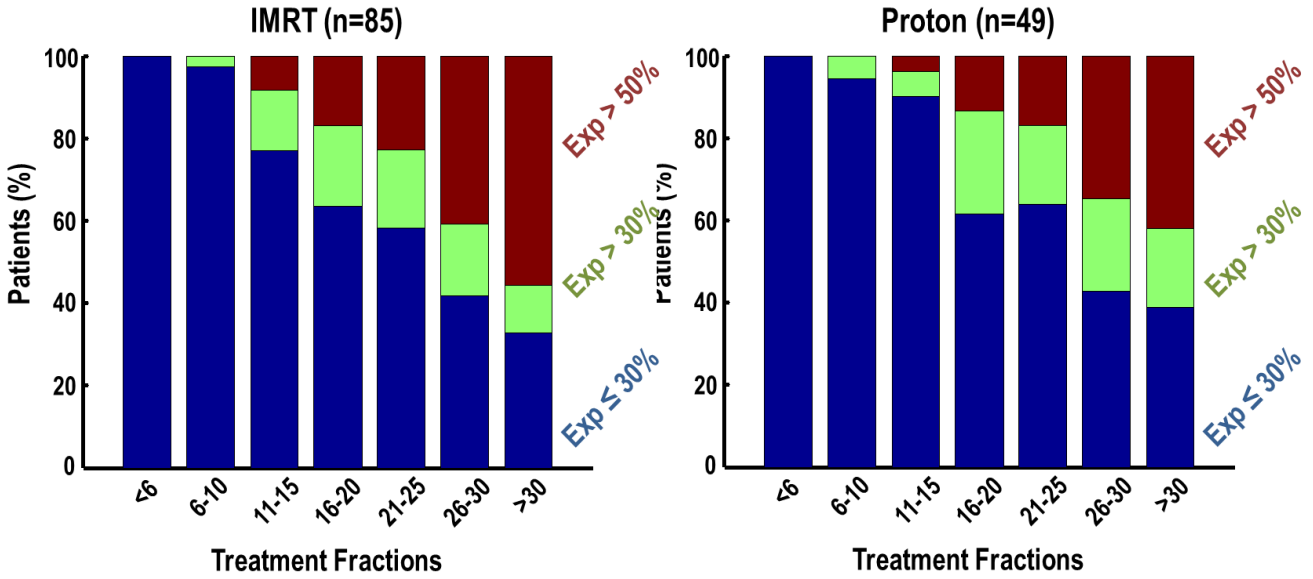

Figure 7.4: *Stacked bar chart of distribution of esophagitis MaxExp1 in 5 fraction time intervals throughout treatment for patients receiving IMRT (left) and PSPT (right), in the expansion biomarker analysis, with threshold expansion levels of 30%, and 50%. Blue represents ≤ 30%, green > 30%, and red represents > 50% MaxExp1.*

The results of the statistical analysis of the expansion biomarker and toxicity for the 49 PSPT patients are shown in Table 7.5. Similar to the IMRT patients, MaxExp1 and LenExp30% are the highest performing biomarkers. The Kruskal-Wallis test between expansion biomarker was then utilized on MaxExp1 and LenExp30%. There was significant difference in MaxExp1 for the ≥ grade 2 and ≥ grade 3 complication groups, for different treatment modality. This is illustrated in Figure 7.5. LenExp30% showed significant difference in biomarker value between treatment modality in the ≥ grade 3 complication group. For normalized uptake, $nSUV_{AxMax1}$ and $nSUV_{Len40\%}$ were the highest performing FDG-based biomarkers. The Kruskal-Wallis test was

utilized for $nSUV_{AxMax1}$ and $nSUV_{Len40\%}$ and did not yield any significant difference between values in the IMRT and PSPT cohorts. The results of the Kruskal-Wallis tests for biomarker value between modality are shown in Table 7.6.

Table 7.5: *Statistical analysis of esophageal expansion biomarkers and esophagitis grade, for patients receiving PSPT (n=49). Rs is the Spearman rank correlation coefficient, P is the p-value, and AUC is the area under the curve from ROC analysis.*

Expansion Metrics	**≥Grade 2**			**≥Grade 3**		
	AUC	***P***	***Rs***	***AUC***	***P***	***Rs***
MeanExp	0.831	0.002	0.444	0.819	0.004	0.429
MaxExp1	0.869	<0.001	0.496	0.828	0.003	0.414
MaxExp3	0.817	0.006	0.425	0.700	0.066	0.268
MaxExp5	0.822	0.006	0.432	0.694	0.066	0.261
MaxExp7	0.825	0.005	0.436	0.708	0.066	0.280
MaxExp9	0.831	0.005	0.444	0.708	0.072	0.280
MaxExp11	0.828	0.004	0.440	0.706	0.072	0.276
LenExp20%	0.858	0.001	0.482	0.840	0.003	0.457
LenExp30%	0.892	<0.001	0.565	0.863	0.003	0.469
LenExp40%	0.875	<0.001	0.530	0.839	0.003	0.490
LenExp50%	0.775	0.001	0.428	0.797	0.028	0.437
LenExp60%	0.750	0.002	0.394	0.806	0.072	0.461
LenExp70%	0.613	0.049	0.225	0.536	0.139	0.072
LenExp80%	0.563	0.155	0.160	0.576	0.072	0.195
LenExp90%	0.513	0.522	0.068	0.556	0.072	0.304
LenExp100%	0.513	0.522	0.068	0.556	0.072	0.304
PercExp60	0.800	0.002	0.403	0.847	0.003	0.466
PercExp65	0.828	0.002	0.440	0.858	0.003	0.481
PercExp70	0.839	0.001	0.455	0.844	0.004	0.462
PercExp75	0.850	0.001	0.470	0.833	0.003	0.447
PercExp80	0.856	0.001	0.477	0.814	0.003	0.421
PercExp85	0.856	0.001	0.477	0.792	0.005	0.391
PercExp90	0.861	0.001	0.485	0.797	0.005	0.399
PercExp95	0.867	0.001	0.492	0.806	0.004	0.410

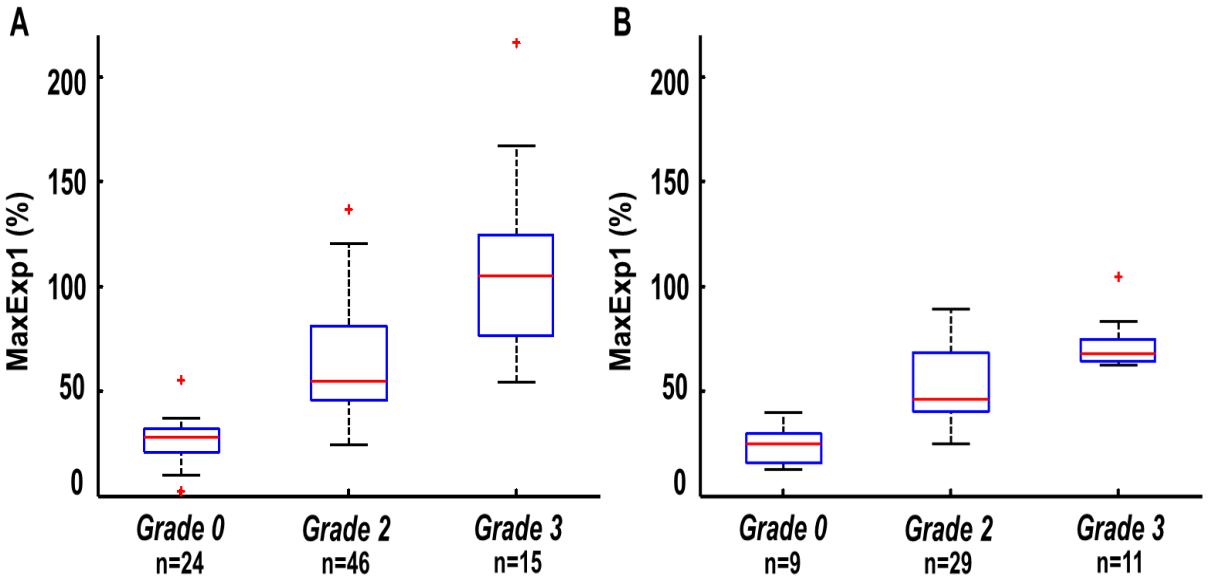

Figure 7.5: *Boxplots of MaxExp1 for either IMRT (A) or PSPT (B) treatment modalities, grouped according to patient maximum esophagitis grade during treatment. Red '+' indicate outliers, and box edges are the 25th and 75th quartiles. Whiskers represent range and the red line within the box represents the median value.*

Table 7.6: *p-values of Kruskal-Wallis ANOVA tests for biomarker values between patients treated with PSPT and IMRT for the expansion (n=134) and FDG-PET (n=79) cohorts. Significance level is p<0.05.*

Kruskal-Wallis ANOVA

	Expansion		**Normalized Uptake**	
Endpoint	***MaxExp1***	***LenExp30***	$nSUV_{AxMax1}$	$nSUV_{Len40\%}$
All	0.392	0.483	0.182	0.100
≥Grade 2	0.049	0.093	0.182	0.100
≥Grade 3	0.016	0.036	0.815	0.616
EUD ≥ 32 Gy	0.131	0.455	0.828	0.713

The relationship between MaxExp1 biomarker value and esophageal equivalent uniform dose for the given treatment modality and esophagitis severity is shown in Figure 7.6. The relationship between expansion and dose in a 9 slice subvolume centered at the slice of maximal expansion for the given treatment modality and esophagitis severity is shown in Figure 7.7. In either case, there exists no clear difference in dose-response between PSPT and IMRT. EUD does not show a strong delineation of toxicity severity for increasing EUD above 32 Gy. Below approximately 32 Gy, grade 3 esophagitis does not occur for either treatment cohort. However, after this dose level severity does increase but several grade 0 patients have similar EUD to grade 3 patients. We do see a strong separation of grade 0 and grade 3 patients around MaxExp1 equal to 50%. A similar dose and expansion-response trend is observed for the subvolume analysis. There is no severe toxicity until about 35 Gy and expansion discriminates between severe toxicity and patients that are asymptomatic, at around 40% mean subvolume expansion. Additionally, many patients that are asymptomatic have similar subvolume mean doses as patients who experience grade 3 esophagitis.

The distribution of MaxExp1 for patients with esophagus equivalent uniform dose above 32 Gy, grouped according to treatment modality, is shown in Figure 7.8. It is evident that the values of MaxExp1 have similar values near the median, regardless of treatment modality. Additionally, there was no significant different between treatment modality for either expansion or FDG-based biomarker values for patients within this dose range, according to the Kruskal-Wallis ANOVA test. The results of the statistical tests are shown in the last row of Table 7.6.

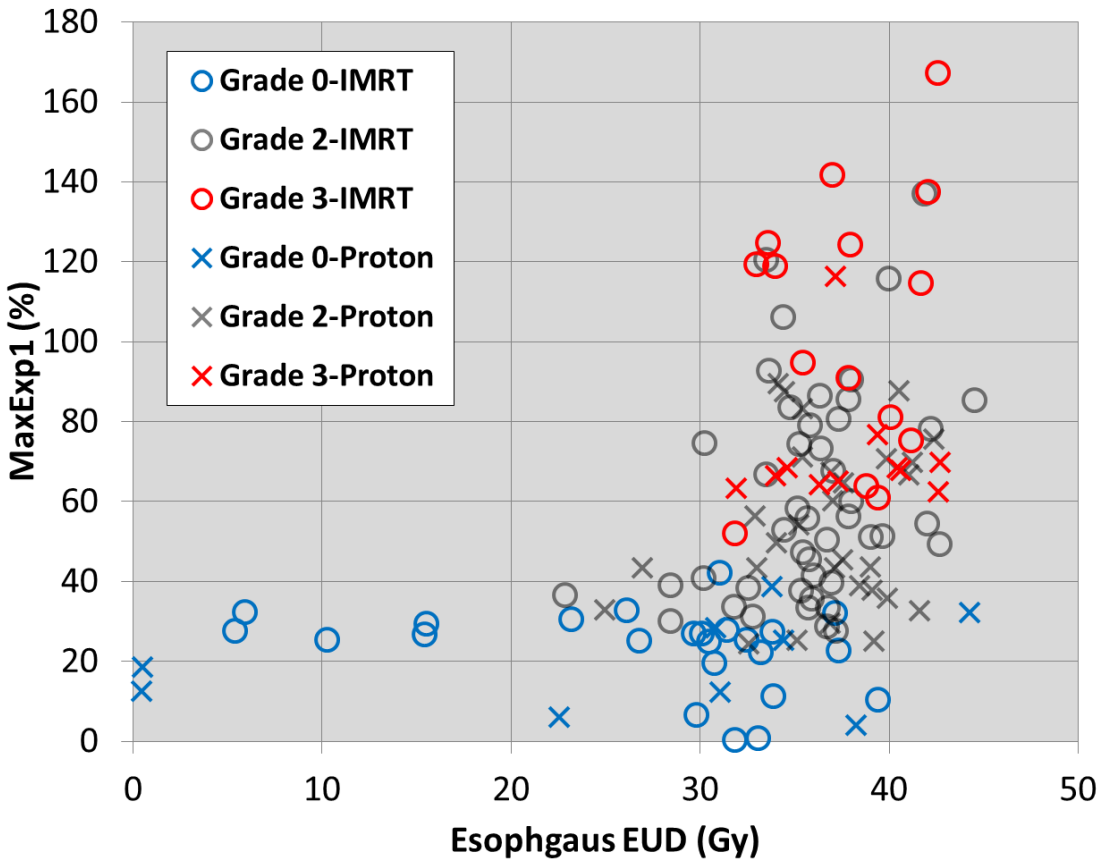

Figure 7.6: *Plot of MaxExp1 for a given esophageal EUD, with data points grouped according to treatment modality ('O' for IMRT and 'X' for PSPT), as well as maximum esophagitis grade during treatment (blue = grade 0, black = grade 2, and red = grade 3).*

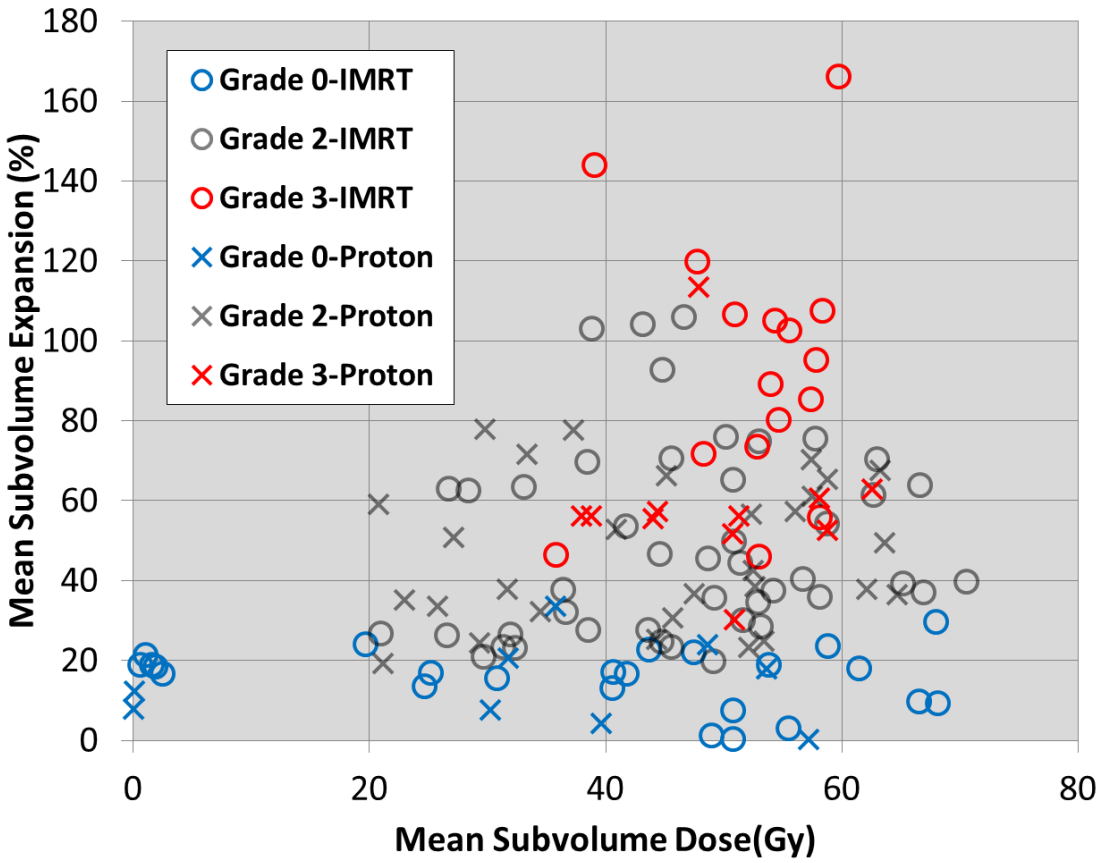

Figure 7.7: *Plot of 9 slice subvolume mean expansion and mean dose for 134 study patients, with data points grouped according to treatment modality ('O' for IMRT and 'X' for PSPT), as well as maximum esophagitis grade during treatment (blue = grade 0, black = grade 2, and red = grade 3).*

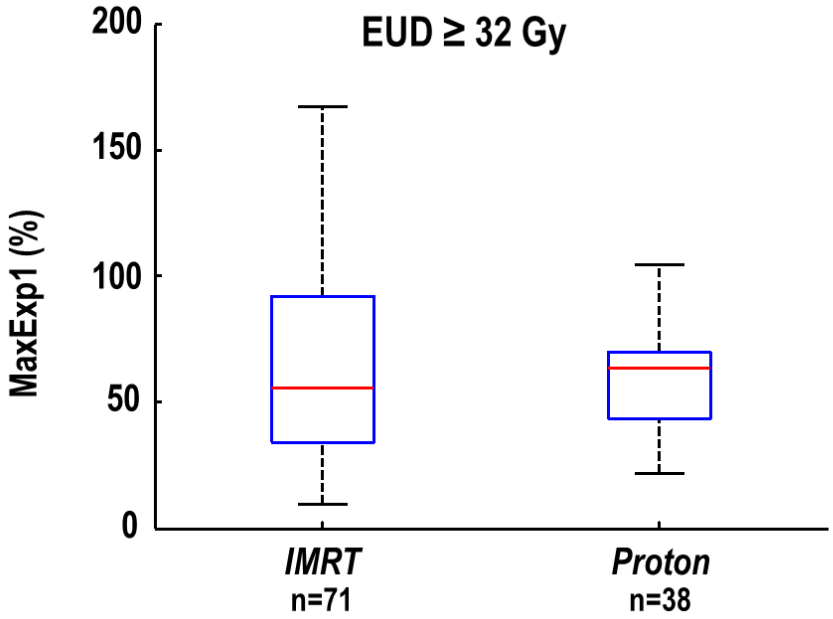

Figure 7.8: *Boxplot of the distribution of MaxExp1 values for patients with esophageal EUD ≥ 32 Gy, grouped according to treatment modality. The edges of the box represent the quartile values of expansion, with the red line within each box representing that groups median expansion value. The range of values is represented by the black whiskers and the red '+' denotes outliers (values beyond 1.5 times the interquartile range from the edge of the box).*

7.2.2 Expansion Random Forests Regression Modelling

The results of the calibration of Random Forests regression models, both with and without modality as a predictor variable, are shown in Figure 7.9A. A reasonable fit of the training is obtained with goodness-of-fit around 0.64 on average, for both model types. The predictive regression model performance is similar for both model types with and without modality as a predictor variable. This is shown in Figure 7.9B. The RMSE is similar for both model types with RMSE approximately 43% on average. The out-of-bag-permutation error ranked modality as 32nd out of 50 predictors in the modality predictor model types. This indicated modality is not influential in the response model.

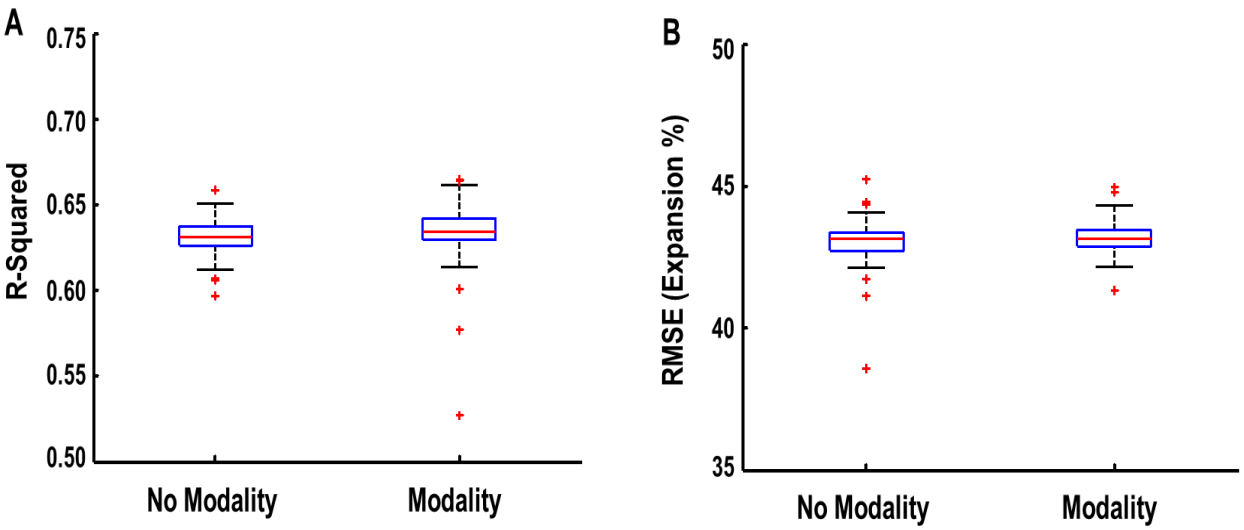

Figure 7.9: *Boxplot of Random Forests regression models goodness-of-fit (A) and RMSE (B) of MaxExp1 for 134 patients. Modality indicates models were constructed with modality information as predictor variable. Red '+' indicate outliers, and box edges are the 25th and 75th quartiles. Whiskers represent range and the red line within the box represents the median value.*

7.2.3 LASSO NTCP Modelling and Modality

Results for the LASSO NTCP model construction procedure is listed in Table 7.7. For both model types with and without modality as a predictor the model fit and predictive performance were similar, as indicated by $AUC_{Training}$ and AUC_{Test}, respectively. The distribution of both metrics between either model types had similar standard deviations. Additionally, the highest recurring model predictors were nearly identical, with mean esophageal dose being the highest recurring predictor regardless of model type.

Table 7.7: *Results of the LASSO logistic regression NTCP model construction process for 49 PSPT and 85 IMRT patients, for 1000 iterations of cross-validation. Models constructed with treatment modality as a predictor for ≥ grade 3 esophagitis are listed on the row titled 'modality'. The 5 highest recurring predictors from all iterations of the model construction procedure are listed from highest to lowest recurring. Standard deviation of AUC values are listed in parentheses.*

Lasso Classification of ≥ Grade 3 Esophagitis (n=134)

Model Type	*$AUC_{Training}$ (S.D.)*	*AUC_{Test} (S.D.)*	*Recurring Predictors*
No modality	0.845 (±0.053)	0.727 (±0.087)	MED, LE60100%, Tumor location, V60, Smoking Status
Modality	0.843 (±0.051)	0.725 (±0.088)	MED, LE60100%, V65, Smoking Status, Tumor Location

7.2 Chapter Discussion

In this chapter, the influence of treatment modality, whether from passive scatter proton therapy or photon-based IMRT, was addressed using the objective esophageal radiation response biomarkers. One of the purported benefits of proton-based radiotherapy over photon-based treatments is a reduction in normal tissue dose and thereby a reduction in normal tissue toxicity.[79-87] This work utilizes the FDG and expansion biomarkers to probe what difference, if any, exists in observed response with photon or proton treatment. For the repeat analysis of expansion biomarkers and toxicity with the patients treated with proton therapy, the results were consistent with the analysis of the IMRT patients in chapter 3. Both MaxExp1 and LenExp30% were the highest performing biomarkers in terms of classifying esophagitis severity between either treatment modality.

The FDG-based biomarkers first investigated in chapter 5 were reanalyzed by separating patients based on IMRT or PSPT treatment, to determine if there is any difference in FDG-response for patients based on treatment modality. No significant difference was found for in the distribution of FDG-response for either the $nSUV_{AxMax1}$ or $nSUV_{Len40\%}$ biomarkers between patients treated with IMRT or PSPT. For both expansion and FDG-uptake response, this shows robustness between either imaging biomarker types and also verifies the initial analysis of the relationship between the biomarkers and toxicity found in chapters 3 and 5, respectively.

Comparison of dosimetric differences between modality cohorts showed some differences in the expansion biomarker patient group, with PSPT patients having higher dose-volumes. While this was shown to have statistical significance, this did not show a difference in the timing of response or toxicity. The distribution of MaxExp1 for grade 3 patients showed a

smaller standard deviation and less extreme maximum values, while having similar population mean and median biomarker values. This difference in MaxExp1 response can be explained by the smaller sample size of the PSPT cohort compared to IMRT. Removing the three IMRT patients that had the most extreme MaxExp1 values within the grade 3 maximum esophagitis group made the distribution of MaxExp1 statistically similar to that of the PSPT patients (for Kruskal-Wallis test, the p-value becomes > 0.05). Furthermore, the previously identified MaxExp1 > 50% threshold as a surrogate for severe esophagitis is consistent between both modality cohorts. There was no significant difference in dose between modality cohorts in the FDG-based biomarker analyses.

The Random Forests regression modelling showed quite similar model fits regardless of whether or not patient modality information was included in the model construction process. The results of the goodness-of-fit show reasonable calibration for both model types. The similarity of the root mean squared errors from the independent test sets are similar for both model types. The RMSE values were approximately 43% on average in both modelling scenarios, indicating respectable model performance for a multivariate regression model. Furthermore, the regression models that included modality information did not benefit from treatment type as a predictor. The out-of-bag permutation error showed a low ranking of importance for modality as a predictor. These results indicate there is no strong influence of modality on expansion response in the esophagus.

The influence of modality on esophagitis grade was examined with the LASSO NTCP modelling analysis. These had similar results as the Random Forests regression modelling process, where modality did not show a strong influence as a model predictor for the grade 3

esophagitis complication. With the strong correlation of expansion biomarker and toxicity, this result should be expected. Both model types were quite similar in performance, as well as the most recurring predictors in the iterative model construction process.

This work was not without limitations. The dosimetry of proton-based radiation therapy is still an active area of research.[88-94] For this study we used proton dose calculated from the treatment planning system, which can differ from more robust dose calculation methods such as Monte Carlo.[95-98] This is of particular interest in the normal esophagus and NSCLC treatment as the distal end of the proton beam is typically proximal to the esophagus. Further studies examining the impact of miscalculation of proton dose and any resultant effect on the esophagus would be of interest. Another limitation is that the study patients were treated with PSPT. Newer proton radiation treatment techniques, specifically spot-scanning are replacing the passively-scattering technique, as the standard treatment where protons are utilized.[99-100] While spot-scanning shows reduced normal tissue dose, it is unknown if this would have any effect on response in the normal esophagus.

In conclusion, the previously identified highest performing radiation response biomarkers were verified with the patient cohort receiving proton therapy. This strengthens the conclusions from chapters 3 and 5 on which biomarkers were the most robust. Additionally, there is not any significant difference in biomarker response for patients treated with either PSPT or IMRT, despite higher dose-volume for the PSPT patients in the expansion analysis. Inclusion of modality information did not have an effect on toxicity prediction modelling.

Chapter 8

Esophageal Expansion to Identify Patient Radiosensitivity

In previous chapters, esophageal expansion was shown to be a robust measure of patient-specific radiation-response in the esophagus. It was also shown that the expansion-response varied patient to patient, even for patients with similar delivered doses. These previous results hinted at the possibility of expansion-response as an *in-vivo* quantification of individualized radiosensitivity in the esophagus.

In this chapter, the utility of the esophageal expansion-response as a patient-specific quantification of inherent radiosensitivity was assessed. Study patients were clustered according to expansion and corresponding delivered dose to the esophagus to identify subpopulations of inherent radiation sensitivity, using Gaussian mixture modelling. This clustering was calculated for expansion-response at the week of maximal expansion for a given patient, as well as expansion-response toward the end of treatment at approximately fraction 30. Radiosensitive patients were classified for the clusters with high expansion and a corresponding lower dose. This radiosensitivity tag was used in the toxicity prediction modelling process to improve model performance.

8.1 Methods and Materials

8.1.1 Patient Population

The patient population for this chapter (project 3.3) contains the same patient cohort used for the expansion-based analyses from chapter 7 (project 3.2). This was a combination of patients treated with IMRT (n=85), and PSPT (n=49), for a total of 134 study patients.
All patients were selected from the prospective clinical trial at University of Texas-MD Anderson Cancer Center for stage III NSCLC, described in previous chapters. Esophagitis scoring was conducted in accordance to Common Terminology Criteria for Adverse Events version (CTCAE) 3.0. Table 8.1 lists the demographics and distribution of maximum esophagitis grades during treatment for patients used for the analysis in this chapter. Our study was approved by the University of Texas-MD Anderson Cancer Center Institutional Review Board and was compliant with HIPAA regulations.

CT scans were acquired on General Electric Lightspeed Discovery ST or Lightspeed RT16 (GE Healthcare, Waukesha, WI) or Philips Brilliance 64 (Philips Healthcare, Bothell, WA) CT scanners operated at 120 kV. Voxel dimensions were 0.98x0.98x2.50 mm^3 in the right-left direction, anterior-posterior, and superior-inferior direction, respectively, with a 512×512-pixel area. Patient treatment planning and segmentation was conducted using the Pinnacle treatment planning system (Phillips Healthcare), with esophageal contours segmented from the cricoid cartilage to the gastroesophageal junction, in the axial plane, with Pinnacle version 9.8.

Table 8.1: *Demographics of study patients (n=134).*

Characteristic	Datum
Median age (range)	
All	66 (38-85)
Male	66 (43-85)
Female	65 (38-80)
Sex	
No. of Males	75
No. of Females	59
Histologic findings	
Squamous cell carcinoma	47
Adenocarcinoma	75
Large cell carcinoma	5
Other	7
Smoking history	
Current smoker	44
Former smoker	79
Never smoked	11
Stage	
IIa	5
IIb	9
IIIa	59
IIIb	56
IV	5
Treatment dose, Gy	
74	88
66	38
60	8
Maximum Esophagitis Grade	
Grade 0	33
Grade 2	95
Grade 3	26

8.1.2 Quantification of Expansion-Response

The expansion-response for a given patient was quantified as the mean expansion and corresponding mean delivered dose, to the subvolume of the esophagus comprised of 9 axial slices of the esophagus, centered at the slice location of MaxExp1. Expansion is quantified at the week of maximal expansion as well as at the end of treatment, around fraction 30. This region of analysis was chosen for two reasons. First, the 9 axial slice region is consistent in size for all study patients. Second, the size of analysis region must be large enough to quantify uncertainty in dose, yet sufficiently small enough to localize response.

Delivered dose was quantified as voxel dose at the time of expansion quantification. This can be less than planning dose, especially for analysis examining time of maximum expansion, as maximal expansion occurs many fractions before treatment completion for most patients, and incorporates a temporal element in the response quantification. The combination of expansion value and corresponding delivered dose at the time of maximum expansion is the expansion-response for a given patient.

8.1.3 Radiosensitivity Clustering of Patients

The expansion dose-response quantified at the week of maximal expansion as well as at the end of treatment around fraction 30, were clustered separately using multiple clustering methods including: Gaussian mixture model using expectation-maximization (GMM-EM), Bayesian Gaussian mixture model (Bayesian GMM), and K-Means mixture model (K-Means). All three methods are variations of clustering using Gaussian mixture modelling, which is a process of identifying membership of the data to a finite number of unique clusters, based on the

assumption that the observed data distribution is a collection of multiple Gaussian distributions. These unique underlying Gaussian distributions are representative of the data clusters we seek to identify as indicative of patient radiosensitivity.

The method by which each of the three clustering algorithms derives the Gaussian distributions, and thereby data clusters, are different. K-Means clustering is commonly used technique where squared Euclidean distance is used as a dissimilarity measure.[62,101,102] Minimization of dissimilarity for data points in a given number of clusters is used to find the solution. Gaussian mixture modeling with expectation-maximization is a similar procedure to K-Means clustering.[62] One notable difference between these two methods is that covariance of data and probability of a data point belonging to a particular cluster is used to find a solution (Expectation). After expectation of a given data point is calculated, maximization calculates model parameters based on the means of membership for data in the clusters. This is an iterative process that is repeated until convergence is achieved. Bayesian GMM is similar to the GMM-EM, with the critical difference that in this method, all parameters are considered random variables where the priors are used in the calculation.[103]

The underlying premise in utilizing clustering to identify patient sub groups that are radiosensitive is that a particular cluster must have a proportionally higher expansion per delivered dose than other clusters. Based on the previous assumption, we assume that the 3 following clusters should be observed based on radiosensitivity: the radiosensitive cluster, which has the highest expansion per delivered dose; the radio-insensitive cluster, which has high delivered dose, but proportionally lower expansion than the radiosensitive group; and

third the radio-normal cluster, which has lower expansion and delivered dose than the two other clusters.

Before clustering was calculated, patients with esophageal subvolume doses of less than 20 Gy were excluded from the analysis (n=8). This is because there was insufficient dose to determine what group of radiation sensitivity these patients would belong. All 8 of these excluded patients had grade 0 maximum esophagitis. After clustering the 126 remaining patients using the three different techniques, the radiosensitive patient cluster was identified and then used in the NTCP modelling process.

8.1.4 LASSO NTCP Modelling and Radiosensitivity

In chapters 4 and 7, LASSO penalized logistic regression was utilized as a multivariate analysis method. In this chapter, we similarly used LASSO logistic regression to examine if the radiosensitivity cluster membership substantially improves esophagitis prediction modelling. First, LASSO NTCP models were constructed with the 126 patients that were clustered in a repeated cross-validation procedure, for 1000 iterations, which was previously described in chapters 4 and 7. To reiterate, predictor variables in the form of dosimetric and clinical factors were used as covariates to create NTCP models for ≥ grade 3 esophagitis, according to CTCAE version 3.0.

The same dosimetric and clinical factors described in chapters 4 and 7 were utilized as covariates in the models reported in this chapter (Table 8.2). Models were trained and tested on separate data in each iteration of the cross-validation procedure. Each iteration used randomly drawn training and test sets comprised of 75% and 25% of the study patient population,

respectively. Model performance was quantified with Brier scores and AUC from ROC analysis on both the training (fit performance) and test (prediction performance) data sets. The recurrence of model features was quantified by recording variables in each model, for every iteration of the cross-validation procedure.

Table 8.2: *Predictor variables used in the NTCP model construction process.*

Predictor Index	Predictor	Predictor Index	Predictor
1	Smoking Status	28	$LE60_{25\%}$
2	Induction Chemotherapy	29	$LE50_{25\%}$
3	GTV	30	$LE40_{25\%}$
4	Histology-other	31	$LE30_{25\%}$
5	Histology-Large Cell	32	$LE20_{25\%}$
6	Histology-Adenocarcinoma	33	$LE10_{25\%}$
7	Histology-Squamous Cell	34	V70
8	Nodal Involvement	35	V65
9	Stage-IV	36	V60
10	Stage-IIIB	37	V55
11	Stage-IIIA	38	V50
12	Stage-IIB	39	V45
13	Stage-IIA	40	V40
14	Tumor Location-Left Lateral	41	V35
15	Tumor Location-Right Lateral	42	V30
16	Tumor Location-Left Medial	43	V25
17	Tumor Location-Right Medial	44	V20
18	Tumor Location-Left Upper	45	V15
19	Tumor Location-Right Upper	46	V10
20	Gender	47	Mean Esophagus Dose
21	Age	48	Max Esophagus Dose
22	$LE60_{100\%}$	49	Prescription Dose
23	$LE50_{100\%}$	50	Radiosensitivity Tag
24	$LE40_{100\%}$		
25	$LE30_{100\%}$		
26	$LE20_{100\%}$		
27	$LE10_{100\%}$		

The model construction procedure was then repeated with radiosensitivity as an additional covariate. The previously described clustering technique was used to identify patients that had proportionally higher expansion-response then other study patients, and this information was quantified as a dichotomous variable (1 for radiosensitive patient, 0 otherwise), in the LASSO NTCP model construction process. Model performance was assessed and recurring model predictors were cataloged for every iteration of the model construction process. The results of both model construction scenarios (with and without the radiosensitivity predictor) were compared.

8.1.5 Analysis of IMRT versus Protons for Radiosensitive Patients

As will be shown in the results section, clustering of expansion-response can identify radiosensitive patients. This information was used to reanalyze patients that were radiosensitive to determine if radiation modality affects expansion-response for this subset of patients. As will be shown later in the result section, Bayesian GMM based clustering leads to the highest predictive performance in multivariate modelling, and thus will be the cluster method utilized for this modality analysis. Expansion-response was analyzed at the end of treatment, to keep time of analysis approximately constant.

8.2 Results

8.2.1 Expansion-Response

The expansion-response for a given patient's week of maximal expansion, for all 134 study patients, is shown in Figure 8.1. The expansion-response towards the end of treatment is shown in Figure 8.2. For both time points, an overall trend of increasing dose and toxicity severity is observed, but this has high patient-to-patient variability. Additionally, the expansion per delivered dose is quite variable. The distribution of expansion in 10 Gy dose partitions, from 20 Gy up to 70 Gy, is shown in Figure 8.3, for both time points of expansion-response. Here a high variance of expansion is observed for patients with similar doses. The standard deviation of expansion in a given dose partition is also shown, with a standard deviation of expansion value of 30% being typical.

The 8 pateints excluded from the clustering and NTCP analyses can be observed in Figures 8.1 and 8.2 as the patients with mean subvolume doses under 20 Gy. The resultant clustering from each of the three methods at the two different expansion-response time points is shown in Figure 8.4. The radiosensitive cluster is red, the radio-normal cluster is blue, and the radio-insensitive cluster is black, in each of the three clustering methods. For all three methods, the assigned clusters' radiation sensitivity met the necessary assumptions of expansion-reponse described in the methods section.

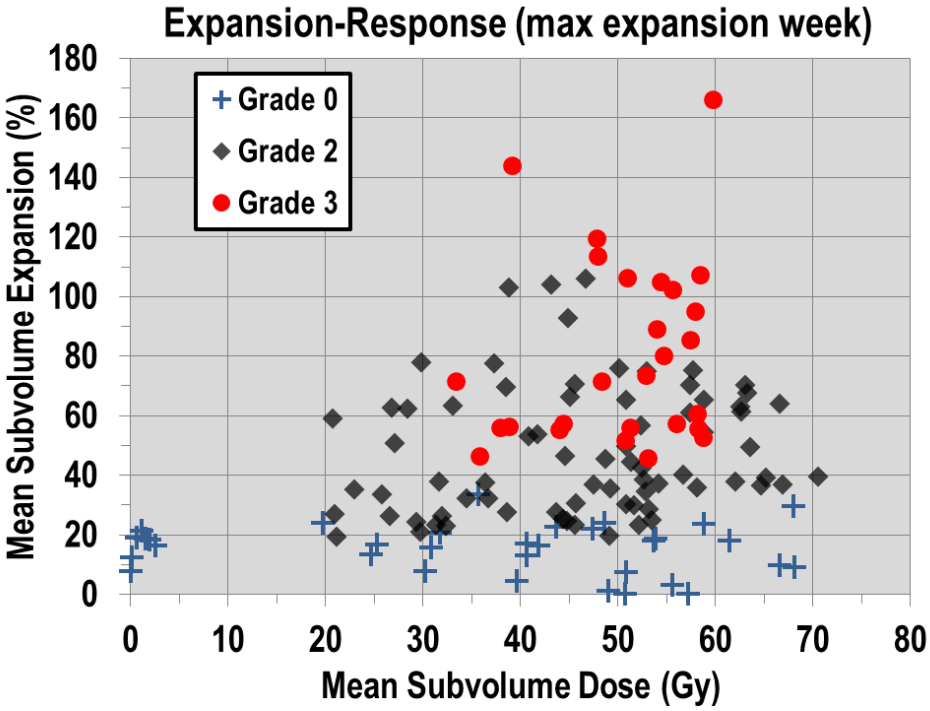

Figure 8.1: *Plot of expansion-response at patient's week of maximal expansion in the analyzed subvolume of the esophagus for 134 study patients. Patient markers denoted maximum esophagitis grade during treatment.*

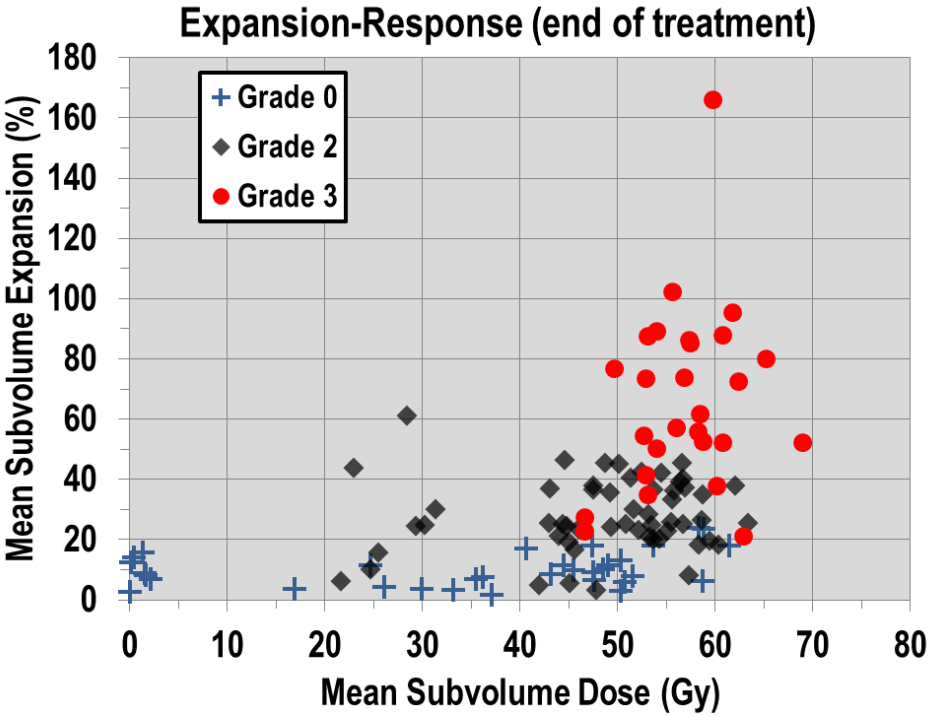

Figure 8.2: *Plot of expansion-response at the end of radiation therapy (approximately fraction 30) in the analyzed subvolume of the esophagus for 134 study patients. Patient markers denoted maximum esophagitis grade during treatment.*

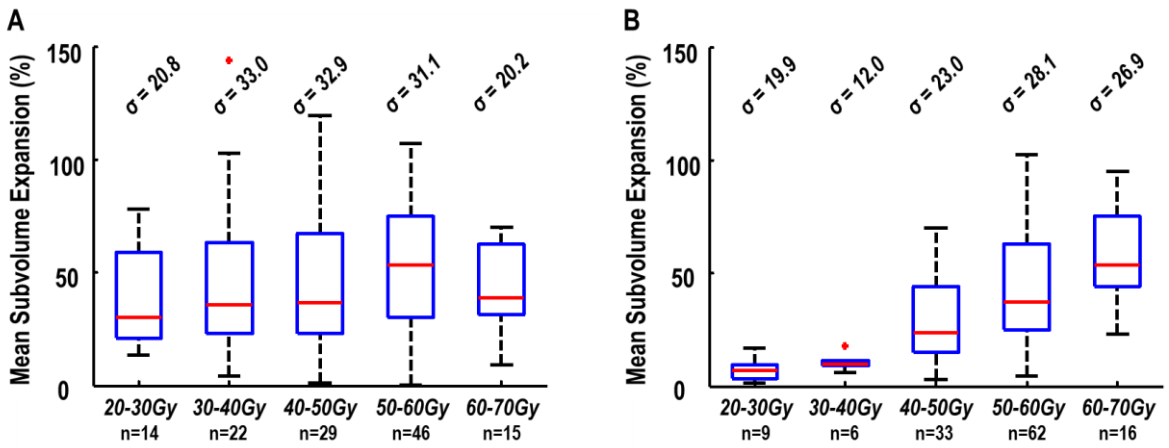

Figure 8.3: *Boxplot of the distribution of mean subvolume esophageal expansion grouped according to mean subvolume doses of 20 to 30 Gy, 30 to 40 Gy, 50 to 60 Gy, and 60 to 70 Gy, for 126 study patients, for expansion-response quantified at the patient's week of maximal expansion (A) and at treatment fraction 30 (B). The standard deviation of expansion in each dose group is shown above each box. The edges of the box represent the quartile values of expansion, with the red line within each box representing that groups median expansion value. The range of values is represented by the black whiskers and the red '+' denotes outliers (values beyond 1.5 times the interquartile range from the edge of the box).*

8.2.1 Radiosensitivity Clustering

Distributions of cluster membership, as well as esophagitis grade within each cluster is given in Table 8.3 for expansion-response quantified for the week of maximal expansion, and in Table 8.4 for expansion-response quantified at the end of radiation therapy. For all three clustering methods, no grade 0 patients were found in the radiosensitive cluster, but most grade 3 patients were. However, many grade 2 patients were also found in the radiosensitive cluster. GMM-EM and K-Means showed the most similarity in patient cluster membership, with Bayesian GMM showing consistency with the other two methods. The radio-insensitive (black) cluster contained the most patients regardless of clustering method, and all 3 esophagitis grade endpoints were observed within this cluster.

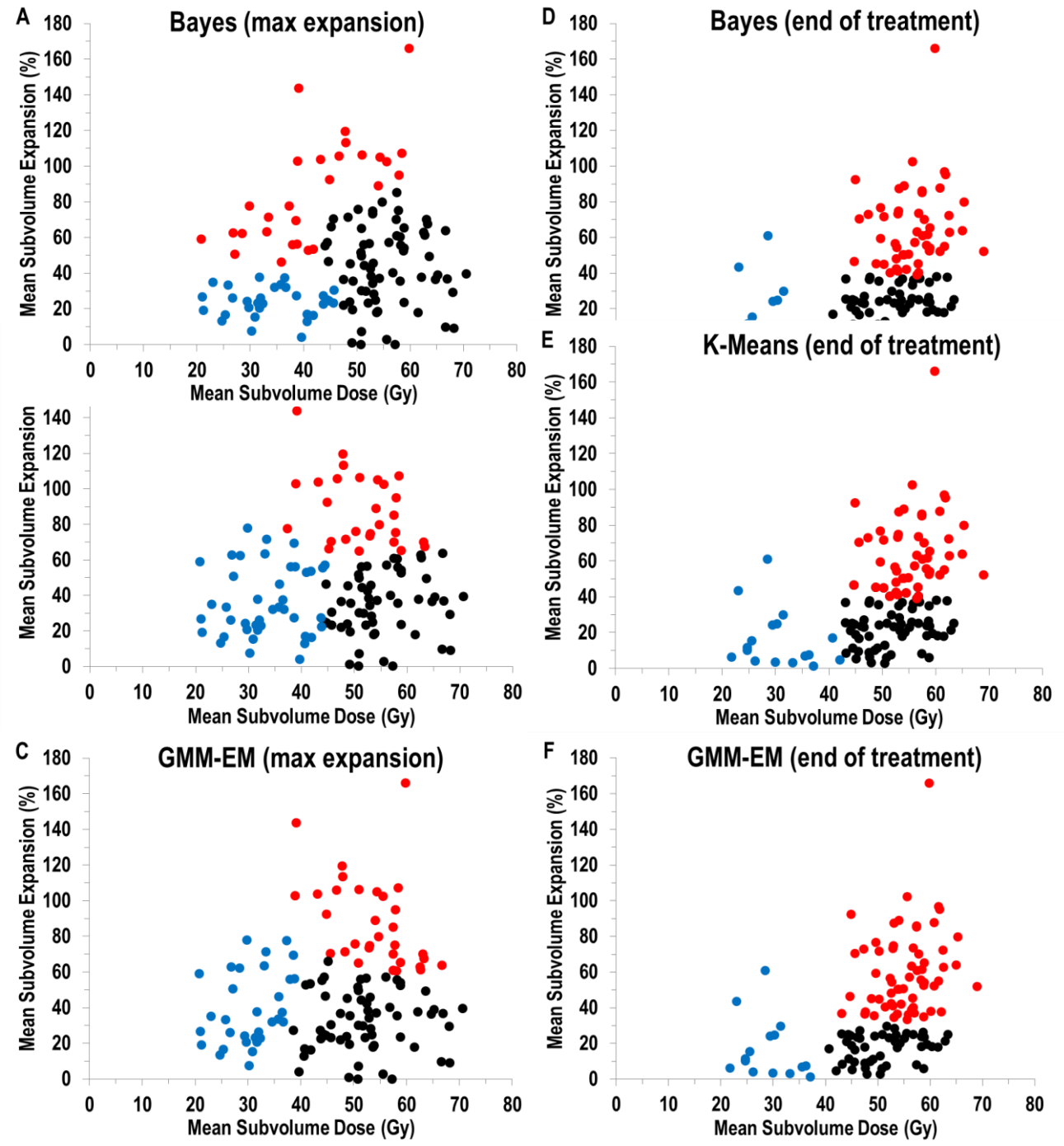

Figure 8.4: *Plots of patient clustering of expansion-response measured at the week of a patient's maximal expansion by Bayesian Gaussian mixture model (A), for cluster size of 3, K-Means clustering (B) for cluster size of 3, and Gaussian mixture model using expectation-maximization clustering (C) for cluster size of 3, and expansion-response measured near treatment fraction 30 by Bayesian Gaussian mixture model (D), for cluster size of 3, K-Means clustering (E) for cluster size of 3, and Gaussian mixture model using expectation-maximization clustering (F) for cluster size of 3, for 126 patients in all scenarios. Cluster membership is denoted by color of expansion-response data point for a given patient.*

Table 8.3: *Characteristics of clustering membership for Bayesian GMM, K-Means, and GMM-EM clustering methods for expansion-response at the week of maximal expansion for the 126 patients analyzed in the cluster analysis.*

Bayesian GMM					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	14	14	28	22.2
Blue	11	20	0	31	24.6
Black	14	41	12	67	53.2
n =	25	75	26	126	100.0
K-Means					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	16	13	29	23.0
Blue	11	25	6	42	33.3
Black	14	34	7	55	43.7
n =	25	75	26	126	100.0
GMM-EM					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	18	14	32	25.4
Blue	6	22	4	32	25.4
Black	19	35	8	62	49.2
n =	25	75	26	126	100.0

8.2.3 LASSO NTCP Modelling and Modality

The LASSO NTCP model construction procedure had similar distributions of recurring model predictors, even for NTCP models not using radiosensitivity as a predictor variable. These results are shown in Figures 8.5 and 8.6, for expansion-response at the week of maximal expansion, and in Figures 8.7 and 8.8, for expansion-response calculated at the end of radiation therapy. For models constructed with the radiosensitivity variable, this predictor was the most recurring variable and was chosen in over 99% of the 1000 iterations of model construction. Mean esophageal dose was the second most recurring predictor in the radiosensitivity information inclusive models, as well as the most recurring predictor in the models not including the radiosensitivity information.

Table 8.4: *Characteristics of clustering membership for Bayesian GMM, K-Means, and GMM-EM clustering methods for expansion-response at the end of radiation therapy for the 126 patients analyzed in the cluster analysis.*

Bayesian GMM					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	29	21	50	39.7
Blue	7	8	0	15	11.9
Black	18	38	5	61	48.4
n =	25	75	26	126	100.0

K-Means					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	19	20	39	31.0
Blue	8	7	0	18	14.3
Black	17	46	6	69	53.8
n =	25	75	26	126	100.0

GMM-EM					
Cluster	**Grade 0**	**Grade 2**	**Grade 3**	**n =**	**% of Cluster**
Red	0	37	23	60	47.6
Blue	7	8	0	15	11.9
Black	18	30	3	51	40.5
n =	25	75	26	126	100.0

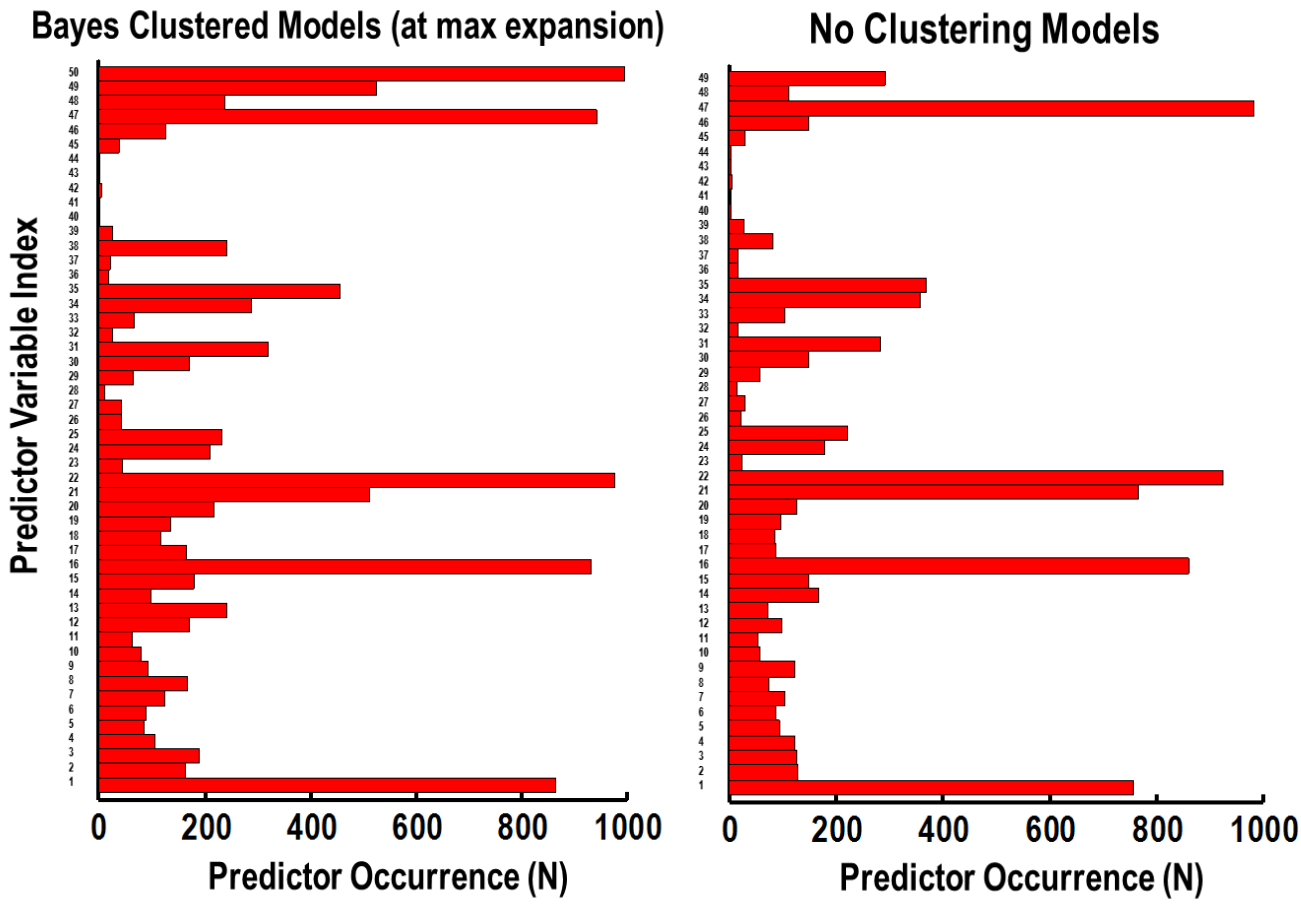

Figure 8.5: *Bar charts of the occurrence of predictors for all 1000 iterations of the LASSO logistic regression NTCP modelling construction process. Models using the radiosensitive tag variable from Bayesian GMM clustering of expansion-response at the week of maximal expansion are shown in the left bar chart, and models created without the radiosensitivity predictor is shown in the right bar chart. The predictor index number identifies the specific predictor variable from Table 8.2.*

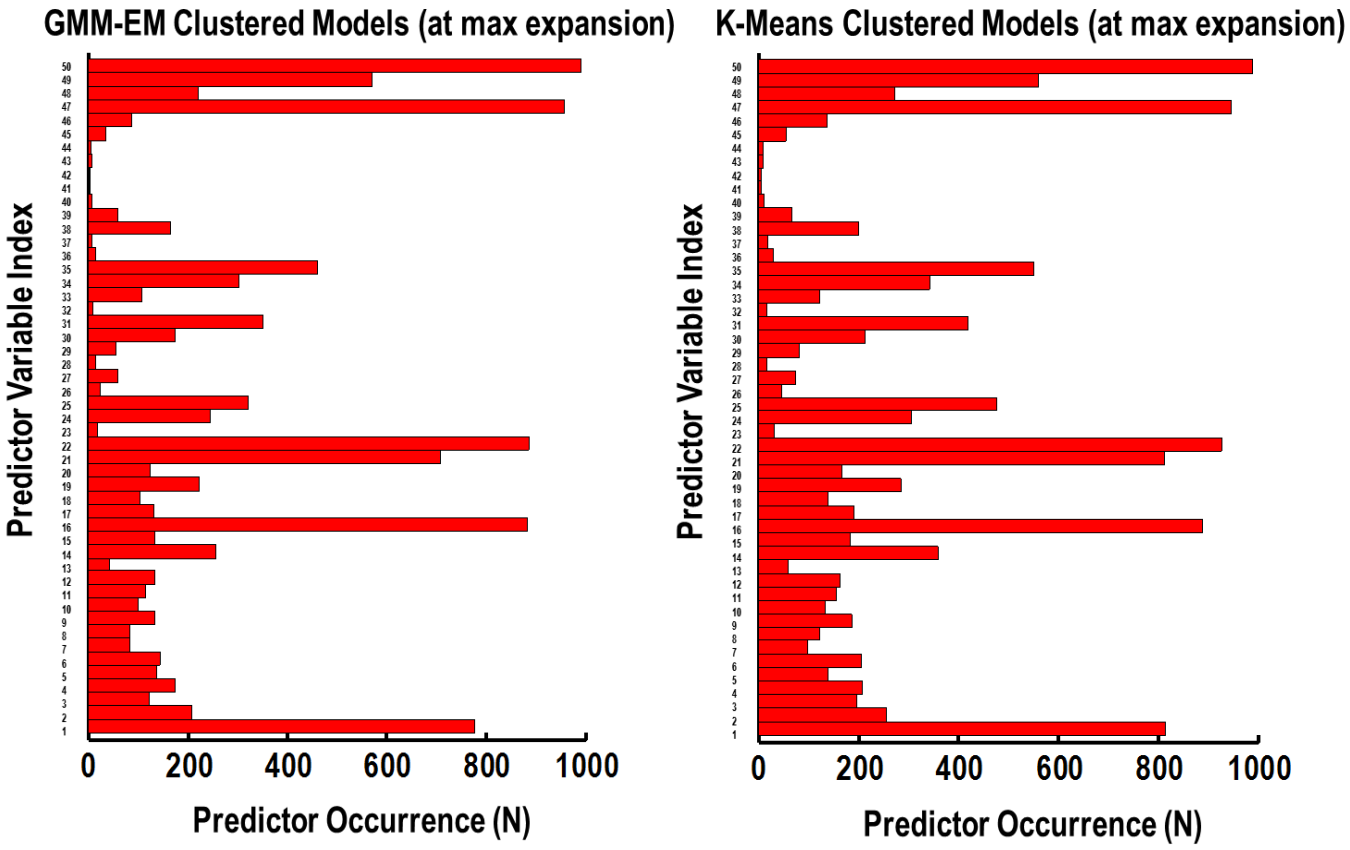

Figure 8.6: *Bar charts of the occurrence of predictors for all 1000 iterations of the LASSO logistic regression NTCP modelling construction process. Models using the radiosensitive tag variable calculated using expansion-response at the week of maximal expansion and from GMM-EM clustering are shown in the left bar chart, and from K-Means clustering are shown in the right bar chart, respectively. The predictor index number identifies the specific predictor variable from Table 8.2.*

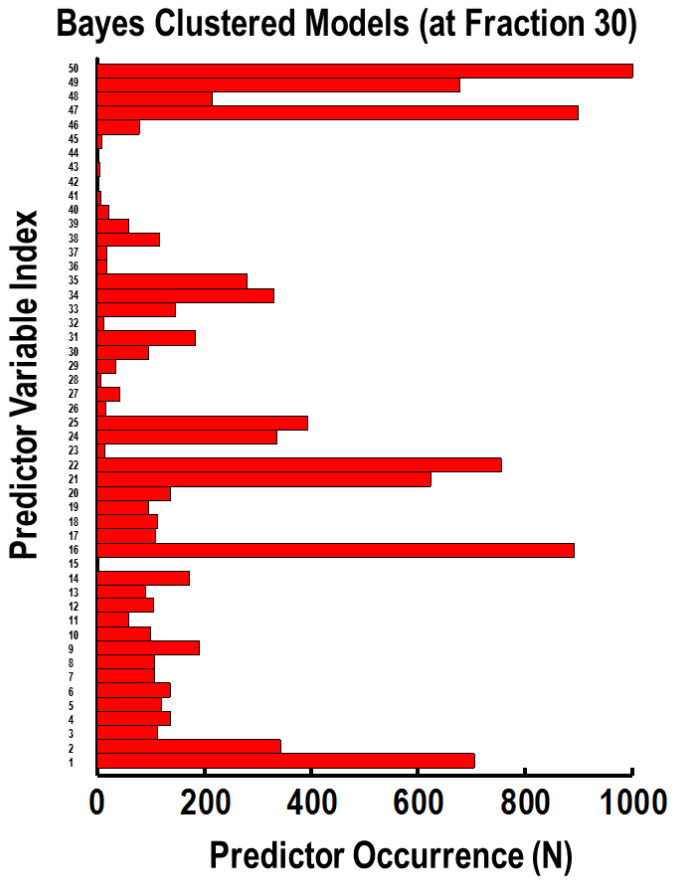

Figure 8.7: *Bar charts of the occurrence of predictors for all 1000 iterations of the LASSO logistic regression NTCP modelling construction process. Models using the radiosensitive tag variable from Bayesian GMM clustering of expansion-response at the end of radiation therapy are shown in the left bar chart. The predictor index number identifies the specific predictor variable from Table 8.2.*

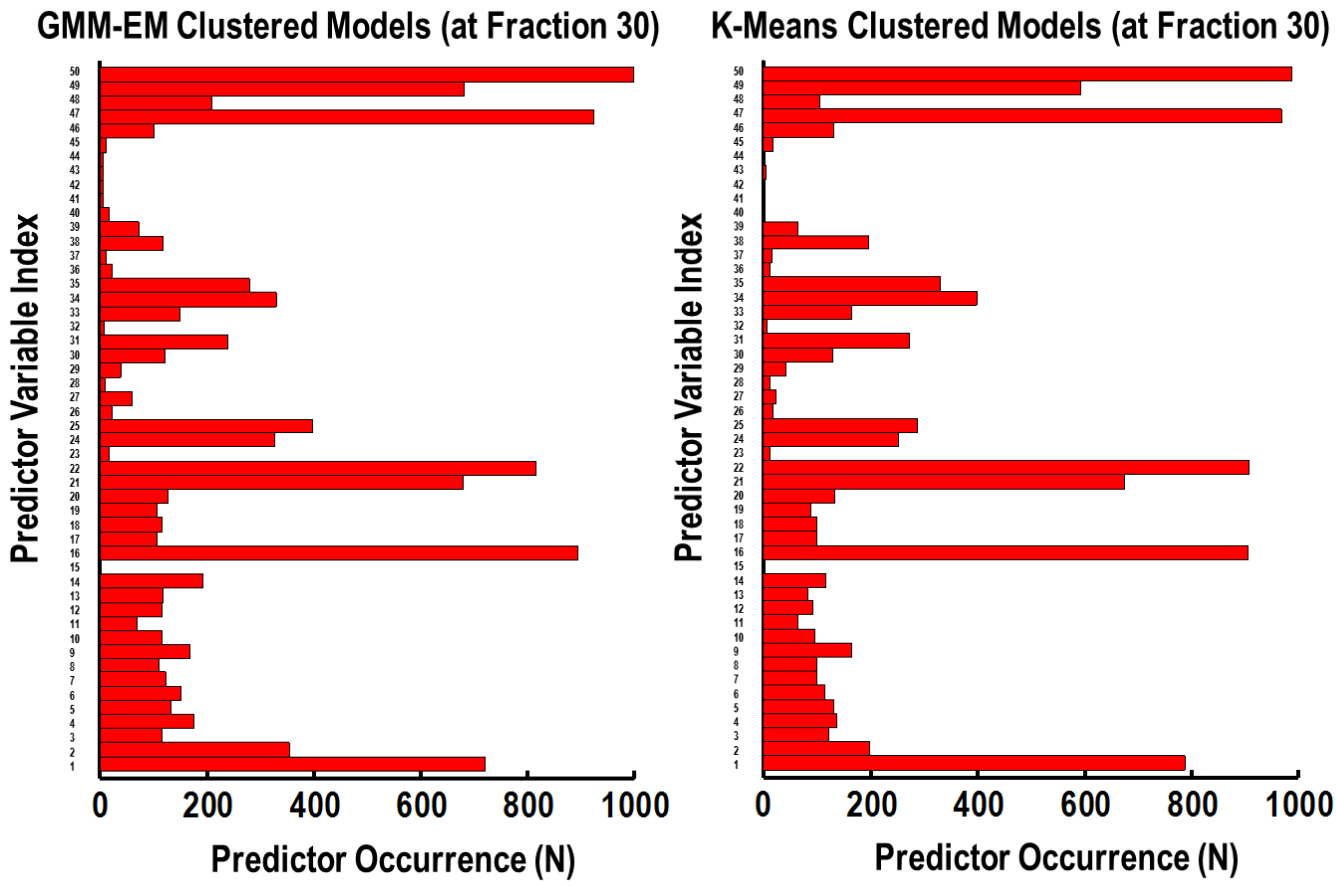

Figure 8.8: *Bar charts of the occurrence of predictors for all 1000 iterations of the LASSO logistic regression NTCP modelling construction process. Models using the radiosensitive tag variable calculated using expansion-response at the end of radiation therapy and from GMM-EM clustering are shown in the left bar chart, and from K-Means clustering are shown in the right bar chart, respectively. The predictor index number identifies the specific predictor variable from Table 8.2.*

The NTCP model performance for all four investigated scenarios is summarized for expansion-response at the week of maximal expansion in Table 8.5, as well as the expansion-response at the end of radiation therapy in Table 8.6. All three model types using clustering to identify radiosensitive patients outperform modelling without radiosensitivity information (first row of Table 8.5). The training and predictive performance of models using clustering have significantly higher $AUC_{Training}$ and AUC_{Test} (according to a paired T-test on AUC values for each corresponding iteration of cross-validation, $p<0.05$). Both types of Brier scores were optimal (lower value) for the three clustering/radiosensitivity model types compared to no radiosensitivity information models. The models using the Bayesian GMM clustering had the highest performance of any of the four model types.

Table 8.5: *Results of the LASSO logistic regression NTCP model construction process using multiple clustering methods to identify radiosensitive patients, from expansion-response quantified at the week of maximal expansion, for a total of 126 study patients. The highest recurring predictors from all 1000 iterations of the model construction process are listed from highest to lowest recurring. Standard deviation of AUC values are listed in parentheses.*

Model	$AUC_{Training}$ (S.D.)	AUC_{Test} (S.D.)	Brier Score	Scaled Brier (%)
No Clustering	0.842 (±0.065)	0.693 (±0.099)	0.175 (±0.020)	18.2 (±12.0)
Bayes Clustering	0.903 (±0.043)	0.790 (±0.090)	0.141 (±0.021)	7.8 (±11.2)
GMM-EM Clustering	0.893 (±0.051)	0.756 (±0.092)	0.150 (±0.021)	13.0 (±12.4)
K-Means Clustering	0.907 (±0.055)	0.763 (±0.094)	0.148 (±0.022)	14.3 (±12.9)

Model	Top Recurring Predictors
No Clustering	MED, $LE50Gy_{100\%}$, Left Medial, $LE60Gy_{100\%}$, Smoking Status
Bayes Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Left Medial, Smoking Status
GMM-EM Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Smoking Status, Left Medial, Age
K-Means Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Smoking Status, Left Medial, Age

Table 8.6: *Results of the LASSO logistic regression NTCP model construction process using multiple clustering methods to identify radiosensitive patients, from expansion-response quantified at the end of radiation therapy, for a total of 126 study patients. The highest recurring predictors from all 1000 iterations of the model construction process are listed from highest to lowest recurring. Standard deviation of AUC values are listed in parentheses.*

Model	$AUC_{Training}$ (S.D.)	$AUC_{Prediction}$ (S.D.)	Brier Score	Scaled Brier (%)
Bayes Clustering	0.906 (±0.046)	0.792 (±0.090)	0.140 (±0.020)	14.6 (±11.9)
GMM-EM Clustering	0.900 (±0.048)	0.773 (±0.089)	0.144 (±0.020)	15.3 (±11.7)
K-Means Clustering	0.885 (±0.053)	0.753 (±0.094)	0.151 (±0.019)	12.1 (±11.2)

Model	Top Recurring Predictors
Bayes Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Left Medial, Smoking Status
GMM-EM Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Smoking Status, Left Medial, Age
K-Means Clustering	RS Tag, MED, $LE60Gy_{100\%}$, Smoking Status, Left Medial, Age

8.2.4 Modality and Radiosensitivity

Patients clustered into the radiosensitive group using the Bayesian GMM did not show a significant difference in expansion-response. Figure 8.9a shows the expansion-response of these patients and their corresponding modality. A boxplot of the distribution of subvolume expansion is shown in Figure 8.9b. Kruskal-Wallis ANOVA determined a p-value of 0.09, which is not statistically significant ($p<0.05$).

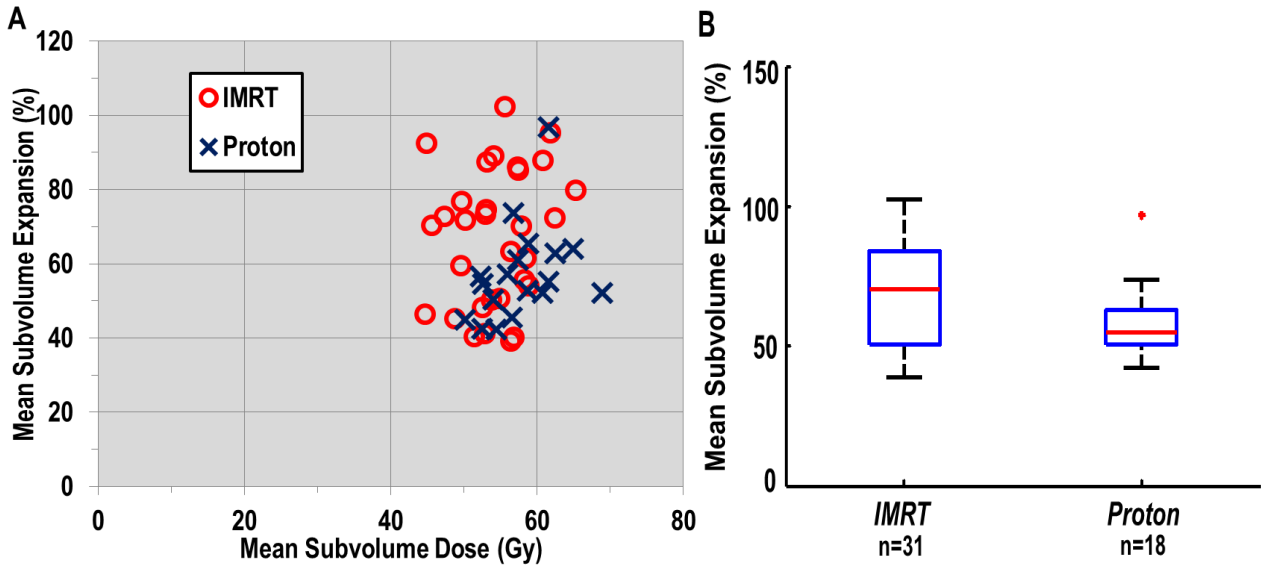

Figure 8.9: *(A) Expansion-response at the end of radiation therapy for radiosensitive clustered patients according to Bayesian GMM and grouped according to modality. (B) Boxplot of subvolume expansion for patients in (A).*

8.3 Chapter Discussion

In this chapter, three different clustering methods (K-Means, Gaussian mixture model using expectation-maximization, and Gaussian mixture model using Bayesian methods) were utilized to identify patients' inherent radiosensitivity from their respective expansion-response. The full expansion patient dataset which was utilized in chapter 7 is used in this chapter. Clustering was carried out for expansion-response at the week of a given patient's maximal expansion, as well as expansion-response towards the end of treatment with the expansion quantification closest to fraction 30. This information was then used to label patients as either radiosensitive or not radiosensitive. This label was then converted to a dichotomous variable and used in the NTCP modelling process in an attempt to improve esophagitis prediction models. In a similar analysis to chapter 4 and chapter 7, LASSO logistic regression was utilized

in a repeated cross-validation procedure to test model performance when including this radiosensitive variable in the model construction process.

The expansion-response of these patients was highly variable for both time points of expansion-response quantification. For similar subvolume doses, many patients had vastly different amounts of expansion, as well as esophagitis grade. This shows the pitfall of toxicity prediction modelling without accounting for inherent radiation sensitivity, where variability of patients' response outweighs the study population's average observed response. Additionally, in chapters 6 and 7 we did not observe expansion being influenced from either dose-geometry or modality, respectively. The variability of response for patients with similar delivered dose may render detecting such effects arduous if patient radiosensitivity is not considered. A reanalysis of the question of modality influencing expansion was carried out for the patients within the radiosensitive cluster. This analysis confirmed our previous findings that radiation modality, in the form of IMRT and proton therapy, does not influence expansion response.

The results of the different clustering methods were consistent between the three techniques. Radiosensitive clusters were observed in all three methods that met the assumed trend of proportionally higher response for a lower delivered dose, compared to the other two cluster types. The other two clusters met their assumed trends, with a low response with a corresponding low delivered dose for the radio-normal cluster type, and a lower response with a corresponding higher dose, compared to the radiosensitive cluster, for the radio-insensitive cluster type. These were consistent regardless of clustering method or for time points of expansion-response quantification.

NTCP models using any of the three clustering methods to create radiosensitivity predictor variables outperformed NTCP models not utilizing radiosensitivity information, for a grade 3 maximum esophagitis endpoint. The Bayesian GMM-based radiosensitivity models had the highest overall performance out of the four modelling scenarios. GMM-EM and K-Means radiosensitivity derived models still outperformed models without radiosensitivity information. The performance metrics are even more impressive, as 8 low-dose, low-response, and asymptomatic patients were excluded from the model construction process. In typical modelling situations, these types of patients are easily classified by the model and contribute to higher model performance, which would be reflected in any quantification of predictive ability. By not including these patients, the modelling situation is more difficult to classify esophagitis, and this translates into a more robust model.

Model performance for the expansion-response at the end of radiation therapy was similar to modelling using the expansion-response at the week of maximal expansion. This is a very impactful result, as simply quantifying expansion-response towards the end of treatment allows for direct application in outcome assessment. This lends to the potential for expansion-response quantification at the end of therapy, in conjunction with the clustering methodology, to be used as a framework to objectively assess outcomes and quantify variability of inherent radiation-response within a study cohort. More on this will be described in the project discussion in chapter 9.

This work was not without limitations. The clustering process is unsupervised in terms of esophagitis outcome, and therefore requires some assumptions for interpretation. As described in the methods section, the cluster assignment of radiosensitivity was determined based on the

assumptions of relative expansion-response within the study population. It is vital to validate these findings on an external dataset, as it would be interesting to see if cluster assignment and shape would change with new patient data. Another limitation was that the radiosensitivity information was only used dichotomously (radiosensitive or not radiosensitive). It would be of interest to analyze the utility of not just the radiosensitive clusters, but also patients labelled as radio-normal and radio-insensitive. The radio-insensitive cluster in particular would be of interest in dose-escalation studies.

In conclusion, clustering techniques can be applied to the expansion-response mechanism to determine patient radiosensitivity in the esophagus. This radiosensitivity information can be used in the NTCP modelling process to improve toxicity prediction performance. Patient inherent radiosensitivity can be assessed towards the end of radiation therapy. These results should be verified on an external dataset.

Chapter 9

Discussion

9.1 Summary of Findings

In this dissertation, novel imaging biomarkers of physiological radiation-response were derived, validated, and applied to advance the knowledge of dose-response and radiation injury in the esophagus. The previous paradigm for understanding dose-response in the esophagus, in our context of a side effect from the treatment of lung malignancy, can be summarized as delivering radiation for the treatment of non-small-cell lung cancer, which induces normal tissue toxicity quantified by subjective physician grading criteria. The subjectivity in quantifying response, together with the lack of ability to localize response to precise sub-regions of the esophagus, served as the impetus for this PhD work. A summary of the specific aims, hypotheses, projects, and results, can be found in Table 9.1.

The work in chapter 3 showed that esophageal expansion can be calculated from planning and corresponding intra-treatment 4DCT scans to quantify relative, localized volume change in the esophagus. Metrics of esophageal expansion were derived as objective imaging biomarkers of radiation-response in the esophagus. We then showed these imaging biomarkers were highly correlated to esophageal toxicity. It was also shown in this chapter that expansion peaks in the last few weeks of treatment, indicating optimal assessment of the expansion radiation-response towards the conclusion of radiation therapy.

Table 9.1: *Summary of dissertation specific aims, hypotheses, and results from individual projects.*

Specific Aim	Hypothesis	Projects	Results
SA1: Analysis of CT-based Esophageal Expansion to Quantify Radiation Response in the esophagus during radiation therapy	CT-based esophageal expansion is a biomarker of radiation response in the esophagus and can be used to improve outcome modelling of radiation-induced esophagitis	Project 1.1	Expansion is a radiation response biomarker and highly correlated to toxicity
		Project 1.2	NTCP models with expansion biomarker as endpoint outperform traditional grade endpoint NTCP models
SA2: Analysis of FDG-PET to Quantify Esophageal Radiation Response in the esophagus during radiation therapy	FDG-PET uptake is a biomarker of radiation response in the esophagus and can be used to predict symptom progression during radiation therapy	Project 2.1	Normalized uptake is a radiation response biomarker and highly correlated to toxicity
		Project 2.2	Normalized uptake can predict toxicity progression during radiation therapy
SA3: Analysis of Esophageal Dose-Response Using Radiation Response Biomarkers	Esophageal expansion will identify if dose-geometry or radiation type contribute to radiation injury in the esophagus, and that expansion can be used to quantify patient-specific radiosensitivity	Project 3.3	Dose-geometry does not have a detectable influence on expansion
		Project 3.2	IMRT and Proton therapy do not have substantial differences in expansion
		Project 3.3	Expansion can identify radiosensitive patients and this knowledge can improve toxicity prediction models

Next, the two expansion-based biomarkers most highly correlated to maximum treatment toxicity, MaxExp1 and LenExp30%, were shown to be robust endpoints in the toxicity prediction modelling process (chapter 4). Three different modelling construction techniques (forward, stepwise logistic regression; LASSO penalized logistic regression; and Random Forests classification) all showed a higher predictive performance for models created with either of the two expansion-based biomarker endpoints, compared to models created with esophagitis grade as the model endpoint. Furthermore, the repeated cross-validation procedure effectively created 1000 individual prediction models, though created from a similar distribution of

samples, and showed the robustness of this result was not due to random partition of the model training data.

The other biomarker type, based on normalized FDG uptake, was studied in chapter 5. FDG uptake from a single mid-treatment scan was normalized to the low dose region of an individual patient's esophagus, providing a patient-specific, radiation-response quantification. Metrics of normalized uptake were derived in a similar fashion as the expansion-based metrics, with many shown to be highly correlated to esophageal toxicity. Interestingly, the two highest toxicity-correlated FDG-based biomarkers, $nSUV_{AxMax1}$ and $nSUV_{Len40\%}$, were similar in form to the highest performing expansion-based biomarkers (MaxExp1 and LenExp30%). This is an intuitive finding, as swelling and inflammation are related physiological processes.

FDG uptake was also shown to be predictive of toxicity progression for patients asymptomatic at the time of the FDG-PET scan. The magnitude of the FDG-based biomarker could predict which patients would develop esophagitis. This was not information obtainable from the radiation dose, indicating unique and clinically beneficial information in the FDG-PET study for the purpose of understanding radiation-response in the esophagus.

Chapter 6 was the first of three chapters addressing specific aim 3, and examined whether or not dose-geometry influences expansion radiation-response. The primary interest of this project was to investigate if particular dose-geometries, particularly reducing cross-sectional coverage of dose across the axial plane of the esophagus, allows for partial sparing and a reduction in response of the esophagus. Utilizing the expansion biomarker allowed precise quantification of localized dose-response, not previously possible. It was shown that

there was no clear trend between the coverage of dose across the cross-sectional area of a given slice of the esophagus and expansion-response.

Next, patients treated with either IMRT or Proton therapy were analyzed to determine if modality had a strong influence on esophageal radiation-response (chapter 7). FDG-based biomarkers did not show any significant difference in biomarker value when grouped by treatment modality and toxicity outcome. Analysis of biomarker magnitude for comparable equivalent uniform dose did not show a dependence on radiation therapy modality. Analysis of expansion-based biomarker grouped according to treatment modality and esophagitis grade outcome did show a lower response for patients treated with proton therapy compared to IMRT, for patients who had grade 3 esophagitis. However, removing the outliers in the IMRT group then showed no significant difference between response and modality. Furthermore, the proton patients that become grade 3 esophagitis during treatment would still be classified as high-response patients when using the MaxExp1 ≥ 50% threshold criterion. Analysis of expansion biomarker value as a function of equivalent uniform dose did not show a difference between treatment modalities. When examined using multivariate modelling, modality was not influential in either Random Forests regression of MaxExp1 value, or LASSO logistic regression classification of the ≥ grade 3 esophagitis outcome, despite strong prediction performance for models in both scenarios.

In chapter 8, it was shown that esophageal expansion can quantify patient radiosensitivity, and this result improved esophagitis prediction modelling. Clustering patients based on either their maximum expansion and corresponding delivered dose to the sub-volume of the esophagus with the highest response (expansion), or expansion-response and

corresponding dose quantified at the end of radiation therapy, showed patients with similar doses can have markedly different response, as well as toxicity. This highlights a paramount concern that previous dose-response studies could not overcome: the inherent variability of an individual patient's radiosensitivity. By identifying the patients in the radiosensitive cluster, we were able to substantially improve esophagitis prediction modelling by including this radiosensitivity information in the model. By being able to quantify the radiosensitivity at a single time point towards the completion of therapy lends itself to be an impactful tool for outcome assessment. When put into context with projects 3.1 and 3.2, the results of the radiosensitivity analysis in chapter 8, shows that patient radiosensitivity is a dominant variable in esophagus dose-response. Furthermore, future studies of dose-response in the esophagus need to consider patient variation in radiosensitivity. Otherwise, any resulting trend, association, or lack of either, can be overshadowed by the distribution of patient radiosensitivity within the study population. Additionally, this result of the degree of variability in radiosensitivity leads to many future directions, which will be discussed in the following sections.

9.2 Limitations and Future Work

While overall there are many impactful findings from this work, there exists some limitations. This patient cohort was unique in that patients had weekly 4DCT acquisitions during treatment, which is required to quantify esophageal expansion. Similarly, many patients also had a single mid-treatment FDG-PET/CT scan in the treatment position with immobilization, which is necessary to calculate the FDG-based biomarkers. In both types of imaging, this is not commonplace in radiation oncology clinics.

While this work suggests clinical benefit to acquire either type of scan to quantify expansion or FDG uptake, the logistics and cost are prohibitive in the typical clinical setting. This reality restricts the clinical application of imaging biomarker application to academic/research hospitals until either the additional benefit of these imaging acquisitions is increased or the cost of imaging is reduced. However, image guidance is being utilized at an increasing pace in radiation oncology. In the case of FDG-PET, the imaging dose and cost are even more prohibitive than 4DCT.

Another limitation in this work is the dosimetry, particularly for patients treated with proton therapy. Dose calculations for proton treatments are still an active area of research, with uncertainty in dose at the distal end of the proton beam. This is of particular concern in dose to the normal esophagus, as NSCLC tumor location is typically such that the distal end of the proton beam lies near the esophagus. It would be interesting to note any difference in proton dose calculated using more robust methods, such as Monte Carlo simulations, to the treatment planning software calculated dose in this work. Furthermore, it would be of interest if any presumed differences in dose have an impact on dose-response in the esophagus.

Toxicity prediction modelling was a common application in many stages of this work. However, one limitation is the fact we did not have an external dataset to validate the modelling results. Unfortunately, the lack of suitable external validation datasets is more often than not the reality in many novel toxicity prediction studies. However, cross-validation is a suitable technique to approximate model performance in lieu of an external validation dataset. It would of interest to validate these results as image guidance is utilized in radiation oncology at an ever increasing pace.

Another limitation that is inherent in outcome assessment studies is the fact that we are using retrospective data from patients already affected by the toxicity of interest. However, one very impactful result from this could seek to shift this paradigm, which we describe in the next section.

9.3 Application to Clinical Trials

In the age of personalized medicine, the uniqueness of patients is being quantified to tailor treatment. Furthermore, radiogenomics and radiomics are playing their part in this schema by attempting to identify pre-treatment biomarkers. This is certainly the case with radiosensitivity and its relationship with normal tissue toxicity. Validating pre-treatment biomarkers can be arduous when toxicity grade is the endpoint used for validation. As mentioned throughout this work, esophagitis grade is subjective.

As we have shown in chapter 8, radiosensitivity has a profound impact on radiation-response and toxicity, such that patients of similar dose can have markedly different response. From these facts raises an important question: how can we validate pre-treatment biomarkers, particularly in the case of toxicity given the subjective outcomes associated with grading criteria? The use of either of the in-vivo imaging biomarkers investigated in this work can serve this crucial purpose.

By using expansion and/or FDG-based imaging biomarkers we can potentially validate (or invalidate) pre-treatment biomarkers of interest in a prospective clinical trial. By quantifying expansion and/or FDG-uptake in the esophagus towards completion of radiation therapy, along

with the corresponding esophagus dose, any pre-treatment biomarker of radiosensitivity in the esophagus can potentially be validated, as illustrated in Figure 9.1.

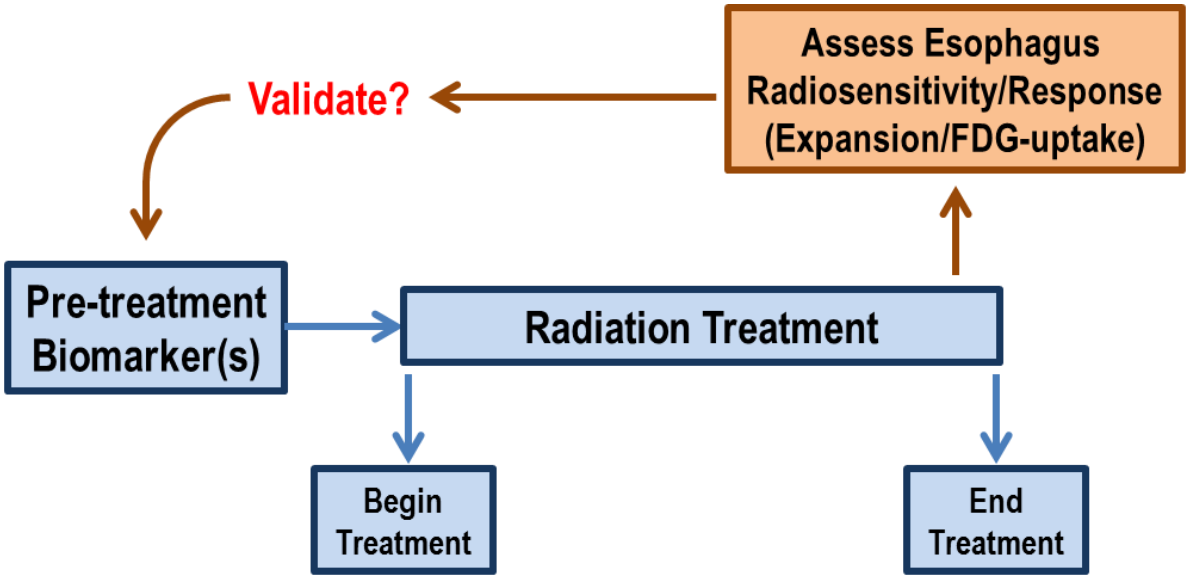

Figure 9.1: *Application of expansion or FDG-based imaging biomarker to validate pre-treatment biomarker.*

While the most logical application of expansion or FDG-based imaging biomarkers is for the validation of pre-treatment radiosensitivity and toxicity biomarkers, this framework could potentially be applied to other scenarios. For example, radioprotectors could also be validated in a similar manner for protection of the esophagus. Another example would be escalation/de-escalation of radiation dose and its potential impact on toxicity. This framework could identify patient-specific response, with consideration to inherent radiation sensitivity. Furthermore, this validation framework could be applied in any scenario that involves comparison of two treatment modalities or techniques, where esophageal response is a concern.

9.4 Conclusions

In conclusion, the individual projects in this work collectively addressed and supported the central hypothesis that esophageal expansion as quantified from 4DCT, and normalized uptake quantified from FDG-PET imaging, are imaging biomarkers of radiation-response in the esophagus that improve the understanding of esophageal radiation injury. The overall objective of this work was to obtain a deeper understanding of dose-response in the esophagus, with the long term goal of increasing treatment outcome and improving patient quality of life. This further understanding of esophageal dose-response included knowing what role, if any, dose-geometry, radiation type, and inherent patient radiosensitivity, can have on response. Probing these specific potential influences on radiation-response was not possible without the use of esophageal expansion or FDG-based biomarkers. The utilization of two different imaging biomarker types as radiation-response metrics, sought to reduce the uncertainty in quantifying response as a symptom grade, as well as provide spatial information of the extent of response in the esophagus.

Chapters 3 and 4 showed expansion is a radiation-response biomarker that can quantify toxicity and improve esophagitis prediction modelling, confirming the hypothesis of Specific Aim 1. Chapter 5 showed FDG-uptake to be an esophageal radiation-response biomarker as well, that can predict symptom progression during radiation therapy, thereby confirming the hypothesis of Specific Aim 2. The influences of dose-geometry and modality on expansion were shown to be non-existent in chapters 6 and 7, respectively. Chapter 8 showed expansion and corresponding dose can quantify patient radiosensitivity, and together with the findings from

chapter 6 and 7 confirm the hypothesis of Specific Aim 3 that expansion can identify what contributes to radiation injury in the esophagus.

Collectively, these findings show the benefit of expansion and FDG-uptake as *in-vivo* radiation-response imaging biomarkers that can be utilized to understand and thereby prevent radiation injury in the esophagus.

Bibliography

1 Bruner DW, Movsas B, Konski A, Roach M, Bondy M, Scarintino C, Scott C, and Curran W. Outcomes research in cancer clinical trial cooperative groups: the RTOG model. *Quality Life Res* **2004** 13(6):1025-1041.

2 Werner-Wasik M, Yorke E, Deasy J, Nam J, and Marks LB. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys* **2010** 76(3):S86-S93.

3 Rose J, Rodrigues G, Yaremko B, Lock M, and D'Souza D. Systematic review of dose-volume parameters in the prediction of esophagitis in thoracic radiotherapy. *Rad Oncol* **2009** 91(3):282-287.

4 Cox JD, Pajak TF, Asbell S, Russell AH, Pederson J, Byhardt RW, Emami B, and Roach M. Interruptions of high-dose radiation therapy decrease long-term survival of favorable patients with unresectable non-small cell carcinoma of the lung: analysis of 1244 cases from 3 radiation therapy oncology group (RTOG) trials. *Int J Radiat Oncol Biol Phys* **1993** 3(27):493-498.

5 Gomez DR, Tucker SL, Martel MK, Mohan R, Balter P, Guerra JL, Komaki R, Cox, JD, and Liao Z. Predictors of high-grade esophagitis after definitive three-dimensional conformal therapy, intensity-modulated radiation therapy, or proton beam therapy for non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* **2012** 84(4):1010-1016.

6 Kwint M, Uyterlinde W, Nijkamp J, Chen C, de Bois J, Sonke, JJ, and Belderbos J. Acute esophagus toxicity in lung cancer patients after intensity modulated radiation therapy and concurrent chemotherapy. *Int J Radiat Oncol Biol Phys* **2012** 84(2):223–228.

7 Werner-Wasik M. Treatment-related esophagitis. *Semin Radiat Oncol* **2005** 32(Suppl 3):60–66.

8 Berkey, FJ. Managing the adverse effects of radiation therapy. *American Family Physician* **2010** 82(4):381–388.

9 Antonadou D, Coliarakis N, Synodinou M, Athanassiou H, Kouveli A, Verigos C, Georgakopoulos G, Panoussaki K, Karageorgis P, and Throuvalas N. Randomized phase III trial of radiation treatment plus/minus amifostine in patients with advanced-stage lung cancer. *Int J Radiat Oncol Biol Phys* **2001** 51:915-922.

[10] Antonadou D. Radiotherapy or chemotherapy followed by radiotherapy with or without amifostine in locally advanced lung cancer. *Semin Radiat Oncol* **2002** 12(Suppl 1):50-58.

[11] Ito H, Meistrich ML, Barkley HT, Thames HD, and Milas L. Protection of acute and late radiation damage of the gastrointestinal tract by WR-2721. *Int J Radiat Oncol Biol Phys* **1986** 12:211-219.

[12] Komaki R, Lee JS, Kaplan B, Allen P, Kelly JF, Liao Z, Stevens CW, Fossella FV, Zinner R, Papadimitrakopoulou V, Khuri F, Glisson B, Pisters K, Kurie J, Herbst R, and Milas L. Randomized phase III study of chemoradiation with or without amifostine for patients with favorable performance status inoperable stage I-III non–small cell lung cancer: Preliminary results. *Semin Radiat Oncol* **2002** 12(1):46-49.

[13] Leong SS, Tan EH, Fong KW, Wilder-Smith E, Ong YK, Tai BC, Chew L, Lim SH, Wee J, Lee KM, Foo KF, Ang P, and Ang PT. Randomized double-blind trial of combined modality treatment with or without amifostine in unresectable stage III non–small cell lung cancer. *J Clin Oncol* **2003** 21:1767-1774.

[14] Werner-Wasik M, Langer C, and Movsas B. Amifostine in chemoradiation therapy for non-small cell lung cancer: review of experience and design of a phase II trial assessing subcutaneous and intravenous bolus administration. *Semin Oncol* **2005** 32:105–108.

[15] Singh AK, Lockett MA, and Bradley JD. Predictors of radiation-induced esophageal toxicity in patients with non-small cell lung cancer treated with three-dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys* **2003** 55:337–341.

[16] Xiao Y, Werner-Wasik M, Michalski D, Houser C, Bednarz G, Curran W, and Galvin J. Comparison of three IMRT inverse planning techniques that allow for partial esophagus sparing in patients receiving thoracic radiation therapy for lung cancer. *Med Dosim* **2004** 29:210–216.

[17] Niedzielski JS, Bluett JB, Williamson RT, Liao Z, Gomez DR, and Court LE. Analysis of esophageal-sparing treatment plans for patients with high-grade esophagitis. *J Appl Clin Med Phys* **2013** (4):163-170.

[18] Al-Halabi H, Paetzold P, Sharp GC, Olsen C, and Willers H. A Contralateral Esophagus-Sparing Technique to Limit Severe Esophagitis Associated with Concurrent High-Dose Radiation and Chemotherapy in Patients with Thoracic Malignancies. *Int J Radiat Oncol Biol Phys* **2014** 92(4):803-810.

[19] Trotti A, Colevas AD, Selser A, Rusch V, Jaques D, Budach V, Langer C, Murphy B, Cumberlin R, Coleman CN, and Rubin P. CTCAE v3.0: Development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* **2003** 13:176–181.

[20] National Cancer Institute. Common Terminology Criteria for Adverse Events v4.0 (NIH publication #09e7473). Bethesda, MD: National Cancer Institute, National Institutes of Health, US Department of Health and Human Services; **2009**.

[21] Xu CJ, van der Schaaf A, Schilstra C, Langendijk JA, and van't Veld AA. Impact of learning methods on the predictive power of multivariate normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* **2012** 82(4):677-684.

[22] Xu CJ, van der Schaaf A, van't Veld AA, Langendijk JA, and Schilstra C. Statistical validation of normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* **2012** 84(4):123–129.

[23] Deasy JO, Bentzen SM, Jackson A, Ten Haken RK, Yorke ED, Constine LS, Sharma A, and Marks LB. Improving normal tissue complication probability models: the need to adopt a 'data-pooling' culture. *Int J Radiat Oncol Biol Phys* **2010** 76:151-154.

[24] Van der Schaaf A, Langendijk JA, Fiorino C, and Rancati T. Embracing phenomenological approaches to normal tissue complication probability modeling: A question of method. *Int J Radiat Oncol Biol Phys* **2015** 91:468-471.

[25] Dieleman EMT, Senan S, Vincent A, Lagerwaard FJ, Slotman BJ, and van Sörnsen de Koste JR. Four-dimensional computed tomographic analysis of esophageal mobility during normal respiration. *Int J Radiat Oncol Biol Phys* **2007** 67(3), 775–80.

[26] Yamashita H, Kida S, Sakumi A, Haga A, Ito S, Onoe T, Okuma K, Ino K, Akahane M, Ohtomo K, and Nakagawa, K. Four-dimensional measurement of the displacement of internal fiducial markers during 320-multislice computed tomography scanning of thoracic esophageal cancer. *Int J Radiat Oncol Biol Phys* **2011** 79(2), 588–95.

[27] Kahn D, Zhou S, Ahn SJ, Hollis D, Yu X, D'Amico TA, Shafman TD, and Marks LB. Anatomically-correct dosimetric parameters may be better predictors for esophageal toxicity than are traditional CT-based metrics. *Int J Radiat Oncol Biol Phys* **2005** 62(3):645-651.

[28] Lopez Guerra JL, Gomez D, Wei Q, Yuan X, Liu Z, Wang LE, Yuan X, Zhuang Y, Komaki R and Liao Z. Association between single nucleotide polymorphisms of the transforming growth factor β1 gene and the risk of severe radiation esophagitis in patients with lung cancer. *Radiother Oncol* **2012** 105(3):299–304.

[29] Kelsey CR, Jackson L, Langdon S, Owzar K, Hubbs J, Vujaskovic Z, Das S, and Marks LB. A polymorphism within the promoter of the TGFβ1 gene is associated with radiation sensitivity using an objective radiologic endpoint. *Int J Radiat Oncol Biol Phys* **2012** 82(2):247–255.

[30] Alsner J, Andreassen CN, and Overgaard J. Genetic Markers for Prediction of Normal Tissue Toxicity After Radiotherapy. *Semin Radiat Oncol* **2008** 18(2):126–135.

[31] Andreassen CN, Alsner J, and Overgaard J. Does variability in normal tissue reactions after radiotherapy have a genetic basis - Where and how to look for it? *Radiother Oncol* **2002** 64(2):131–140.

[32] Belderbos J, Heemsbergen W, Hoogeman M, Pengel K, Rossi M, and Lebesque J. Acute esophageal toxicity in non-small cell lung cancer patients after high dose conformal radiotherapy. *Radiother Oncol* **2005** 75(2):157–164.

[33] Ahn S, Kahn D, Zhou S, Yu X, Hollis D, Shafman TD, and Marks LB. Dosimetric and clinical predictors for radiation-induced esophageal injury. *Int J Radiat Oncol Biol Phys* **2005** 61(2):335–347.

[34] Maguire PD, Sibley GS, Zhou SM, Jamieson TA, Light KL, Antoine PA, Herndon JE, Anscher MS, Marks LB. Clinical and dosimetric predictors of radiation-induced esophageal toxicity. *Int J Radiat Oncol Biol Phys* **1999** 45(1):97–103.

[35] Werner-Wasik M, Pequignot E, Leeper D, Hauck W, and Curran W. Predictors of severe esophagitis include use of concurrent chemotherapy, but not the length of irradiated esophagus: a multivariate analysis of patients with lung cancer treated with nonoperative therapy. *Int J Radiat Oncol Biol Phys* **2000** 48(3):689–696.

[36] Liang Y, Messer K, Rose BS, Lewis JH, Jiang SB, Yashar CM, Mundt AJ, and Mell LK. Impact of bone marrow radiation dose on acute hematologic toxicity in cervical cancer: Principal component analysis on high dimensional data. *Int J Radiat Oncol Biol Phys* **2010** 78(3):912-919.

[37] Deasy JO, Moiseenko V, Marks L, Chao KSC, Nam J, and Eilsbruch A. Radiotherapy Dose-Volume Effects on Salivary Gland Function. *Int J Radiat Oncol Biol Phys* **2010** 76(3):S58-S63.

[38] Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dwason LA, and Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys* **2010** 50(3):695-704.

[39] Mendenhall WM, Mendenhall CM, and Mendenhall NP. Submandibular gland-sparing intensity-modulated radiotherapy. *Am J Clin Oncol.* **2014** 37(5):514-516.

[40] Berkovich GY, Levine MS, and Miller WT. CT findings in patients with esophagitis. *Am J Roentgenol* **2000** 175:1431-1435.

[41] Court LE, Tucker SL, Gomez D, Liao Z, Zhang J, Kry S, Dong L, and Martel MK. A technique to use CT images for in vivo detection and quantification of the spatial distribution of radiation-induced esophagitis. *J Appl Clin Med Phys* **2013** 14:4195.

[42] Kubota K. From tumor biology to clinical PET: A review of positron emission tomography (PET) in oncology. *Annals Nuc Med* **2001** 15(6):471–486.

[43] Mawlawi OR, and Court LE. Four-Dimensional PET-CT in Radiation Oncology. *PET Clin* **2013** 8(1):81–94.

[44] Guerrero T, Johnson V, Hart J, Pan T, Khan M, Luo D, Liao Z, Ajani J, Stevens C, and Komaki R. Radiation pneumonitis: local dose versus [18F]-fluorodeoxyglucose uptake response in irradiated lung. *Int J Radiat Oncol Biol Phys* **2007** 68(4):1030–1035.

[45] De Ruysscher D, Houben A, Aerts H, Dehing C, Wanders R, Ollers M, Dingemans AC, Hochstenbag M, Boersma L, Borger J, Dekker A, and Lambin P. Increased (18)F-deoxyglucose uptake in the lung during the first weeks of radiotherapy is correlated with subsequent Radiation-Induced Lung Toxicity (RILT): a prospective pilot study. *Radiother Oncol* **2009** 91(3):415–420.

[46] Hart JP, McCurdy MR, Ezhil M, Wei W, Khan M, Luo D, Munden RF, Johnson V, and Guerrero TM. Radiation pneumonitis: correlation of toxicity with pulmonary metabolic radiation response. *Int J Radiat Oncol Biol Phys* **2008** 71(4):967–71.

[47] Mac Manus MP, Ding Z, Hogg A, Herschtal A, Binns D, Ball DL, and Hicks RJ. Association between pulmonary uptake of fluorodeoxyglucose detected by positron emission tomography scanning after radiation therapy for non-small-cell lung cancer and radiation pneumonitis. *Int J Radiat Oncol Biol Phys* **2011** 80(5):1365–1371.

[48] McCurdy MR, Castillo R, Martinez J, Najeeb Al Hallack M, Lichter J, Zouain N, and Guerrero T. [18F]-FDG uptake dose-response correlates with radiation pneumonitis in lung cancer patients. *Radiother Oncol* **2012** 104(1):52–57.

[49] Echeverria AE, McCurdy M, Castillo R, Bernard V, Velez Ramos N, Buckley W, Castillo E, Liu P, Martinez J, and Guerrero TM. Proton therapy radiation pneumonitis local dose-response in esophagus cancer patients. *Radiother Oncol* **2013** 106(1):124–129.

[50] Nijkamp J, Rossi M, Lebesque J, Belderbos, J, van den Heuvel M, Kwint M, Uyterlinde W, Vogel W, and Sonke J. Relating acute esophagitis to radiotherapy dose using FDG-PET in concurrent chemo-radiotherapy for locally advanced non-small cell lung cancer. *Radiother Oncol* **2013** 106(1):118–123.

[51] Mesurolle B, Qanadli SD, Merad, Mignon F, Baldeyrou P, Tardivon A, Lacombe P, and Vanel D. Unusual radiologic findings in the thorax after radiation therapy. *Radiographics* **2000** 20(1):67-81.

[52] Wang H, Dong L, Lii MF, Lee AL, de Crevoisier R, Mohan R, Cox JD, Kuban DA, and Cheung R. Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* **2005** 61(3):725-735.

[53] Brock KK. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys* **2010** 76(2):583-596.

[54] Carroll MM. A representation theorem for volume-preserving transformations. *Int J Non-Linear Mech* **2004** 39(2):219-224.

[55] Leow AD, Yanovsky I, Chiang M-C, Lee AD, Klunder AD, Lu A, Becker JT, Davis SW, Toga AW, and Thompson PM. Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE Trans Med Imaging* **2007** 26(6):822-832.

[56] Schraufnagel DE, Michel JC, Sheppard TJ, Saffold PC, Kondos GT. CT of the normal esophagus to define the normal air column and its extent and distribution. *Am J Roentgenol* **2008** 191(3):748-752.

[57] El Naqa I, Bradley J, Blanco AI, Lindsay PE, Vicic M, Hope A, and Deasy JO. Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors. *Int J Radiat Oncol Biol Phys* **2006** 64(4):1275-1286.

[58] Burman C, Kutcher GJ, Emami B, Goitein M. Fitting of normal tissue tolerance data to an analytic function. *Int J Radiat Oncol Biol Phys* **1991** 21(1):123-135.

[59] Jemal A, Siegel R, Xu J, and Ward E. Cancer statistics 2010. *CA Cancer J Clin* **2010** 60:277–300.

[60] Bradley J, Deasy JO, Bentzen S, and el Naqa I. Dosimetric correlates for acute esophagitis in patients treated with radiotherapy for lung carcinoma. *Int J Radiat Oncol Biol Phys* **2004** 58(4):1106–1113.

[61] Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, Bentzen SM, Nam J, and Deasy JO. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys* **2010** 76(3):S10-S19

[62] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference and prediction*. 2nd ed. New York: Springer; **2009**.

[63] Kuhn M, Johnson K. *Applied Predictive Modeling*. 1st ed. New York: Springer; **2013**.

[64] Xu CJ, van der Schaaf A, Schilstra C, Langendijk JA, and Van't Veld AA. Impact of learning methods on the predictive power of multivariate normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* **2012** 82(4):677-684.

[65] Xu CJ, van der Schaaf A, Van't Veld AA, Langendijk JA, Schilstra C. Statistical validation of normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* **2012** 84(4):123–129.

[66] Cella L, Oh JH, Deasy JO, et al. Predicting radiation-induced valvular heart damage. *Acta Oncologica* **2015** 54(10);1796-1804.

[67] Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, Ten Haken RK, and Yorke ED. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys* **2010** 76(3):S3–9.

[68] Jeraj R, Cao Y, Ten Haken RK, Hahn C, and Marks L. Imaging for assessment of radiation-induced normal tissue effects. *Int J Radiat Oncol Biol Phys* **2010** 76(3):S140–144.

[69] Breiman L. Random Forests. *Machine Learning* **2001**;45(1):5-32.

[70] Burman P. A Comparative study of ordinary cross-validation, v-fold cross validation and the repeated learning-testing methods. *Biometrika* **1989** 76(3):503-514.

[71] Steyerberg EW. *Clinical prediction models: A practical approach to development, validation, and updating.* 1st ed. New York: Springer, **2009**.

[72] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* **2010** 33:1-22.

[73] Rohren EM, Turkington TG, Coleman RE. Clinical applications of PET in oncology. *Radiology **2004** 231*(2):305–332.

[74] Blodgett TM, Meltzer CC, Townsend, DW. PET/CT: form and function. *Radiology **2007*** 242(2):360–385.

[75] Yuan S, Brown R, Zhao L, Ten Haken RK, Gross M, Cease KB, Schipper M, Stanton P, Yu J, and Kong FM. Timing and intensity of changes in FDG uptake with symptomatic esophagitis during radiotherapy or chemo-radiotherapy. *Radiat Oncol* **2014** 9:37.

[76] Mehmood Q, Sun A, Becker N, Higgins J, Marshall A, Le LW, Vines DC, McCloskey P, Ford V, Clarke K, Yap M, Bezjak A, and Bissonnette JP. Predicting radiation esophagitis using 18F-FDG PET during chemoradiotherapy for locally advanced non-small-cell lung cancer. *J Thorac Oncol* **2016** 11:213-221.

[77] Chapet O, Kong FM, Lee JS, Hayman JA, and Ten Haken RK. Normal tissue complication probability modeling for acute esophagitis in patients treated with conformal radiation therapy for non-small cell lung cancer. *Radiother Oncol* **2005** 77:176–81.

[78] Zhu J. Zhang Z, Li B, Liu M, Yu J, Luo L, Shu H, and de Crevoisier R. Analysis of acute radiation-induced esophagitis in non-small-cell lung cancer patients using the Lyman NTCP model. Radiother Oncol **2010** 97:449–54.

[79] Holliday EB, Kocak-Uzel E, Feng L, Thaker NG, Blanchard P, Rosenthal DI, Gunn GB, Garden A, and Frank SJ. Dosimetric advantages of intensity-modulated proton therapy for oropharyngeal cancer compared with intensity-modulated radiation: A case-matched control analysis. *Medical Dosimetry* **2016** 41(3), 189–194.

[80] Holliday, EB, Garden AS, Rosenthal DI, Fuller D, Morrison WH, Gunn GB, Phan J, Beadle BM, Zhu XR, Zhang X, Hanna E, Gilsson BS, Hutcheson KA, El-Naggar AK, Hong J, Hung T, Uzel E, Lewis G, and Frank SJ. Proton Therapy Reduces Treatment-Related Toxicities for Patients with Nasopharyngeal Cancer: A Case-Match Control Study of Intensity-Modulated Proton Therapy and Intensity-Modulated Photon Therapy. *International Journal of Particle Therapy* **2015** 2(1):1–10.

[81] Ramaekers BL, Pijls-Johannesma M, Joore MA, van den Ende P, Langendijk JA, Lambin P, Kessels AG, and Grutters JP. Systematic review and meta-analysis of radiotherapy in various head and neck cancers: Comparing photons, carbon-ions, and protons. *Cancer Treatment Reviews* **2011** 37(3):185-201.

[82] Orlandi E, Iacovelli NA, Bonora M, Cavallo A, and Fossati P. Salivary Gland. Photon beam and particle radiotherapy: Present and future. *Oral Oncology*, **2016** 60:146–156.

[83] van de Water TA, Bijl HP, Schilstra C, Pijls-Johannesma M, and Langendijk JA. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: a systematic review of literature. *The Oncologist* **2011** 16(3), 366–377.

[84] Van De Water TA, Lomax AJ, Bijl HP, De Jong ME, Schilstra C, Hug EB, and Langendijk JA Potential benefits of scanned intensity-modulated proton therapy versus advanced photon therapy with regard to sparing of the salivary glands in oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* **2011** 79(4), 1216–1224.

[85] van der Laan HP, van de Water TA, van Herpt HE, Christianen ME, Bijl HP, Korevaar EW, Rasch CR, van'T Veld AA, van der Schaaf A, Schilstra C, and Langendijk JA. The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. *Acta Oncologica* **2012** 561–569.

[86] Dennis ER, Bussiere MR, Niemierko A, Lu MW, Fullerton BC, Loeffler JS, and Shih HA. A comparison of critical structure dose and toxicity risks in patients with low grade gliomas treated with IMRT versus proton radiation therapy. *Technol Cancer Res Treat* **2013** 12(1):1-9.

[87] Kandula S, Zhu X, Garden AS, Gillin M, Rosenthal DI, Ang KK, Mohan R, Amin MV, Garcia JA, Wu R, Sahoo N, and Frank SJ. Spot-scanning beam proton therapy vs intensity-modulated radiation therapy for ipsilateral head and neck malignancies: a treatment planning comparison. *Med Dosim* **2013** 38(4):390-394.

[88] Yepes PP, Randeniya S, Taddei PJ, and Newhauser WD. Monte Carlo fast dose calculator for proton radiotherapy: application to a voxelized geometry representing a patient with prostate cancer. *Phys Med Biol* **2009** 54(1):21-28.

[89] Yepes PP, Eley JG, Liu A, Mirkovic D, Randeniya S, Titt U, and Mohan R. Validation of a track repeating algorithm for intensity modulated proton therapy: clinical cases study. *Phys Med Biol* **2016** 61(7):2633-2645.

[90] Verburg JM, Testa M, and Seco J. Range verification of passively scattered proton beams using prompt gamma-ray detection. *Phys Med Biol* **2015** 60(3):1019-1029.

[91] Matsumoto Y, Matsuura T, Wada M, Egashira Y, Nishio T, and Furusawa Y. Enhanced radiobiological effects at the distal end of a clinical proton beam: in vitro study. *J Radiat Res* **2014** 55(4):816-822.

[92] Schmid S1, Landry G, Thieke C, Verhaegen F, Ganswindt U, Belka C, Parodi K, and Dedes G. Monte Carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional changes in prostate cancer patients. *Phys Med Biol* **2015** 60(24):9329-9347.

[93] Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys. Med. Biol.* **2012** 57:99-117.

[94] España S, and Paganetti H. Uncertainties in planned dose due to the limited voxel size of the planning CT when treating lung tumors with proton therapy. *Phys Med Biol* **2011** 56(13):3843-3856.

[95] Würl M, Englbrecht F, Parodi K, and Hillbrand M. Dosimetric impact of the low-dose envelope of scanned proton beams at a ProBeam facility: comparison of measurements with TPS and MC calculations. *Phys Med Biol* **2016** 61(2):958-973.

[96] Paganetti H, Jiang H, Parodi K, Slopsema R, and Engelsman M. Clinical implementation of full Monte Carlo dose calculation in proton beam therapy. *Phys Med Biol* **2008** 53(17):4825-4853.

[97] Molinelli S, Mairani A, Mirandola A, Vilches Freixas G, Tessonnier T, Giordanengo S, Parodi K, Ciocca M, and Orecchia R. Dosimetric accuracy assessment of a treatment plan verification system for scanned proton beam radiotherapy: one-year experimental results and Monte Carlo analysis of the involved uncertainties. *Phys Med Biol* **2013** 58(11):3837-3847.

[98] Magro G, Molinelli S, Mairani A, Mirandola A, Panizza D, Russo S, Ferrari A, Valvo F, Fossati P, and Ciocca M. Dosimetric accuracy of a treatment planning system for actively scanned proton beams and small target volumes: Monte Carlo and experimental validation. *Phys Med Biol* **2015** 60(17):6865-6880.

[99] Engelsman M, Schwarz M, and Dong L. Physics controversies in proton therapy. *Semin Radiat Oncol* **2013** 23(2):88-96.

[100] Safai S1, Trofimov A, Adams JA, Engelsman M, and Bortfeld T. The rationale for intensity-modulated proton therapy in geometrically challenging cases. *Phys Med Biol* **2013** 58(18):6337-6353.

[101] Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recogn Lett* **2010** 31:651-666.

[102] Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, and Wu AY. An efficient k-Means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Anal Mach Intell* **2002** 24(7):881-892.

[103] Corduneanu A, and Bishop CM. Variational Bayesian model selection for mixture distributions. *Artificial Intelligence and Statistics* ***2001*** 27-34.

Vita

Joshua Scott Niedzielski was born on November 17, 1982, in Prince Frederick, Maryland to Robert and Judy Niedzielski, as the youngest of three siblings including one sister, Rachael Copsey, and one brother, Joseph Niedzielski. After graduating from Calvert High School in Prince Frederick, Maryland, he then relocated to Stevensville, MI in 2003 where he would begin higher education at Lake Michigan College. Eventually, he transferred to Michigan State University in East Lansing, Michigan where he would earn three Bachelor of Science degrees in Astronomy and Mathematics in 2009, and Physics in 2010. He also served as a reserve police officer for Lincoln Township from 2005 to 2009. He then entered the PhD graduate program for Medical Physics at the University of Texas-MD Anderson Center and Graduate School of Biomedical Sciences in 2010. There he conducted his dissertation research under the guidance of Dr. Laurence Court. His research interests include toxicity prediction modelling, outcome assessment, and application of novel imaging biomarkers for outcome modelling.

Permanent Contact Information:

joshniedzielski@gmail.com