Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Tumour Volume Dynamics: Quantification and Prediction in Non-Small Cell Lung Cancer Radiation Therapy

A thesis submitted to

The University Of Dublin, Trinity College

for the degree of

Doctor of Philosophy

2025

Sarah Barrett

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish copyright legislation and Trinity College library conditions of use and acknowledgement. I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Sarah Barrett

March 2025

DEDICATION

To my parents, Bernadette and Joseph.

For your lifelong support of my education.

To my daughters, Ada and Liadh.

I promise to do the same for you both.

To my husband, Jonathan.

'In your presence

Time rode easy, anchored

On a smile'

ACKNOWLEDGEMENTS

It is important to begin by acknowledging the human impact of the subject matter in this thesis. Lung cancer is the most common cause of cancer death worldwide and so many families are bereaved by this disease. I wish to mention my own family members who have died from lung cancer, my aunt, Elsie Gray, who died in December 2024 and my Grandad, Tom Ginty, who died in 1985.

This work was only feasible with the support of many people, and I will be forever grateful for all that help over these past few years.

Firstly, I'd like to thank Professor Laure Marignol who supervised this thesis, your patience, knowledge, encouragement and support made this work possible from start to finish. In addition to your supervision, I'd also like to thank you for your friendship along the way.

Professor Heiko Enderling, my adjunct supervisor, thanks for being so generous with your insight, time and not to mention your mathematical model! Getting to know and work with you has been a real pleasure.

Dr Gerard Walls, your enthusiasm and energy for research is infectious and I am so grateful for your cooperation, proactiveness and for sharing your data with me. I look forward to many fruitful collaborations in the future.

Dr Mohammad Zahid, thank you for patiently showing me how to use your PSI code in MATLAB and for your help when I couldn't get the code to run, which was frequently. Your insight was so beneficial to this project.

Professor Andrew Simpkin, Andy, having a Professor of Statistics at the end of the phone was pretty handy to say the least. Thank you for answering my questions and keeping me on the right path while writing this thesis.

Dr Conor McGarry and Prof Alan Hounsell, thank you for facilitating the research contained in Chapters 5, 7 and 8, and for allowing me access to the Physics office in the Northern Ireland Cancer Centre. Thank you, Conor, for your advice and comments on this work.

Professor Gerry Hanna, your mentorship and input into this thesis not only helped guide the work but also opened many doors for me and this project, thank you so much.

Professor Michelle Leech for supervising while Laure was on leave, thank you for maintaining the momentum of the project.

I am also grateful to all my fellow researchers that shared their data with me, in particular Dr Janita van Timmeren for sharing data which facilitated Chapter 6 of this thesis.

I would like to extend my thanks to the examiners Dr Alan McWilliam and Dr Jonathan McAleese for taking the time to assess this work.

I am very lucky to have fantastic colleagues in the Discipline of Radiation Therapy, that I count as friends and make work feel easy. Thanks for enduring my relentless PhD chat over the last few months, particularly those who were going through it right alongside me, it was great to have company!

To my wonderful friends Karla, Kate, Aoife, Linda, Sarah, Melissa and Siobhan, thank you for putting up with my lack of availability this past year in particular!

Next, I want to sincerely thank my family for making it possible to complete this thesis, I couldn't have done it without you all.

My Donegal family-in-law, thank you for all the support, especially in the last few months with the girls, it is so appreciated.

My sisters, Elaine and Ailish. I am so lucky to have you both. You keep me going and keep me sane every single day!

Mom and Dad, thank you for all the encouragement over the years, not just during "the PhD years". I cannot adequately express my gratitude for all the sacrifices you made to support the three of us.

Finally, Jonathan, Ada and Liadh.
Jonathan, I could not have done any of this without you, and it would have been a lot less fun along the way! Mo ghrá go deo.
Ada and Liadh, you are by far my most important accomplishments during this PhD. You truly are a gift that I am grateful for every day.

SUMMARY

Radiation therapy (RT) is indicated in the management of all stages of non-small cell lung cancer. During radical RT delivery, on treatment image guidance results in the routine acquisition of cone beam CT scans on a daily or weekly basis. These scans are utilised in the geometric verification of the treatment but have the potential to assess early tumour response.

This thesis investigates longitudinal tumour volume kinetics in patients treated with RT for NSCLC. The existing literature on this topic is conflicting and is limited by the retrospective nature of the evaluation of tumour volume changes.

Firstly, this thesis examined the feasibility of generating large scale datasets of tumour volume measurements using a semi-automated delineation tool. Study 1 tested this approach in high quality imaging acquired for research purposes and found the volumes generated to be acceptable based on geometric and dosimetric comparisons to manual contours. Study 2 piloted the methodology validated in Study 1 on cone beam CT images acquired routinely during RT and found the approach to be feasible and efficient in 66.6% of the population.

Secondly the Proliferation Saturation Index (PSI) mathematical model to predict tumour response to RT was examined in this population. Prediction of tumour response to RT could select patients for tailored adaptive strategies in the clinic.
Study 3 trained and tested the model in a cohort of 164 patients across four dose fractionation schedules. The PSI model predicted tumour volume regression during RT with a high degree of accuracy, using early tumour response dynamics. Comparison of the measured vs predicted

volumes resulted in R^2 values ranging from 0.94-0.97, and Pearson Correlation Coefficients from 0.97-0.98 in the four groups. Study 4 externally validated the findings of Study 3 in a cohort of 71 patients treated with definitive RT alone. Furthermore, it examined the relationship between regression with overall survival, identifying in this cohort that there was a negative relationship between these factors. Study 5 determined that the performance of the PSI model using only pre-treatment data was strong (R^2=0.86 and Pearson Correlation Coefficient=0.93) when comparing measured and predicted volumes. An *in-silico* trial of 6 dose fractionation schedules revealed that the impact of prescription on tumour regression at the end of RT varied widely between individuals.

LAY SUMMARY

Radiation therapy is a core treatment modality for patients with a diagnosis of lung cancer. During this treatment, scans are acquired to ensure accuracy of the radiation delivery. On these images we can see that some tumours shrink a lot in response to radiation and some do not shrink at all. The significance of the varying response is currently uncertain.

Firstly, this thesis tests if we can monitor the tumour volume changes on the images acquired, to measure and report how different tumours respond to radiation therapy.

Secondly, it examines a mathematical model which aims to predict that response in individuals, and tests the accuracy of the predictions in patients treated with different radiation therapy prescriptions.

Finally, it investigates the clinical relevance of tumour changes in a group of patients treated with radiation therapy alone.

LIST OF ABBREVIATIONS

2D: 2-Dimensional

3D: 3-Dimensional

3D CRT: 3-Dimensional Conformal Radiation Therapy

4D: 4-Dimensional

4DCT: 4-Dimensional Computed Tomography

4DCBCT: 4-Dimensional Cone Beam CT

AC: Auto Contour

AI: Artificial Intelligence

AIC: Akaike Information Criterion

AJCC: American Joint Commission on Cancer

ALK: Anaplastic Lymphoma Kinase

ANOVA: Analysis of Variance

ART: Adaptive Radiation Therapy

AUC: Area Under Curve

AVIP: Average Intensity Projection

BED: Biologically Equivalent Dose

BH: Breath Hold

BRAF: V-Raf murine sarcoma viral oncogene homolog B

CBCT: Cone Beam Computed Tomography

CCRT: Concurrent Chemotherapy and Radiation Therapy

CFRT: Conventionally Fractionated Radiation Therapy

CHART: Continuous Hyperfractionated Accelerated Radiotherapy

CI: Confidence Interval

COM: Centre of Mass

CT: Computed Tomography

CTV: Clinical Target Volume

CXR: Chest X-Ray

DIR: Deformable image registration

D_{max}: Maximum point dose

D_{mean}: Mean Dose

DNA: Deoxyribonucleic Acid

DSC: Dice Similarity Coefficient

DVC: Dose Volume Constraint

DVH: Dose Volume Histogram

EBUS: Endobronchial ultrasound

EGFR: Epidermal Growth Factor Receptor

EPID: Electronic Portal Imaging Device

EQD2: Equivalent Dose in 2Gy per fraction

EUD: Equivalent Uniform Dose

FB: Free Breathing

FDG PET: Fluorodeoxyglucose Positron Emission Tomography

FFLF: Freedom From Local Failure

FLT PET: ^{18}F fluorothymidine Positron Emission Tomography

GTV: Gross Tumour Volume

Gy: Gray

HR: Hazards Ratio

ICI: Immune Checkpoint Inhibitors

IGRT: Image Guided Radiation Therapy

IMRT: Intensity Modulated Radiation Therapy

IO: Immunotherapy

ITV: Internal Target Volume

IV: Intravenous

KM: Kaplan Meier

KRAS: Kirsten rat sarcoma viral oncogene homolog

KV: Kilovoltage

LA-NSCLC: Locally Advanced Non-Small Cell Lung Cancer

LC: Local Control

LLL: Left Lower Lobe

LP: Local Progression

LPS: Linear Predictor Score

LUL: Left Upper Lobe

MC: Manual Contour

MD: Microscopic Disease

MDT: Multidisciplinary Team

MIP: Maximum Intensity Projection

MLD: Mean Lung Dose

MSE: Mean Squared Error

MV: Megavoltage

MVCT: Megavoltage Computed Tomography

NSCLC: Non-Small Cell Lung Cancer

NTCP: Normal Tissue Complication Probability

OAR: Organ at Risk

OS: Overall Survival

OTT: Overall Treatment Time

PCC: Pearson Correlation Coefficient

PD-L1: Programmed Death Ligand 1

PET: Positron Emission Tomography

PFS: Progression Free Survival

PRV: Planning Organ at Risk Volume

PS: Performance Status

PSI: Proliferation Saturation Index

PSO: Particle Swarm Optimisation

PTP: Predictive Treatment Planning

PTV: Planning Target Volume

QUANTEC: Quantitative Analysis of Normal Tissue Effects in the Clinic

RLL: Right Lower Lobe

RO: Radiation Oncologist

ROC: Receiver Operating Characteristic

RSS: Residuals Sum Squared

RT: Radiation Therapy

RTOG: Radiation Therapy Oncology Group

RTT: Radiation Therapist

RUL: Right Upper Lobe

SABR: Stereotactic Ablative Radiation Therapy

SBRT: Stereotactic Body Radiation Therapy

SCC: Squamous Cell Carcinoma

SIB: Simultaneous Integrated Boost

TCIA: The Cancer Imaging Archive

TCP: Tumour Control Probability

TD50: Dose that would result in 50% probability of developing complications

TNM: Tumour Nodes Metastases

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

TSS: Total Sum of Squares

TTD: Total Tumour Dose

TVR: Tumour Volume Regression

UA-AC: User-Adjusted Auto Contour

VMAT: Volumetric Modulated Arc Therapy

RESEARCH OUTPUTS

Peer Reviewed Journal Articles

Barrett S., Hanna G.G., Marignol L., An overview on personalisation of radiotherapy prescriptions in locally advanced non-small cell lung cancer: Are we there yet? Radiotherapy and Oncology, 2018, 128(3): 520-533.

Barrett S., Simpkin A.J., Walls G., Leech M., Marignol L., Geometric and dosimetric evaluation of a commercially available auto-segmentation tool for gross tumour volume delineation in locally advanced non-small cell lung cancer: a feasibility study, Journal of Clinical Oncology, 2021 Mar;33(3):155-162

Barrett S., Zahid .M.U., Enderling H., Marignol L., Predicting individual tumour response dynamics in locally advanced non-small cell lung cancer radiation therapy; a mathematical modelling study. International Journal of Radiation Oncology, Biology and Physics, 2025, 121(4): 1077-1087

Peer Reviewed Posters

Barrett S., Leech M., Enderling H., Marignol L., User-adjusted auto-contouring of lung tumour volumes can facilitate accurate multiple measurements of lung GTV over the course of radiotherapy, 16th International Congress of Radiation Research, Manchester, UK, 25-29th August, 2019

Barrett S., Leech M., Marignol L., Time trend analysis of target volume auto-contouring in locally advanced NSCLC over the course of RT, ESTRO 2020, Vienna, Austria, 2020

Barrett S., Marignol L., Hanna G.G., McGarry C., Walls G. M., Delineation of non-small cell lung cancer gross tumour volume on cone

beam CT images: a feasibility study including auto-segmentation. ESTRO 2025, Vienna, Austria, 2025

Poster Presentation

Barrett S., Zahid M., McGarry C., Enderling H., Walls G. Marignol L., Predicting tumour volume reduction in non-small cell lung cancer: Independent validation of a single parameter PSI model. ESTRO 2025, Vienna, Austria, 2025

Invited talk

Lung cancer radiotherapy: Pioneering trends and innovations for future practice

ESTRO 2025, Vienna, Austria, 2025

Manuscripts in Preparation

Technical note:

Semi-automated delineation of non-small cell lung cancer gross tumour volume on daily cone beam CT; is it achievable in a real-world setting?

Research article:

External Validation of the Proliferation Saturation Index Mathematical Model in Non-Small Cell Lung Cancer Treated with Definitive Radiation Therapy

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 6 Figures

Chapter 7 Figures

Chapter 8 Figures

LIST OF TABLES

Chapter 8 Tables

CHAPTER 1

INTRODUCTION

1.1 Lung Cancer

1.1.1 Incidence and epidemiology

Lung cancer is both the most commonly diagnosed cancer, and the leading cause of cancer death globally, accounting for 12.4% of the cases diagnosed and 18.7% of the cancer related deaths reported [1].

In Ireland lung cancer is the 3rd most common cancer diagnosis (excluding non-melanoma skin cancer) with approximately 2660 people being diagnosed annually. It is the leading cause of death, responsible for 20% of all cancer related deaths in Ireland [2].

Epidemiological risk factors are a history of tobacco use, genetic predisposition, occupational exposures, air pollution and passive smoke inhalation with tobacco use being the most significant risk factor [3]. While the rate of tobacco smoking is decreasing, and the overall incidence of lung cancer is decreasing, the incidence in never-smokers is on the rise [4]. Those diagnosed with lung cancer, who never smoked, are more likely to be of female sex, owing to innate risk factors related to genetics, metabolism or hormones that are yet to be elucidated [5].

1.1.2 Histological subtypes

Lung cancers are a heterogeneous group of diseases, but can be broadly grouped into Non-Small Cell Lung Cancer (NSCLC), which accounts for approximately 84% of all lung cancer cases [6] and Small Cell Lung Cancer (SCLC) approximately 14% of the cases. Less common subtypes include thymic tumours or mesothelioma which is strongly associated with occupational asbestos exposure. NSCLC can be further subdivided by histology, with adenocarcinomas and squamous cell carcinomas contributing to the majority of diagnoses, with a minority of large cell carcinomas in this group [7].

Within NSCLC, the advent of molecular diagnostics in oncology has led to increased subdivision based on molecular signatures and mutations at a cellular level. Some common mutations identified include Kirsten rat sarcoma viral oncogene homolog (KRAS), V-Raf murine sarcoma viral oncogene homolog B (BRAF) kinase mutations, epidermal growth factor receptor (EGFR) overexpression and abnormal fusion of the anaplastic lymphoma kinase (ALK) oncogene [8]. These mutations can provide therapeutic opportunities in the form of targeted therapies, generally reserved as adjuvant therapy following definitive surgery or concurrent chemotherapy and radiation therapy (CCRT) [9] or in the management of systemic disease [10, 11]. EGFR mutations can be targeted at an oncogene driver level using small molecule Tyrosine Kinase Inhibitors (TKI), such as Osimertinib [12] or monoclonal antibodies (mAb) such as cetuximab [13]. In the Irish setting up to 53% of patients are reported to be eligible for a targeted systemic therapy [8].

Expression of immunotherapy (IO) markers are also considered with the advent of Immune Checkpoint Inhibitors (ICI). ICI drugs targeting programmed death ligand 1 (PD-L1) have improved outcomes for patients with NSCLC [14, 15] without the need to stratify patients based on tumour PD-L1 expression. While some guidelines recommend treatment

only in a population with PD-L1>1%, this is disputed [16]. PD-L1 expression in a tumour can be dynamic and heterogenous and the clinical benefit in an unselected NSCLC population has been demonstrated in the PACIFIC trial [14].

1.1.3 Staging of NSCLC

Comprehensive staging is required to determine the optimal management of a patient diagnosed with NSCLC and clinical outcomes are linked to disease stage [9].

Recommended radiological imaging includes Chest X-Ray (CXR), Computed Tomography (CT) of the thorax and abdomen, and 18-FluoroDeoxyGlucose Positron Emission Tomography (FDG PET) scan. Endobronchial Ultrasound (EBUS) has been shown to improve staging accuracy in the mediastinum compared to PET alone [17]. MRI is indicated for Pancoast tumours which occur in the apex of the lung, or MRI brain if metastases are suspected. Tissue sampling via image guided biopsy or EBUS guided biopsy are recommended to determine histological subtype. Clinical exam and assessment of patient's history, performance status and pulmonary function are routinely carried out.

NSCLC stage is classified using the Tumour Nodes Metastases (TNM) system according to the American Joint Commission on Cancer (AJCC) and the Union for International Cancer Control (UICC) 8th Edition [18]. Primary tumour size defines the T stage, with T1 tumours ≤3cm and T2 between 3-5cm. T3 tumours are > 5cm but ≤ 7cm, or invading chest wall, pericardium, phrenic nerve; or separate tumour nodules in the same lobe. T4 tumours are >7 cm or tumour invading: mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, oesophagus, spine; or tumour nodules in a different ipsilateral lobe.

N0 denotes no lymph node involvement, N1 indicates ipsilateral pulmonary or hilar nodes, N2 ipsilateral mediastinal or subcarinal nodes and N3 denotes contralateral mediastinal, hilar, or supraclavicular nodes. M0 defines no metastatic spread and M1 indicates metastatic disease. This is further subdivided into M1a, M1b and M1c depending on the extent of disease. M1a may be a single nodule in a contralateral lobe and M1c is multiple extra thoracic metastases. The updated 9th edition of TNM staging was implemented in January 2025, with subdivision of N2 status based in single or multiple nodal stations involved, and subdivision of M1c to consider if metastases are in a single organ or multiple organs [19].

TNM allows for grouping of patients in to stages I-IV depending on the extent of disease. Stage I is early stage, T1-2, N0 disease. Stage II are those with slightly larger primary tumours T2b-T3 or N1 disease. Stage III is all other cases with M0 disease, patients with bulky T4 primaries, and all N2-N3 disease regardless of T stage. Stage IV patients have metastatic disease.

1.1.4 Management of NSCLC

Radiation Therapy (RT) is indicated in the management of NSCLC across all stages of disease. Patients presenting with Stage I disease are considered resectable, and definitive surgery is recommended [9], however those who are medically inoperable should receive stereotactic ablative body radiation therapy (SABR) [20]. Stage II or III patients presenting with low burden nodal involvement, N1 and single N2 non-bulky non-invasive nodes, may also be eligible for surgical resection [9].

Patients with larger primary tumours, and multiple N2 nodes are potentially resectable [21], but N2 bulky disease, invasive nodes, or N3 disease are unresectable and referred for definitive RT with concurrent

chemotherapy. For these patients RT is delivered with non-SABR dose fractionation schedules, frequently 2Gy per day for 30-33 fractions. If performance status is poor, or the patient has significant comorbidities sequential chemotherapy is recommended. Platinum based chemotherapy is standard, generally Cisplatin or Carboplatin in combination with Vinorelbine, Gemcitabine or Docetaxel [9]. These patients, with non-metastatic but unresectable disease due to the extent of nodal involvement, are generally Stage III [22] and clinical outcomes for this group remain sub optimal [14, 23, 24]. Standard of care treatment for patients with locally advanced NSCLC (LA-NSCLC) is CCRT, followed by consolidation ICI immunotherapy in the form of anti-PD-L1 drug Durvalumab [9].

The introduction of adjuvant immunotherapy following CCRT has been immensely promising with an improvement in 5-year overall survival (OS) from 33.4% to 42.9% for patients treated with Durvalumab compared to a placebo [14]. However, real world data shows that not all patients receive this regimen with one US based survey reporting only 48% of patients with Stage III NSCLC receiving CCRT, and of those only 55% went on to receive immunotherapy [25]. PACIFIC-2 aimed to build on the promising findings by testing immunotherapy concurrently with CCRT [26], however the data failed to show an improvement in progression free survival (PFS) for patients [27].

Recently, the LAURA trial, in patients with EGFR-mutated Stage III NSCLC, demonstrated that adjuvant Osimertinib following definitive CCRT, resulted in significantly longer PFS than placebo. The percentage of patients alive at 12 months, without progression was 74% (95% CI 65 to 80) with Osimertinib, compared to 22% (95% CI 13 to 32) in the placebo group. While this exciting result is welcomed, mature data on OS is awaited and the optimal duration of drug therapy is to be determined [28].

Local control (LC) is a predictor of OS [29] and for these patients we need to achieve this control of the primary tumour using optimal RT techniques, and manage the systemic potential for failure using a multimodality chemotherapy and immunotherapy approach. While there have been extensive strides made in the systemic therapeutic options in recent years, RT prescriptions have remained unchanged in routine clinical practice.

1.2 Principles of definitive radiation therapy in NSCLC

In Stage II or III, definitive RT alongside concurrent or sequential chemotherapy is recommended, and patients who are medically unfit for this approach are treated using RT alone. Long-established dose fractionation schedules of 60-66Gy in 30-33 fractions (2Gy per fraction) [9, 30] or 55Gy in 20 fractions (2.75Gy per fraction) [31] are commonly used for these patients with LA–NSCLC. While these dose fractionation schedules are common in the clinic, alternative prescriptions have been examined in clinical trials [32].

1.3 Radiobiology of dose and fractionation schedules in definitive RT

Conventionally Fractionated RT (CFRT) delivers daily doses in the region of 1.8Gy-2Gy per fraction, 5 days per week. CFRT aims to exploit radiobiological principles of reoxygenation and redistribution to improve treatment efficacy. By adopting a CFRT regimen, cell kill is increased by allowing reoxygenation of hypoxic tumour cells over the course of treatment. Improving cell oxygenation increases the rate of cell kill due to the oxygen enhancement ratio [33]. Additionally, cells are not equally sensitive to irradiation throughout the cell cycle and are most sensitive during mitosis and most resistant at S phase [34]. Redistribution of cells

over the course of CFRT theoretically allows cells resistant at one time point to move into a more sensitive phase at subsequent fractions of radiation [35]. Simultaneously, CFRT can also lead to a reduction in normal tissue damage by facilitating repair and repopulation in non-tumour cells over the course of RT. In lung radiotherapy, acute toxicity in the oesophagus, an early responding tissue, is often dose limiting. CFRT allows for repair of sub-lethal damage and therefore can reduce normal tissue complication probability [35]. However, the LC rates following this treatment remain low [36] and so alternative treatment options have been studied in this population.

The benefit of RT dose escalation has been explored and a clear dose response relationship has been reported in the setting of curative intent RT for NSCLC [37]. The potential benefit of increasing total dose from 60Gy to 74Gy was investigated in RTOG 0617 and has not translated into a clinical benefit for patients, and was even found to result in poorer outcomes [38]. Patients in the dose escalated arm had a median OS of 20.3 months compared to 28.7 months in the standard dose arm therefore 60Gy/30# remains the current standard of care [38].

This unexpected, and somewhat disappointing, result has been considered extensively by the lung cancer radiation oncology community [39]. In RTOG 0617, total dose was escalated by increasing the total number of fractions to 37. The failure of dose escalation to improve outcomes, may be attributed, in part at least, to the increase in overall treatment time which would allow for accelerated repopulation of tumour cells and it is recommended to avoid prolonged RT delivery [40]. NSCLC exhibits rapid repopulation after 3-4 weeks of radiation therapy and overall treatment durations of < 4 weeks could help to counteract this [41] which would support this hypothesis that repopulation was a factor in the poorer outcomes in the dose escalated group. Additionally, the data suggests that increased heart dose was critical, with better OS seen in tumours located

further from the heart and lower heart dose metrics [42-44]. What was clear from this trial is dose escalation, by increasing the number of fractions, is not a solution to improve local control in this setting.

Accelerated and hyper-fractionated RT have been explored as alternative dose fractionation schedules to counteract tumour repopulation over the course of treatment. A meta-analysis demonstrated that there was a significant survival benefit associated with either accelerated or hyperfractionated regimens compare to CFRT, but this benefit was at the expense of increased oesophageal toxicity [45]. Continuous hyperfractionated accelerated radiotherapy (CHART) has been studied in this population and data demonstrates an advantage both in local control and overall survival when compared with CFRT [46]. A combination of dose escalation and CHART has been explored in a Phase 1 trial with promising results however this is not in conjunction with chemotherapy and acute toxicity remains a concern [47]. Despite the encouraging evidence around the use of accelerated and hyperfractionated regimens in this population, these protocols are not commonly implemented in the clinic and it has been hypothesised that this may be due to the logistical difficulties of delivering accelerated regimes as well as increased acute toxicity concerns [48].

A possible alternative approach to tackling the impact of accelerated repopulation in these tumours is the use of moderately hypofractionated schedules of 2.5-2.75Gy per fraction to reduce the overall treatment duration without the logistical issues of continuous hyperfractionated regimens. This option has become more achievable with the advances of modulated techniques and improved image guidance which previously limited our ability to safely deliver higher daily doses [49]. The SOCCAR Trial demonstrated the tolerability and feasibility of this regimen in the CCRT setting [31].

The possibility of hypofractionated dose escalated RT has also been considered in the clinic for NSCLC, and while this may offset the accelerated repopulation of the tumour, acute toxicity can be severe with grade 5 adverse events reported [50]. Accelerated hypofractionated RT has been investigated in the concurrent chemotherapy setting [51] and shown to have acceptable toxicity, however this study used concurrent gemcitabine as a radiosensitizer, not Cisplatin which is not the standard of care. A more recent study looking at the use of a hypofractionated regime of 60Gy in 24# concurrently with Cisplatin chemotherapy reported encouraging clinical outcomes, but again noted the occurrence of severe oesophageal toxicity [52].

The optimal dose fractionation schedule, which maximises tumour control probability (TCP) while maintaining normal tissue complication probability (NTCP), has not yet been established in this population.

1.4 Personalised RT for NSCLC

Current RT practice consists of standardised treatment protocols for the vast majority of patients. In the case of NSCLC all steps of the definitive RT process from treatment planning, dose fractionation schedules, and on-set geometric verification are carried out using population protocols rather than tailoring each step to the specific patient or the specific tumour. While a number of approaches to personalising RT dose have been considered in the literature [32, 53] none have yet been implemented clinically, and these are examined in detail in Chapter 3 [54]. Clinical outcomes have improved in recent years due to advances in systemic anti-cancer therapies, however, the RT approach has not evolved in parallel. Standard approaches fail to consider the individual biology of such a heterogeneous population. Principles of precision medicine strive towards a personalised treatment and move away from the traditional one-size fits all strategy.

Since RTOG 0617, the research focus has been on dose intensification and personalisation rather than blanket escalation. RTOG 1106 examined dose adaptive RT (ART) based on FDG-PET scan acquired after 40Gy. Following this scan a personalised replan with a daily dose per fraction of between 2.2 and 3.8Gy was designed using isotoxic lung planning [53]. The study did not demonstrate an improvement in local control in the experimental arm but also did not see a detrimental effect in this dose intensification unlike RTOG 0617, and the authors conclude that studies to refine personalised RT should be considered in this era of immunotherapy.

The PET Boost trial randomised patients to either whole tumour boost or a PET avid sub volume boost, defined as the area within the primary tumour that held a SUV $>$ 50 % SUV_{max} on the pre- treatment 18F-FDG-PET scan [32]. The dose per fraction was escalated as high as achievable within a 3-5.4Gy range, again using an isotoxic approach, and the limiting factor was Mean Lung Dose (MLD) as determined on a conventional plan for that patient. The median escalated dose was 3.3Gy and 3.5Gy and excellent freedom from local failure (FFLF) was achieved with rates of 97% and 91% at 12 months, in the whole tumour and sub volume groups respectively. However, unexpectedly high rates of Grade 5 toxicity (18%) were reported and the authors suggest this was in part linked to tumour encasement of the great vessels [55].

More recently NARLAL2 trial reported its first results in loco-regional control at ESTRO 2024. This trial randomised patients to standard 66Gy/33# or a dose escalated arm where the aim was to deliver mean doses of 95Gy (tumour) and 74Gy (lymph nodes) again using FDG PET avidity as a guide and MLD and Lung V_{20Gy} as isotoxic dose constraints [56]. The primary endpoint was local regional control rate, and they report early results of a significant improvement in local control at 3 years, and mature data is awaited. Additionally, no increase in early toxicity was

identified, and only 1 Grade 5 event was reported, which occurred in the standard arm [57].

SABR assisted escalation in the LA-NSCLC population is being considered in the literature. An early phase, non-randomised trial recently reported their outcomes following a study which sought to identify if a post CCRT SABR boost was safely deliverable [58]. This group treated patients with 40Gy/10# followed by a 5# SABR boost to FDG active tissue with an additional dose of either 25Gy, 30Gy or 35Gy all with concurrent weekly chemotherapy (carboplatin/paclitaxel).They report excellent 2 year LC of 74.1%, 85.7%, and 100.0% and 2 year OS 30.0%, 76.2%, and 55.6% in the low-, intermediate-, and high-dose cohorts, respectively. Two Grade 5 events occurred in the high dose group, and acceptable toxicity was reported in the other groups.

Not all trials report escalation strategies, and some aim to de-intensify RT for these patients. Ohri et. al reported a Phase II study allocating patients to either a high or low dose group based on the metabolic tumour volume defined on FDG PET [59]. The low-risk group received either 57Gy or 52.5Gy in 25 fractions, and the high risk 65Gy/25#. The study met its primary endpoint which was the absence of high residual metabolic activity on FDG PET 12-16 weeks post RT. The authors report that the low rates of local progression in the low risk group support the suggestion that we can de-escalate in selected lesions [60].

We see from these studies that there is a real ambition among the clinical lung cancer community to drive change and personalise the dose for patients, however the strategies remain somewhat generic for this heterogeneous population. Most aim to escalate as much as possible and the element of personalisation is determined the dose limitations of surrounding normal tissue. We have yet to derive a method to elucidate which patients may benefit from escalation, or which of the many

potential dose fractionation schedules is the optimal choice for a given patient.

1.5 Role of biomarkers in NSCLC RT

Clinical biomarkers, both prognostic and predictive, are of interest to the radiation oncology community generally, and the NSCLC community specifically. For patients receiving systemic drug therapy the option to personalise treatment using biomarkers such as EGFR or PD-L1 expression exist to help match the optimal systemic therapy to each individual case. This approach is also desirable to optimise RT, and has been investigated widely in the literature but not yet adopted to routine clinical practice [61]. Various classes of biomarkers exist such as tumour tissue biomarkers identified on tumour biopsies, circulating tumour biomarkers such as circulating tumour DNA, immunogenic biomarkers and non-invasive imaging biomarkers [61].

Imaging biomarkers such as radiomics can be extracted from routine clinical imaging, acquired for diagnosis or RT planning, without an additional burden on either the patient or the health service. Radiomics is the extraction of quantitative data from medical images and analysing these data to gain insight beyond the scope of the human review of the images [62, 63]. In lung cancer this has been largely evaluated on pre-treatment imaging where studies have shown radiomics features to be prognostic for clinical outcomes [64, 65]. While there are many studies and promising signals in the literature, the variation in scan acquisition, reconstruction and methodological rigour in these studies indicates that the process remains challenging [66]. Delta radiomics where longitudinal imaging, such as on-treatment Cone Beam CT (CBCT) scans, are evaluated to identify changes in features over time has been investigated in NSCLC and may provide a signal during RT to instigate ART or modify the plan [67]. The optimal approach may well be the integration

of multi-omic features into a predictive model to gain the maximum insight from all available data.

1.6 Tumour volume regression as emerging biomarker

Tumour volume regression during the course of RT is not a classic clinical marker predictive of outcomes, however in recent years regression has been increasingly studied as an emerging biomarker, potentially indicative of clinical response.

In 2014 Brink et al investigated the relationship between local control (LC) and tumour regression as seen on CBCT and found that those with more pronounced regression had poorer LC and overall survival (OS) [68]. Conflicting findings were reported in 2015 where a study of 38 patients treated with CCRT were analysed and regression during RT was predictive of OS, and reported that for every 10% decrease in Gross Tumour Volume (GTV) from day 1 to day 43, the risk of death decreased by 44.3%. This study separated patients based on a GTV regression of 39.3% and found a significant difference in OS of 31 months versus 10 months for those with regression >39.3% vs those with regression <39.3%. Wald et al. supported this data with a study that demonstrated that larger relative volume reduction during the first 4 weeks of treatment (>50%) was significantly associated with improved distant control, progression free survival (PFS), and OS [69].

Kanzaki et. al [70] evaluated this concept in 2016 where stage III NSCLC patients were treated with a shrinking field approach. They reported that a higher tumour regression rate calculated after delivery of 40-44Gy was associated with better OS and PFS in patients with a squamous cell carcinoma, and conversely associated with poorer OS in the adenocarcinoma subgroup.

Further work aiming to validate Brink et al.'s findings, separated patients based not only on histology (adenocarcinoma vs non-adenocarcinoma) but also chemotherapy status [71]. This group reported that chemotherapy was a confounding factor in elucidating the relevance of tumour volume regression (TVR) as a predictive biomarker. In this study large tumour regression in patients with non-adenocarcinoma histology being treated with CCRT was a negative indicator of LC. The opposite effect was identified in patients not receiving chemotherapy, with larger regression being associated with improved clinical outcomes, significantly improved OS and improved but non-significant LC. A 2018 systematic review on the predictive and prognostic value of tumour volume changes during radical RT highlighted that the group of patients being studied are heterogeneous. Additionally, within the included studies there is further heterogeneity between methods of GTV measurement, time of evaluation of regression, initial stages and treatment choices [72]. The authors highlight that future work should address these confounding variables.

An international retrospective observational cohort study aimed to elucidate the relevance some of these variables [73]. The study analysed n=347 patients treated in 21 centres across 5 European countries who were treated with definitive RT or CCRT and investigated the prognostic relevance of GTV. Within this group a subset of n=176 patients had a replanning GTV measurement following delivery of 40-50Gy. The study found that the absolute volume reduction between GTV pre-treatment and GTV during RT (following 40-50Gy) was significantly associated with OS. Absolute tumour volume at start of RT or relative GTV reduction was not associated with OS [73]. The authors conclude that the tumour volume during RT is a more robust prognostic factor than tumour volume pre-treatment. This is a finding echoed by In Lee et al. who report that relative GTV regression at week 4 of RT was the most promising predictor for disease progression, compared to absolute regression at this time point or tumour volume pre-RT [74]. In addition, they noted that the absolute

GTV volume at week 4 exhibited a stronger association with survival outcomes than the initial GTV pre CCRT. These patients received a total dose of 60 -75Gy in 2.0/2.2Gy fractions. The TORCH study, a retrospective observational study evaluating the relevance of TVR mid-RT on clinical outcomes, reported that tumour reduction after 40-50Gy of RT was associated with improved survival in patients who when on to receive standard of care Durvalumab IO [75].

Conversely a large analysis of 394 LA-NSCLC patients found no link between TVR and OS [76]. This study reported baseline GTV as the best predictor of outcomes and the authors suggest that earlier studies may be underpowered to adequately test the hypothesis.

1.7 Prediction of tumour volume regression

While TVR is emerging as a potential biomarker in NSCLC RT, the research thus far focuses on retrospectively assessing measured regression and correlating that to clinical outcomes. Prospective prediction of TVR would be valuable to facilitate informed decision making at an earlier timepoint and limited research has been published in this field.

Brink et al predicted tumour volume in a cohort of NSCLC patients treated with 60-66Gy in 30-33# at the timepoints of one third, two thirds and at the end of RT, by fitting measured volumes on CBCT with an exponential function of time and solving it forwards [68]. Good agreement was observed using two thirds of treatment data to predict TV at end of RT with R^2 of 0.95, and less agreement was reported using the first one third of RT data (R^2=0.73).

A more complex model, forecasting both tumour volume and vascular responses, was developed and tested in animals using functional MRI assessments. The selected model, from 10 tested, resulted in tumour

volume differences of less than 12% between the measured and simulated volume and a strong spatial agreement based on a Dice Similarity Coefficient (DSC) > 0.74 [77].

Prediction of changes in volume and shape during RT were reported by Amungongo et al. Their approach extracted intensity values from target volume voxels on CBCTs acquired during weeks 1 and 2 of RT, and extrapolated the time domain of 4 potential models (Gaussian, quadratic, cubic and linear) to predict the intensity values in week 3 and 4 of RT. The linear model performed best at predicting patients who may require adaptive RT in week 3 due to tumour shrinkage [78].

1.8 Proliferation Saturation Index model

The Proliferation Saturation Index (PSI) model is a mathematical model which aims to predict tumour regression over the course of RT delivery. The model, originally proposed by Prokopiou et al. [79-83], utilises longitudinal tumour volume measurements to calculate a patient-specific PSI value, and uses this index to simulate and predict response to RT. The PSI model allows us to simulate tumour response to RT using an initial dose fractionation schedule, or test response to alterative prescriptions if modelled response is suboptimal. We can then theoretically optimise the RT prescription for each patient as demonstrated for dose personalisation in head and neck cancer [84].

1.8.1 Model summary

The PSI model is a patient-specific, single variable response model [79]. The model is based on the ecological concept that tumours exist in host environments with limited resources and therefore each environment has a specific carrying capacity *(K)*. This carrying capacity is patient specific as it is relevant only to that tumour and its microenvironment [80]. PSI

is defined as the volume of the tumour at the start of RT divided by the carrying capacity *(K)*. It follows that tumour volumes close to their carrying capacity, with a high PSI, would have a small proportion of proliferating cells, due to reduced availability of resources. These high PSI tumours are anticipated to show a shallow response to RT, as radiation sensitivity is related to proliferation [34]. Conversely, tumours with a low PSI would have ample access to resources, be highly proliferative, and therefore respond quickly to irradiation and show greater regression.

The PSI model can dynamically evaluate the change in tumour volume over a given time for specific RT schedules. The model parameters are calibrated using tumour volume measurements from complete courses of RT. Using this information, the PSI model can determine the population growth rate, λ [unit 1/day], and a radiation sensitivity value, α [1/Gy]. The only free parameter for each patient is then PSI. For each subsequent patient, the model is solved on tumour volume data acquired early during RT using the population level growth and radiation sensitivity rates to determine the individual PSI and predict the individual tumour volume regression in response to the remainder of the treatment course.

1.8.2 Mathematics of the PSI model

In the PSI model [79-81], tumour growth is modelled as logistic growth using the differential equation $\frac{dV}{dt} = \lambda V\left(1 - \frac{V}{K}\right)$ [81], where λ [1/day] is the intrinsic growth rate of the tumour, V is the gross tumour volume [cc] and K is defined as tumour carrying capacity [cc]. PSI is defined as the ratio of tumour volume at the start of RT (V_0) to the carrying capacity (K) $PSI = \frac{V_0}{K}$. PSI is therefore a unitless value between 0 and 1, and is representative of the proportion of the tumour volume not proliferating. Response to RT is modelled following the Norton-Simon hypothesis that RT response is proportional to the tumour growth rate [85], using the

linear quadratic model $SF(d) = e^{-\alpha d - \beta d^2}$ where *SF(d)* is the surviving fraction of cells following a given dose of radiation, *d* [Gy], and α [1/Gy] and β [1/Gy^2] are radiation sensitivity parameters. Finally, the change in volume following a dose of RT is modelled as an instantaneous reduction $V_{post\ RT} = V_{pre\ RT} - V_{pre\ RT}\,\gamma(1 - \frac{V_{pre\ RT}}{K})$ where γ is the volumetric death term per RT dose and is linked to the LQ model $\gamma = 1 - SF(d)$. For NSCLC an α/β ratio of 10 Gy is recognised and so the model only estimates the α parameter. Of note, α is the volumetric radiosensitivity, and thus not comparable with *in vitro* estimates of radiosensitivity values.

Data required to determine these values are longitudinal tumour volume measurements, the associated timepoints for each measure in days since start of RT, and the RT prescription. The PSI model can determine individual values of λ, α, and PSI (given a set α/β ratio), or alternatively set values for λ and α can be input to only allow PSI as a patient specific individual parameter. Finally, tumour regression can be modelled beyond the input data to simulate tumour response to RT from an early timepoint.

1.8.3 Potential clinical use of PSI model

Whilst change in tumour volume is a vast oversimplification of the underlying tumour growth mechanisms and radiobiological response, this simplicity is also the appeal of the PSI model, as there is only a single patient specific variable. While a plethora of more complex mathematical models have been developed to simulate radiation response [86-93], it is clinically infeasible to calibrate multiple interdependent parameters and make patient-specific predictions with limited uncertainty from few clinical measurements [94, 95]. The appeal of the simple PSI model is that it can predict tumour response to RT early in treatment, which could be a method of triggering plan or dose adaptation. The model enables

simulation of response to a variety of prescriptions and could inform selection of the optimal prescription for individual patients.

The PSI model has been evaluated in a small retrospective cohort of NSCLC patients [79, 80] and has been explored more thoroughly in the head and neck RT setting where it is shown to accurately predict tumour volume dynamics during RT [81, 82, 96]. A prospective clinical trial in head and neck RT has demonstrated promising results in using PSI to optimise prescription and clinical outcomes [97]. Robust evaluation of the PSI model in the NSCLC setting is warranted to evaluate its potential in this disease site. The Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement provides a clear framework to complete this process [98].

1.9 Lung tumour delineation

One difficulty in implementing a model such as PSI is the requirement of repeated delineation of the GTV to measure volumetric changes. Target volume definition remains a challenge in modern radiation therapy [99] and in NSCLC these challenges are exacerbated by breathing motion, proximity to mediastinal structures of similar density to tumour and regions of atelectasis [100]. For treatment planning purposes these challenges are somewhat alleviated by the incorporation of 4-Dimensional CT (4DCT) [101], PET imaging [102] and peer review processes [103]. However, to train the PSI model repeated, longitudinal measures of GTV are required and so GTV must be defined multiple times for each patient. It is not feasible, from either a patient intervention or resource perspective, to replicate this intensive target volume delineation process repeatedly over the course of conventionally fractionated RT.

Studies that have investigated tumour regression, which required repeated GTV delineation describe 3 main approaches to overcome this difficulty.

Firstly, acquisition of additional diagnostic or planning CT's at weekly intervals or at a pre-specified timepoint to measure regression have been reported [74, 104] . While this may generate the most accurate volumes, this leads to additional imaging dose for the patient and is resource intensive for a department. A second approach is the manual delineation on repeated CBCTs by a Radiation Oncologist (RO) [105, 106], but again this is resource intensive and not feasible beyond the context of a clinical trial procedure. Finally, in-house deformable registration solutions have been proposed to deform the original GTV onto the relevant CBCT to generate a new volume [71]. This solution while innovative, is not readily available to all departments or researchers. To accurately train and test models, large datasets are required, and a more pragmatic solution is required.

Radiation Therapists (RTTs) are experienced delineating on planning scans and while their role traditionally involved delineation of OARs, it has expanded into target volume delineation [107] with data showing clinically acceptable agreement between RO's and RTTs in a number of clinical sites and imaging modalities [108-110] including lung GTV [111]. With the move towards automation in delineation RTTs can also competently adjust manual contours to achieve clinically acceptable volumes [112]. One solution to the challenge of gathering large quantities of data from routine CBCT could be the use of a commercially available tool to automatically delineate the target volume and for manual adjustments to this volume be carried out by the treating RTT.

1.10 Central hypothesis of this thesis.

The key hypothesis in this thesis is that early response dynamics can be extracted from routine imaging and used to predict tumour volume regression at the end of treatment and identify patients at risk of poor prognosis. This would enable the personalisation of treatment.

CHAPTER 2

THESIS AIM AND METHODOLOGY

2.1 Overall aim

The aim of this thesis was to establish if it is possible to both quantify and predict tumour volume dynamics longitudinally during RT as a prerequisite to the optimisation of treatment approaches for each patient. A brief narrative summary of each study is noted below with full aims and methods outlined in the relevant chapters

2.1.1 Literature review

An Overview of Personalised Prescriptions in Locally Advanced Non-Small Cell Lung Cancer; Are we there yet?

First the literature was analysed to examine the potential of personalised prescriptions in the treatment of LA-NSCLC and establish the approaches being reported. It included both clinical studies and research proposing methods to personalise the dose fractionation schedule. A systematic search of the literature was carried out in PubMed and Embase databases to identify relevant studies for inclusion.

2.1.2 Study 1

Geometric and dosimetric evaluation of a commercially available auto-segmentation tool for gross tumour volume delineation

Next the potential of semi-automated GTV delineation was examined. Study 1 sought to ascertain the accuracy of a target volumes generated

with a semi-automated process on high-quality images acquired over multiple timepoints during RT. Data were acquired from an anonymous, open access database, The Cancer Imaging Archive (TCIA) [113]. Lung tumour GTVs were delineated on weekly 4DCT datasets, and the automated contours were compared to manual contours, for both geometric agreement and dosimetric differences over the course of treatment.

2.1.3 Study 2

Monitoring lung tumour volume on CBCT

Following testing of the delineation process on 4DCT, Study 2 piloted the GTV delineation methodology described in Study 1 on CBCT datasets. CBCT images are inherently of poorer quality than 4DCT and in this chapter the potential of delineating on CBCT is explored and a list of exclusion criteria that prevent delineation of GTV on CBCT is generated. In those patients where delineation was feasible, efficiency of the process was recorded as time required to adjust the auto-contour, and the structures were audited by a clinician to evaluate the accuracy of the volumes.

2.1.4 Study 3

Training and testing the PSI model in non-small cell lung cancer

Study 3 reports on the training and testing of the PSI model across 4 different dose fractionation schedules. In line with TRIPOD statement this work is a Type 1a study, describing the development of a prediction model [98] . This study determined the model's performance at fitting data in these prescription groups, with one or two free parameters, and assessed if it accurately predicts TVR in these groups. Data fitting and predictions were assessed using the Coefficient of Determination (R^2),

Pearson Correlation Coefficient (PCC), Mean squared error (MSE) and absolute volume difference between simulated and measured volumes.

2.1.5 Study 4

External validation of the PSI model performance in predicting tumour volume regression in NSCLC

Study 4 was a validation study in an external dataset of the PSI model in NSCLC, a TRIPOD Type 4 study [98]. Model performance was evaluated using R^2 and PCC. Additionally, this chapter examines the clinical relevance of TVR in the studied population using Kaplan Meier (KM) plots, Univariable and Multivariable Cox Proportional Hazards Regression. Receiver Operating Characteristic (ROC) curves and associated Area Under Curve (AUC) values were utilised to determine the predictive power of clinical and volumetric metrics in relation to OS.

2.1.6 Study 5

Developing the PSI model in the pre-treatment setting in NSCLC

Finally Study 5 examined the potential of the PSI model in the pre-treatment setting. Furthermore, the clinical relevance of simulated TVR was examined, and *in-silico* testing of alternative dose fractionation schedules was performed.

2.2 Patient cohorts and research ethics

This thesis included 5 experimental studies, which used 3 different cohorts of patients with NSCLC.

Study 1 utilised data from the TCIA database and as such did not require ethical approval [114, 115]. The TCIA database, a repository of radiology and histopathology imaging, was formed in 2011 and is funded by the National Cancer Institute Cancer Imaging Programme [116].

Studies 2,4 and 5 utilised a pseudonymised subset of the NI-HEART dataset [117]. Governance approvals were provided, and ethical approval waived, by the Belfast Health & Social Care Trust (BHSCT) for these studies (IRAS 293181). A data sharing agreement between Trinity College Dublin and BHSCT to facilitate the transfer of pseudonymised tumour volume and clinical data was put in place (Appendix 10.7) and a data protection risk assessment was carried out and approved by the TCD Data Protection Officer, Appendix 10.8.

Study 3 utilised two fully anonymised datasets containing tumour volume measurements for 162 patients with NSCLC abstracted from prior publications [71, 118] and as such did not require ethical approval.

2.3 Source of funding

A research grant from Varian Medical Systems, (grant number 011871), partially funded this work. Varian Medical Systems had no input in this study design, data analysis or thesis preparation.

CHAPTER 3

LITERATURE REVIEW

This literature review was published in 2018 in Radiotherapy and Oncology.

Barrett S., Hanna G.G., Marignol L., An overview on personalisation of radiotherapy prescriptions in locally advanced non-small cell lung cancer: Are we there yet? Radiotherapy and Oncology, 2018, 128(3): p520-533, See Publications for a copy of the published article.

3.1 Introduction

NSCLC is the most common form of lung cancer and with poor treatment outcomes and an age-standardised 5-year survival reported to be 10-20% world-wide [23]. Early stage disease treatment for medically inoperable patients has been revolutionised with the advent of stereotactic ablative body radiotherapy (SABR), which has become a standard of care for patients with medically inoperable NSCLC and which has comparable outcomes to surgery for Stage I disease [119, 120]. Outcomes in the locally advanced (LA) setting remain poor with little improvement in OS over the past 40 years [24]. RT is a mainstay of treatment for LA-NSCLC and is delivered either as radiotherapy alone or in conjunction with chemotherapy either concurrently or sequentially. An established dose fractionation of 60-66 Gy in 30-33 fractions remains the RT standard of care [30] for patients with LA –NSCLC.

The benefit of RT dose escalation has been explored and a dose response relationship has been reported in the setting of curative intent RT for NSCLC [37]. Theoretically, dose escalation is key to improving LC, PFS

and OS. The randomised control trial RTOG 0617, sought to compare outcomes following standard dose RT of 60Gy versus a higher dose of 74Gy in LA-NSCLC [38]. Dose escalation was found to be associated with worse outcomes and poorer OS. A more recent meta-analysis investigating RT dose response relationships in NSCLC reported survival benefits for escalation in the RT alone setting and no survival benefit in the CCRT setting [121].

It has been hypothesised that any potential benefit of dose escalation has not translated clinically in the findings of the RTOG 0617 due, at least in part, to the increased toxicity induced from both the RT and chemotherapy regimens as well as potential variation in RT delivery and quality assurance [122, 123]. The overall extension of treatment duration in a conventional 2Gy per day fractionation schedule, may be contributing factor to the poorer outcomes given that longer treatment times may facilitate tumour repopulation. As a result adapting the dose fractionation schedule, without extending overall treatment duration, has been suggested as an alternative approach [124].

While the evidence is somewhat conflicting, it seems clear that universal dose escalation in a linear fashion is not the solution for this patient cohort. It thus follows that adopting a personalised, patient specific approach to modifying the radiotherapy prescription may permit safe RT dose escalation [124].

While recent advances in RT technology have been vast, the evolution is now moving away from increasing the geometric capabilities of the equipment and towards personalising and individualising treatment parameters to a patient's specific needs and requirements [125]. The poor OS associated with the treatment of LA-NSCLC with RT call for an urgent need to transform outcomes for patients. Personalisation of

radiotherapy dose prescription and safe dose escalation may be one strategy to address this.

Isotoxic treatment planning has been suggested as a method of personalising radiotherapy for each patient based on their individual treatment plan [126]. In isotoxic RT planning, the prescribed dose is increased incrementally until an organ at risk (OAR) maximum tolerated dose is reached. As a result, each patient will have a personalised prescription. However, this approach of dose escalation is not based on any patient or disease specific features.

Several methodologies have been reported in the literature to modify LA-NSCLC prescriptions, which may prove more successful than RTOG 0617 in improving outcomes for these patients. This review aims to assess both the on-going clinical work being carried out in developing modified dose fractionation schedules, as well as evaluate emerging novel proposals to personalise LA-NSCLC radiotherapy prescriptions.

3.2 Search strategy methodology

A systematic search of the literature was carried out in PubMed and Embase databases using the key words 'Non-small cell lung cancer' 'Personalised Radiotherapy' 'Dose Optimization' and associated Mesh Terms. See Appendix 10.1.1 for complete search strategy. A scoping search of the literature was also performed.

Inclusion criteria were defined as personalised radiotherapy studies relating to NSCLC (including a pre-treatment change in dose per fraction, escalation, de-escalation based on TCP or NTCP, modelled personalised radiotherapy studies and studies published after January 2000. Exclusion criteria were defined as any other site (including SCLC), adaptive studies based on mid-treatment response rather than pre-treatment personalisation, change to chemotherapy prescription, no change to RT

prescription, stereotactic radiotherapy studies, protons/carbon ion studies, genetic or biomarker-based studies and non-EBRT escalation studies. Studies that reported prescription modification for stratified groups of patients were excluded and only those that individually adjusted the prescription on a per patient basis were included.

Due to heterogeneity between the study types and a limited sample size secondary analysis of the data was not feasible. As a result, the findings were evaluated in a qualitative manner. Nineteen relevant publications were identified (Figure 3-1) [50, 79, 127-143].

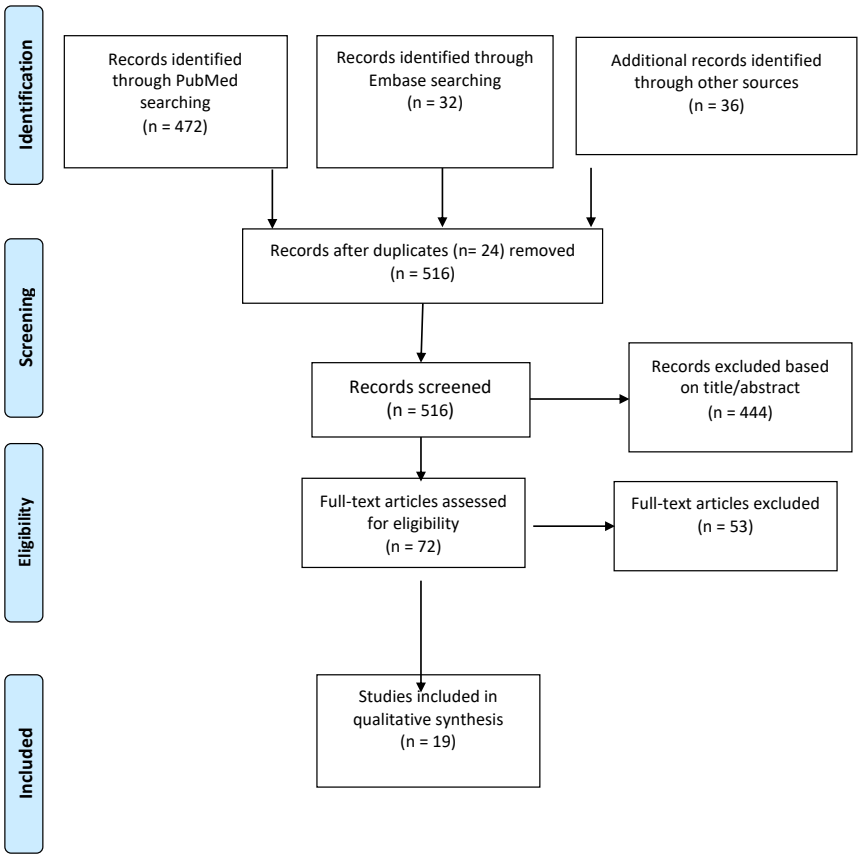


Figure 3-1 Prisma Flow Diagram

3.3 Results

3.3.1 Feasibility of dose modification in LA-NSCLC – *in-silico* studies

Five retrospective treatment planning studies [127-131] were identified which evaluated the feasibility of adapting the NSCLC RT prescription dose, based essentially on isotoxic planning and established OAR dose volume constraints. Findings from these five studies are summarised in Table 3-1.

Two studies [127, 128] compared the effectiveness of RT techniques for the purpose of dose modification by evaluating 3D-Conformal RT (3DCRT), Intensity Modulated RT (IMRT), simultaneous integrated boost (SIB) and adaptive RT (ART) techniques. The remaining three [129-131] focused solely on the dose adaptation using their current institutional planning techniques. All five studies concluded that a form of prescription personalisation via uniform dose escalation or SIB is achievable while adhering to accepted dose volume constraints on surrounding normal tissue. Mean Lung Dose (MLD) was the most frequently reported dose limiting parameter.

While all five planning studies [127-131] reported that dose modification was feasible in most patients with NSCLC Stage Ib-III, this was a small cohort of patients, n=101 in total. Additionally, a variety of technologies (3DCRT, IMRT, Volumetric Modulated Arc Therapy (VMAT) and helical IMRT), techniques (ART, SIB), and dose modification approaches (1.8Gy twice daily, 2.5Gy-5.4Gy daily) were reported making these findings difficult to interpret in a clinically meaningful way. While the treatment prescription was personalised to the patient based on their normal anatomy and target volumes, the overarching aim was to escalate the dose for all patients rather than ascertain differences between those

that may or may not benefit from a modified dose prescription. RTOG 0617 has demonstrated that this may not be beneficial to all patients [38].

3.3.2 Clinical evidence for personalised radiotherapy prescriptions

Following the establishment of feasibility, prospective clinical trials were implemented to evaluate the use of personalised prescriptions in a patient population. Five prospective clinical trials [50, 132-135] were identified that utilised isotoxic treatment planning techniques to individually prescribe RT for LA-NSCLC. Findings from these five trials are summarised in Table 3-2.

There was sizeable heterogeneity between these trials. The study by van Baardwijk et al. (2010) and Wanet et al. were prospective single arm studies [132, 135] whilst the remaining three studies were Phase I and Phase II clinical trials [50, 133, 134]. The patient population inclusion criteria varied. Two studies [50, 132] included only patients receiving RT, whereas patients receiving concurrent CRT were included in the remaining three studies [133-135]. Disease stages I-IV were included over the five studies. There was heterogeneity reported in the RT techniques utilised. Three studies [132-134] report the use of 3DCRT techniques and Electronic Portal Imaging Device (EPID) imaging in contrast with the more advanced techniques of IMRT and tomotherapy reported in two trials [50, 135]. RT prescription was modified for each patient based on their individual plan features, mainly normal tissue dose volume constraints or FDG PET uptake values (see Table 3-2 for details). The resultant prescriptions ranged between 1.5 Gy-1.8 Gy twice daily and 2.28 Gy-4.8 Gy daily, with treatment duration ranging from 25 to 40 days.

Interpreting these trials in a clinically meaningful way is difficult due to the vast heterogeneity observed between patient and disease characteristics as well as the technical approaches taken. One of the main sources of clinical variation between these studies is the parameters used

to determine the personalised dose. Standard normal tissue dose volume constraints (DVCs) were used as dose limiting parameters in two studies [132, 133] which had adopted a twice-daily accelerated fractionation schedule. One study [50] reported the mean fraction-normalized lung dose divided by the normalized prescription dose as the dose limiting constraint. One study reported DVCs in Equivalent 2Gy per fraction (EQD2) doses [135]. The remaining trial [134] consisted of 2 trial arms, one with an escalating oesophageal dose limiting constraint. In the other treatment arm dose limiting parameters were modified based on the oesophageal toxicity outcomes from the first cohort. As a result, any direct comparative analysis between these studies is difficult and potentially unreliable.

The clinical outcomes reported in these trials were generally positive, with three of the five trials reporting an increased rate of OS with increased dose, which resulted from personalising the prescription. One study [50] found no significant difference in OS between the 57Gy and 63.25Gy cohorts. One study did not report OS in relation to dose modification [135]. Varying levels of toxicity were reported, ranging from mild and transient to Grade 5 fatal toxicities, depending on target location and dose delivered.

The findings presented suggests that dose escalation is not only clinically feasible, but also potentially desirable in terms of improved outcomes (Table 3-2). An increase in OS was associated with an increase in dose in three of the five studies. However, significant toxicity was reported in all studies with Grade 5 adverse events also reported in four of the five studies. Cannon et al [50] reported that long term follow-up revealed unacceptable late toxicity and recommended a maximum tolerated hypofractionated dose of 63.25Gy in 25 fractions. This toxicity may well be due to the lack of established evidence-based DVCs in the moderately hypofractionated setting. QUANTEC guidelines caution the user on the

limitations of applying standard DVCs to modified fractionation schedules and highlighted the need for further work in this area [144]. A review of DVCs utilised in moderately hypofractionated RT for NSCLC highlights the lack of consensus in the clinical setting [145]. Additionally, it has recently been shown that heart dose is a key factor in predicting OS, however, the optimal DVC is yet to be determined [146].

One potential method of reducing this toxicity could be improving the dose conformity to the target using more modulated techniques. In these trials, a number of 3DCRT planning approaches were reported. It is possible that the use of isotoxic planning techniques which require the increased conformity that comes with IMRT or VMAT, could translate to reduced clinical toxicity and the published outcomes of the isotoxic IMRT trial are eagerly awaited [126]. However, two studies that reported Grade 5 haemoptysis utilised modulated techniques but studies observed this toxicity in patients with centrally located tumours [50, 135].

All trials summarised here modified the RT prescription to increase the dose delivered per fraction or per day, rather than extending the prescription in a linear 2 Gy per fraction manner. In spite of the negative outcome of RTOG 0617 [122, 123], these results are encouraging and warrant further investigation. However, these studies do not consider RT dose personalisation based on the molecular or genomic characteristics of the tumour thus there is potential for a further personalised medicine approach to the treatment of NSCLC with RT.

3.3.3 Future directions of personalised radiotherapy prescriptions

Nine studies [79, 136-143] were identified that outlined novel concepts which may prove to be a future basis for personalised RT prescriptions, findings are summarised in Table 3-3. These studies can be broadly

categorised depending on the patient specific feature they are utilising as a basis for personalisation.

3.3.3.1 Spatial relationship between target and OARs

Exploiting the patient specific spatial relationship or distance between target volumes and OARs in order to individually modify the dose prescription was the focus of two publications [136, 137]. Petit et al. [136] utilised retrospective planning data to model a patient specific maximum achievable dose, based on PTV to OAR distance. The model performed well in predicting the max dose in 79% of patients (n=89).

A similar concept utilised a spatio-temporal model [137]. This study considered the distance between target and OARs but also integrated factors, such as tissue α/β ratios and the linear quadratic model of response, to select an optimal prescription for an individual. This incorporated both the physical features of the patient's anatomy but also tumour doubling and lag times to predict the optimal fractionation schedule. It concluded that the biologically effective dose (BED) could be increased. However, in some patients, there was no significant difference in the tumour BED achieved with the model-determined optimal fractionation and the standard of care approach. The use of this model may result in a more personalised approach for patients, as it was successful in distinguishing between those who may or may not benefit from a dose prescription modification, albeit only in relation to BED. This may lead to improved stratification of patients and potentially reduce the toxicity observed in the clinical trials.

3.3.3.2 Targeting sub-volumes of disease

These studies focused on identifying a sub volume within the tumour that may represent a radio-resistant portion of disease. Once identified, a boost was prescribed to this volume only. One approach attempted to generate a mathematical model that would predict visible tumour response to

radiation therapy during treatment and consequently also predicted a residual disease volume that would remain at the end of RT delivery. This was generated from a database of treated patients was the focus of both a feasibility study [138] and a follow up analysis [139]. Disease response to treatment was predicted prior to treatment delivery and an individualised simultaneous integrated boost (SIB) to this 'residual' volume would then be prescribed at the outset. Dose escalation to the residual volume was determined to be feasible [138] but the cohort of patients was small (n=5). The follow up study n=10 [139] found that use of this predictive approach facilitated mean dose escalation of 10.4Gy, to the actual residual tumour volume. One limitation of these studies, similar to the planning studies and clinical trials, is that they assume that dose escalation is the most desirable strategy. Modification of the prescription is optimised to the patient's anatomy and the highest achievable dose, rather than predictive disease features.

3.3.3.3 Hypoxia targeting

Tumour hypoxic regions are known to be an area of radioresistance and as such they make an attractive target for dose modification studies. A radiobiology focused personalised model was proposed [140] which recommended a dose fractionation schedule that was based on a reoxygenation model, revised to also account for repopulation and repair. Changes in tumour hypoxia over the duration of treatment delivery was considered in this model, which lead to a heterogeneous fractionation approach, with the dose per fraction increasing after short breaks, such as weekends. This mathematically modelled study identified that a theoretical benefit in TCP was only observed in well-oxygenated tumours. In a hypoxic setting, the heterogeneous dose schedule was not comparable to the extreme hypofractionation as is commonly seen in the SABR protocols. At present, extreme hypofractionation does not seem a feasible approach for patients with LA-NSCLC, but this method of hypoxia targeting may still be clinically applicable in this cohort of patients. This

methodology of modelling tumour hypoxia would require routine hypoxic imaging, perhaps in a longitudinal setting, to accurately determine tumour oxygenation status, which is not widely clinically available.

A recent study [141] modelled individualised dose escalation based on either hypoxic sub-volume targeting or whole tumour dose escalation. This study evaluated the uniform vs. non-uniform dose escalation in mathematically modelled population exhibiting varying ranges of hypoxia. It was reported that uniform dose escalation was more beneficial at the early stages of treatment, followed by a hypoxia-targeted dose escalation at a later stage. It was postulated that this is due to the hypoxic regions changing rapidly in the early stages of RT delivery. Again, routine hypoxia imaging would be required to implement such a technique.

3.3.3.4 TCP/NTCP models

Two studies [142, 143] incorporated radiobiological modelling into the treatment planning process in order to select the most appropriate dose fractionation schedule for the individual case. The basis for the selection was TCP and NTCP models [142] and a novel 'Bifurcation Number' [143]. Both studies reported the proposed methodologies to be beneficial in selecting an optimal prescription.

BioSuite, a novel software program, was developed and was the subject of a feasibility study [142], which incorporates a number of TCP and NTCP models into the RT prescription process. The method proposed dictates that an optimal treatment plan is created and exported to BioSuite. Based on the patient specific anatomy and the individual's treatment plan BioSuite identifies the optimal dose fractionation considering TCP and NTCP models to maximise the therapeutic window for the patient. The concept was used in the treatment of two patients with NSCLC and utilised lung TD_{50} and oesophageal max dose as the dose limiting constraints. The software calculates a desirable dose

fractionation schedule suited to the individual. The individualised approach is clear here, however it is reliant on an optimal dosimetric plan being produced to determine what can be achieved. As such is potentially limited by the skill of the individual dosimetrist. This uncertainty may be mitigated by the introduction of automatic planning techniques, which have been shown to produce high quality treatment plans in patients with late stage NSCLC [147].

The novel concept of a Bifurcation Number has been developed by Keller et al. [143], and is defined as a dose ratio between the tumour and one OAR. The Bifurcation Number was then compared to the α/β ratios between normal tissue and tumour, to determine the optimal fractionation schedule for an individual. This methodology begins to take into account patient specific features as well as known radiobiological principles. Furthermore, this option indicates the optimal dose modification strategy, which is in keeping with the desire to tailor treatment to the patient. However, as only one OAR is considered in these calculations, this may be a limiting factor in the clinical application of this concept.

3.3.3.5 Modelled proliferation rate

A patient specific Proliferation Saturation Index (PSI) was the focus of one study [79], which is generated from tumour volume changes over time and the host organ's tumour carrying capacity. The carrying capacity is an ecological concept that defines the maximum ecosystem (tumour) that a host environment can sustain (patient) and is often used in mathematical modelling of cancer dynamics. It was hypothesised that a high PSI indicates a tumour volume is close to the host's maximum carrying capacity and hence it would have a low proportion of cells rapidly proliferating. This in turn would indicate a poor sensitivity to RT and therefore a need for a modified prescription. The concept was tested on a mathematically modelled population (n=500) and four clinical cases. It was reported that the PSI correlated with tumour volume reduction

during RT when using standard fractionation schedules. While promising, this study was limited by the lack of patient data. Further work on this concept has been encouraging in predicting treatment early response in patients with oropharyngeal carcinoma [148] however, clinical evaluation is necessary to ascertain the potential benefit of this concept in alternative fractionation schedules. The most recent work published on this concept highlights the dependence of the model of tumour growth law utilised for the calculations [149]. One advantage of the PSI is that it is calculated from routinely acquired CT datasets that are used for diagnostic and planning purposes. As a result, this may have little impact on departmental resources or the patient pathway through treatment.

3.4 Discussion

In the literature reviewed, a variety of methodologies have been proposed and investigated to personalise RT prescriptions in patients with LA-NSCLC, however, most are not yet being tested in the clinical setting. One clear reason for this lack in clinical progression is the lack of robust *in-silico* evidence. Many of the proposed models have been tested on mathematically generated populations in an academic setting. Further work is required to translate these potential models into clinical RT treatments.

The retrospective planning studies reviewed have shown that dose modification is feasible and hence implementation of this concept is clinically possible. Some clinical evidence has been gained from a small number of trials looking at the alternative dose fractionation schedules, though these are largely focused on dose escalation, mainly exploiting the spatial relationship between targets and OARs. To an extent, this approach is personalised to the individual's gross anatomy. However, to truly personalise the treatment we should also target disease features that are identifiable at a cellular level.

Furthermore, a true person-centred approach is lacking in this research, with the majority of studies and proposals aiming to simply escalate dose without a focus on the patient characteristics. A recent publication [150] highlights the vast heterogeneity within this population and the myriad of factors that are not yet considered in our treatment approach. This may well provide a starting point for improved patient stratification and thus personalisation. Whilst our efforts should not move entirely away from personalising the RT plan, we cannot continue to do this without also giving equal consideration to patient related factors.

One potential barrier of implementing personalised radiotherapy prescriptions is the resource burden of such approaches. Clinical

departments do not have unlimited resources available to deliver an entirely personalised approach to every patient at present. Those concepts and theories which have a minimal impact on workflow and resources, are most likely to be the first introduced to clinical practice, if shown to be beneficial

3.5 Conclusions

Personalised medicine is fast becoming a basic requisite in the modern era of cancer treatment. The treatment of LA-NSCLC with RT is an area of on-going investigation in the pursuit of a personalised approach. Much of the data identified in this review was retrospective planning studies and this limits the strength of any conclusions. The current clinical evidence, although promising in terms of improving outcomes, is lacking sufficient patient numbers or focus on personalisation to lead to a change in clinical practice and results from ongoing trials are eagerly awaited. Several encouraging strategies have been identified but do not yet provide the level of evidence required to inform clinical practice and a renewed effort to personalise RT dose prescription in the treatment of NSCLC is needed.

3.6 Tables

Tables legend
2D: 2 Dimensional
3DCRT: 3-Dimensional Conformal Radiation Therapy
3DCT: 3 Dimensional Computed Tomography
4DCT: 4 Dimensional Computed Tomography
ART: Adaptive Radiation Therapy
BED: Biologically Equivalent Dose
CBCT: Cone Beam Computed Tomography
CT: Computed Tomography
Dmax: Maximum point dose
Dmean: Mean dose
DV: The absorbed dose in Gray that covers a specified percent volume V
DVC: Dose Volume Constraint
DVH: Dose Volume Histogram
EPID: Electronic Portal Imaging Device
EQD2: Equivalent Dose in 2 Gy per fraction
EUD: Equivalent Uniform Dose
FDG PET: Fluorodeoxyglucose Positron Emission Tomography
GTV: Gross Tumour Volume
Gy: Gray
HR: Hazard Ratio
HX4: 3-[18F]fluoro-2-(4-((2-nitro-1H-imidazol-1-yl)methyl)-1 H-1,2,3-triazol-1-yl)propan-1-ol (Hypoxia tracer)
IGRT: Image Guided Radiotherapy
IMAT: Intensity Modulated Arc Therapy
IMRT: Intensity Modulated Radiation Therapy
ITV: Internal Target Volume
MLD: Mean Lung Dose
NSCLC: Non-Small Cell Lung Cancer
NTCP: Normal Tissue Complication Probability
NTDmax: maximum normalized total dose
NTDmean: mean normalized total lung dose: MLD, adjusted for the dose per fraction
OAR: Organ at Risk
OS: Overall Survival
OTT: Overall Treatment Time
PET-CT: Positron emission tomography- CT
PFS: Progression Free Survival
PRV: Planning Organ at Risk Volume
PSI: Proliferation Saturation Index
PTP: Predictive treatment planning
PTV: Planning Target Volume
rNTDmean: mean fraction-normalized lung dose divided by the normalized prescription dose
RT: Radiation Therapy
RTPN: Radiation Therapy Pneumonitis
SBRT: Stereotactic Body Radiation Therapy
SIB: Simultaneous Integrated Boost
TBR: Tumour-to-Background Ratio
TCP: Tumour Control Probability
Td: Tumour doubling time
TD50: dose that would result in 50% probability of developing complications
TED: Tumour Effective Dose
Tk: Tumour lag time
TTD: Total Tumour Dose
VD: The percentage volume that receives a dose of at least D in Gray.
VMAT: Volumetric Modulated Arc Therapy
WHO PS: World Health Organisation Performance Status

Table 3-1 Feasibility of dose modification in LA-NSCLC – *in-silico* studies

Author, Year (ref)	**Study Type**	**Participants n=()**	**Participants stage**	**Target Definition**	**Target Volumes**	**Dose Escalation Limiting Parameter**	**Maximum Allowed Dose**	**RT Technique**	**Dose Achieved**	**Dose Achieved in biologically equivalent 2Gy fractions (EQD2,T)**	**Study Conclusions**
Warren, 2014 (14)	Retrospective planning study (3DCRT vs. IMRT for isotoxic planning	n=20	II-III	union of GTV delineated on all phases of a 4DCT	GTV and involved nodes	Spinal cord PRV<50 Gy to 1 cm3, Lungs (excluding GTV) MLD <20 Gy, Brachial plexus PRV<66 Gy to 1 cm3, Proximal bronchial tree and great vessels PRV <75.1 Gy to 1 cm3	79.2Gy/ 1.8Gy per fraction twice daily	3DCRT vs. Inverse-planned CRT vs. IMRT (step and shoot)	70.2Gy/ 39 fractions, twice daily RT.	69.0 Gy EQD2	Use of IMRT allows for higher isotoxic dose escalation than 3DCRT
Guckenberger, 2012 (15)	Retrospective planning study (Only Isotoxic dose escalation section included)	n=13 n=1 Radiotherapy, n= 12 Concurrent chemo RT	III	Planning CT and FDG PET	GTV and involved nodes (no elective nodal irradiation)	MLD level achieved with a prior 3DCRT plan	not stated	IMRT ±SIB ± ART	average for n=7 planned with IMRT + SIB+ ART 78.8Gy ±2.7Gy SD	not reported	Isotoxic SIB adaptive replanning is a promising strategy to safely escalate dose
Even, 2015 (16)	feasibility study comparing various boost volumes	n=10	Iib-IIIb	Mid ventilation phase of 4D FDG PET (Boost volumes SUV50% max and HX4 TBR > 1.4 or whole tumour)	GTV and involved nodes	Lungs Dmean < 20 Gy (corrected to EQD2); spinal cord D0.1 < 51 Gy (EQD2 < 52 Gy); oesophagus V36 < 80%; brachial plexus D0.1 < 66 Gy (EQD2 < 66 Gy); whole heart Dmean < 46 Gy (EQD2 Dmean < 46 Gy); planning organ-at-risk volume mediastinal	129.6Gy/24 fractions to PTV boost volume	VMAT SIB to FDG and HX4 volumes	The average prescribed doses to the boost volume were 87.1 ± 10.1 Gy (PTV Prim), 107.3 ± 20.6 Gy (PTV FDG) and 117.6 ± 15.2 Gy (PTV HX4).	not reported	Statistically higher prescriptions achieved for boost volume plans than primary alone. No significant difference in OAR DVHs

						structures (OAR + 5-mm margin) D0.1 < 76 Gy (EQD2 < 94 Gy), (D0.1 the dose delivered to 0.1% of the OAR)					
Hoffmann, 2012 (17)	retrospective planning study (increasing dose per fraction ± the number of fractions, up to 33, based on normalised total doses)	n=38	III	FDG PET CT	not reported	NTDmean ≤19 Gy, NTDmax, (Dmax in 2-Gy fractions using EQD2) oesophagus NTDmax 80 Gy, NTDmean 34 Gy;, spinal cord NTDmax 50 Gy; brachial plexus NTDmax 66 Gy; heart NTDmean 26 Gy	82.5Gy in 33 fractions	Step and Shoot IMRT	median total tumour dose of 74.2Gy (Range:67.5-79.9Gy) in 33 fractions	64.9Gy (Range 57.0-71.9Gy) TED in 2-Gy fractions corrected for overall treatment time. *note greater TED was achieved using 15 fraction schedule (67.0Gy (57.0-75.4 Gy)	Individualised prescription enabled dose escalation and theoretical therapeutic gain in 79% of cases
Van Elmpt, 2012 (18)	Planning analysis of Phase II Trial Randomized to Arm A boost whole GTV, Arm B, boost >50% SUVmax region only	n=20, prescription modification successful in n=15	I-IIIb	Mid ventilation planning CT dataset in conjunction with co-registered FDG PET CT	GTV only for dose escalation (nodal volumes given non-modified dose of 66Gy/24 #	MLD 20Gy	129.6 Gy/24 fractions once daily (max 5.4 Gy per fraction)	Step and Shoot IMRT	Average prescribed dose Arm A: 78.4 ± 7.6Gy [72.2–102.7Gy] for 24 fractions Arm B: 87.2 ± 13.1Gy [76.8–129.6Gy] for 24 fractions	For whole population: average 77 Gy in 24 fractions were achieved (EQD2 of 85 Gy, α/β 10 Gy) for the entire primary tumour, or for the boost region up to 87 Gy in 24 fractions (EQD2 of 99 Gy, α/β 10 Gy)	Dose-escalation by boosting the primary tumour or the high FDG uptake region within the primary tumour as feasible in 75% of the patients analysed

Table 3-2 Clinical evidence for personalised RT prescriptions

Author, Year (ref)	**Trial Type**	**Participants n=()**	**Participants stage**	**Target Definition Dataset**	**Target Volumes**	**Dose Escalation Limiting Parameter**	**Maximum Allowed Dose**	**RT Technique**	**IGRT technique**	**Dose Achieved**	**Dose Achieved in biologically equivalent 2Gy fractions (EQD2,T)**	**Outcomes (LC/OS)**	**Outcomes (Toxicity)**
van Baardwijk, 2010 (19)	Prospective, single arm study	n=116, no concurrent chemoRT	I,& II if inoperable, III	Fused PET-CT with planning CT	GTV and involved nodes (no elective nodal irradiation)	MLD dependant on lung function between 10 and 19 Gy, Cord Dmax 54Gy ±0.5Gy, Great Vessels or Bronchi Dmax 70.02Gy, brachial plexus Dmax 66Gy	79.2Gy. 1.8Gy per fraction, bidaily treatment, 8 hours apart	3D CRT	EPID	Median prescribed TTD = 64.8±11.4Gy with OTT 25 ± 5.8 days (range, 2 to 50 days)	66.0 ± 7.1 Gy (range, 51.9 to 73.1 Gy)	OS was higher with a higher EQD 2,T. Median OS was 21.0 months (95% CI, 15.8 to 26.2 months)	Acute toxicity was mild and transient (4.8% experienced grade 3 dysphasia, Late toxicity: Grade 3 and Grade 4 was reported in 10% of patients, mainly those with dyspnoea prior to treatment
van Baardwijk, 2012 (20)	Phase II Trial	137, Concurrent chemo RT	Stage III (Stage II n=1)	Fused PET-CT with Mid ventilatio	GTV and involved nodes (no elective	MLD 19.0 ± 1.0 Gy, Spinal cord Dmax 54.0 ±	69Gy (due to proximity to central	3D CRT	EPID	Median prescribed dose 65.0 ± 6.0 Gy (51–	53.9 ± 3.9 Gy (range 43.1–63.1 Gy),	OS was better with a WHO-PS of 0 (p = 0.009),	acute Grade 3 dysphagia 25.5%, Late Grade 3

				n planning CT dataset	nodal irradiation)	0.5 Gy, and Brachial Plexus Dmax 66 Gy, Oesophagus Dmax 75Gy	structures) 1.5 Gy fractions twice daily up to 45 Gy with an interfractio n interval of at least 8-hours, followed by once daily fractions of 2 Gy based on the ESPATU¨ phase III trial scheme			69 Gy) OTT of 35 ± 5.7 days (18–48 days).		smaller nodal volume (GTV-2; p = 0.015) and a higher EQD2,T (p = 0.037). Median OS 25.0 months (95%CI 19.8–30.3 months)	dysphagia 4.6%. Acute Grade3/4 dyspnoea 2.9% and 0.7% respectively , Late grade 3 dyspnoea 2.3%, 0.8% (n=1) grade 5 pneumonitis
Cannon, 2013 (21)	Phase I	n=79, n=75 evaluable per protocol patients (no concurrent chemo)	All stages eligible Stage I/II =9%, Stage IIIA=27%, stage IIIB =44%, Stage IV=13%, Recurrent patients = 8%	CT ±PET	GTV and involved nodes (no elective nodal irradiation)	rNTDmean. At a given dose level, patients with higher rNTDmean (and therefore higher bin number) were predicted to be at higher risk for radiation pneumonitis	Dose per fraction escalation to 57 Gy at 2.28 Gy/ fraction , 63.25 Gy at 2.53 Gy/fraction , 69.25 Gy at 2.77 Gy/fraction , 75 Gy at 3.0 Gy/fraction , 80.5 Gy at 3.22 Gy/fraction	IMRT/Helica l tomotherapy	not reported	median delivered dose was 57 Gy (range, 22.8 to 85.5 Gy) in 25 fractions.	60, 70, 80, 90, 100, and 110 Gy, respectively (reported in relation to the various dose escalation groups)	median overall survival rate was 16.0 months, no significant difference in local control in patients treated to 69.25 Gy and higher compared with those treated at the 57 and 63.25 Gy dose levels	no incident of Grade 3 acute or late oesophageal toxicity, no incidents of grade 3 or greater radiation pneumonitis . 6 incidents of grade 4-5 toxicity that were possibly related to RT (fatal haemoptysi

							, and 85.5 Gy at 3.42 Gy/fraction (all delivered in 25Fractions , once per day, 5 days per week)						s and lung abscess) found to relate to bronchial tree high dose regions and central tumours
Landau, 2016 (22)	Phase I/II	n= 84 with concurrent chemo (n= 82 evaluable patients	Stage II & II	3DCT or 4DCT	reported as 'gross tumour volume' no distinction made as to nodal volumes	Split into 2 groups 1: an escalating oesophageal dose constraint (progressively raised from 65 Gy to 68 Gy and then 71 Gy to 1cm3 of oesophagus). Group 2:lung and other normal tissue dose constraints (MLD EQD2mean 18.2 Gy, Heart D100% <45 Gy, D67% <53 Gy, D33% <60 Gy, Cord D0.1cc ≤47	73Gy/30 fractions, once daily over 40 days	3DCRT(n= 81) and VMAT (n=3)	not reported	overall mean of 67.7 Gy, with means of 68.9 Gy and 66.8 Gy for groups 1 and 2, respectively	not reported	median OS was 36.9 months (95% CI 31.7-42.1 months) he risk of death was lower for the higher tumour dose subgroup (HR 0.53, 95% CI 0.28-1.02, P=.06).	Grade 2 to 5 RTPN was seen in 30.5% of patients (n=3 Grade 3, n=0 Grade 4 or 5. The grade 2 to 5 esophagitis rate overall was 82.9% (n=5 Grade 3, n= 0 Grade 4, n=1 Grade 5) Oesophageal dose was limited to 68 Gy following Grade 5 toxicity in 71Gy cohort.

						Gy, Brachial Plexus D30%≤ 60 Gy, D0.1cc ≤ 65 Gy, Oesophagus initially 63Gy but increasing to 65Gy then 68 Gy as trial progressed							
Wanet, 2017 (23)	Prospective, single arm study	N=13	Stage II-III	Contrast enhanced 3DCT and 4D FDG PET to create PET ITV for boost	GTV and involved nodes, individualized boost prescribed to PTV_{PET} only (62.5Gy/25# prescribed to remainder of PTV)	Reported in EQD2: Lung- GTV, MLD dependent on FEV_1 (<10-19Gy,) V20 Gy <30%, V13 Gy<40%, V10 Gy<45%, V5 Gy <65%. Esophagus V40 Gy <50%,V50 Gy <30%, V70 Gy <20%. Cord D2% <50 Gy. Brachial plexus D2% <60 Gy. Large Vessels D2% <94 Gy.	4.8Gy/25# to PTV_{PET}	Helical IMRT or IMAT	Daily online correction using MV-CT or CBCT	PTV_{PET} reached 82.1 ± 17.9Gy for all patients, with values of 89.2 ± 24.0 and 75.0 ± 3.2Gy for peripheral and central tumors respectively	90.8 Gy, 100.8 Gy and 81.3 Gy for all patients, peripheral and central tumours respectively.	OS at 36 months was 52.8%, Overall PFS at 36 months was 34.6%, Local PFS at 36 months was 52.8% (from supplementary material)	Acute toxicity: grade 3 toxicities included esophagitis (23.1%), dyspnoea (23.1%), fatigue (30.8%), dermatitis (15.4%), anorexia (15.4%) and nausea (7.7%). Late toxicities: five patients presented grade 1 pulmonary fibrosis (38.5%), four

						Bronchi D2% <94 Gy. Whole Heart Dmean <46 Gy, V40 Gy <50%.							patients developed grade 2 pneumonitis (30.8%). 2 patients developed grade 5 haemoptysis.

Table 3-3 Future directions of personalised RT prescriptions

Author, Year (ref)	**Proposed parameter for individualisation (model based on?)**	**Brief overview of Concept**	**Participants n=()**	**Participants stage**	**Target Definition**	**Target Volumes**	**Dose Escalation Limiting Parameter**	**Maximum Allowed Dose**	**RT Technique**	**Reported Conclusions**
Petit, 2015 (24)	Patient anatomy and prior plans (OAR to PTV distance)	Optimal achievable DVHs were predicted based on OAR distance to PTV, based on this max achievable dose escalation is predicted	Training cohort n= 50, Validation Cohort n=39	II-III	not reported	GTV and involved nodes	OAR dose: MLD excluding GTV <20 Gy; plexus brachialis Dmax <70 Gy; heart Dmean<46 Gy; Dmax to the heart, mediastinal envelop and oesophagus <76 Gy; D0.1cc spinal cord <54 Gy.	69Gy/2Gy per fraction, once daily	VMAT	The method accurately predicted the mean lung dose within -0.11 ± 1.9 Gy (mean ± SD) and the maximum achievable dose escalation to the PTV, within one 2 Gy fraction for 79% of the patients.

Kim, 2015 (25)	Spatiotemporal optimization	Model to predict optimal dose based on spatial relationship of target to OARs and incorporating Td and Tk	16	n/a	n/a	n/a	Heart Dmean ≤ 45 Gy, lungs D mean ≤20 Gy, cord Dmax ≤45 Gy, oesophagus Dmax ≤63 Gy, and unspecified tissues D05 ≤60 Gy,	60Gy/30 fractions EUD	IMRT	Suggest that personalizing a fractionation schedule could potentially deliver a larger BED and EUD to tumours without exceeding the normal tissue BED DVCs with a large Td. For tumours with a small Td, there is no significant difference in tumour BED between an optimal fractionation and a standard fractionation. The results also showed that an optimal N ∗ (number of fractions) does not depend on Tk as long as N ∗ is larger than Tk .
Zhang, 2014 (26)	Predicted residual tumour based on prior patients treatment response (feasibility Study)	The predictive atlas should identify the location of residual disease following treatment, this volume is then prescribed an escalated personalised SIB	n=5	locally advanced	CBCT and planning CT	Tumour only	in house dosimetric OAR constraints	56-70Gy	IMRT	PTP incorporating a tumour regression model from the start, is technically feasible. PTP represents a new approach to increase tumour dose without increasing toxicities
Zhang, 2017 (27)	Predicted residual tumour based on prior	The predictive atlas should identify the location of residual disease following treatment, this volume is	n=10	locally advanced	RO and Medical physicist delineated	GTV Tumour	local standard DVCs (Lung - GTV Mean dose	no max	IMRT + SIB	The error in predicting the tumour shrinkage rate averaged 6% ± 13%

	patients treatment response	then prescribed an escalated personalised SIB			on Planning CT and last fraction CBCT		<20Gy, Cord Dmax < 50Gy			(mean ± standard deviation) compared to the true shrinkage rate. PTP maintains the same OAR DVHs, but yielded an increased PTVresd D95 and PTVresd Dmean to the actual residual tumour of 6.5 ± 1.8 Gy and 10.4 ± 2.9 Gy, respectively
Lindblom, 2015 (28)	Radiobiology modelling on *in silico* population	inclusion of repopulation and repair into a model that considers reoxygenation. A heterogeneous fractionation schedule was considered (increasing dose per fraction after short breaks)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Heterogeneous fractionation has only been found to be effective in the case of an overall well-oxygenated tumour. For the hypoxic tumour, a schedule employing 2- and 4 Gy-fractions could not compete with the SBRT-like schedules in terms of achieving high levels of TCP
Chvetsov, 2017 (29)	hypoxic areas as seen on functional imaging (uniform vs. non-uniform dose distributions) *in silico* trial	modelling of uniform vs. non uniform dose escalation using mathematical formulae accounting for hypoxia and re-oxygenation	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Uniform dose escalation is more efficient at the beginning of treatment when the density of clonogens is nearly uniform and the hypoxic tumour is subject to change quickly. Hypoxia-

										targeted dose escalation will be more effective when applied later, after the density of clonogens becomes non-uniform.
Uzan, 2012 (30)	BioSuite software (utilizing a number of TCP/NTCP models)	creating a standard plan, exporting to BioSuite and modelling TCP/NTCP based on the patient specific anatomy	n=2 (NSCLC)	I & III	not reported	GTV or GTV and involved nodes	NTCP parameters Lung TD50 = 29.2 Gy, Oesophagus Dmax = 72 Gy	not reported	3DCRT	Dose Response curves generated by BioSuite could indicate underutilised therapeutic windows. Radiobiologically guided prescription dose and fraction number customisation of external beam radiotherapy can be easily performed using BioSuite. The isotoxic dose and fractionation or 2D optimisation effectively convert all the various effects on TCP and NTCP into a single dose response curve
Keller, 2013 (31)	Bifurcation Number(B) 'a generalized dose ratio between the normal tissue and the tumour'	Calculate the bifurcation number to determine optimal fractionation 'If B is smaller than the ratio of alpha/beta ratios between normal tissue and tumour, then a single fraction is optimal; otherwise the	n=46	not stated (Conventional, hypofractionated and SBRT all included)	not reported	GTV and involved nodes	n/a	n/a	n/a	The bifurcation number may assist in selection of an appropriate fractionation protocol once the dose distribution has been optimized.

		optimal treatment is an infinite number of doses. These fractionation schedules correspond clinically to hypo- and standard/hyperfractionation,'								
Prokopiou, 2015 (32)	Proliferation Saturation Index (PSI)	A PSI is generated based on tumour volume changes between diagnosis and CT simulation, predicting the rate of growth and proliferation, therefore the radiosensitivity of the tumour. This may indicate the need for alternative fractionation schedules	n=500 in a modelled population(MATLAB), n=4 (patient data from the literature)	n/a	n/a	GTV tumour (modelled and reported)	n/a	n/a	n/a	Use of the PSI during the initially planning stage could facilitate the simulation of a variety of fractionation schedules and the treatment protocol can dynamically be adapted based on this.

CHAPTER 4 STUDY 1

GEOMETRIC AND DOSIMETRIC EVALUATION OF A COMMERCIALLY AVAILABLE AUTO-SEGMENTATION TOOL FOR GROSS TUMOUR VOLUME DELINEATION

This chapter was published in Clinical Oncology in 2021.

Barrett S., Simpkin A.J., Walls G., Leech M., Marignol L., Geometric and dosimetric evaluation of a commercially available auto-segmentation tool for gross tumour volume delineation in locally advanced Non-Small Cell Lung Cancer: a feasibility study, Journal of Clinical Oncology, 2021 Mar;33(3):155-162

See Publications for a copy of the published article.

4.1 Introduction

Accurate definition of a GTV on a CT scan is a key step in the RT process, however, it remains challenging and prone to inter- and intra-observer error [99, 151]. In lung cancer RT, consensus guidelines on contouring lung GTV recommend the use of a 4DCT, IV contrast and multimodality imaging in the form of a recent FDG PET scan [103]. 4DCT consists of a CT scan taken in synchrony with the patients breathing cycle. For each slice acquired in a traditional 3DCT, 10 slices are acquired in a 4DCT, throughout an entire inhale and exhale. In clinical practice, this GTV delineation approach is taken by the RO when defining the target volume for treatment planning purposes. However, this must

be repeated on each of the 10 breathing cycle associated phased scans which is time and resource intensive [152].

Auto segmentation tools can be employed to automatically create contours using a variety of methods, often based on statistical modelling of size and shape of normal anatomy, prior knowledge of anatomical structures from an atlas of reference images or a combination of both. Consequently, much of the research on auto-segmentation focuses on normal tissue and auto-delineation of Organs at Risk (OARs) as this tissue is somewhat predictable normal anatomy. While these tools are very powerful, with the ability to create contours in minutes, limitations exist and research shows that manual review of these structures remains essential to avoid errors [153]. The introduction of artificial intelligence (AI) may reduce the need for this manual adjustment [154].

Auto-segmentation tools have also been suggested as a method of improving both delineation efficiency and accuracy of lung GTV contours [152, 155]. True automatic-contours (AC) can be generated with no user input, however these are often subsequently manually adjusted to create a semi-automatic user-adjusted auto-contour (UA-AC) [112, 156]. This option is attractive in the setting of 4DCT due to the large number of datasets which may require delineation as this is a time-consuming task for the RO.

In addition to the contouring burden of 4DCT datasets, clinical advances have led to repeated imaging over the course of radiotherapy treatment delivery, increasing the imaging data available for each patient. Analysis of GTV changes on repeated imaging during the course of RT may identify changes to the target volume that require plan adaptation. While adaptive RT based on changing target volumes can ensure accurate dose delivery [157], one study demonstrated that physician inter-observer variability defining GTV was increased during the course of RT [158]. To

facilitate rapid and repeated definition of target volumes on both planning CTs and subsequent imaging, a readily available, automated approach would be desirable.

The Proliferation Saturation Index (PSI) model [79] is a mathematical model which predicts GTV regression in response to RT, and to train this model longitudinal tumour volume measurements are necessary. Therefore, this model requires repeated delineation of the GTV over the course of RT, which is not standard clinical practice. The process is time consuming and potentially prone to inter-observer variation, and so an automated approach to define these volumes is attractive.

Many of the auto-segmentation tools evaluated in the literature are locally available systems which were not widely available to clinical departments [155]. Furthermore, many available auto-segmentation approaches utilise deformable registration to propagate a physician defined volume onto a subsequent image, meaning clinician input remains central to the process.

This study investigates the feasibility of using a commercially available auto-contouring system (Smart Segmentation® Varian Medical Systems) in the delineation of lung GTVs without the requirement of propagation of a physician defined Manual Contour (MC).

To complete this analysis, we accessed a unique dataset of NSCLC patients with both a pre-treatment 4DCT and weekly research 4DCT scans acquired over the course of RT delivery. This intensive imaging schedule allowed a rare opportunity for a longitudinal analysis of auto-contouring on a high-quality image at multiple timepoints during RT.

4.1.1 Aims and Objectives

The aims of this chapter are:

1. To establish the geometric accuracy of automated lung tumour gross target volumes produced by a commercially available auto contouring tool
    a. To determine the accuracy of AC and UA-AC in comparison to a gold standard MC by comparing Dice Similarity Coefficient (DSC), shift in structure Centre of Mass (COM) and absolute volume difference (cm^3)
    b. To evaluate differences between the AC and UA-AC agreement with the MC by a ranked Wilcoxon test and evaluation of Bland Altman plots and ascertain if the AC or UA-AC produces the more reliable contour
    c. To determine if tumour proximity to non-lung normal tissue impacts on auto-contouring accuracy.

2. To evaluate the clinical reliability of auto-contours compared to physician defined contours
    a. To determine if there is a statistical difference between target volume dose metrics on MC or UA-AC volumes by comparing key dosimetric indices on MC GTV, CTV and PTV vs. UA-AC GTV, CTV and PTV using a multilevel model.
    b. To compare the clinical acceptability of dose metrics of both MC and UA-AC volumes over the course of RT.

3. To ascertain if geometric accuracy of auto contoured volumes varies over time
    a. To assess the clinical reliability of UA-AC over the course of RT, by comparing the geometric accuracy measures from the first scan to the last scan of each patient using a Wilcoxon Signed Rank Test.

4. To assess the reliability of GTV generated on Average Intensity Projection Image (AVIP) reconstructions of 4DCT and 4DCBCT using the UA-AC approach.
    a. To ascertain if an acceptable contour can be generated on a 4DCT AVIP by plotting MC as measured on a single-phase image versus the UA-AC produced on the AVIP and calculating R^2 and PCC. COM, DSC and Absolute volume will be compared to the MC GTV.
    b. To compare the GTV produced on CT vs CBCT by plotting GTV as measured on a 4DCT vs 4DCBCT AVIPs and calculating R^2, PCC and comparison of absolute volumes

4.2 Materials and Methods

4.2.1 Dataset

A research dataset of 20 LA-NSCLC patients with 4DCT scans uniquely acquired longitudinally during their radiotherapy treatment delivery, was identified at The Cancer Imaging Archive [113, 115] and freely accessed. This atypical dataset described by Hugo et. al [114], consisted of a series of 4DCT scans acquired weekly and 4D Cone Beam CT (CBCT) scans acquired daily over the course of conventionally fractionated radiotherapy treatment delivery. Physicians delineated primary GTV, nodal target volumes and organs at risk on each phase of the 4DCT datasets. No GTV were delineated on any 4DCT reconstruction, or on the 4DCBCT datasets. The primary GTV was assessed in this study and nodal target volumes were not included.

Of the 20 patients, 7 were excluded as only one 4DCT with contoured target volumes was available. The remaining 13 patients were evaluated for inclusion in the study. 5 were excluded due to the primary tumours being located within a region of atelectasis and a lack of associated FDG PET scan for discernment of tumour, see Appendix 10.2.1 for examples of excluded patients). Eight patient datasets were included, with a total of 44 physician-delineated 4DCTs. The GTV location was reviewed and categorised as either abutting normal soft tissue in the thorax or not. See Table 4-1 for details on included patients. All patients were planned to receive 62-70.2Gy in 31-39 once-daily fractions.

Table 4-1 Demographics of included cases
TNM: Tumour Nodes Metastases, LUL: Left Upper Lobe, RUL: Right Upper Lobe, LLL: Left Lower Lobe, RLL: Right Lower Lobe

Pt ID	*T N M*	*Stage*	*Tumour Location*	*Total dose (Gy)*	*Number of #*	*Dose per fraction (Gy)*	*Tumour volume (cc)*	*Tumour Abutting normal tissue*
107	2 2 0	IIIA	LUL	63	35	1.8	18	No
108	2 2 0	IIIA	RUL	70.2	39	1.8	12	No
110	2 3 0	IIIB	RLL	62	31	2	55	No
112	3 2 0	IIIA	RUL	63	35	1.8	31	Yes
113	1 1 0	IIA	LLL	66	33	2	78	Yes
114	2 2 0	IIIA	RLL	66	33	2	179	Yes
115	1 3 0	IIIB	RUL	66.6	37	1.8	7	No
117	3 2 0	IIIA	RUL	66	33	2	10	No

4.2.2 Contour Methods

4.2.2.1 Manual Contours

Each 4DCT scan had an accompanying RT structure set, which included a physician delineated MC, created on each phase of the 4D dataset. The CT 50% phase dataset was chosen for evaluation, as this is a stable expiration phase. The delineation of the MC was supervised by a single experienced RO [114] and the MC was deemed to be the gold standard contour for the purpose of evaluation.

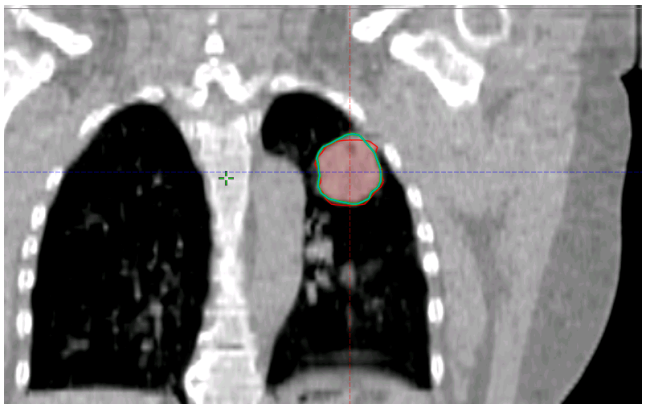
4.2.2.2 Auto-Contours

For each 4DCT an AC was generated on the same 50% phase. The structure set was opened with the corresponding CT series, a lung window was utilised to locate the approximate mid-volume slice. Automatic delineation was performed using the Smart Segmentation (SS) function on Varian Eclipse™ V15.5 by following the manufacturer's instructions (Appendix 10.2.2).

4.2.2.3 User-Adjusted Auto-Contours

The AC was duplicated, and manually adjusted by an experienced Radiation Therapist (RTT) to exclude gross extensions into adjacent normal tissues only, resulting in a UA-AC for each CT50%. This process was audited by a Clinical Oncologist with significant experience in treating lung cancer. See Figure 4-1 for examples of contours generated and adjustments made.

(a)



(b)

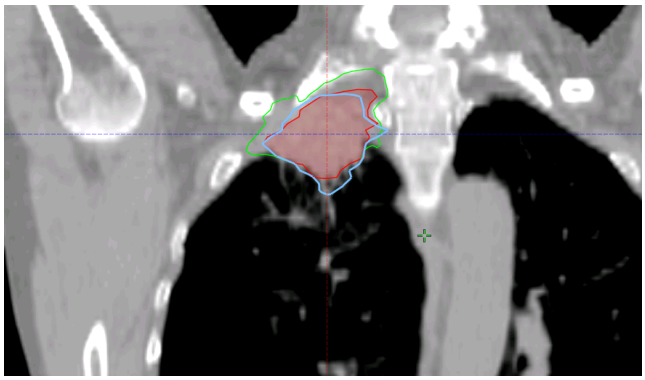


Figure 4-1 Examples of contours created
Manual Contour in Red, Auto Contour in Green, User Adjusted-Auto Ccontour in Blue.
(a), no user adjustment was required
(b) increased user adjustment was required after auto-contouring as tumour abuts normal tissue (chest wall)

4.2.3 Geometric comparison of software-aided and manual contours

To quantify the geometric agreement between the MC and automatically generated contours, 3 metrics were utilised: DSC, shift in structure COM and Volume difference in cm^3 in line with published recommendations [159].

DSC measures the spatial overlap between 2 contours and is the volume of the union normalised by the mean of the 2 volumes. A DSC of 1 indicates perfect overlap and a DSC of 0 indicates no overlap. DSC values greater than 0.7 are generally considered to show good agreement [160]. DSC was measured in Radialogica fullReport ™ V 1.4.5.12438.

COM coordinates in x, y and z directions for each contour were extracted from Varian Eclipse™ V15.5 Statistics. They were converted into an overall COM Shift from the MC using the Euclidian distance formula. Volume difference between AC and UA-AC to the MC was recorded.

To evaluate differences between the AC and UA-AC agreement with the MC a ranked Wilcoxon test was performed. A p-value of less than 0.05 was assumed to be statistically significant.

Bland-Altman plots [161, 162] a graphical method to compare two measurement techniques where the differences between each method are plotted against the averages of both methods, were created to evaluate the agreement between contouring methods in determining target volume. The geometric agreement analysis was repeated excluding 3 patients from this analysis, as they were considered potentially poor candidates for auto-contouring due to tumour proximity to normal soft tissue. See Table 4-2 for a summary of contour methods and comparison metrics used in this study.

Table 4-2 Summary of contour methods and comparison metrics used in this study.

Contours generated

Acronym	***Name***	***Brief Description***
MC	Manual Contour	Physician defined manual contour, generated under the supervision of a single experienced Radiation Oncologist
AC	Auto-Contour	Generated using the commercially available software with no manual corrections
UA-AC	User-Adjusted Auto-Contour	Adjusted auto-contour to correct gross errors where contour extends into normal tissue

Agreement measures

Acronym	***Name***	***Brief Description***
DSC	Dice Similarity Coefficient	Geometric measure of the overlap between 2 structures
COM	Centre of Mass	Centre of mass coordinates difference from gold standard manual contour to evaluate positional differences between contours
Vol	Volume	Absolute volume difference between auto-contours and the gold standard manual contour

4.2.4 Dosimetric comparison of software-aided and manual contours

On the earliest scan for all 8 patients from the MC, a Clinical Target Volume (CTV) and a Planning Target Volume (PTV) were created using an expansion of GTV + 5mm for CTV and CTV + 0.9cm superior/inferior and 0.7cm laterally to create the PTV [126]. A clinically acceptable treatment plan in line with the protocol outlined by Haslett et al. [126] was devised using a VMAT technique. On all subsequent CT scans a MC and UA-AC CTV and PTV were created using the same formulae, and the original plan was copied onto each subsequent scan and recalculated using the fixed monitor units. Renormalisation was not performed.

For each scan the dosimetric indices of minimum dose, maximum dose, mean dose D2%, D98%, D50%, D95% and V107% on MC GTV, CTV and PTV vs. UA-AC GTV, CTV and PTV were measured and compared using a multilevel model to account for the clustering of the responses (dose, volume) within patients, structures and plan [163]. Once this clustering has been allowed for using random effects, there is a fixed effect for whether the contouring was MC or UA-AC as our main variable of interest. Whether the tumour was deemed to be abutting normal soft tissue or not was also controlled for in each model. Since there were eight responses (minimum dose, maximum dose, mean dose, D2%, D50%, D95%, D98% and V107%) Bonferroni correction was applied to account for multiple testing, such that statistical significance is taken to be differences between clinician and automated delineation with an associated p-value below $0.05/8 = 0.00625$.

In addition, the planning requirement of PTV D95% $> 90\%$ and V107% $< 1\text{cm}^3$ was reviewed on all subsequent scans to assess compliance with original plan for both MC and UA-AC.

4.2.5 Time trend analysis of software-aided contours

Differences in the geometric accuracy measures from the first scan to the last of each patient were evaluated using a Wilcoxon Signed Rank Test to assess the reliability of the UA-AC over the duration of RT.

4.2.6 Average intensity projection analysis of software-aided contours

No physician defined MC were available on the AVIP reconstructions of either CT or CBCT. Therefore, to test the software performance on AVIPs, initially the UA-AC generated on a 4DCT AVIP reconstruction was compared to the corresponding CT50% phase MC to ascertain if there was a correlation between the volumes generated. A scatter plot was

created and R^2 and PCC calculated along with DSC and COM. Patient 113 was excluded from this analysis as only 2 CBCTs existed for this patient resulting in 7 patients and n=39 4DCTs.

Secondly, the 4DCBCT AVIP was considered. To compare the CT AVIP volume with CBCT, the most closely acquired 4DCBCT (usually within 1 day, max 2 days) was reconstructed as an AVIP and a UA-AC of the tumour volume was derived.

This volume was compared to the corresponding 4DCT AVIP Volume using a scatter plot, R^2 and PCC. For this analysis only CBCTs acquired within 2 days of a CT were included which resulted in n= 32 4DCBCTs analysed.

4.3 Results

4.3.1 Geometric comparison of software-aided and manual contours

The geometric comparison metrics of the raw and user-adjusted contours to the MC are displayed in Figure 4-2. UA of auto-contours significantly improved geometric agreement to MC compared to AC alone (Table 4-3). Clinician audit of the user adjustments demonstrated complete agreement for 87.5% of the cases and satisfactory agreement for derivation of UA-AC in >90% cases sampled. As the UA-AC contours were shown to be geometrically superior to the AC alone, the remainder of the analysis focused on UA-AC metrics compared to MC.

The Bland Altman analysis identifies improved agreement between UA-AC and MC when the tumour is smaller (Figure 4-3), however, the larger tumours also abutted adjacent normal tissues.

The impact of the anatomy surrounding the tumour on auto-contouring accuracy was evaluated next. The exclusion of the 3 patients whose tumour was abutting normal tissue (n=28 4DCTs included), further improved all of the geometric agreement metrics measured (Table 4-4). Improved agreement between MC and UA-AC was also observed on Bland Altman analysis when these 3 cases were excluded (Figure 4-3) .

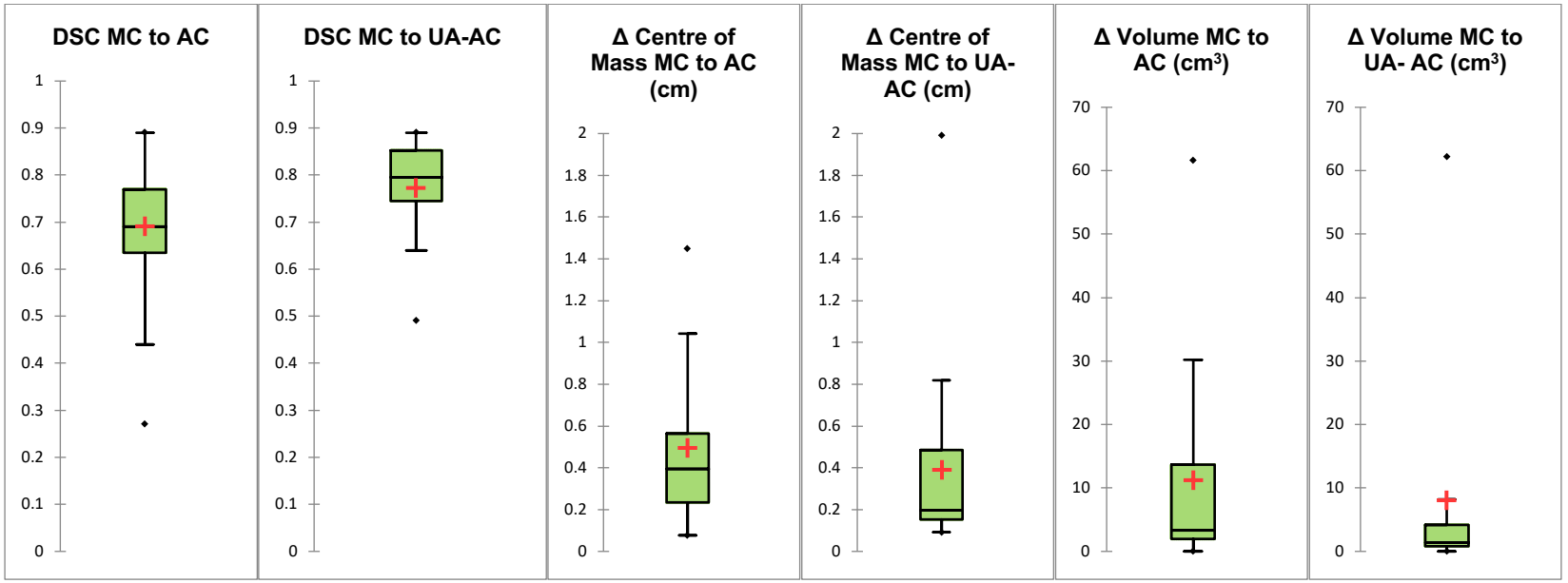


Figure 4-2 Box plots comparing the AC and UA-AC to the MC.

Box plots of geometric metrics measured on 44 datasets comparing the AC and UA-AC to the MC showing an improvement in all geometric measures when user-adjustment was incorporated. All improvements were statistically significant ($p < 0.05$, Wilcoxon Signed rank test)

DSC: Dice Similarity Coefficient, Δ:Delta, MC: Manual Contour, AC: Auto-Contour, UA-AC: User-Adjusted Auto-Contour, Volume differences (cm^3) are absolute

Red Cross= mean

Table 4-3 Geometric comparison to manual contour (all patients)

DSC = Dice similarity coefficient; COM = centre of mass; MC= Manual Contour; AC= Auto-Contour; UA-AC= User-Adjusted Auto-Contour ∗Median, Means and COM offset are provided in centimetres, volume variance is provided in centimetres squared. P values generated from Wilcoxon Signed rank test

	AC congruence with MC		***UA-AC congruence with MC***		
Metric	**Median (range)**	**Mean (SD)**	**Median (range)**	**Mean (SD)**	***P*** $^{\sim}$
Volume-based metrics					
DSC	0.69 (0.27-0.89)	0.69 (0.122)	0.8 (0.49-0.89)	0.77 (0.099)	<0.0001
Absolute volume difference	3.35 (0-61.6)	11.15 (15.48)	1.4 (0-62.2)	8.05 (16.498)	0.005
Distance-based metric					
COM offset	0.39 (0.08-1.45)	0.49 (0.377)	0.20 (0.09-1.99)	0.39 (0.409)	0.0003

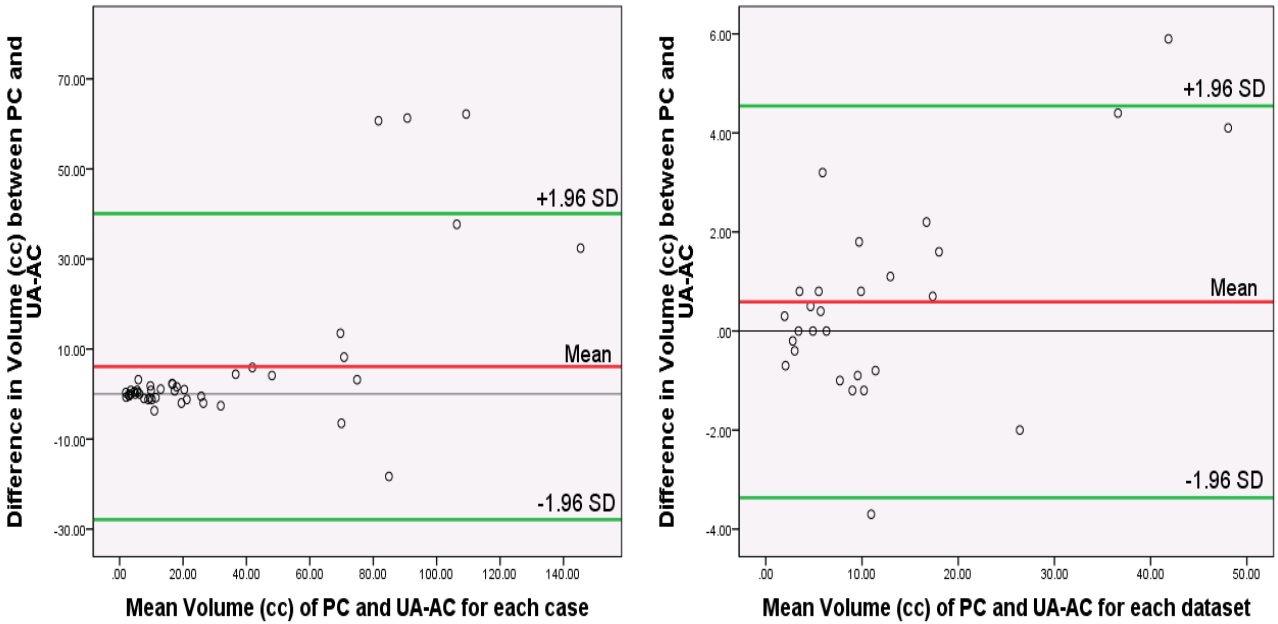


Figure 4-3 Bland-Altman comparison of MC and UA-AC

Bland-Altman comparison of MC and UA-AC for all patients on left. Bland-Altman comparison excluding 3 large tumours abutting normal tissue on right demonstrating improved agreement. PC: Physician Contour, UA-AC: User-Adjusted Auto-Contour.

Table 4-4 Geometric comparison to manual contour (excluding 3 patients)

Geometric comparison to Manual Contour (MC) of User Adjusted-Auto Contour (UA-AC) excluding 3 patients with tumour volumes abutting normal soft tissue

DSC = Dice similarity coefficient; COM = centre of mass; MC= Manual Contour, UA-AC= User-Adjusted Auto-Contour

	UA-AC congruence with MC excluding 3 cases	
Metric	**Median (range)**	**Mean (SD)**
Volume-based metrics		
DSC	0.8 (0.49-0.89)	0.78 (0.091)
Absolute volume difference	0.85cm^3 (0-5.9)	1.45 cm^3 (1.5)
Distance-based metric		
COM offset	0.16cm (0.09-0.69)	0.21cm (0.14)

4.3.2 Dosimetric comparison of software-aided and manual contours

There were no statistically significant differences in dose metrics measured from the MC volumes and the UA-AC volumes, after Bonferroni correction. There were two nominally significant findings in D50% (mean difference -0.06%, 95% confidence interval for difference [-0.1, -0.01]; p = 0.009) and V107% (mean difference 0.14cc, 95% confidence interval for difference [0.03, 0.24]; p = 0.012). All data are available in Appendix 10.2.3 and Appendix 10.2.4.

All 8 original plans met planning constraints of PTV D95% > 90% and V107% < 1cm^3. Subsequent scans (n= 36) were assessed to determine if these metrics were breached in either the MC PTV or the UA-AC PTV. The PTV D95% > 90% dose constraint was still met in 17/36 re-plans using MC and in 21/36 re-plans using UA-AC. V107% < 1cm^3 planning

constraint was met in 29/36 re-plans using both the MC and the UA-AC (Appendix 10.2.4).

4.3.3 Time trend analysis of software-aided contours

Between 5 and 8 4DCT scans sets were available for each case. Scan dates ranged from 26 days pre-RT to day 47 of RT. Median duration between first and last scan was 47.5 days (range 27-65 days). The median DSC decreased from 0.82 (range 0.76-0.89) to 0.72 (range 0.49-0.86) (p= 0.035) between first and last scan. Median COM shift from MC at first scan was 0.22 cm (range 0.09-0.71 cm), at last scan was 0.23 cm (range 0.09-1.99 cm). Mean COM shift increased from 0.31 cm to 0.52 cm (p=0.058). Median volume difference between MC and UA-AC at first scan was 3.15 cm^3 (range 0.8-32.4 cm^3), at the last scan was 1.0 cm^3 (range 0.7-60.7 cm^3). Mean Volume difference increased from 6.6 cm^3 to 9.9 cm^3 (p=0.866).

4.3.4 Analysis of software-aided contours on AVIP datasets

Scatter plot comparing the absolute volumes of the MC vs the UA-AC created on the 4DCT AVIP demonstrate good agreement between the 2 measures R^2=0.89 and PCC =0.94 (Figure 4-4). Median DSC=0.75 (0.54-0.91), median COM shift=0.26cm (0.06cm -1.97cm) and median volume difference=2.1 cm^3 (range 0 cm^3- 66.3 cm^3) (Figure 4-5).

Thirty-two patients had a 4DCBCT within 2 days of a 4DCT and absolute volume of the UA-AC structures created on the AVIPs of 4DCT and 4DCBCT showed excellent agreement (Figure 4-6) with R^2= 0.99 and PCC of 0.99.

Table 4-5 displays the absolute volumes of the MC, the 4DCT AVIP and the 4DCBCT AVIP, as well as the ratio of the volumes.

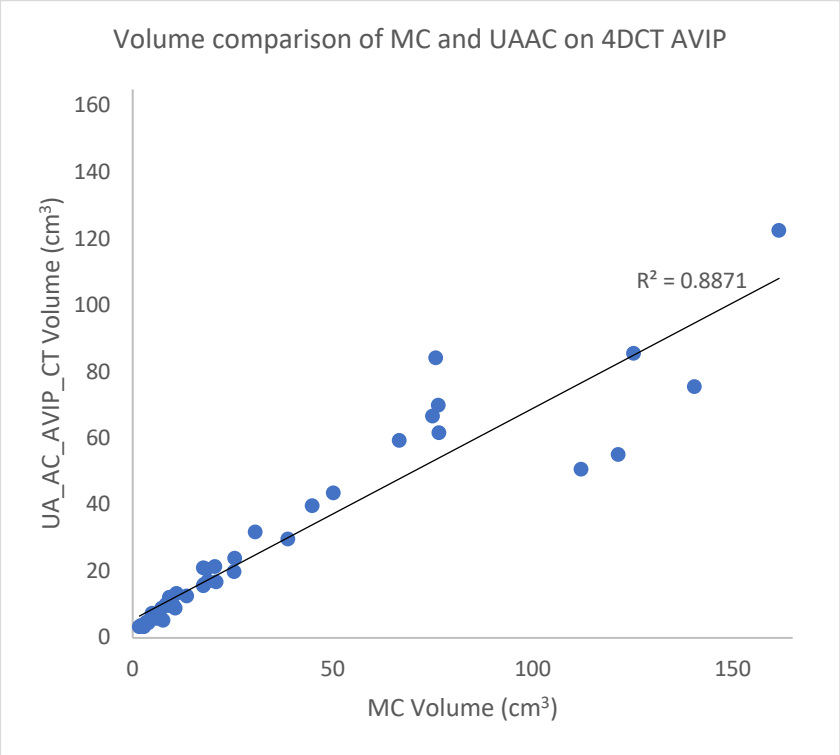


Figure 4-4 Scatter plot comparing MC GTV volume (cm^3) with UA-AC volume(cm^3) generated on the 4DCT AVIP reconstruction.

X axis displays Manual Contour, Y axis displays UA-AC on AVIP

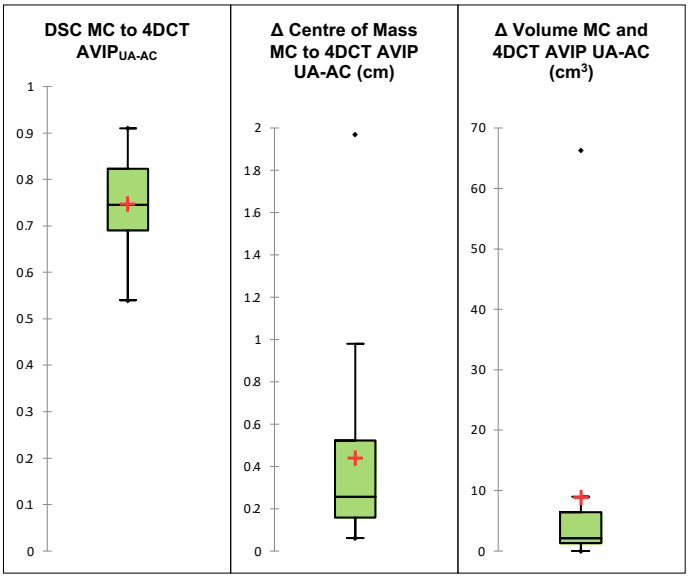


Figure 4-5 DSC, COM, Absolute Volume MC to 4DCT AVIP UA-AC

Box plots of geometric metrics measured on 39 4DCT AVIP datasets comparing the UA-AC to the MC. 4DCT: 4Dimensional CT, AVIP; Average Intensity Projection, DSC: Dice Similarity Coefficient, Δ: Delta, MC: Manual Contour, UA-AC: User-Adjusted Auto-Contour Red cross= mean

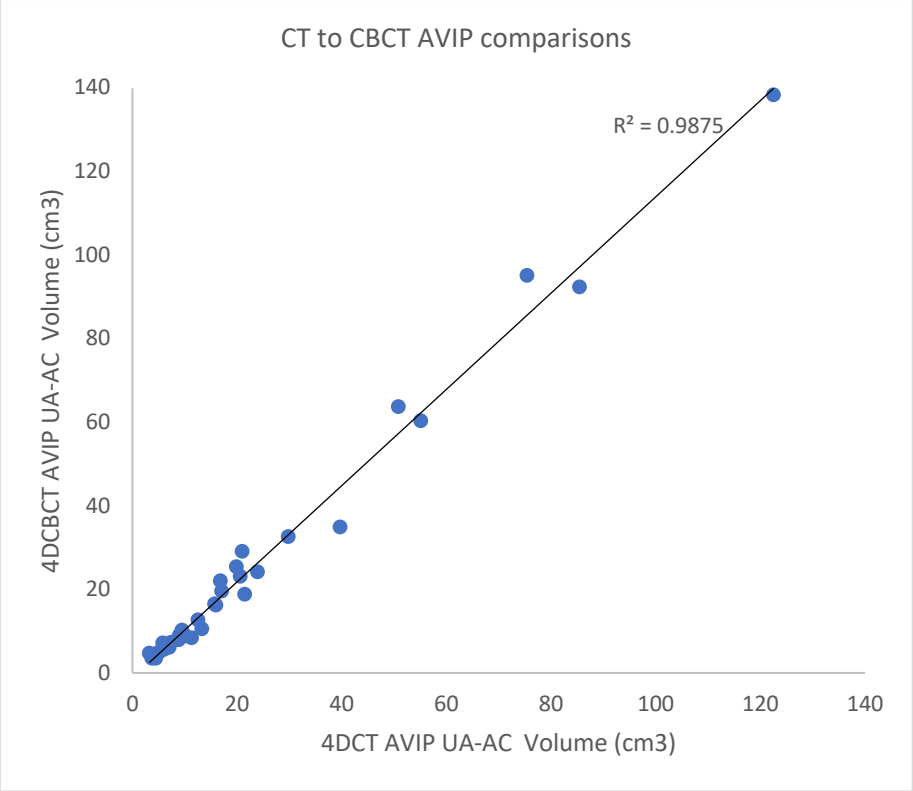


Figure 4-6 Scatter plot comparing UA-AC volumes generated on 4DCT and 4DCBCT AVIP Reconstructions

CT; Computed Tomography, 4DCT; 4Dimensional CT, CBCT; Cone Beam CT AVIP; Average Intensity Projection

Table 4-5 Absolute tumour volumes and ratio of tumour volumes across scan types.

	Volume [cm^3]			*Ratios*		
Pt ID	MC on CT 50%	4DCT AVIP	4DCBCT AVIP	CT AVIP: 50%	CBCT AVIP: 50%	CT: CBCT AVIP
P107	18.8	17.5	19.6	0.93	1.04	1.12
P107	17.8	16.9	16.3	0.95	0.92	0.96
P107	13.5	12.3	12.7	0.91	0.94	1.03
P107	10.6	9	9.1	0.85	0.86	1.01
P107	6.3	5.8	5.7	0.92	0.90	0.98
P108	11	13.5	10.6	1.23	0.96	0.79
P108	10.3	9.5	10.2	0.92	0.99	1.07
P108	7.2	8.8	8	1.22	1.11	0.91
P108	9.6	11.2	8.4	1.17	0.88	0.75
P110	38.8	29.1	32.7	0.75	0.84	1.12
P110	44.8	38.6	35	0.86	0.78	0.91
P110	25.4	19.9	25.4	0.78	1.00	1.28
P110	17.7	15.8	29.2	0.89	1.65	1.85
P114	161.6	126.3	138.5	0.78	0.86	1.10
P114	125.2	89.8	92.5	0.72	0.74	1.03
P114	140.4	76.3	95.2	0.54	0.68	1.25
P114	121.4	54.5	60.4	0.45	0.50	1.11
P114	112	50.8	63.8	0.45	0.57	1.26
P115	7.5	4.9	5.1	0.65	0.68	1.04
P115	5.9	5.3	7.2	0.90	1.22	1.36
P115	4.9	9.5	6.1	1.94	1.24	0.64
P115	3.9	5.7	3.6	1.46	0.92	0.63
P117	8.4	11.8	8.9	1.40	1.06	0.75
P117	4.9	9	7.4	1.84	1.51	0.82
P117	3.4	4.9	4.8	1.44	1.41	0.98
P117	2.7	3.4	4.8	1.26	1.78	1.41
P117	2.1	4.2	3.5	2.00	1.67	0.83
P112	25.5	35.9	24.2	1.41	0.95	0.67
P112	18.5	43.3	23.2	2.34	1.25	0.54
P112	20.6	32.9	18.9	1.60	0.92	0.57
P112	20.9	31.5	22.1	1.51	1.06	0.70
P112	17.7	17.4	16.5	0.98	0.93	0.95
Median	13.50	13.50	12.70	0.93	0.95	0.98
Max	161.60	126.30	138.50	2.34	1.78	1.85
Min	2.10	3.40	3.50	0.45	0.50	0.54

4.4 Discussion

This study investigated the feasibility of using a commercially available auto-contouring software to delineate lung GTVs, by evaluating both the geometric and dosimetric reliability of the contours generated over the duration of RT. While the number of cases was small, this rare image intensive dataset uniquely supported an in-depth longitudinal evaluation of the contours generated. In addition, the stability of the auto contouring process over time was investigated.

4.4.1 Geometric comparison of software-aided and manual contours

The study has identified that user-adjustment of auto-contours significantly improved geometric agreement in all metrics, in line with prior work on this topic evaluating non-commercially available segmentation and deformable registration [112, 155, 160, 164]. Consensus guidelines highlight the ongoing requirement for manual review and adjustment of auto-contoured structures [165]. Similarly, in OAR delineation, user adjustment of software-generated auto-contours was reported to be a viable strategy to produce clinically acceptable contours [112]. Following user-adjustment, this study reports good agreement between contours comparing to the MC with mean and median DSC $>$ 0.7 (0.77 and 0.8 respectively), which is generally accepted as good agreement between gold standard and assessed contour [160, 164]. DSC, while commonly reported within RT literature, has limitations when comparing contours. DSC is sensitive to the absolute volume of the structure being evaluated. In smaller structure volumes, minor discrepancies can result in poor DSC scores, whereas larger volumes dilute the effect [166]. It is therefore recommended that more than one geometric metric is employed to more comprehensively assess overlap. The addition of a distance-based metric, such as the Hausdorff distance would have been beneficial in this analysis. AI informed auto-segmentation tools may reduce this need for user-adjustment [154], however the algorithm tested

in this study is commercially and readily available making it an attractive option for clinical use.

The median volume difference between UA-AC and the MC, was 1.4 cm^3 (range 0-62.2 cm^3). While the median volume difference was low, large individual differences were observed between the UA-AC and the MC, up to 62.2cm^3. We hypothesised that large variation may occur when tumours were adjacent to normal structures within the thorax, such as the mediastinum or chest wall, which have similar density to tumour. A subgroup analysis was completed, excluding 3 cases where the tumour abutted normal tissue, which led to a considerable reduction in the median (0.85cm^3) and range (0-5.9cm^3) of differences. These data highlight that some tumours may be more amenable to auto-contouring than others and appropriate case selection may be a key step in implementing this technology in routine clinical practice.

Related to discernment of tumour and normal tissues, prior publications have highlighted the difficulty in generating target volumes in nodal regions [160] or in regions of atelectasis where defining the tumour boundary with CT alone is difficult [99]. While this study excluded tumours located within atelectasis, other studies have enacted a restriction on cases based on the percentage of tumour surrounded by lung parenchyma [68]. One limitation of this study is that it didn't include node-positive cases, for which target delineation is known to be more difficult [160], however the auto-segmentation tool investigated is designed for primary tumour volume delineation only, so it was not appropriate to assess its performance on nodal volumes.

Exclusion of the 3 cases in which the tumour was abutting normal soft tissue resulted in improved geometric agreement with a median DSC of 0.8, median COM shift of 0.16cm and median volume difference of 0.84cm^3. The mean DSC of 0.78 is comparable to the value of 0.84

reported recently evaluating inter-observer variability in CTV delineation [167]. This, in conjunction with the high level of agreement in COM and structure volume, also demonstrates that appropriate case selection increases geometric accuracy of adjusted auto-contours.

4.4.2 Dosimetric comparison of software-aided and manual contours

Clinical acceptability of UA-AC-based RT planning was considered in this study. A VMAT treatment plan was generated for each patient and used to evaluate the dosimetric differences between the dose metrics measured from the physician-defined MC and target volumes generated using the UA-AC method.

UA-AC volumes performed as well as the MC in relation to target coverage over the course of treatment in this population with no statistically significant differences between the measured dose metrics on either structure over the course of treatment. However, over the course of RT, the plan quality decreased and dose coverage of the PTV was compromised in both the MC and UA-AC target volumes. Raw data from plan analyses is available in Appendix 10.2.4.

Qualitative review of the plans revealed there was significant tumour regression, with magnitudes similar to that reported in previous studies [168, 169], during RT which may be responsible for the poor target coverage as RT progresses. This may indicate a need to introduce ART for LA-NSCLC RT to ensure planned dose coverage is delivered, which has been reported previously [157]. The LARTIA Trial reported on local control and toxicity for patients undergoing ART and reported approximately 25% patients required repeated planning due to tumour regression during treatment [104]. Some data suggests that ART improves OS in this population [170], and so the dosimetric deterioration in plan quality observed in this study may lead to a deterioration in clinical

outcomes also. However, more recent work by this group note that the improvement in clinical outcomes following ART may be histology dependant [171], with an improvement in OS seen in patients with squamous cell carcinomas, but no OS increase in those with adenocarcinoma. Despite the lack of OS improvement, ART still benefited all patients with a reduction in radiation pneumonitis [171].

Repeated planning for ART requires repeated contouring, and so the UA-AC method could be a useful tool in this setting for saving time. In this study user adjustment was performed by an experienced RTT. The high level of oncologist agreement of the user adjusted contours indicates the applicability of the software across trained staff groups i.e. RTTs, dosimetrists and oncologists. The role of the RTT in delineation traditionally has stopped at normal tissue delineation, however with the emergence of online ART there is a move for Advanced Practice RTTs to expand their role into target volume delineation [172]. This process has been successfully trialled in magnetic resonance guided prostate cancer RT [173]. With the advent of more online ART, the method described here to generate UA-AC of the target volume could prove to be a useful approach either to perform an efficient dose calculation check and trigger an offline adaptive plan, or to facilitate online ART for patients with a breach of dose coverage metrics.

Other factors beyond TVR, such as patient contour changes or set-up uncertainties, may have contributed to the observed variation in plan dosimetry [174]. This study aimed to test if the dosimetry reported on MC volumes or UA-AC volumes were statistically different but causes of dosimetric changes were not investigated. Future work should determine if the variation observed is a direct result of TVR and include nodal volumes as well as the primary tumour.

4.4.3 Time trend analysis of software-aided contours

While ART may improve target coverage in this patient population, it requires repeated target delineation over the course of RT. It has been shown that inter-observer variability in target delineation increases over the course of RT in the ART setting [158]. This analysis identified that while 2 of 3 geometric congruence metrics declined between the first and last timepoints, only DSC decline reached statistical significance. Both this study and Apolle et al's [158] analysis on MC inter-observer variability was carried out on a limited dataset (6 observers delineating on 5 cases). While less agreement was observed between MC and UA-AC over the course of RT, Apolle et al have shown that MC reliability decreases during RT, and so further research is required in relation to both manual and auto-contours to elucidate which contour more accurately represents the true target volume.

4.4.4 Analysis of software-aided contours on AVIP datasets

In addition to evaluating the auto-segmentation process on single phase CT data, we further examined its performance in AVIP reconstructions of 4DCT and 4D CBCTs. AVIPs generated from 4D datasets most closely match a 3D CBCT which is routinely acquired for most patients for image guidance purposes [175]. For PSI modelling delineation of tumour volumes on CBCT data is required. While this dataset did not include any 3DCBCTs it did contain longitudinal 4DCBCT data however, physician generated contours were not available on the AVIP or the individual phases of the 4DCBCTs.

Comparison of the UA-AC generated on a CT AVIP demonstrated good agreement to the physician defined MC with R^2=0.89 and PCC=0.94 (Figure 4-4) and it would not be expected for these structures to match exactly due to the fundamental difference in the image. The DSC and COM metrics measured here (Median DSC =0.75, median COM shift = 0.26cm) were comparable to prior work looking at delineation on AVIP

projections where the authors reported DSC of 0.7 ± 0.1 and COM shift of 3.1mm ± 1.1mm comparing 4DCT to 4DCBT target volumes [176]. More relevant in evaluating the performance of the software on CBCT data was the comparison between contours defined on 4DCT AVIP vs 4D CBCT AVIP. In this scenario there was excellent agreement with both R^2 and PCC of 0.99 (Figure 4-6). These data suggest that the auto-contouring process tested is equally as reliable on CT as CBCT datasets. To generate longitudinal tumour volume kinetics data, it is essential to define the tumour on the routinely acquired CBCT images. While other methods are proposed in the literature, they are often in house developed solutions or reliant on AI and not readily available [155, 177, 178]. This method utilises a widely available commercial solution to generate GTV structures.

4.4.5 Relevance for PSI modelling

The Proliferation Saturation Index (PSI) model utilises longitudinal tumour volume measurements to predict tumour volume kinetics in response to RT [79]. While volumetric imaging is acquired routinely over the course of RT in the form of CBCTs, these scans are not delineated as part of standard practice. To test the use of the PSI model on larger datasets a semi-automated approach would be required to ensure that volume measures can be generated in a timely manner.

The method of auto-segmentation of lung tumour volume evaluated in this study is appealing as it utilises a commercially available auto-contouring solution. Additionally, it has been shown that the user-adjustment by an experienced RTT generated reliable tumour volumes. This demonstrates the potential for a non-oncologist expert to contour larger datasets for the purposes of training the PSI model.

4.4.6 Limitations

While the findings from this study are promising, we recognise that our analysis was limited to a single MC comparison ('gold standard') and comparison to another physician's contours may have yielded slightly different results. The relatively small sample size limits the applicability of our findings however the methodology was robust in that the analysis evaluated geometric agreement, dosimetric reliability and uniquely assessed trends with serial 4DCT scanning.

4.5 Conclusions

This feasibility study suggests that UA-AC can produce clinically and geometrically accurate GTV contours for NSCLC when appropriate cases are selected. Our findings suggest that adaptive strategies are merited in NSCLC RT and that automated techniques could augment adaptive RT workflows. Further work to validate the findings on larger datasets and incorporation into prospective studies of adaptive radiotherapy is warranted prior to clinical implementation.

CHAPTER 5 STUDY 2

MONITORING LUNG TUMOUR VOLUME ON CBCT

5.1 Introduction

Delineation of a reliable GTV in lung cancer is a challenge due to breathing motion, proximity of similar density structures, and atelectasis within the lung [100]. When defining a clinical GTV for treatment planning these challenges are circumvented with the use of 4 Dimensional Computed Tomography (4DCT), Intravenous (IV) contrast and ^{18}Fluorodeoxyglucose Positron Emission Tomography (FDG PET) CT [100].

To evaluate TVR in response to RT, longitudinal definition of the GTV is required on datasets that are not used routinely for clinical target volume definition, such as the AVIP reconstruction of a 4DCT and the on-treatment CBCTs.

Some studies investigating TVR over time are interventional trials, where additional mid-treatment planning CT scans are acquired. While these provide gold-standard image quality, they are a significant resource burden to departments and add additional dose to patients [74]. Delineation on daily CBCT alleviates this issue, however it also presents challenges. Image quality is inherently limited compared to CT and the resulting image also lacks IV contrast.

Retrospective studies evaluating CBCT delineation frequently report on in-house algorithms, that deform the clinical GTV onto the CBCT and a RO is required to edit the resultant volume and create a GTV_{CBCT} [71]. This solution is resource intensive and makes generating GTV_{CBCT} volumes on a large scale unrealistic. Deformable image registration (DIR) is the process of minimising variation between two images, or contours by the deformation of one structure to match the target image. DIR based contouring of lung GTV on AVIP images has been shown to be reasonable when tumour motion is limited to less than 10mm, but caution is advised in this approach where larger motion exists [179]. DIR algorithms have been integrated into commercial treatment planning systems and are used in research to evaluate dose accumulation to good effect [180]. However, many uncertainties remain which limits their widespread integration into clinical practice as evaluation metrics are not universally agreed upon and geometric errors can be large and difficult to quantify [181].

Auto contouring (AC) tools can be employed to repeatedly define the GTV and other structures on CT [182], but the utility of AC in defining lung GTV_{CBCT} is under explored with most research focusing on CT or OARS [155]. To bring repeated GTV delineation to a larger scale in the clinic a method is required which is both accessible to many departments and is not reliant on RO manual input which leads to a significant bottleneck in a workflow [183].

The semi-automated method tested in Study 1 produced geometrically and dosimetrically acceptable contours with AC edits by the RTT rather than RO on a single-phase image from 4DCT. This is a stable image, with minimal motion artefact. CBCT and AVIP scans have poorer image quality and increased motion artefact, as well as lacking IV contrast which is employed at treatment planning to aid identification of the target volume. RTT led auto contouring on CBCT and AVIP data would reduce

the workload on the RO and allow for larger scale longitudinal delineation of target volumes.

This study seeks to pilot semi-automated RTT led GTV delineation on CBCT and AVIP reconstructions of 4DCT, using Varian Lung Tumour Smart Segmentation tool.

5.1.1 Aims and Objectives

1. To assess the feasibility of generating a GTV contour on CBCT data using Varian Smart Segmentation™ lung tumour segmentation tool.
    a. To identify the characteristics and percentage of cases of lung tumours amenable to auto contouring on CBCT.
    b. To determine the efficiency of semi-automated GTV delineation by measuring the time required to adjust the resulting AC and ascertain if the duration is correlated with tumour motion or tumour intra-thoracic location.
    c. To quantify the level of adjustments required by volume difference between auto contour and adjusted autocontour and scoring of the revisions required on a scale of 1-4.
    d. To independently audit the GTV contours generated, and establish the level of agreement between the RTT and a Clinical Oncologist

2. To explore longitudinal tumour volume kinetics during the delivery of radical RT in NSCLC.
    a. To analyse the rate of TVR in this population.

5.2 Materials and Methods

5.2.1 Patient population

Governance approvals were provided, and ethical approval waived, by the Belfast Health & Social Care Trust (IRAS 293181) for this study.

The NI-HEART database of 478 patients with NSCLC treated with radical RT was accessed to select patients for CBCT delineation [117]. This cohort consisted of patients treated with radical RT for NSCLC, at the Northern Ireland Cancer Centre, Belfast City Hospital, between January 1, 2015, and December 31st, 2020. Node positive patients were excluded as this study sought to examine GTV delineation only.

Initial screening to identify patients with T1, 2, 3 primary tumours and N0 M0 disease excluded 352 cases. A further 12 were excluded due to miscoding of T or N stage, or missing DICOM data.
114 patients were anonymised and the planning CT AVIP reconstruction of the 4DCT, and day 1 CBCT, were screened for suitability to include as per criteria noted in Table 5-1. For each included patient a pre-treatment planning 3DCT or 4DCT were included, as well as CBCT days 1-3 and a weekly CBCT thereafter (minimum 3 further CBCTs).

Planning CT scans were acquired on a GE large bore Optima 580, Discovery CT590 RT (GE Healthcare, Waukesha, WI) or a Siemens SOMATOM. CBCT scans were acquired on Varian Truebeam imaging panels and the default Varian Lung protocol.

5.2.2 Contouring method

For each patient the planning AVIP or 3DCT was accessed and the clinical Internal Target Volume (ITV) reviewed in both lung and thoracic window settings.

On the planning scan, an auto-contour (AC) was generated. A lung window was utilised to locate the approximate mid-volume slice and automatic delineation was carried out using the lung tumour Smart Segmentation function on Varian Eclipse™ V18.0 according to manufacturer's instructions.

The AC was duplicated and manually adjusted to exclude gross extensions into adjacent normal tissues to generate a UA-AC for the 3DCT (GTV3DCT) or the AVIP (GTVAVIP) following the methodology outlined in Chapter 4 [182]. This process was repeated for each subsequent CBCT resulting in an AC and UA-AC of the target volume for each CBCT (GTVCBCT).

For those patients with a 4DCT planning scan, GTV had been clinically contoured on all 10 phases of the 4DCT. To calculate value representative of tumour motion the distance between the centre of mass of GTV on the 50% phase (end expiration) and GTV 0% phase (end inspiration) was determined using the Euclidian distance formula, $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$.

5.2.3 Contour Evaluation

For each tumour it was noted if it adjoined normal soft tissue structures (Chest Wall, Diaphragm, Mediastinum (above heart chambers), Mediastinum (below heart chambers), Apex of Lung or Other). An estimated percentage of GTV touching normal tissue was recorded on a 4-point scale (<25%, 25-50%, 50-75%, >75%).

Scoring of the extent of revision required for each AC were recorded on a 4 point Likert scale (1= no revision 2=minor revisions 3= major revisions 4=not suitable for revision), in line with the most frequently reported qualitative assessment scales in the literature [184]. Actual and absolute

volume differences between the AC and the UA-AC were measured to give a quantitative indication of the level of revisions required.

5.2.4 Efficiency analysis

Time required to adjust each AC volume was recorded in seconds. To ascertain if a time trend existed in contouring the first to the last CBCT, a single factor ANOVA was performed. A paired t-test compared the time required to adjust the AC on planning CT to day 1 CBCT to evaluate if AC adjustment time was significantly different between CT and CBCT.

5.2.5 Contour audit

A selection of the UA-AC volumes were audited by a Clinical Oncologist with significant experience in treating lung cancer. An informal stratified sampling approach was used to select contours for independent review to evaluate UA-AC on planning CT, CBCT and at various timepoints during the during the course of RT. Contours were rated with a 4-point scale (1=no revisions, 2=minor revisions, 3=major revision, 4=unsuitable contour) [184].

No physician defined manual contour were available on the planning CT AVIP that could facilitate a geometric evaluation of contouring accuracy. Therefore, correlation between the UA-AC volume and the volume of the structure delineated on a single phase (50%) was evaluated by the PCC.

5.2.6 Tumour volume response dynamics

Rates of TVR, or progression, in this population were calculated as the relative volume reduction at the final image, divided by the timepoint of image acquisition in days. Comparisons were made between TVR in patients treated with 3DCRT versus a modulated technique, and between those with Adenocarcinoma, Squamous Cell Carcinoma or a Clinical/Other diagnosis.

5.3 Results

5.3.1 Patient population

Following screening of 114 cases n=38 (33.3%) patients were excluded, resulting in n=76 patients included in the analysis. The most common rational for exclusion was the GTV boundary was not distinguishable on CBCT due to motion blur and/or tumour location. See Table 5-1 for exclusion criteria, and numbers in each exclusion criterion, and Figure 5-1 for image examples of excluded patients.

Table 5-1 Inclusion and exclusion criteria for CBCT GTV delineation

GTV: Gross tumour Volume, 3DCT: 3-Dimensional Computed Tomography 4DCT: 4-Dimensional Computed Tomography, CBCT: Cone Beam Computed Tomography, AVIP: Average Intensity Projection

Inclusion Criteria	Exclusion Criteria (and number excluded)	
T 1, 2, 3 where T3 did not indicate satellite nodules	GTV located within region of extensive consolidation	(10)
Visible GTV on planning CT (4DCT AVIP or 3DCT) and on day 1 CBCT	GTV boundaries not visible on CBCT due to motion/dense lung tissue or attachment to diaphragm	(12)
Mild atelectasis permitted provided tumour was easily discernible	GTV has considerable contact with mediastinal structures or great vessels and boundaries not visible on CBCT	(6)
	Infection mid RT which altered target volume	(1)
	RT replanned due progression visible on day 1 CBCT	(1)
	Endobronchial GTV	(2)
	T3 with satellite nodules	(2)
	Tumour with large cavity making ‘volume of disease’ unreliable	(1)
	Post-surgical target volume	(3)

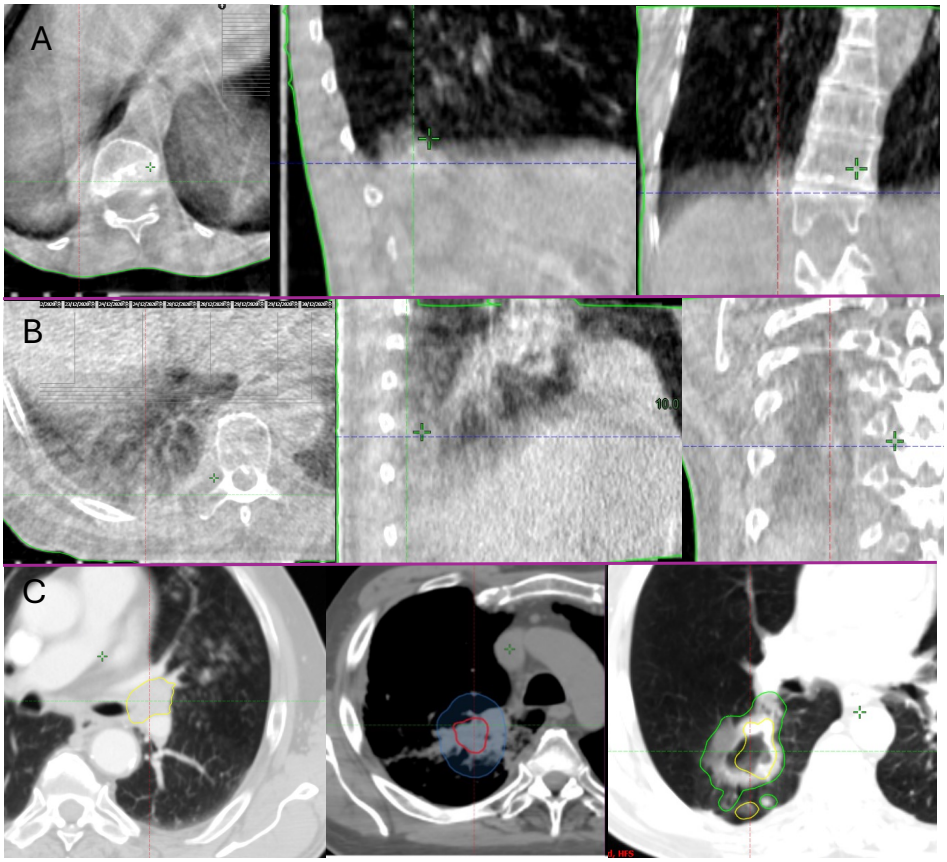


Figure 5-1 Examples of patients excluded due to imaging features listed in Table 5-1

Panel A displays CBCT of a tumour in right lower lobe with large motion making superior border indistinct and attached to diaphragm resulting in unclear inferior border.

Panel B displays CBCT of a tumour in right lower lobe with large motion and dense lung tissue making all borders indistinct

Panel C displays from left to right, planning CT scans with a tumour in left hilum where IV contrast aided delineation at planning CT, a tumour located within extensive consolidation and a tumour with central necrosis and satellite lesions.

Delineation of GTV on the included patients resulted in n=553 scans contoured, n=72 AVIP from 4DCT, n=4 3D planning CT and n=477 CBCT. See Table 5-2 for demographic details on included patients.

Table 5-2 Demographics of patients included in CBCT delineation
3DCRT: 3-Dimensional Conformal Radiation Therapy, IMRT: Intensity Modulated Radiation Therapy, VMAT: Volumetric Modulated Arc Therapy, PS: Performance Status

Patient Demographics	***N (%)***
Age (median, range) years	73.5 (50-90)
Gender	
Female	45 (59%)
Male	31 (41%)
Performance Status	
0	5 (6.6%)
1	27 (35.5%)
2	39 (51.3%)
3	5 (6.6%)
T-stage	
1	38 (50%)
2	21 (27.6%)
3	17 (22.4%)
Subtype	
Squamous cell carcinoma	25 (32.9%)
Adenocarcinoma	14 (18.4%)
Clinical	34 (44.7%)
Other	3 (4%)
Dose Fractionation Schedule	
55Gy/20#	71 (93.4%)
52.5/20#	1 (1.3%)
60–66Gy/30–33#	4 (5.3%)
Radiotherapy planning	
3DCRT	24 (31.6%)
IMRT	14 (18.4%)
VMAT	38 (50%)

Patients had an average of 7 scans (range 6-8). Of the 72 patients who had a 4D planning CT, median tumour motion was 0.3cm (range 0-2.4cm).

5.3.2 GTV segmentation and evaluation

71% (n=55) of tumours were attached to normal soft tissue (other than lung), with breakdown of the anatomy adjacent to the tumour and the percentages in Figure 5-2.

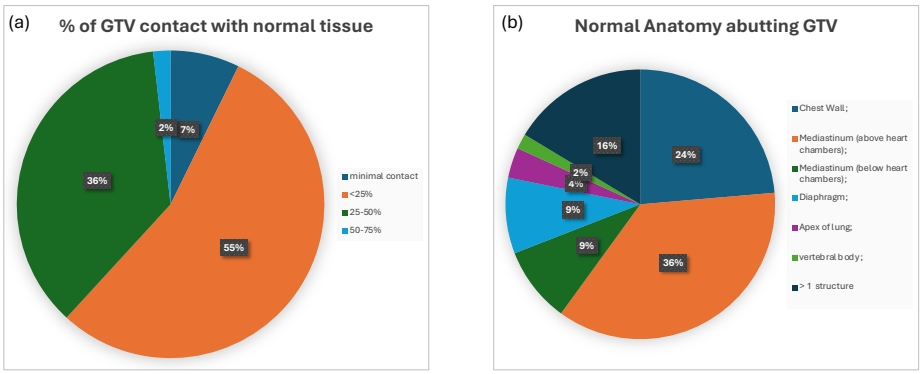


Figure 5-2 Summary of tumour contact with normal soft tissue structures as noted on planning CT

Panel (a) showing the percentage of the GTV in contact with soft tissue. Panel (b) notes the breakdown of various anatomical structures in contact with the GTV.

No revision of the AC was required in 8.1% (n=45) of the scans, minor revisions were necessary in 59.1% (n=327), major revisions in 30.7% (n=170) and 2% (n=11) of AC volumes were deemed unsuitable for editing. See Figure 5-3 for examples a patient where no revisions were required and where the contour generated was deemed unsuitable.

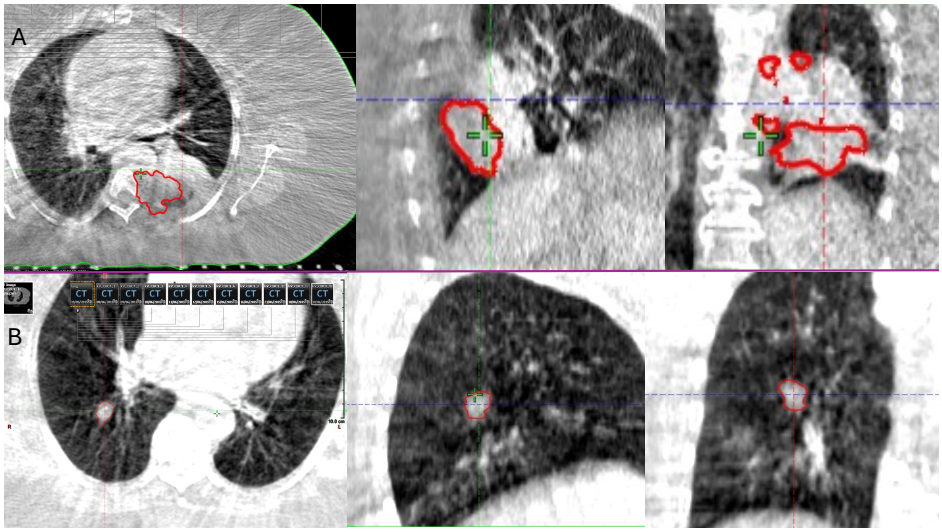


Figure 5-3 Examples of Auto Contours derived by Smart Segmentation on CBCT datasets

Panel A displays CBCT of a tumour where autocontouring failed to identify the tumour due to artefact on the scan due to large body habitus.

Panel B displays CBCT of a tumour Example where autocontouring performed well and no adjustments were required.

Median volume difference between AC and UA-AC was 0.9cm^3 (range 0-100cm^3), as seen in Figure 5-4.

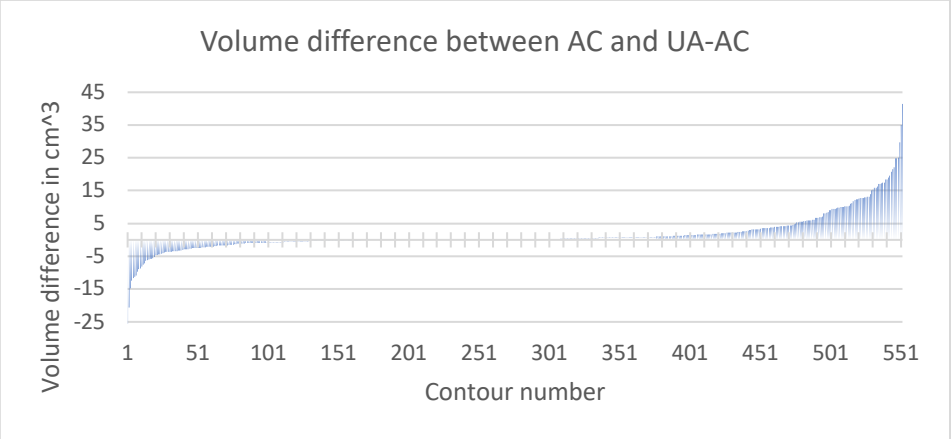


Figure 5-4 Volume difference between auto-contour and adjusted auto contour for each scan

Arranged by magnitude difference (excluding one outlier where the difference was -100cm^3)
X axis displays each scan, Y axis displays the volume difference

5.3.3 Efficiency

Median time to adjust the AC on planning CT was 59 seconds (range 0-501 seconds) and on CBCT was 83 seconds (0-460 seconds). Day 1 CBCT contour adjustment took significantly longer than planning CT, median of 92 seconds (range 84-460 seconds), $p< 0.0001$.

There was no significant difference in contour adjustment time from first CBCT to last, median 98 seconds (range 92-298 seconds) (p=0.79 ANOVA).

Median adjustment time for Adenocarcinoma and Squamous Cell Carcinoma subtypes were 86.5 seconds (range 6.6-166.4 seconds) and 112.7 (33.9-292.1 seconds) respectively, (p=0.123).

For each patient an average time to adjust the AC was calculated and this was compared to tumour motion, and the percentage of the GTV

adjacent to soft tissue. Weak correlation between an increased contour adjustment time with each factor were observed (Figure 5-5).

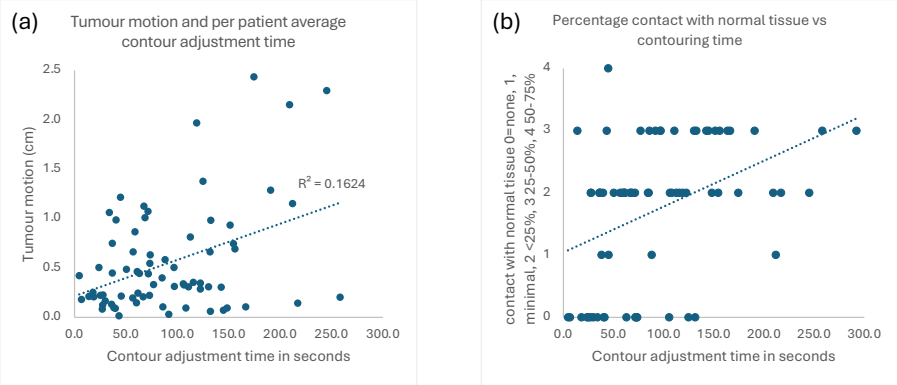


Figure 5-5 Contouring adjustment time in seconds

Scatter plots demonstrating the relationship between contour adjustment time and tumour motion in panel (a), and tumour contact with normal tissue in panel (b)

17% (n=102) of the contours were independently audited, and excellent agreement was observed between the RO and the RTT (Table 5-3). Only 2 volumes required major revisions, and these occurred in the same patient where a small extension of tumour was omitted by the auto contouring system and not included in the revised volume.

Table 5-3 Independent clinician audit of user-adjusted auto contours on CBCT. Agreement rated on a 4 point scale.

	(n)	*%*
total reviewed	**102**	**100**
no revisions	87	85.3
minor revisions	13	12.7
major revisions	2	2.0
Unsuitable contour	0	0.0

When comparing the UA-AC generated on the AVIP, to the manual contour defined on the 50% phase image of the same scan, GTV_{AVIP} was

smaller in 87.5% (n=63) of cases. Median ratio of $GTV_{50\%}$ to GTV_{AVIP} was 0.74 (range 0.25-3.28) and Pearson Correlation Coefficient = 0.94.

5.3.4 On-treatment tumour volume dynamics

For each patient TVR in response to RT was determined from the CBCT acquired on days 1-3 and weekly thereafter. See Figure 5-6 for the timepoint of each scan acquisition.

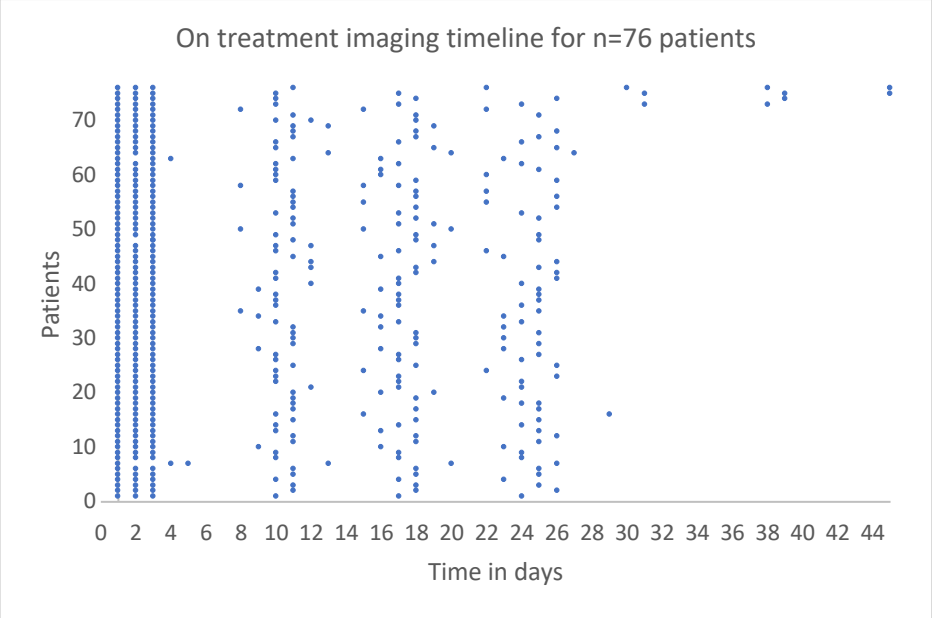


Figure 5-6 CBCT scan acquisition timepoints for each patient.

Blue dots represent a CBCT acquired, each row is an individual patient

Median tumour volume on day 1 of RT was 13.5 cm^3 (range 0.7-119.9 cm^3), Figure 5-7. The final CBCT was acquired on median day 25 (range 17-45 days) and median volume at this timepoint was 9.3 cm^3 (range 0.3-95.5cm^3), Figure 5-8.

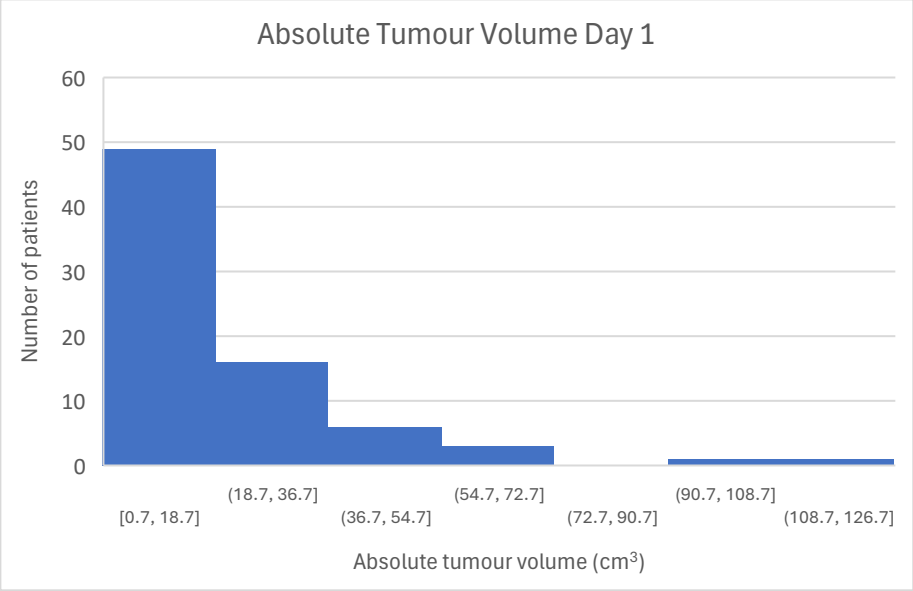


Figure 5-7 Histogram displaying range of tumour volumes day 1 RT of n=76 patients.

Volume bins along the x axis and number of patients on y axis.

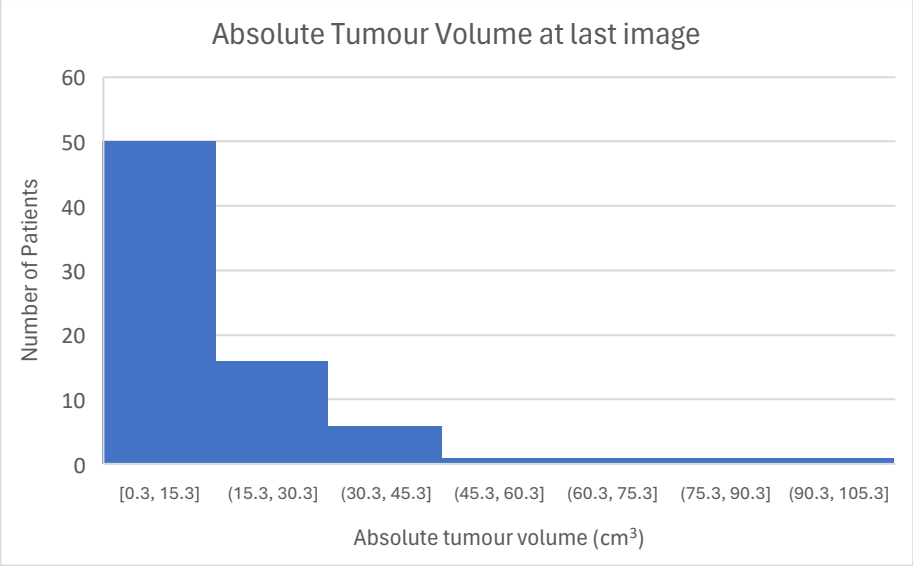


Figure 5-8 Histogram displaying range of tumour volumes at last imaging timepoint of n=76 patients.

Volume bins along the x axis and number of patients on y axis.

5.3.5 Tumour volume regression

Sixty-two (81.6%) patients had TVR by the end of RT. No trend was observed linking initial volume to magnitude of regression, Figure 5-9 and Figure 5-10.

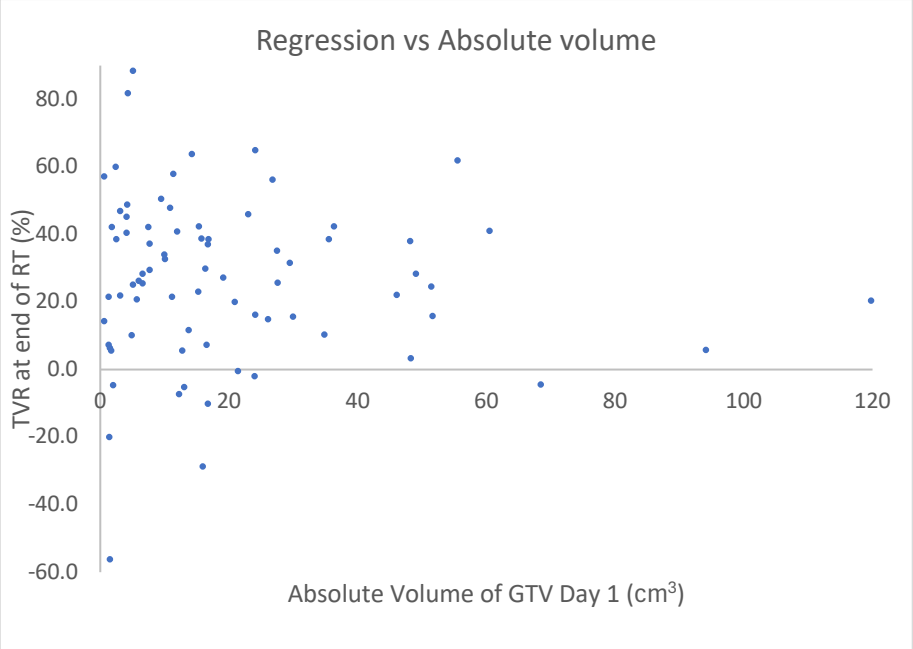


Figure 5-9 Scatter plot displaying TVR against Day 1 absolute volume.

Blue dots represent each individual patient.

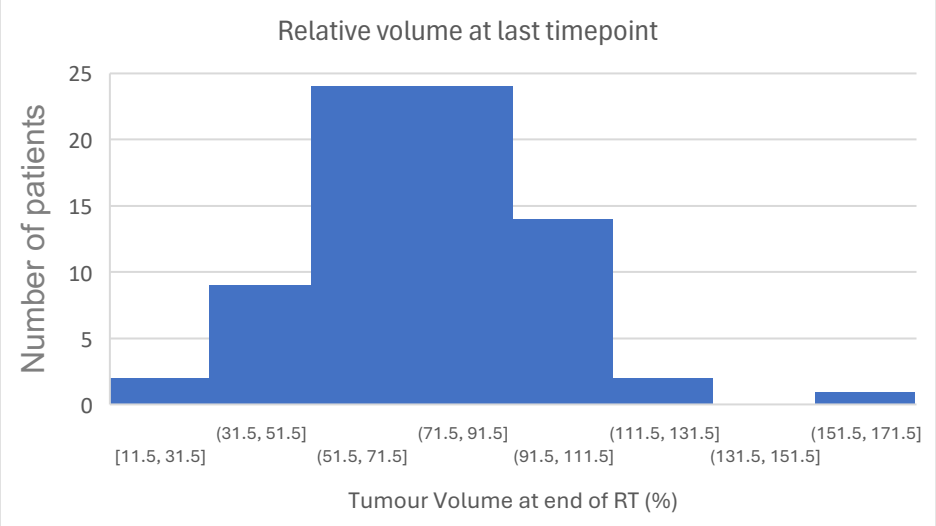


Figure 5-10 Histogram displaying the range of tumour volumes at end of RT (%).

Volume bins along the x axis and number of patients on y axis.

In those patients with TVR during RT, a median rate of regression of 0.67% per day (range 0.04-2.59%), was observed. The median regression from day 1 to final CBCT was 15.83% (range 0.78-64.71%).

Two (2.6%) patients remained stable, and 12 (15.8%) showed tumour volume progression with a median progression rate of 0.52% per day (range 0.04-1.67%), and a median increase in volume of 13.19% (range 0.83-38.46%). Relative response in all patients can be seen in Figure 5-11 and individual tumour response dynamics are displayed in Appendix 10.3.1 for each patient.

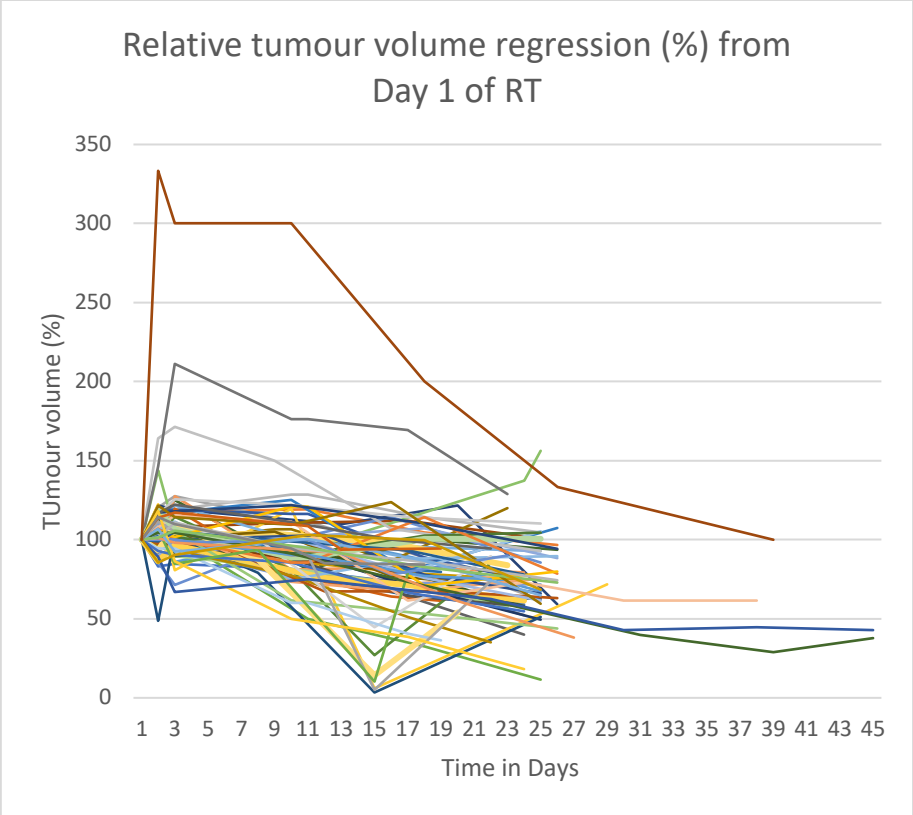


Figure 5-11 Relative tumour volume changes over the course of RT.

X axis displays the time in days with Day 1 being the start of RT. Y axis displays the tumour volume as a percentage relative to volume on day 1.

Rates of change in patients treated with modulated vs 3DCRT techniques, and between histological subtypes were compared. All data were normally distributed as per Shapiro-Wilk test. No significant variation was detected between 3DCRT and IMRT (p=0.24, t-test) Figure 5-12, or between Squamous, Adenocarcinoma or Clinical subtypes (p=0.13 ANOVA), Figure 5-13, Table 5-4.

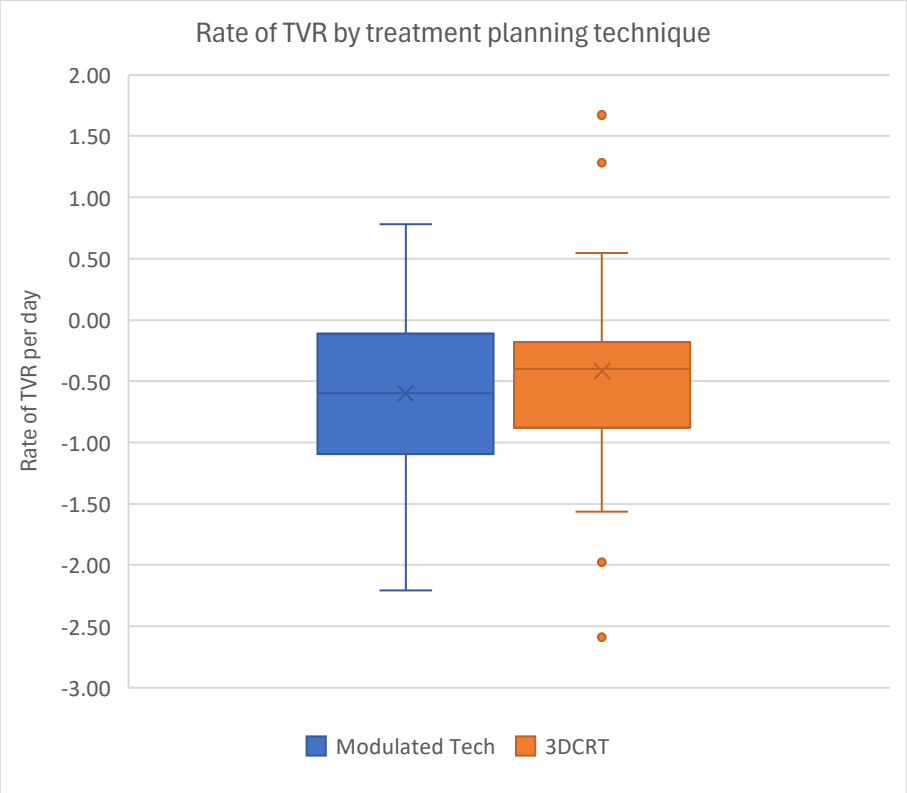


Figure 5-12 Box and Whisker plot displaying TVR rates (% TVR per day) in patients treated with 3DCRT or modulated techniques

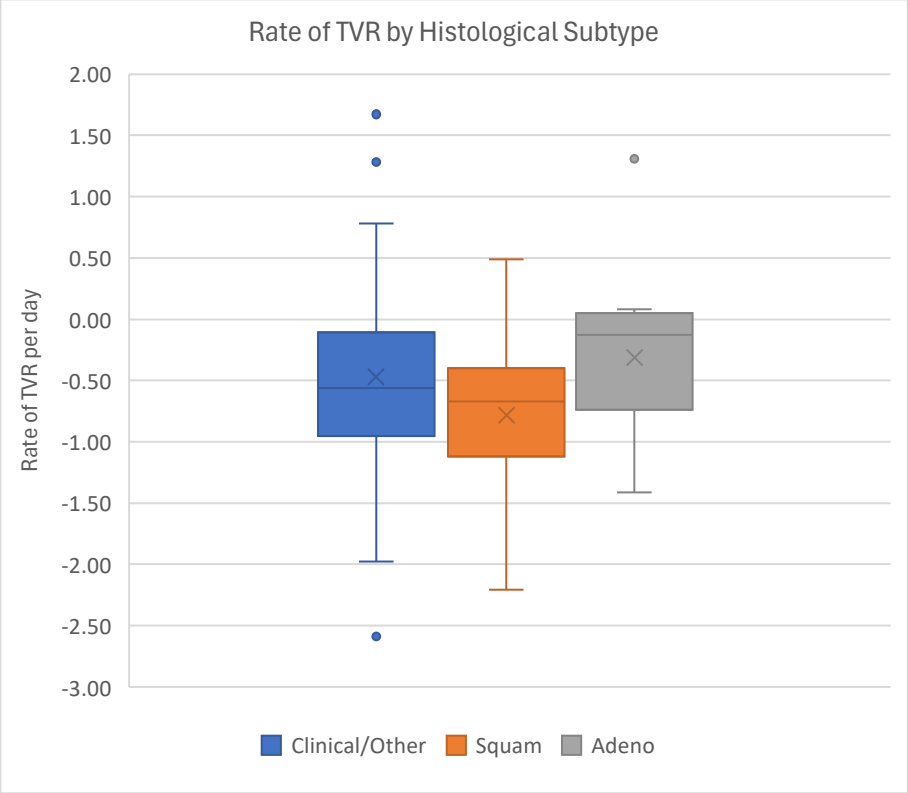


Figure 5-13 Box and Whisker plot displaying TVR rates (percentage TVR per day) in AC, SCC or Clinical/Other subtypes.

Table 5-4 Summary of tumour volume dynamics in response to RT

IMRT: Intensity Modulated Radiation Therapy, VMAT: Volumetric Arc Radiation Therapy, 3DCRT: 3-Dimensional Conformal Radiation Therapy

	Regression	***Progression***
	Median rate per day (n)	**Median rate per day (n)**
All patients	0.67% (62)	0.49% (14)
IMRT/VMAT	0.68% (43)	0.36% (9)
3DCRT	0.58% (19)	1.28% (5)
Squamous	0.68% (23)	0.25% (2)
Adeno	0.41% (10)	0.08% (4)
Clinical/Other	0.69% (29)	0.61% (8)

5.4 Discussion

This study sought to pilot semi-automated GTV delineation of NSCLC cases on AVIP and CBCT images and to elucidate relevant factors which may influence delineation. Image quality prevented the use of this approach in 33.3% of patients and the explicit factors that led to exclusion are outlined in Table 5-1. In the 76 patients deemed suitable for AC, the process was found to be efficient, with median adjustment time of 59 and 83 seconds respectively for AVIP and CBCT scans, and effective with 98% of contours suitable with only minor revisions.

One third of patients (33.3%) were unsuitable for CBCT delineation primarily due to challenges defining the tumour volume on non-contrast enhanced imaging. The benefit of both IV contrast [103] and multimodality imaging such as FDG PET [102] are well established recommendations to improve GTV delineation in planning CT scans. The lack of these imaging aids in CBCT comparatively restricts the interpretation of these scans.

Eliminating motion artefact in the form of a breath hold CBCT, for example, could improve image quality and consequently delineation accuracy [185]. Breath hold is uncommon when treating patients with NSCLC outside of SABR but up to 87% of patients are compliant with achieving breath hold [186] and increasing the use of this motion management techniques would improve AC quality on CBCT data.

Effacement of normal soft tissue structures did not prevent GTV delineation on CBCT, except for select tumours attached to the diaphragm or located within the hilum. Many lung AC algorithms rely on the density difference between the tumour volume and the low density of lung tissue. Tumours adjacent to soft tissue structures are therefore more challenging to define due to weak contrast between the structures [100]. While a trend for increased contouring time was observed in those cases,

the trend was weak. Similarly, the increased contour adjustment time noted for tumours with greater motion was not significant. A strength of this study is the measurement of the time required to manually edit the AC, which is lacking in most studies reporting on AC implementation [184, 187]. It is essential that the time required by a member of the multidisciplinary team (MDT) to review and edit a contour is reported to ascertain if the addition of an AC process is clinically helpful. In all cases, the overall time required to manually adjust the AC was low, median 83 seconds on CBCT, suggesting that this workflow may be suitable for adoption in select cases on busy treatment units.

Furthermore, the process was carried out by an RTT, which could facilitate more widespread adoption of GTV delineation compared to the requirement for the RO to repeatedly delineate the GTV. RTT led target volume delineation has been adapted in other sites such as prostate, bladder and breast [173, 188-190]. While lung cancer may pose more challenges as the target is a GTV not a clinical target volume of an anatomical organ, RTT led delineation of lung SABR volumes has been reported [111].

Accuracy of the UA-AC volumes was considered, although quantitative analysis was constrained by the lack of a clinical GTV on either AVIP or CBCT, which would facilitate geometric comparison of the volumes. Delineation of NSCLC tumour volumes is subject to uncertainties due to the impact of cardiopulmonary motion and CBCT limitations [175, 191] and, as a result a clinician's manual contour could be a flawed benchmark. Interobserver variability between clinicians has shown to increase over the course of RT [158], indicating that repeated delineation of the GTV by the RO for comparison purposes could introduce uncertainty. A pragmatic approach of auditing the RTT adjusted contours by an expert clinician was taken here, and demonstrated excellent agreement between the RO and RTT, with only 2 contours requiring a major revision. Prior work

using this contouring methodology on 4DCT data demonstrated both geometric and dosimetric acceptability of the GTV volumes generated (Chapter 4) [182]. One limitation of this study is a single observer adjusted all volumes, and while this eliminates inter-observer variability, to strengthen this analysis both repeated delineation by the single observer, and inclusion of multiple observers would be desirable to measure the intra- and inter-observer variability of the method.

Physician interobserver variability in repeated lung GTV delineation resulted in volume changes ranging from -86% to +109% [192], but this was measured on CT imaging, not CBCT which could have a larger uncertainty. A similar study evaluated a semi-automated contouring tool in peripheral lung tumours, and found comparable accuracy of the contouring approach in both planning CT and CBCT datasets, but a trend of poorer interobserver agreement was reported as RT course progressed [177]. Although only evaluating manual delineation, one group reported that inter-observer variability in CBCT GTV delineation was greater than inter-scan variability, but the observers in this case were a consultant RO, a resident training RO and a medical student. This implies at least one of the observers was in-experienced in lung GTV delineation [193]. These studies evaluated peripheral lung tumours in the SABR setting, which may not be comparable to this population where 71% of tumours were adjacent to soft tissue structures. There is a paucity of literature on CBCT delineation in this scenario. One study reporting on automated and semi-automated GTV delineation for ART in the locally advanced NSCLC setting found the automated contours to be generally acceptable, with minimal changes required to the AC [194]. Larger modifications were required when there was greater tumour regression, and in an individual who had an oesophageal stent placed min-treatment. This highlights that the process of supervising the AC system and making clinical adjustments remains critical to ensure quality contours. This

study however, used a deformable algorithm to deform the planning GTV onto the daily CBCT, compared to creating a new AC for each scan.

A qualitative review of the contour accuracy is somewhat limited as it is a subjective evaluation of the structure, however, this approach is commonly reported in the literature evaluating AC ([184, 195]. Here we found that 98% of contours were deemed acceptable with no revisions or minor revisions, which compares favourably to other studies which reported 79% and 88.75% of AI based auto-segmentations of lung GTV were acceptable with minor modifications [196] [197].

The second aim of this study was to describe the tumour volume dynamics in response to RT. In this population 81% of patients had TVR measured on the last image, and the median rate was 0.67% per day. This is smaller than that reported by Michienzi et al. who reported a rate of 1.5% per fraction in 90% of patients (n=27) [191] and Lim et. al., reporting a TVR rate of 1.46% per fraction in all patients evaluated (n= 31) [198]. In both studies, the patients were treated with CCRT with a dose fractionation schedule of 60Gy/30# and >45Gy respectively. In this population the patients received definitive RT alone with a more accelerated approach of 55Gy/20#. Both studies reported heterogeneity in tumour stage and histological subtypes similar to the population investigated here.

The median regression measured here was 15.83%, which is less that the rates reported in other studies 39.3%, 39%, 25% and 18.5% [74, 76, 105, 138]. Similarly, these studies report on patients treated with CCRT. One study reported a mean reduction of 38% in a population treated with adjuvant chemotherapy followed by RT alone, however the data was gathered on MVCT [106] which has lower resolution the KV CBCT and may not be generalisable to higher quality imaging.

A further potential confounding factor is tumour histology. In this dataset a large proportion of patients have a Clinical or Other diagnosis, indicating that no histology information is available. Histology may influence the rate of TVR as response to RT is linked to proliferation rates and this varies between subtypes [34, 199]. Furthermore, the relationship between regression and outcomes can vary between adenocarcinoma or squamous cell carcinoma [70].

Potential applications of this work are in online ART workflows [200, 201] and delta radiomics analyses [66, 202], which require repeated GTV delineation. A 2020 European survey highlighted that the increased use of ART in lung cancer RT is a priority for many centres, but technical limitations were identified as a barrier to implementation [203]. One such technical limitation is CBCT image quality limiting GTV definition. This study has identified the factors that can aid patient selection for CBCT GTV delineation, and in these appropriately selected patients, we have shown this commercially available solution produced acceptable GTVs that could be generated in a short timeframe.

The delineation of longitudinal GTV volumes facilitates training and testing of the utility of the PSI model on large datasets of routinely collected imaging. Prior work was limited by either additional imaging requirements or reliance on the RO to manually edit structures, and large scale delineation is desirable to facilitate research [204]

5.5 Conclusion

Semi-automated delineation using a commercial solution was feasible in appropriately selected patients. Tumour volume regression was observed in 81% of patients at a rate of 0.67% per day. Tumour volume kinetics were derived and recorded for 76 patients which will facilitate external validation of the PSI model.

CHAPTER 6 STUDY 3

TRAINING AND TESTING THE PSI MODEL IN NON-SMALL CELL LUNG CANCER RADIATION THERAPY

This chapter was published in the Internation Journal of Radiation Oncology Biology and Physics in 2025.

Barrett S., Zahid .M.U., Enderling H., Marignol L., Predicting Individual Tumour Response Dynamics in Locally Advanced Non-Small Cell Lung Cancer Radiation Therapy; a Mathematical Modelling Study. International Journal of Radiation Oncology, Biology and Physics, 2025, 121(4): p1077-1087

See Publications for a copy of the published article.

6.1 Introduction

GTV changes during RT for NSCLC are common with up to one third of all patients having significant tumour shrinkage, but the ability to predict which tumours may shrink remains elusive [205]. The rate of GTV regression for individuals varies, and has been reported in this population at a rate of 1.5% decrease per fraction resulting in a mean regression of 43.1% (+/-23.1SD) from the start to end of RT [191].

TVR in response to RT is also an emerging imaging biomarker, which may be predictive of clinical outcomes such as LC or OS [68, 71, 72] and research is ongoing in this area [72, 73]. Current research focuses on retrospective analyses of tumour volume dynamics over time and how

these relate to clinical outcomes. In this context, it would be desirable to predict the tumour regression at an earlier timepoint, and if the response is predicted to be suboptimal, dose-adapt RT to improve that response.

This study evaluates the performance of the PSI mathematical model to predict tumour regression over the course of RT delivery. The model, originally proposed by Prokopiou et al. [79-83], utilises longitudinal tumour volume measurements to calculate a patient-specific PSI, and uses this index to simulate and predict response to RT. The PSI model allows us to simulate tumour response to RT using an initial dose fractionation schedule, or test response to alterative prescriptions if modelled response is suboptimal. We can then theoretically optimise the RT prescription for each patient as demonstrated recently for total dose personalization for head and neck cancer [84].

The PSI model has been tested in a small retrospective cohort of NSCLC patients [79] [80] and has been explored more thoroughly in the head and neck RT setting where it is shown to accurately predict tumour volume dynamics during RT [81, 82, 96].

This study examines the utility of the PSI modelling approach across a range of dose fractionation schedules in NSCLC. As outlined in the TRIPOD statement this work is a Type 1a study, describing the development of a prediction model [98]. This study tests the robustness of the model's performance across 4 conventionally fractionated dose prescriptions commonly used in treatment of LA-NSCLC. It determines the model's performance at fitting data in these prescription groups, with one or two free parameters, and it will assess if it accurately predicts tumour response to RT in these patients. Finally, we ascertain the model sensitivity to input parameter variation.

6.1.1 Aims and Objectives

1. To establish the ability of the PSI model to fit tumour volume kinetics data for patients receiving 4 different dose prescriptions for NSCLC RT
    a. To establish how well the model fits the data in each prescription group, by plotting measured vs simulated volumes on a scatter plot and determine the R^2 value. MSE and Akaike information criterion (AIC) will be calculated for each prescription group.
    b. To ascertain if PSI or α (radiation sensitivity parameter) is more patient specific by plotting values in Box and Whisker plots and evaluating range and standard deviations.

2. To examine the feasibility of using a single variable model
    a. To derive the optimal α for each prescription group (α_g) and establish if these values vary significantly between groups and if no significant difference is observed to derive a population α value (α_p)
    b. To establish how well the model fits the data in all 4 prescription groups using α_p by plotting measured vs modelled volumes on a scatter plot and determine the R^2 value and calculate MSE.
    c. To quantify potential change in model quality by comparing AIC in each prescription group to prior values.

3. To ascertain if the PSI model can accurately predict TVR over the course of RT.
    a. To model tumour regression for the course of RT using only the first 3 tumour volume measurements acquired during RT and the α_p value.

b. To establish the accuracy of the modelled volumes using scatter plots, R^2 and PCC.
c. To determine if PSI outputs significantly vary using reduced datasets compared to full dataset using Box and Whisker plots and t-test comparison.
d. To perform a parameter sensitivity analysis to establish robustness of the model to variations in the input parameters.

6.2 Materials and Methods

6.2.1 Patient population

A de-identified dataset containing tumour volume measurements for 141 patients with NSCLC was abstracted from a prior publication [71]. This dataset is described fully previously [71], but briefly, contained tumour volume measurements acquired longitudinally during a course of radical RT or Chemo/RT delivered at Maastro Clinic, Maastricht, the Netherlands. Patients within the dataset were heterogeneous in relation to stage, histology and use of chemotherapy, however, all patients were treated with an RT dose >45Gy. Tumour volume was measured on free breathing CBCT. Patients were treated with several distinct dose fractionation schedules. For the purposes of this study included patients from this dataset were those treated with 3 prescriptions; 23–24 fractions × 2.75Gy daily (n= 28), 32–42 fractions × 1.8Gy daily (n= 27), and 30 fractions × 1.5Gy Bidaily followed by 5–12 fractions × 2Gy daily (n=84), resulting in n= 139 patients.

An additional dataset was abstracted from a second publication [118]. This dataset consisted of 25 patients with LA- NSCLC, treated with daily RT on the Hi-Art helical Tomotherapy unit at the London Regional Cancer Program, Ontario, Canada, from 2005 to 2007 and is most fully described in Woodford et al (2007) [106]. Of the 25 patients 2 were excluded as they were not treated with a radical dose prescription (n=23). These data were previously included in a PSI evaluation study [80], however in this work the PSI model algorithm has been updated to Particle Swarm Optimisation (PSO) from a genetic algorithm which may alter the model performance. A prescription dose of 30-32 fractions x 2 Gy per fraction was used for these patients and all were treated with neo-adjuvant chemotherapy [118]. GTV of the primary tumour only was measured on daily megavoltage CT (MVCT) images. Each dataset

included tumour volume measurements and the associated timepoint it was acquired (daily or weekly imaging) from day 1 of RT.

6.2.2 Model parameters

The PSI model has 3 modifiable parameters, a tumour intrinsic growth rate (λ), a radiation sensitivity value (α) and the patient specific PSI. Custom scripts in MATLAB (MathWorks) were generated to simulate TVR with inputs of dose fractionation schedule, λ, α, and PSI. A variety of optimisation scripts were written using the PSO function within MATLAB Global Optimisation Toolbox. The optimal parameter values are found by minimising the MSE across all the data therefore finding the values for λ, α and PSI which result in the best fit to the data. As the data for this study contained no pre-treatment information to calculate unperturbed tumour growth, the published $\lambda = 0.012$ 1/day was utilised for all patients [80], α and PSI were derived as per methods below.

6.2.3 PSI model data fitting with 2 individual parameters

For each prescription group an optimal fit of the data was derived using MATLAB PSO. The model was allowed to optimise parameter values for α and PSI for each individual patient (α_i, and PSI_i), by minimising the residual MSE between measured and simulated tumour volumes over time. Lower bounds of 0 and upper bounds of 0.99 were set for both parameters. Scatter plots of measured vs simulated tumour volumes were generated for each prescription group. The R^2 and MSE values were calculated as a measure of how well simulated values matched measured values.

R^2 was calculated using the formula $1 - \frac{RSS}{TSS}$, where RSS= Residuals Sum Squared (measured volume - simulated volume)^2 and TSS = Total Sum of Squares (Sum of measured volumes- mean of measured volumes)^2.

MSE for each patient, i, was calculated as the normalized square error at each observation, j, between the measured Volume, $V_{m,j}$, and the simulated volume, $V_{s,j,}$ averaged over the number of observations, n_m:

$$MSE_i = \frac{1}{n_m} \sum_{j=1}^{n_m} \frac{(V_{m,j} - V_{s,j})^2}{V_{m,j}},$$

and MSE per group (MSE_g) was calculated as the average of MSE_i of all patients in the group:

$$MSE_g = \frac{1}{n_{patients}} \sum_{i=1}^{n_{patients}} MSE_i.$$

Akaike information criterion (AIC) was calculated as a measure of model quality using the formula (2*k)+n*ln(SSE_i/n) where k is the number of free parameters, n is the number of data points and SSE_i = MSE_i *n_m is the Sum Squared Error for each patient, i. AIC per group is calculated as the average of AIC of all patients in the group When comparing models using AIC values, it is assumed that the model with the lowest AIC (AIC_{min}) results in the least information lost from the true data. A change in AIC (Δ AIC) from AIC_{min} ≤ 2 is substantial evidence that there is no loss in model quality, Δ AIC ranging between 4 and 7 have considerably less support than the AIC_{min} model, and models with Δ AIC > 10 have essentially no support [206].

The range of $\alpha_{i,}$ and PSI_i values generated for each patient were visualised with Box and Whisker plots. Mean and Standard Deviations were calculated for each group. The range of $\alpha_{i,}$ values derived between prescription groups were compared using single factor ANOVA and Box and Whisker plots to estimate the biological volume reduction effect of each dose fractionation schedule.

6.2.4 PSI model data fitting with a single patient specific parameter

An optimal α for each prescription group (α_g), was determined by testing a range of α values in each group. The value which generated the best fit to the measured data, assessed by the highest R^2 achieved, was selected.

The model then derived the optimal fit to the data using α_g and was allowed the freedom to generate only an individual PSI parameter (PSI_i). Following data fitting with α_g and PSI_i, scatter plots were generated and R^2, MSE_g, and AIC in each prescription group were calculated. Finally, data fitting with a population α (α_p) was tested using the mean of the group α values. Data were compared to prior values generated using α_i, and PSI_i.

6.2.5 Prediction of tumour response to RT

To test the predictive power of the model, only the initial 3 tumour volume measurements for each patient were inputted into the model, along with the population α value, and the tumour regression during the remainder of RT was simulated. The simulated volumes were compared to the measured volumes, excluding the initial 3 measures, using scatter plots, R^2 and PCC. A sensitivity analysis on the model parameters was carried out to quantify the model-robustness, by varying the input parameters by $\pm 20\%$ and comparing resulting R^2 and PCC values to those generated with the optimal parameters. The individual PSI values derived using only 3 data points were compared to those derived using all data using Box and Whisker plots and paired t-test.

6.2.6 Data analysis

Model simulations and data analyses were carried out in MATLAB (MathWorks), Microsoft Excel and SPSS.

6.3 Results

6.3.1 Description of patient population and available tumour volume measurements

162 patients with 1474 tumour volume measurements were included in this study. Patients in the 2Gy per fraction group had a median of 30 tumour volume measurements (range 15-33) measured on MVCT. For the remaining 3 prescription groups a median of 6 tumour volume measurements were available for each patient over the course of treatment, range 5-7, with 2 patients having 7 CBCTs, and 24 having 5 CBCTs. All patients had CBCT day 1 of RT. See Figure 6-1 for overview of scan acquisition timepoints. Median (range) tumour volumes on day 1 for each group were 115.30cc (3.2-830.8cc), 32.5cc (0.6-198.3cc), 29.6cc (1.3-327.8cc) and 43.4cc (0.7-318.1cc) for the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively (Figure 6-2). Volumes in the 2Gy dataset were significantly larger than the other 3 ($p< 0.0001$) but no significant difference was seen between remaining 3 groups ($p=0.667$). Demographic and clinical data for each patient were unavailable.

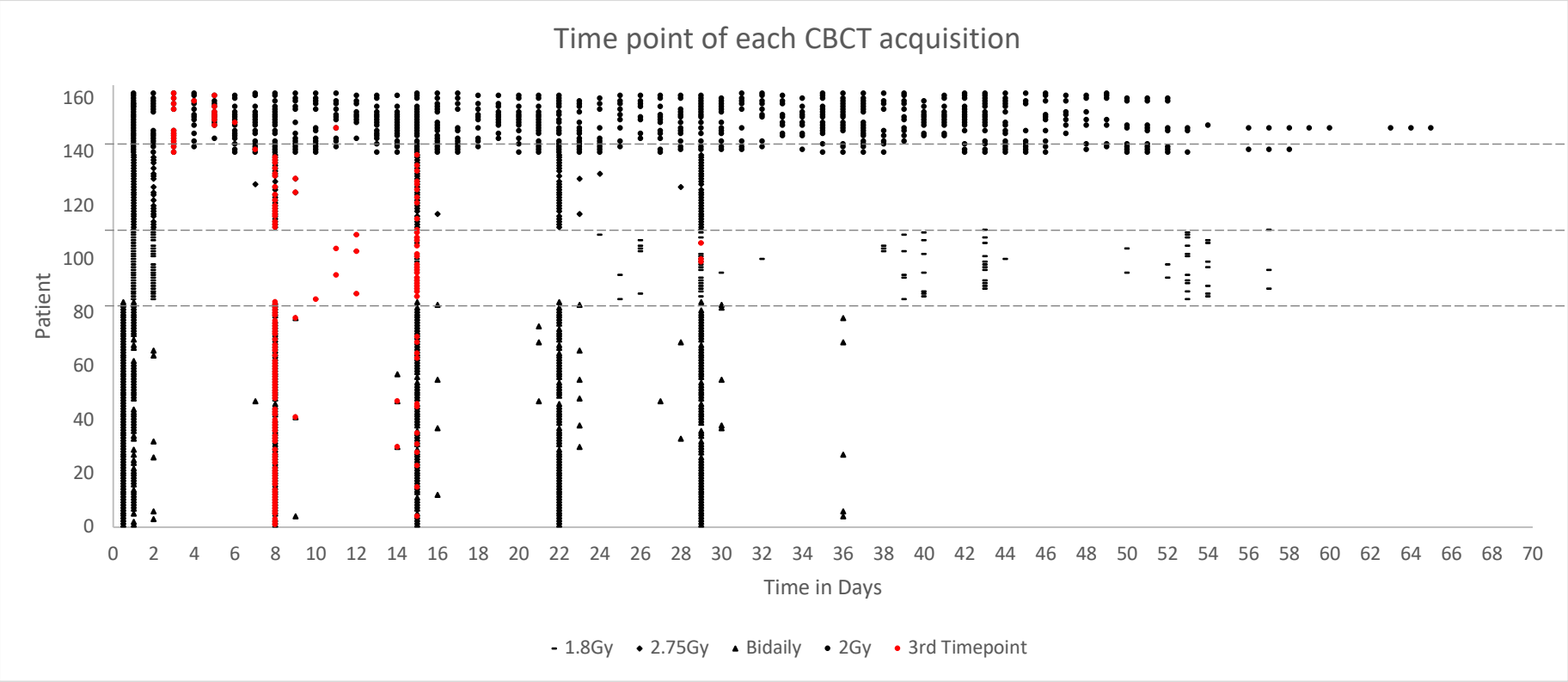


Figure 6-1 CBCT acquisition timepoints

For each patient, the timepoint of each CBCT and associated tumour volume measurement in days since RT started is displayed in black. The timepoint of the 3rd measurement for each patient highlighted in red. Patients in the bidaily group had more than one image on day 1 and so these are coded as day 0.5 and day 1.

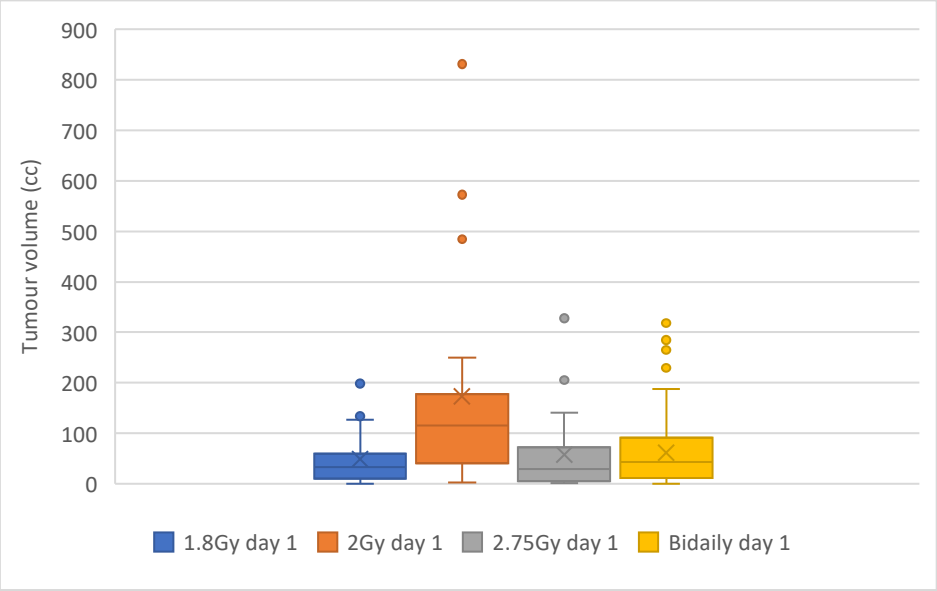


Figure 6-2 Box and Whisker plot displaying the range of tumour volumes (cm^3) Day 1

6.3.2 PSI model fits data with individual patient parameters

We first tested that the PSI model was able to fit the tumour volume kinetics data with individual α and PSI parameters. The PSI model, with α_i and $PSI_{i,}$ fit the data in each prescription group with high accuracy. Scatter plots for each group show excellent agreement between measured and simulated tumour volumes (Figure 6-3) as indicated by R^2 values of 0.993, 0.995, 0.997 and 0.956 and MSE_g of 0.625, 0.142, 0.211, 2.071 in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively. The range and standard deviation for derived PSI_i values were larger than α_i in all 4 prescription groups (Figure 6-3). The range of α_i values derived for each prescription group (Figure 6-4) were compared using single factor ANOVA and were not found to be significantly different, p=0.0845. Mean (±SD) α_i values were 0.028 (±0.018), 0.035 (±0.022), 0.024(±0.021) and 0.035 (±0.024) in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively.

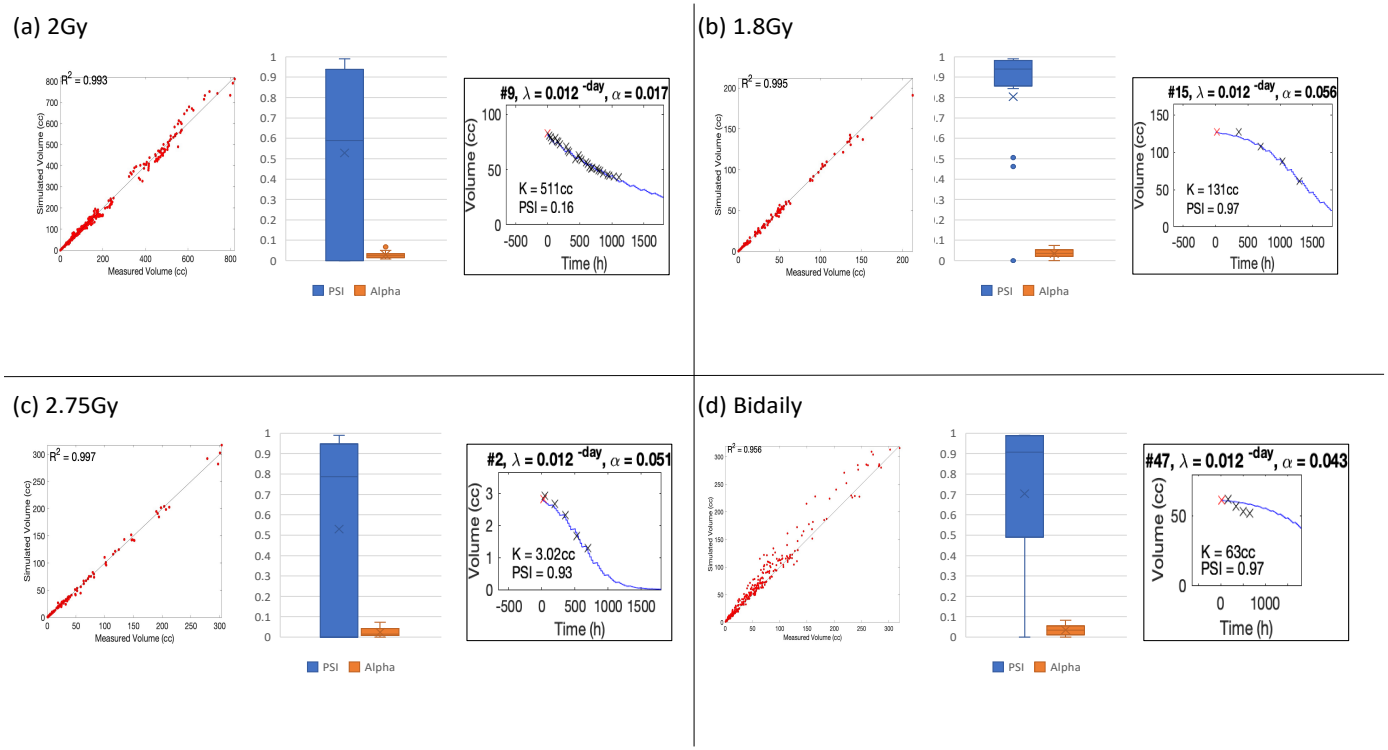


Figure 6-3 PSI model data fitting and parameter output values

Panel (a), (b), (c) and (d) show scatter plots of data fitting using α_i and PSI_i across the 2Gy, 1.8Gy, 2.75Gy and Bidaily prescription groups respectively Box and Whisker plots display the values derived for PSI_i and α_i in each group, additionally included is a sample of individual patient fit in each prescription group with measured volumes marked as x and model simulated volume displayed as blue curve.

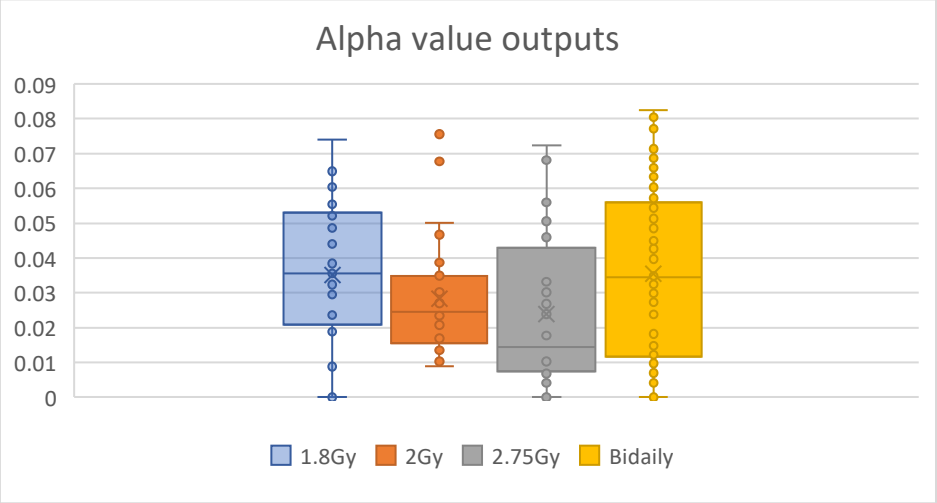


Figure 6-4 Box and Whisker plot displaying the range of α_i values derived in each prescription group

6.3.3 PSI model fits the data with a single patient specific parameter.

We next examined whether data fitting performance of the PSI model was significantly impacted if we use a prescription specific α parameter, or a population α parameter.

The α parameter for each prescription group (α_g), were estimated at 0.02, 0.03, 0.022 and 0.048 in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively. Using the relevant α_g and allowing individual PSI (PSI_i), excellent fit of the data was achieved as shown in scatter plots (Figure 6-5). R^2 values were found to be 0.99, 0.991, 0.995 and 0.992 and MSE_g values were 1.202, 0.231, 0.289, 0.314 respectively. An improvement in data fit was observed in the Bidaily group with an increase in R^2 from 0.956 to 0.992, and a decrease in MSE_g from 2.071 to 0.314.

Mean AIC values were calculated and compared to original data fitting with α_i, and PSI_i to examine if there was a loss in model quality moving from 2, to 1 free parameter. In the 1.8Gy and 2.75Gy groups marginal changes in mean AIC values were observed, with a Δ AIC increase of +0.37 and Δ AIC decrease of -0.08 measured in the single parameter

models respectively. A moderate increase in mean AIC (Δ AIC +6.12) was measured in the 2Gy group indicating a moderate loss of data fit quality when reducing to one free parameter in this group. A substantial decrease in mean AIC (Δ AIC -9.06) was observed in the Bidaily group indicating that the single parameter model resulted in a truer fit to the data in patients with this prescription.

Finally, a population α parameter, (α_p) was tested using the mean of the group α values, $\alpha_p = 0.03$.

Again we found excellent fit of the data with R^2 values of 0.988, 0.991, 0.995 and 0.992 and MSE_g values of 0.878, 0.23, 0.293 and 0.324, in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively (Figure 6-5). Minimal changes in AIC values were noted moving from (α_g) to (α_p) with the same trends as those reported above. In the 1.8Gy and 2.75Gy groups marginal changes in mean AIC values were observed, Δ AIC +0.37 and -0.25 respectively compared to the 2 parameter model. In the 2Gy group Δ AIC was +6.12 and in the bidaily group again the Δ AIC decreased substantially by -9.11.

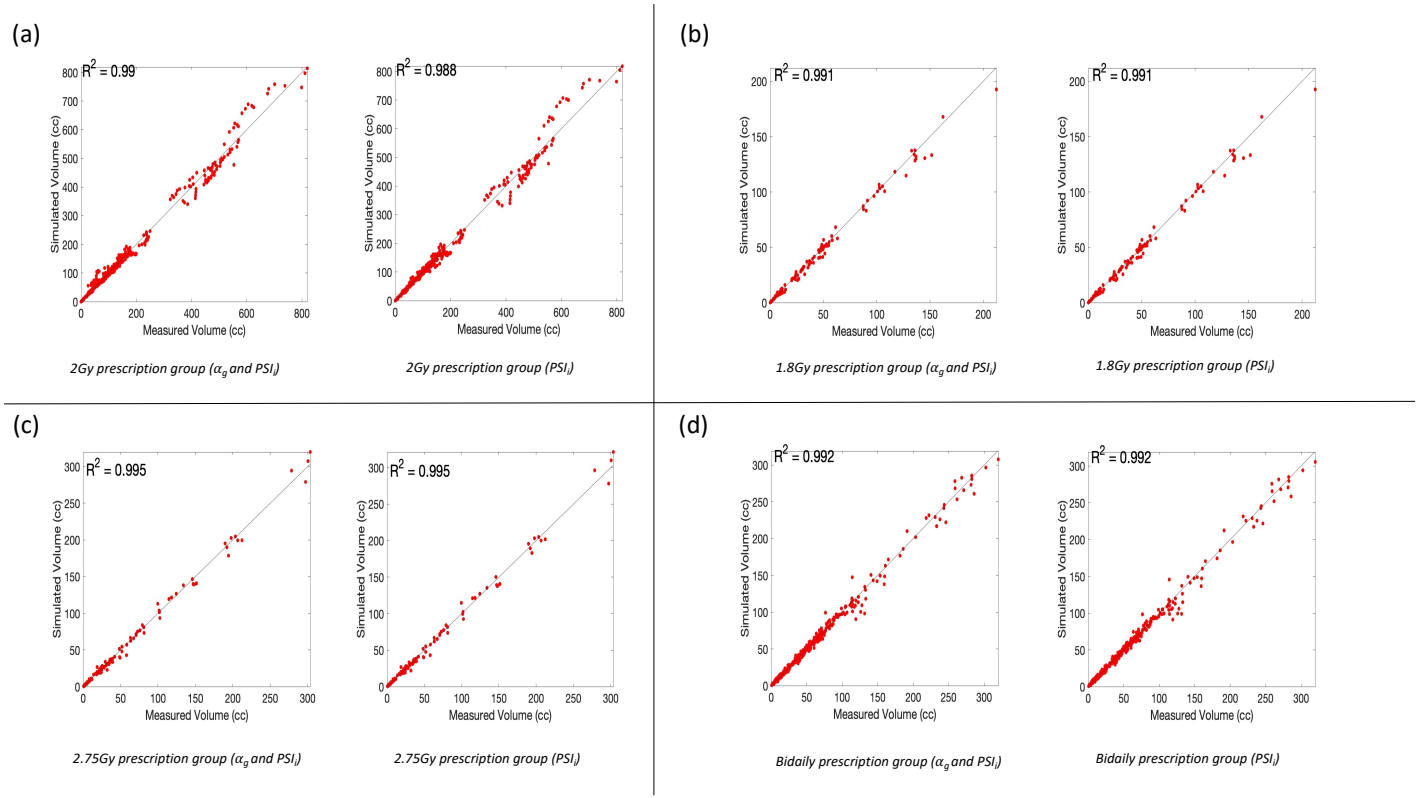


Figure 6-5 Scatter plots displaying data fitting using group and population α values.

X axis displays measured volumes (cm^3) and Y axis model simulated volumes (cm^3)Panels (a), (b), (c) and (d) show scatter plots of data fitting performance in single variable model using prescription group α (α_g, PSI_i) and population α (α_p, PSI_i) across the 2Gy, 1.8Gy, 2.75Gy and Bidaily prescription groups respectively

6.3.4 PSI model accurately predicts tumour volume regression during RT

Finally, we tested the ability of the PSI model to predict tumour volume kinetics based on the initial 3 measurements of volume acquired during RT using population λ and α. PSI model predicted tumour volume regression during RT with a high degree of accuracy, as seen in the scatter plots demonstrating measured vs predicted volumes (Figure 6-6), with values for R^2 = 0.968, 0.954, 0.968, 0.937 and PCC = 0.984, 0.977, 0.984, 0.968 in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively. Individual patient plots of model prediction vs measured data are available in Figure 6-7, Figure 6-8Figure 6-9Figure 6-10.

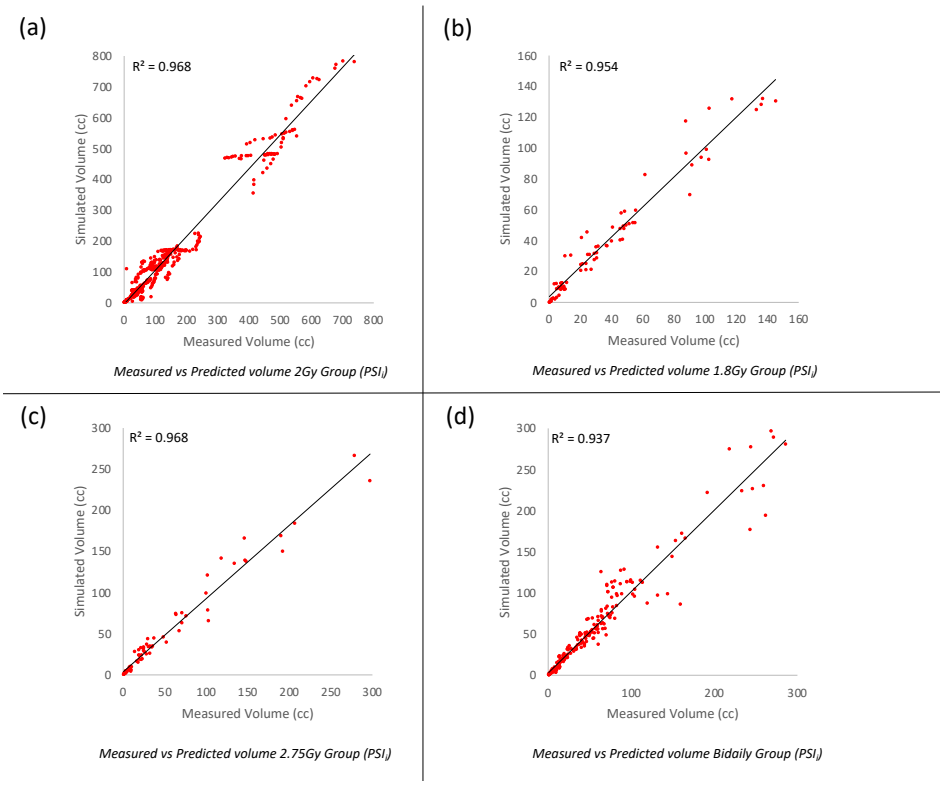


Figure 6-6 Data prediction scatter plots

Scatter plots display measured vs predicted tumour volumes, in each prescription group single variable model (α_p and PSI_i), excluding the first 3 tumour volume measurements. Panels (a), (b), (c) and (d) displaying the values in the 2Gy, 1.8Gy, 2.75Gy and Bidaily groups respectively

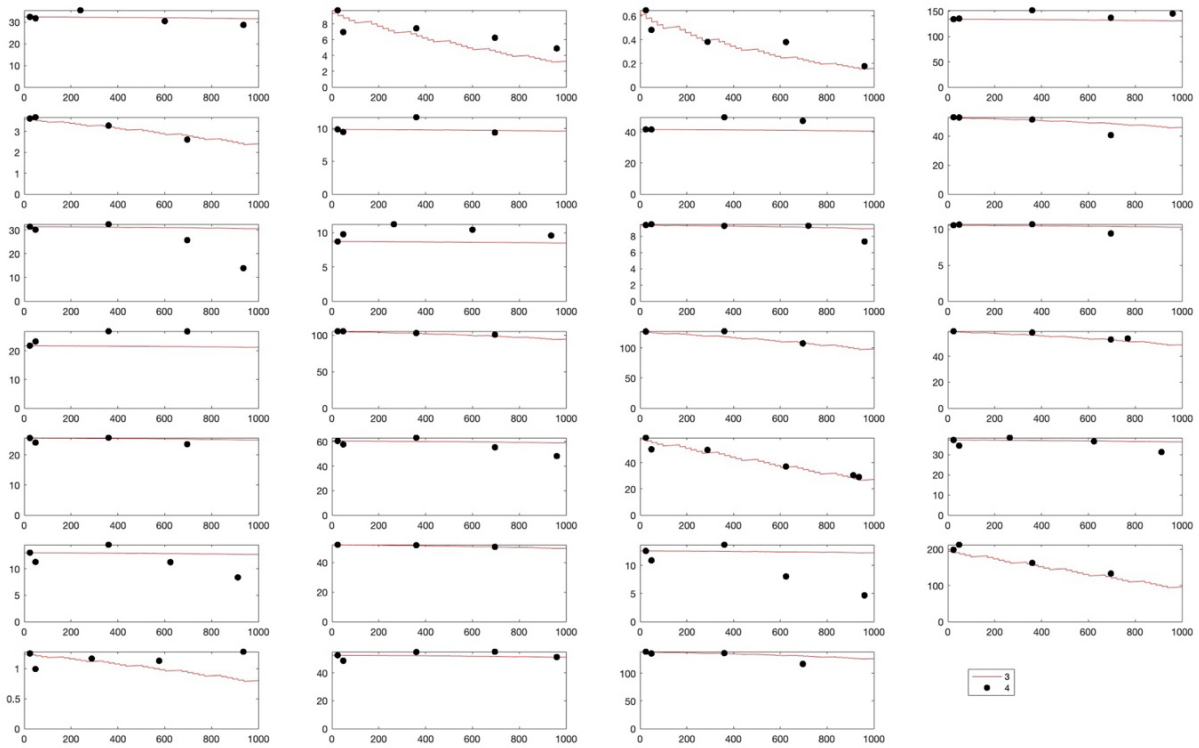


Figure 6-7 1.8Gy Group measured vs predicted volumes in individual patients

X axis denotes time in hours, y axis is tumour volume in cc

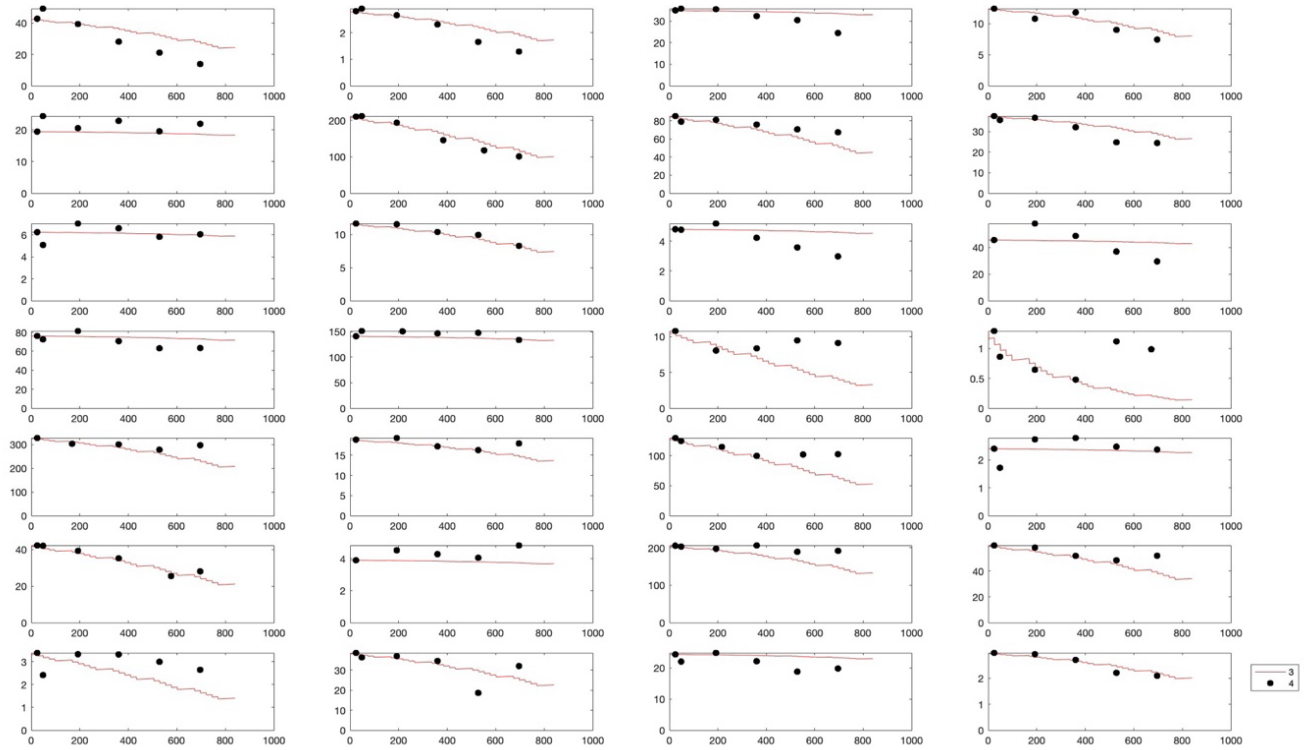


Figure 6-8 2.75Gy Group measured vs predicted volumes in individual patients

X axis denotes time in hours, y axis is tumour volume in cc

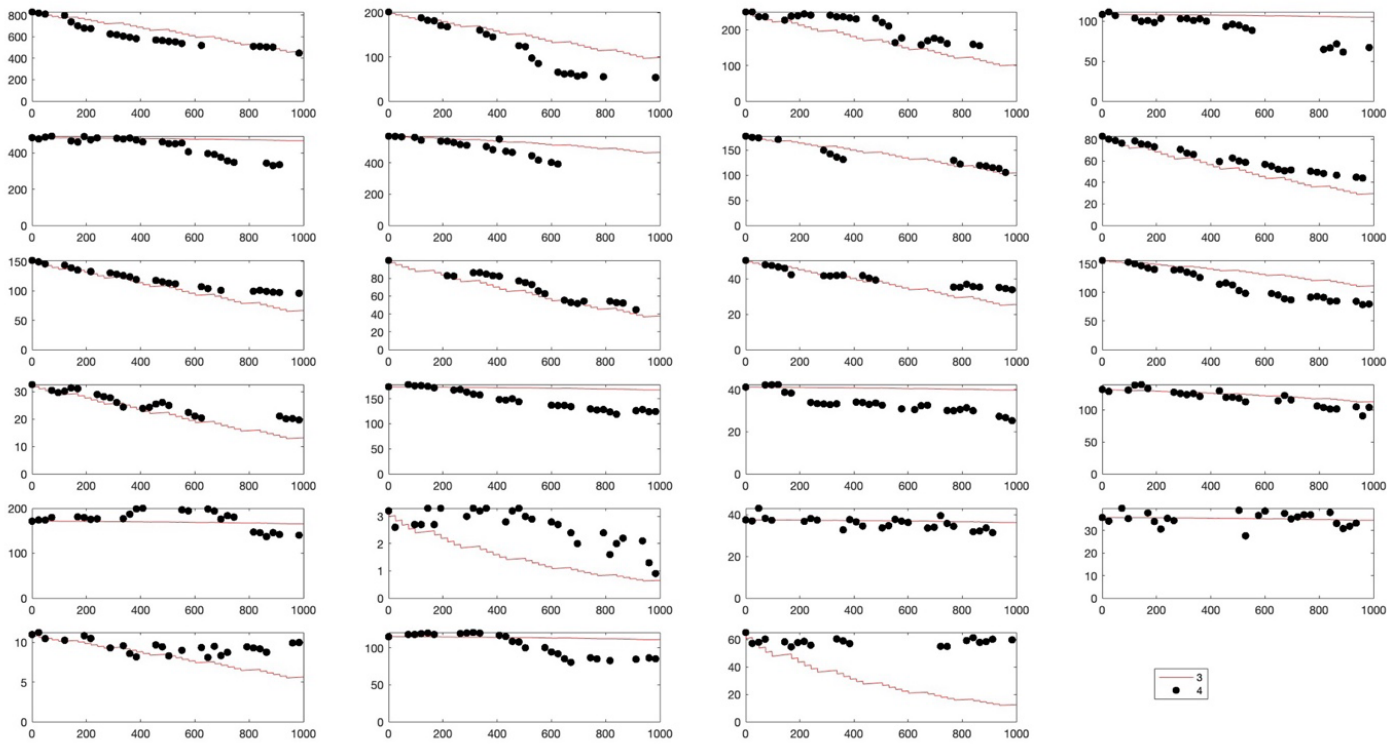


Figure 6-9 2Gy Group measured vs predicted volumes in individual patients

X axis denotes time in hours, y axis is tumour volume in cc

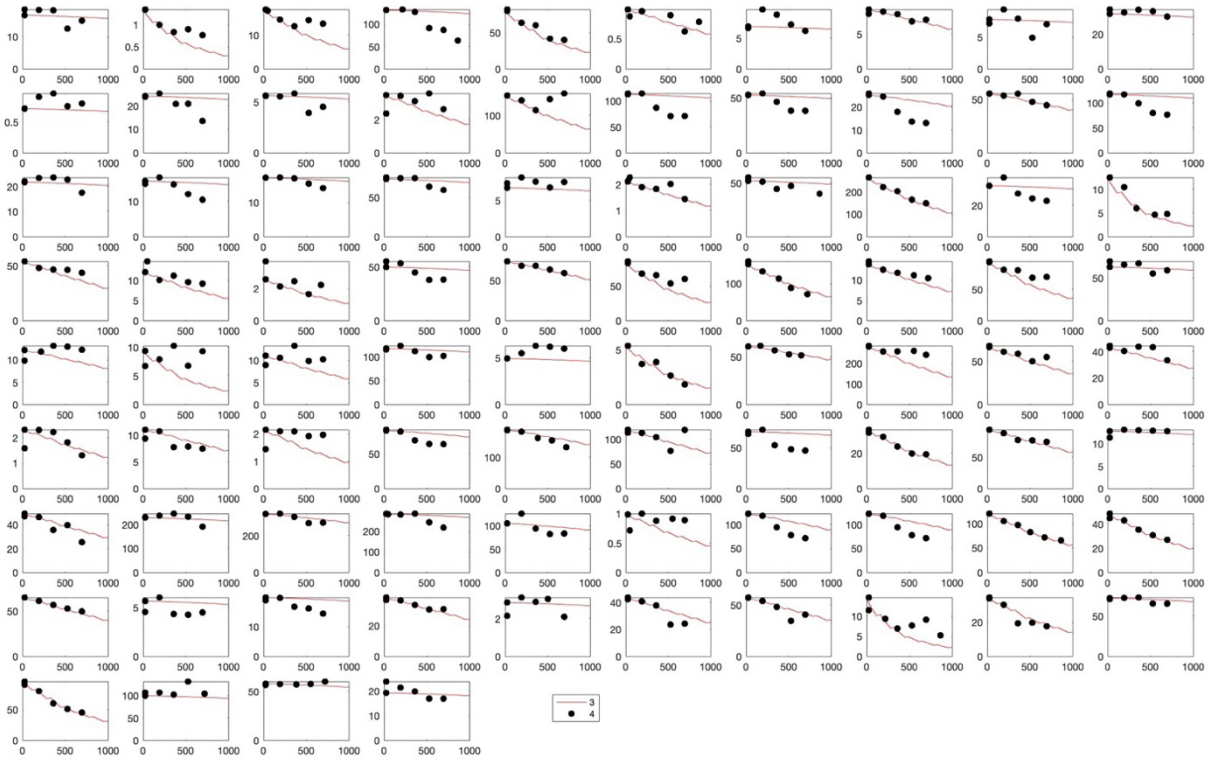


Figure 6-10 Bidaily Group measured vs predicted volumes in individual patients

X axis denotes time in hours, y axis is tumour volume in cc

Patient specific PSI values derived with only 3 data points were compared to those generated with the complete dataset. Those derived from the Bidaily Group were significantly different with mean PSI in the complete dataset = 0.631 and mean PSI derived from 3 data points = 0.889 ($p < 0.00001$) (Figure 6-11). The model was tested for sensitivity to parameter changes by varying the α and λ parameters by $\pm$ 20%. No substantial change in performance was observed, with the largest percentage difference in either R^2 or Pearson Correlation Coefficients being 2.15%. See Table 6-1 for all values.

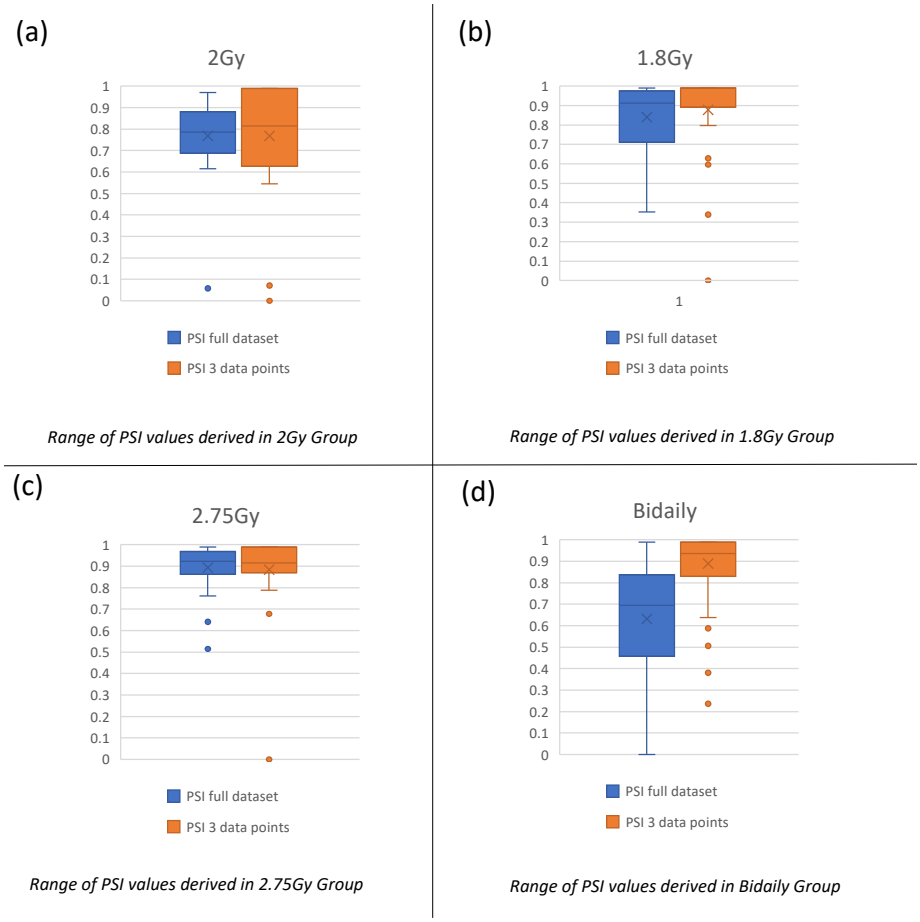


Figure 6-11 Box and Whisker plot displaying the range of PSI values derived from full dataset and from first 3 tumour measures

Table 6-1 Parameter Sensitivity Analysis

Table shows the R^2 and Pearson Correlation Coefficient for data prediction after varying the optimal derived input parameters (α 0.03, λ 0.012) by +/- 20% in each prescription group. R^2= Coefficient of Determination, PCC= Pearson Correlation Coefficient, % Δ = Percentage Difference

Prescription	***α 0.03, λ 0.012***		***α, λ -20%***				***α, λ +20%***			
Group	R^2	PCC	R^2	$\% \Delta R^2$	PCC	$\% \Delta PCC$	R^2	$\% \Delta R^2$	PCC	$\% \Delta PCC$
2Gy	0.969	0.984	0.972	-0.27%	0.986	-0.14%	0.964	0.50%	0.982	0.25%
1.8Gy	0.954	0.977	0.934	2.15%	0.966	1.08%	0.952	0.26%	0.975	0.13%
2.75Gy	0.968	0.984	0.969	-0.13%	0.984	-0.07%	0.965	0.25%	0.982	0.12%
Bidaily	0.937	0.968	0.941	-0.38%	0.970	-0.19%	0.930	0.75%	0.965	0.37%

6.4 Discussion

The PSI model is a mathematical tool that simulates and predicts volumetric tumour response dynamics during RT. This study examined the PSI model performance fitting data in 4 distinct prescription groups, derived a population radiation sensitivity parameter (α_p) and established the performance of the model predicting tumour response to RT in each group.

This study showed that the PSI model can fit longitudinal data from patients treated for LA-NSCLC across a variety of dose prescriptions to a high degree of accuracy (R^2 range 0.956-0.997, Figure 6-3) using a population wide tumour growth rate parameter (λ_p) derived from the literature. This is an improvement on the previous reported results in both NSCLC [80] and oropharyngeal cancer [81]. Sunassee et al [80] previously examined the PSI model in the patient group treated with 2Gy per day also included in this study. Here we found an improvement in the models performance with a previous R^2 of 0.91 reported compared to R^2 of 0.99. This prior work employed an earlier iteration of the PSI model which did not use PSO to fit the data. PSO is a computational method that aims to iteratively solve a problem by having a number of potential solutions which move around the search space to find an optimal solution. The prior work employed a genetic algorithm to derive a combination of parameters that best fits patient data [80, 207] and this may be susceptible to converging to a local minima instead of the global minimum when solving for optimal parameters. Improvement was also seen in the models predictive performance with this study reaching excellent agreement between measured and simulated volumes in this group, $R^2 = 0.97$ compared to $R^2 = 0.84$ in the previous work [80] which may be explained by both the change in algorithm and the omission of 2 patients which may have been incorrectly identified as having radical RT. It is notable also that the prior work predicted regression from week 2 of RT, but this

analysis used only the first 3 tumour measures, which for the majority of patients (n= 19/23) were acquired by day 5 of RT indicating this improved performance was at an earlier timepoint in treatment.

We allowed the model freedom to create an individual α and PSI for each patient to maximise data fit. The range of values derived are shown in (Figure 6-3). PSI was more variable between patients than α, with PSI ranging from 0 – 0.99, and a ranging from 0 – 0.08. While this was previously investigated in the 2Gy per fraction group [80] it was not tested in a variety of dose fractionation schedules. These findings support the hypothesis that PSI is more variable and patient specific, than the α parameter across all prescription groups.

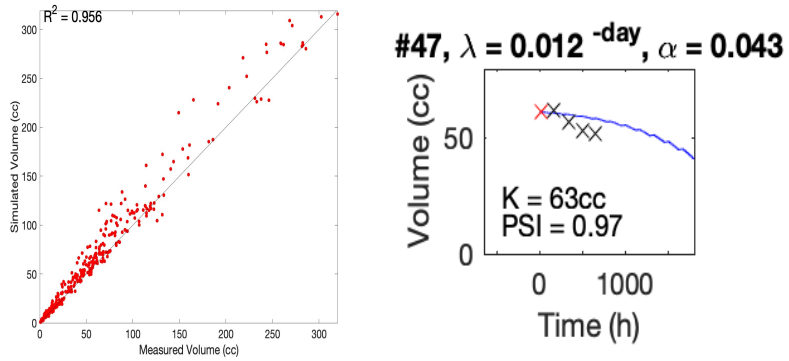
We next investigated an optimal α value for each prescription group by testing a range of α values and establishing which value resulted in the best fit to data, and the difference in individual α values did not reach significance in this small sample (p=0.0845, Figure 6-4). Finally this study confirms that a population α_p value, regardless of dose per fraction, has minimal impact on data fit compared to α_g, with negligible variation in the R^2 (Figure 6-5).

We hypothesised that there could be a loss in model quality when reducing from 2 to 1 free parameters. Model quality was assessed using AIC formula. AIC is a mathematical method to evaluate how well a model fits the data [208], and in this case, we calculated AIC for both single parameter models (α_g/α_p and PSI_i) and the model with 2 free parameters (α_i and PSI_i).

Interestingly AIC_{min} was observed in the single parameter model (α_p, PSI_i) in two groups (2.75Gy and Bidaily) indicating that a single parameter model is preferable. Δ AIC was marginal in the 2.75Gy group (0.25) but was larger in the Bidaily group (9.11) indicating much less support for

the 2-parameter model in this dose fractionation schedule. On review of the scatter plots and individual data fit there was also an observable improvement in the performance of single parameter model in the Bidaily group (Figure 6-12). This unanticipated improvement in the model was robust to adjustments of individual PSO parameters, and therefore could not be attributed to poor optimisation within the model. We anticipate that future examination of a larger dataset will elucidate these results. This prescription was the most complex with Bidaily followed by daily fractions and this finding warrants further investigation in future work.

(a) Bidaily Group, (α_i PSI_i)



(b) Bidaily Group, (PSI_i)

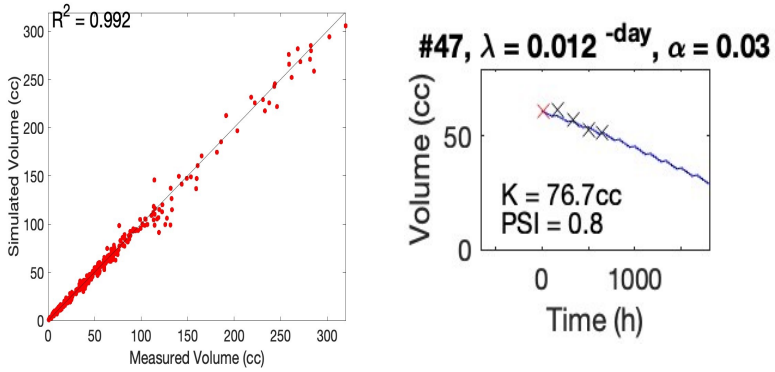


Figure 6-12 Bidaily prescription group variation in fit between single- and 2-parameter models

Panel (a) shows scatter plot of data fitting using α_i and PSI_i in the Bidaily prescription group and a sample of individual patient fit with measured volumes marked as x and model simulated volume displayed as blue curve.

Panel (b) demonstrates the same data and individual patient using α_p and PSI_i

There was no significant loss in model performance, either in AIC or in R^2 values observed when using α_p compared to using α_i. This suggests that we can reasonably use the single variable model in differing prescriptions. Again, these results are in line with the prior head and neck and NSCLC papers which evaluated the PSI model [79-81] which also found reasonable data fitting allowing only PSI as a patient specific parameter within the model. However, this study was the first to evaluate the model in a number of dose fractionation schedules for NSCLC.

Finally the ability of the model to predict tumour response based on early tumour volume dynamics was tested. The first 3 tumour volume measurements were used to predict the remainder of RT and these measures were acquired at varying timepoints in the 4 groups. In the 2Gy group the majority of patients had a 3rd volume by day 5 of RT. In the 1.8Gy and 2.75Gy group 3rd measure was acquired, in the majority of cases, by days 8-15. In the Bidaily group the majority of 3rd tumour volume measures were acquired by day 8. Therefore the prediction of tumour regression was taking place from days 5 to 15 in the most cases (Figure 6-1). Performance of the model stratified by image acquisition timepoint was not assessed in this study as we aimed to test the prediction at the earliest possible timepoint and therefore used the first 3 measures for all patients. Future work in a homogeneous population could examine the improvement in performance when adding more data to the prediction. We found that the single variable PSI model predicted tumour volume regression across the entire population with a high degree of accuracy (Figure 6-6) with R^2 values ranging from 0.937 to 0.969, and all PCC values > 0.968. The findings demonstrate that the model can predict RT response in varying dose fractionation schedules, a pre-requisite if we wish to use the PSI model to simulate response to alternative prescriptions. Prior work by Sunassee et al. [80] only evaluated the model performance in patients receiving 2Gy per fraction, which may not be representative of the current approaches in lung cancer RT.

While we see from these values that the model performed well across the population, the scatter plots reveal some patients where a poor correlation was observed between measured and predicted tumour volume. This is most notable in the 2Gy group (Figure 6-6(a)). 2 patients had predicted volumes that changed very little over the course of RT but the actual tumour volume reduced. On closer review of these 2 cases, one showed a small volume growth in the first 3 measures (477.7cc to 492.9cc) and the second had only a very slight volume decrease (570.7cc to 564.5cc). In both cases the model predicted little to no tumour volume regression during RT. These patients demonstrate that while on a population level the model predicted to a high degree of accuracy, in individual scenarios where tumour volume does not decrease there may be difficulty in predicting the remainder of treatment. Of note, both cases occurred in the 2Gy data group, which had daily MVCT images meaning the model was using data from just the first 3-5 days to simulate RT response for the remainder of treatment.

Plots showing residual absolute volume difference between measured and predicted volumes (Figure 6-13Figure 6-14Figure 6-15Figure 6-16) also demonstrate very good agreement for the majority of patients, with less agreement observed for several patients in the 2Gy group.

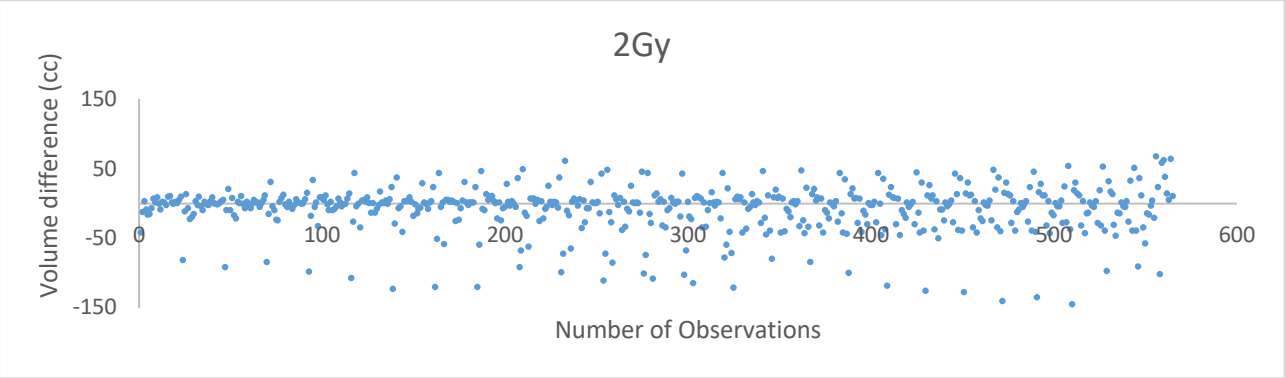


Figure 6-13 Residual volume differences (cm^3) between measured and predicted tumour volume, 2Gy

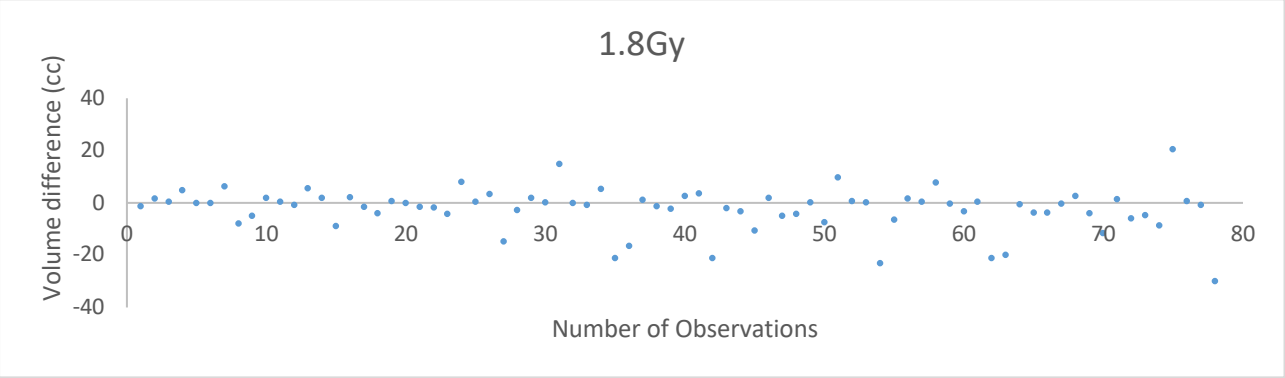


Figure 6-14 Residual volume differences (cm^3) between volume and predicted tumour volume, 1.8Gy

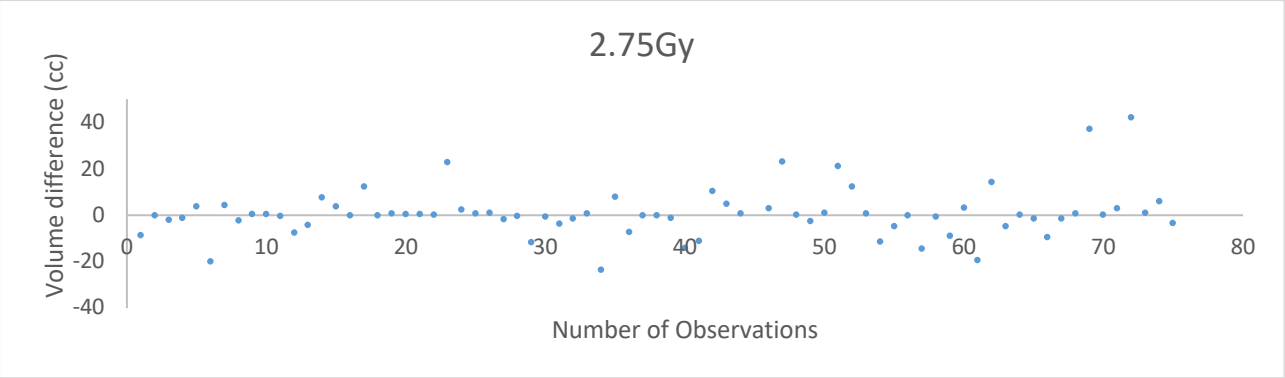


Figure 6-15 Residual volume differences (cm^3) between measured and predicted tumour volume, 2.75Gy.

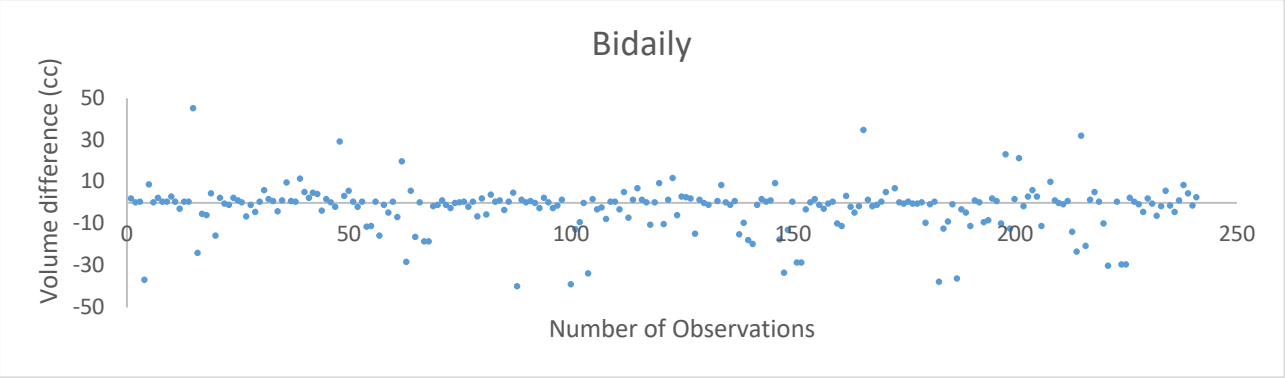


Figure 6-16 Residual volume differences (cm^3) between measured and predicted tumour volume, Bidaily.

Reviewing the individual fitting plots for all patients further enforces that the model may underperform in a subset of individual cases (Figure 6-7Figure 6-8Figure 6-9,Figure 6-10) Examples can be seen in Figure 6-17 where the model performed well and poorly. In the case where the model failed to predict the tumour regression, the first 3 data points, on which the prediction is based, show little change in tumour volume. In these cases it would be possible to add more datapoints to the model to improve the prediction, however this comes at the cost of predicting further into RT delivery, which may be less clinically relevant. Figure 6-18 demonstrates the improved performance in this case when adding further data to the model. Further correlation with biological proliferation markers such as Ki-67 would be of interest, however needle biopsy samples taken for histological diagnosis are not routinely tested for Ki-67 and so this is not feasible outside of a prospective clinical trial.

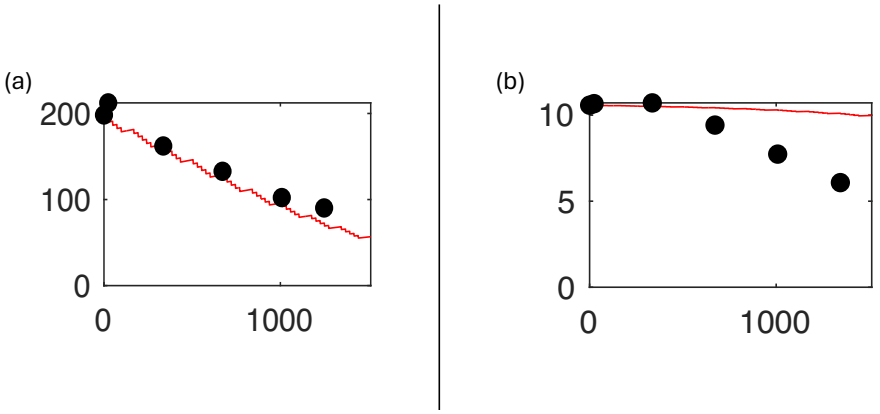


Figure 6-17 Example of good and poor individual predictions

Black dots are tumour volume in cc vs time in hours over the course of RT. Red curve is model predicted tumour regression over the course of RT based on the first 3 tumour measures.

Panel (a) displays a patient where modelled and actual volume agreed, panel (b) demonstrates a patient where the model failed to predict the tumour regression.

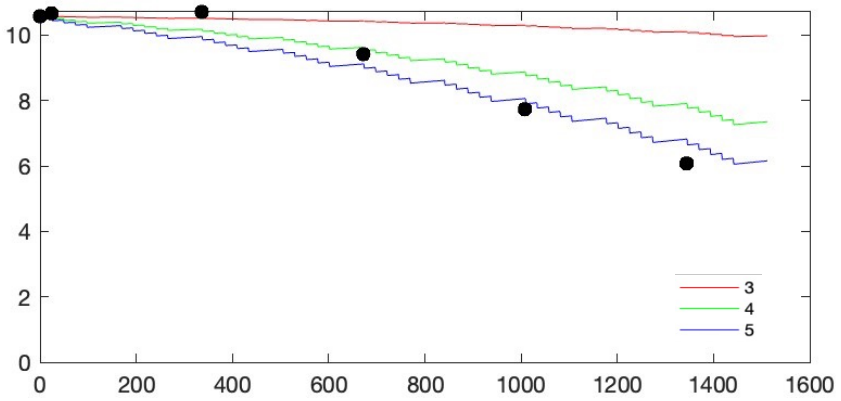


Figure 6-18 Improvement in individual prediction with the inclusion of additional imaging data

The model failed to predict the tumour regression using 3 datapoints and improved prediction with increased data input.

Black dots are tumour volume in cc vs time in hours over the course of RT.

Red curve is model predicted tumour regression over the course of RT based on the first 3 tumour measures. Green curve is predicted tumour regression using 4 data points, blue curve is predicted regression using 5 datapoints.

Accurate prediction of tumour regression during RT has several potential benefits. We know that the RT plan created prior to treatment is rarely an exact representation of the dose delivered by the end of treatment due to inter and intra-fraction variation. Significant changes in anatomy can alter the planned dose distribution and may result in increased dose to normal tissue or decreased dose delivered to the target volume [209, 210]. Predicting which tumours will significantly reduce in size could facilitate for pre-planned ART as opposed to reactive ART, which can cause delays in workflow both at the treatment unit and for the patient. Increased use of ART in lung cancer RT is a priority for many centres [203]. One study found that 90% of patients treated with CCRT displayed GTV regression during their treatment, with a mean regression rate of 1.5% per fraction [191]. In that study no trends were observed correlating tumour regression to age, histology, nodal status initial tumour size or chemotherapy [191]. This highlights the challenge in predicting those that may have clinically relevant regression, and yet the PSI model tested here, performed well

across a heterogeneous population where these factors were unaccounted for.

An emerging clinical interest is the potential link between TVR to clinical outcomes such as LC or OS [68, 71, 73, 211]. While this link has not been clearly established, a 2018 meta-analysis highlights time of assessing GTV reduction as one of several potential confounding factors [211]. More recently a retrospective observational cohort study found GTV volume following 40-50Gy RT was a more robust predictor of overall survival than initial GTV [73] a finding echoed in a second retrospective analysis [212].

The potential benefit of the PSI model in this scenario is twofold. Firstly, if regression is shown to be a reliable clinical predictor, advance knowledge of the regression and the subsequent clinical response, could facilitate a more informative conversation with the patient around the potential outcome and progress of their treatment. Secondly, as this work has shown that the model predicts regression accurately across a variety of prescriptions, future work can model an individual's response to an alternative prescription to test if the predicted response can be improved. Mid-treatment tumour volume reduction has shown to predict for clinical outcomes in NSCLC [73]. PSI based dose adaptation has been considered in head and neck setting [81, 84] and a PSI optimised prescription could improve tumour regression and consequently increase loco-regional control in retrospective data [97]. Response adapted RT is being investigated using other predictive approaches including AI [97] but the appeal of the PSI model is its simplicity and utilisation of data routinely collected in clinical practice.

While these findings are encouraging further work is necessary to validate the model in this group. As stated, this is TRIPOD Type 1a study describing model development [98], as no external validation dataset was

available, and although the model has been tested in new data, this study established a new α_p parameter value. Additionally the change in model algorithm to PSO warranted a re-optimisation of the model parameters, rather than validation of the previous version of the model. Internal validation using leave-one-out cross validation is an option when no independent dataset is available, however, it is subject to limitations such as high variance in performance estimate or unstable predictions, and may underestimate model performance [213]. External evaluation of these findings in a truly independent dataset is essential to ensure meaningful validation.

The data included was gathered from two previous publications [71, 80] . Delineation of tumour volumes in lung are subject to uncertainties due to the impact of respiratory motion and the scan acquisition parameters [175, 191] and no ground truth GTV delineation on CBCT or MVCT is available to carry out a statistical analysis of accuracy in these modalities. The GTV delineation was either manual by an expert observer using planning CT as a guide [106], or deformation of the GTV onto the planning CT with manual adjustments by an expert observer [71]. For each group in this study the imaging modality and delineation methodology was consistent for all measures and so any biases in GTV calculation would be systematic and the relative volumes in cc would remain acceptable to test these hypotheses. All datasets were anonymous tumour volume measurements, timepoints and RT prescriptions only, and therefore lacked associated clinical or demographic information. There may be confounding clinical factors or unknown biases when evaluating the relevance of tumour regression in particular the use, or lack thereof, of concurrent chemotherapy may influence the α parameter. In this analysis we only sought to evaluate the performance of the PSI model and not evaluate the clinical importance of the predicted response. As no clinical data were available regarding individual patients our findings

indicate that the PSI model works well in both the CCRT and RT alone setting as we did not stratify patients based on this criterion.

6.5 Conclusion

The proliferation saturation index model can predict, with a high degree of accuracy, NSCLC tumour volume regression in response to radiation therapy in 4 distinct dose fractionation schedules.

CHAPTER 7 STUDY 4

EXTERNAL VALIDATION OF THE PSI MODEL PERFORMANCE IN PREDICTING TUMOUR VOLUME REGRESSION IN NON-SMALL CELL LUNG CANCER RADIATION THERAPY

7.1 Introduction

Model validation is an essential process in the development and testing of any model to assess its generalisability to new data. Validation assesses the accuracy of a model and can highlight potential weaknesses or biases undetected during the training phase.

The TRIPOD initiative was published in 2015 to provide clear standards on reporting of predictive models in health care settings [98]. The TRIPOD statement denotes that where only a single dataset is available, training and validation of the model can be carried out using a variety of approaches including resampling or splitting the data into training and validation cohorts. Resampling methods include Leave One Out Cross Validation which entails retraining the model on the cohort of n -1 patients and testing performance on the single omitted patient. The process is repeated iteratively for the entire cohort and an average performance metric reported [214]. Splitting the data either randomly or non-randomly into a training and validation cohort is also commonly reported in radiation therapy literature [215]. However, training and

testing of a model in the same dataset may give a misleading impression of high performance [216].

External validation, the process of evaluating model performance in new data, is a more robust method to determine if a model is valid or useful in a clinical setting [98]. External validation tests both the reproducibility of model prediction and transferability of a model to new patients. Prospective validation takes this process further by prospectively deploying a model in a clinical trial setting to quantify its performance without the inherent uncertainties of retrospective data [217].

This chapter evaluates the PSI model performance in an independent dataset, treated for NSCLC with definitive RT which is defined as a TRIPOD Type 4 study [98]. Additionally, the clinical relevance of TVR in this population is examined.

7.1.1 Aims and Objectives

1. To test the PSI model prediction of TVR over the course of RT in an independent dataset, using early tumour response dynamics
    a. To model tumour regression for the course of RT using only the early tumour response dynamics measured on CBCT.
    b. To establish the agreement between the modelled and measured volumes using scatter plots, R^2, PCC and Bland-Altman plot.
    c. To determine if individual PSI values significantly vary using reduced datasets compared to full dataset by Box and Whisker plots and ANOVA.
    d. To perform a sensitivity analysis by varying the parameter input values by +/- 20% to quantify impact on agreement metrics.

2. To examine the clinical relevance of TVR in this cohort
    a. To assess if TVR predicts a significant difference in the clinical outcome measures of LC or OS on Univariable Cox Regression Analysis
    b. To determine if any significant volumetric variables identified persist on Multivariable Cox Proportional Hazards analysis including clinical features
    c. To test the hypothesis that including volumetric variables in a clinical model will improve its performance by comparison of ROC curves and associated AUC values.

7.2 Materials and Methods

7.2.1 Patient population

Governance approvals were provided, and ethical approval waived, by the Belfast Health & Social Care Trust (IRAS 293181) for this study.

A total of 76 patient scans were accessed from the NI HEART dataset, a cohort of 478 patients treated with radical radiation therapy (RT) for NSCLC at the Northern Ireland Cancer Centre, Belfast City Hospital, between January 1, 2015, and December 31st, 2020 [117]. For each patient tumour volume dynamics over the course of RT were derived as per the methods outlined in Chapter 5.

Of the 76 patients in the dataset generated, n=71 were treated with a prescription of 55Gy/20#/4 weeks and, to eliminate dose per fraction as a confounding factor, model validation was completed on these patients. The excluded 5 patients were treated with 66Gy/33# (n=3), 60Gy/30# (n=1) and 52.5Gy/20# (n=1). Demographics for the included patients can be seen in Table 7-1.

For all patients volume of GTV (cm^3) were extracted from the planning CT scan, CBCT days 1-3 and 3 subsequent weekly CBCTs and recorded. In addition to tumour volume measures, clinical outcome data were also available for overall survival and local progression.

Table 7-1 Demographics of patients included in external validation.

IMRT: Intensity Modulated Radiation Therapy, VMAT: Volumetric Arc Radiation Therapy, 3DCRT: 3-Dimensional Conformal Radiation Therapy

Patient Demographics	***N (%)***
Age (median, range) years	74 (54-90)
Gender	
Female	40 (56.4%)
Male	31 (43.6%)
Performance Status	
0	5 (7.04%)
1	25 (35.2%)
2	36 (50.7%)
3	5 (7.04%)
T-stage	
1	35 (49.3%)
2	20 (28.2%)
3	16 (22.5%)
Histology	
Squamous Cell Carcinoma	24 (33.8%)
Adenocarcinoma	13 (18.3%)
Clinical Diagnosis	31 (43.7%)
Other/Not otherwise specified	3 (4.2%)
Radiotherapy Planning	
3DCRT	24 (33.8%)
IMRT	14 (19.7%)
VMAT	33 (46.5%)

7.2.2 Validation of model performance using early tumour response dynamics

The published λ=0.012 [80], and α=0.03, derived as per the methods outlined in Study 3 were used as model parameter values for all patients.

To validate the predictive power of the model, the initial 3 tumour volume measurements for each patient were inputted into the model and the tumour regression for the remainder of RT was simulated. The simulated volumes were compared to the measured volumes using scatter plots, R^2 and PCC. Residual absolute difference between measured and predicted

tumour volumes were calculated and descriptive statistics of median difference and range of variation were recorded. The process was repeated with 4 and 5 volume measure inputs.

A sensitivity analysis on the model parameters was carried out to quantify the model robustness, by varying the input parameters by ±20% and comparing resulting R^2 and PCC values to those generated with the original parameters.

PSI_i values generated for each patient with 3, 4 and 5 inputs were visualised with Box and Whisker plots. The range of PSI_i values between the various data input groups were compared using an ANOVA test to establish if there were significant differences in the values when the data inputs were reduced. Model performance as a test to predict poor TVR at the end of RT (TVR ≤10%) was evaluated using sensitivity, specificity, positive predictive and negative predictive values [218].

7.2.3 Clinical relevance of TVR

Kaplan Meier (KM) plots were generated to establish if significant differences existed between T Stage or histological subtype in relation to OS or LP.

Univariable Cox Proportional Hazards Regression analysis was performed to establish the potential impact of Age, Sex, WHO Performance Status (PS) or histology on both OS and LP. In addition to these clinical factors, the following tumour volume measures were analysed:

- absolute volume at day 1 of RT in cm^3 (V_{Day1RT})
- absolute volume at the final image RT in cm^3 (V_{EndRT})
- Percentage of TVR on final image RT ($V_{\%REG}$, calculated as $[(V_{Day1RT} - V_{EndRT}) / V_{Day1RT}] * 100$)
- Tumour growth pre-treatment expressed as a percentage of the volume, ($V_{\%PRE\_RT}$, calculated as $V_{Day1RT}/V_{PCT}*100$).

The statistically significant variables identified from the univariable analysis were tested for multicollinearity, using a correlation matrix in SPSS, and Multivariable Cox Regression Analysis of the significant variables performed.

Following Multivariable Cox Regression, a Linear Predictor Score (LPS) variable was generated from all included variables, which is the computed value of the linear combination of predictor variables. Receiver Operating Characteristic (ROC) curves were generated for 3-year OS using the LPS for clinical variables, significant volumetric variables and the combined clinical and volumetric variables. Area Under the Curve (AUC) values were compared to ascertain if the addition of volumetric data added to the predictive model.

7.3 Results

7.3.1 Patient population

Seventy-one patients were included in the external validation. Each patient had 6 CBCTs, as indicated on Figure 7-1 which outlines the timepoint of each scan acquisition.

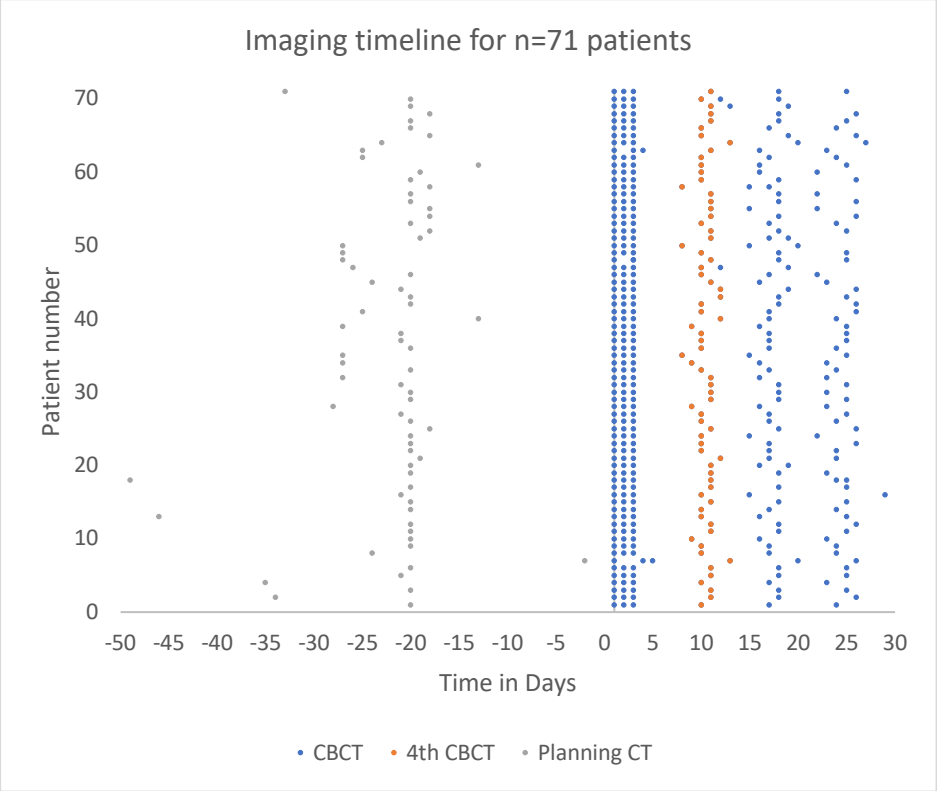


Figure 7-1 Scan acquisition timepoints for each patient in external validation

The start of RT delivery is marked as Day 1 and Planning CT scans are marked as grey dots. CBCTs are blue the 4th CBCT is highlighted as orange.

Median tumour volume on day 1 of RT was 14.3cm^3 (0.7-119.9cm^3). The 4th CBCT was acquired on median of day 10 (range 8-13 days), final image was acquired on median day 25 (17-29 days) and median volume at this timepoint was 10.4cm^3 (0.6-95.5cm^3).

Sixty-one patients had tumour regression by the end of RT and no trend was observed linking initial volume to extent of regression Figure 7-2.

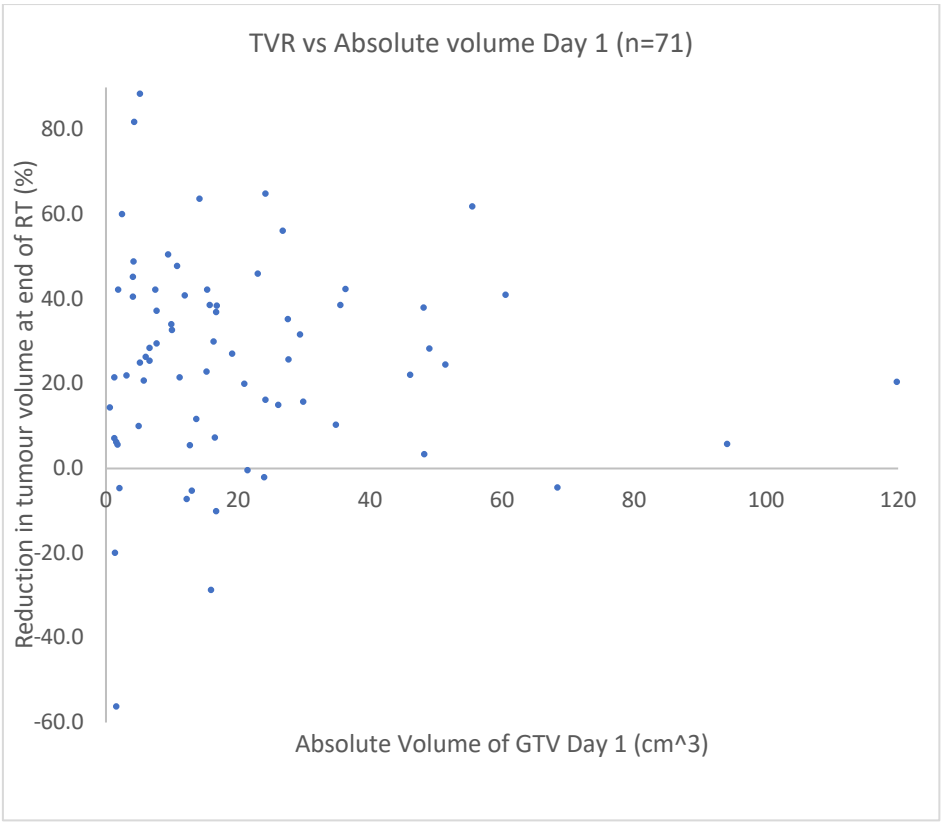


Figure 7-2 Scatter plot displaying TVR against the absolute volume Day 1 GTV.

Blue dots represent each individual patient.

7.3.2 Validation of model performance using early tumour response dynamics

Prediction of TVR using the initial 3 volume measures showed reasonable agreement across all datapoints, R^2=0.81, PCC=0.9 (Figure 7-3), however review of the agreement at the final timepoint revealed weaker agreement with R^2= 0.72, PCC=0.85 (Figure 7-4).

The inclusion of V_{CBCT4} improved performance, R^2=0.91, PCC=0.95 (Figure 7-3) and R^2=0.87, PCC=0.93 for the final image (Figure 7-4). The addition of V_{CBCT5} image further increased R^2 to 0.94 and PCC to 0.97 (Figure 7-3).

Individual prediction curves for each patient based on 3, 4 or 5 images can be seen in Figure 7-5.

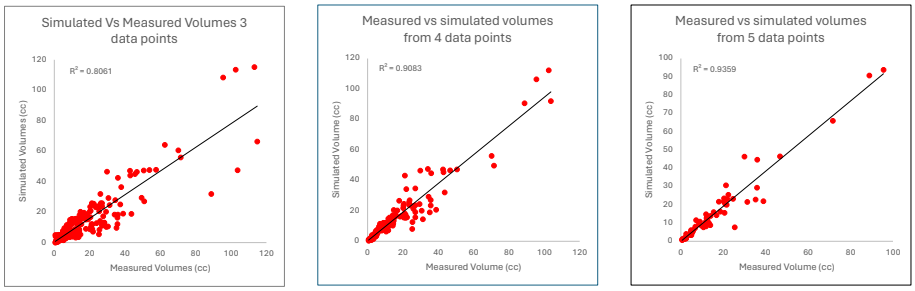
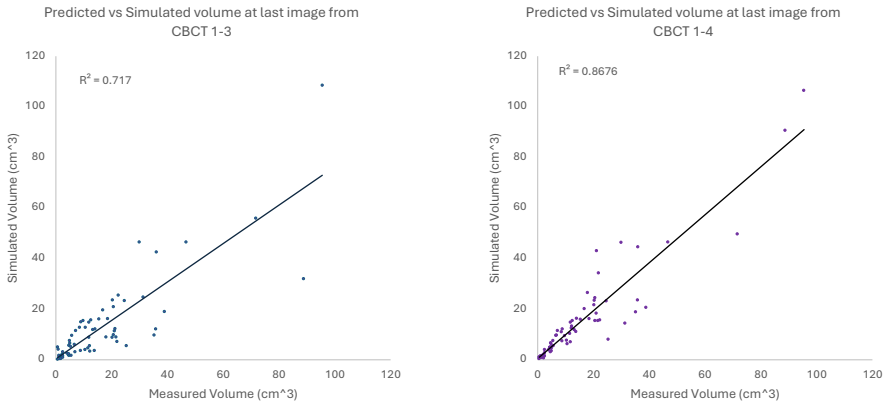


Figure 7-3 Scatter plot displaying measured vs simulated volume using data from 3, 4 and 5 CBCT images

X axis displays measured volume (cm^3) and Y axis the simulated volume (cm^3).



.

Figure 7-4 Scatter plot displaying measured volume vs. model-simulated volume at the final imaging timepoint for each patient.

X axis displays measured volume (cm^3) and Y axis the simulated volume (cm^3).

Panel (a) displays agreement using only data acquired on days 1-3, panel (b) displays agreement when also including CBCT 4.

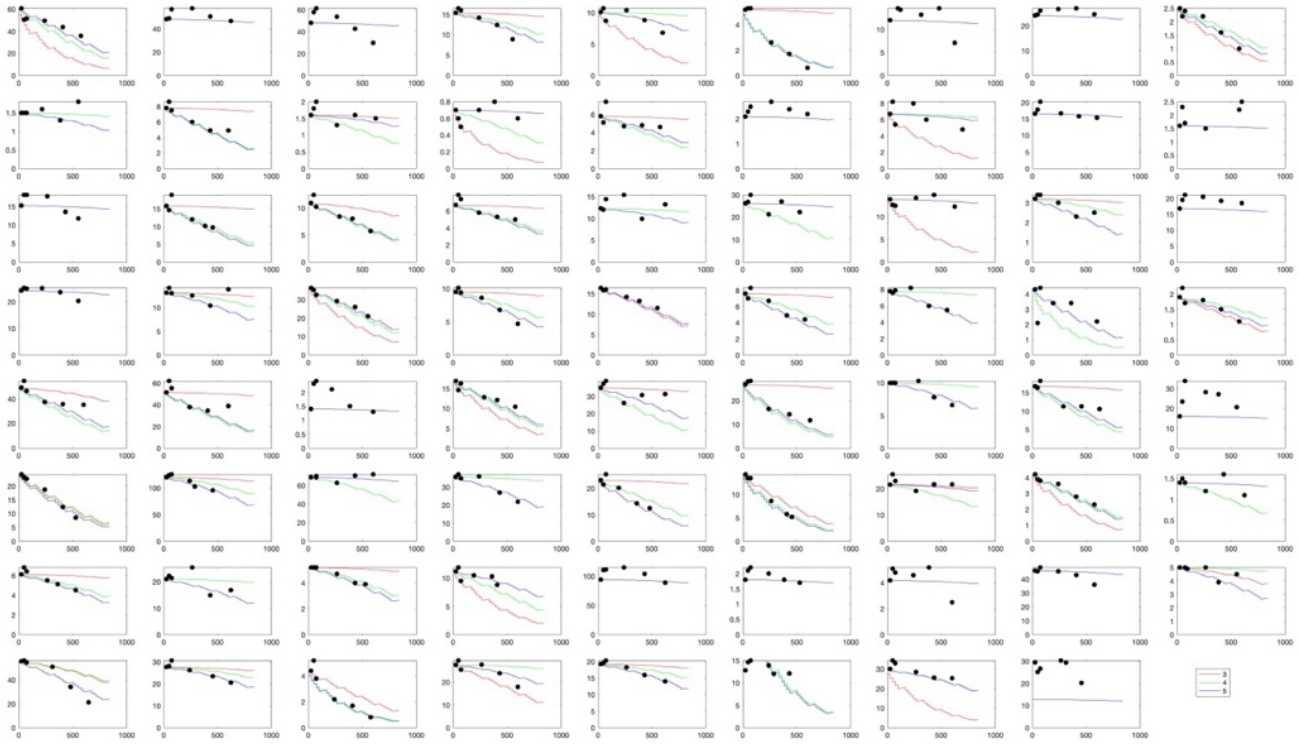


Figure 7-5 Individual plots of actual vs model simulated TVR

X axis displays time in hours, y axis displays absolute tumour volume in cm^3. Measured tumour volumes are black dots. The model predicted regression following 3, 4 and 5 data inputs are overlayed as red, green and blue curves respectively

The absolute volume difference between the last measured tumour volume and the model predicted volume using 4 CBCTs was median 1.5cm^3 (range 0-22.0 cm^3)(Figure 7-6 and Table 7-2).

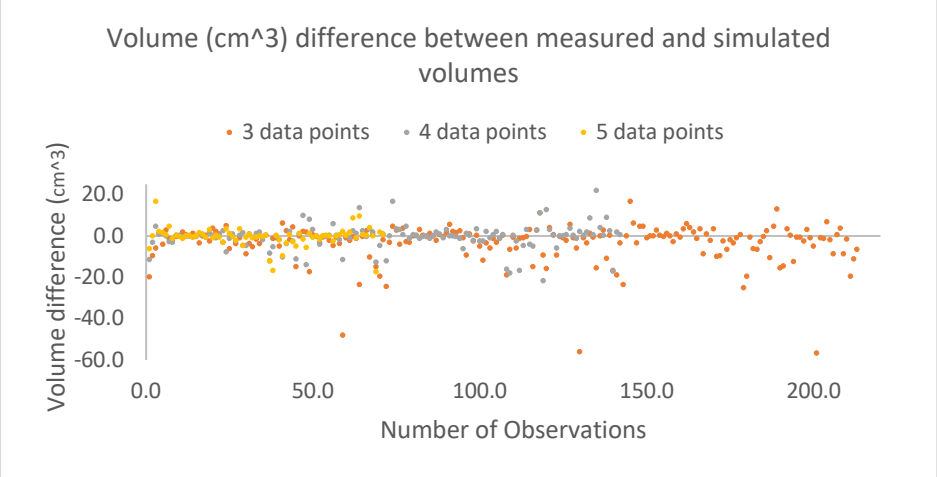


Figure 7-6 Residual volume difference at end of RT between measured and simulated volumes.

X axis displays the number of observations and y axis is the difference in cm^3. Simulated volumes based 3, 4 and 5 CBCT are displayed in orange, grey and yellow respectively.

Table 7-2 Summary of absolute volume differences (cm^3) at end of RT between measured and simulated volumes (V_{day1RT} noted for reference) for each prediction using 3, 4 or 5 inputs.

	V_{Day1RT} (cm^3)	3 data points (cm^3)	4 data points (cm^3)	5 data points (cm^3)
mean	20.4	6.0	3.9	2.5
median	14.3	3.1	2.0	1.1
max	119.9	56.8	22.0	16.9
min	0.7	0	0	0

A Bland-Altman plot was generated to ascertain if there was bias in the agreement between the measured and simulated volumes using 4 CBCTs. A mean difference of -0.68cc was identified, indicating there is limited bias in the prediction. Agreement decreases as the tumour volume increases, but the majority of data points fall between the 95% CI limits (Figure 7-7).

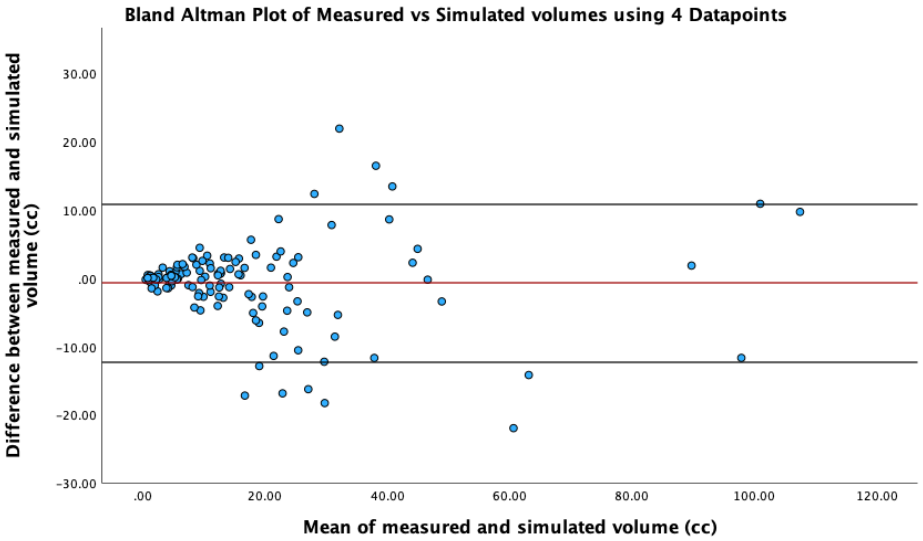


Figure 7-7 Bland-Altman plot comparing measured and model-simulated volumes (cm^3) predicted using CBCT 1-4 for prediction.

Relative tumour volume at the end of RT was compared to the actual percentage volume and is displayed in Figure 7-8, which indicates that the model always predicts a degree of TVR for each patient.

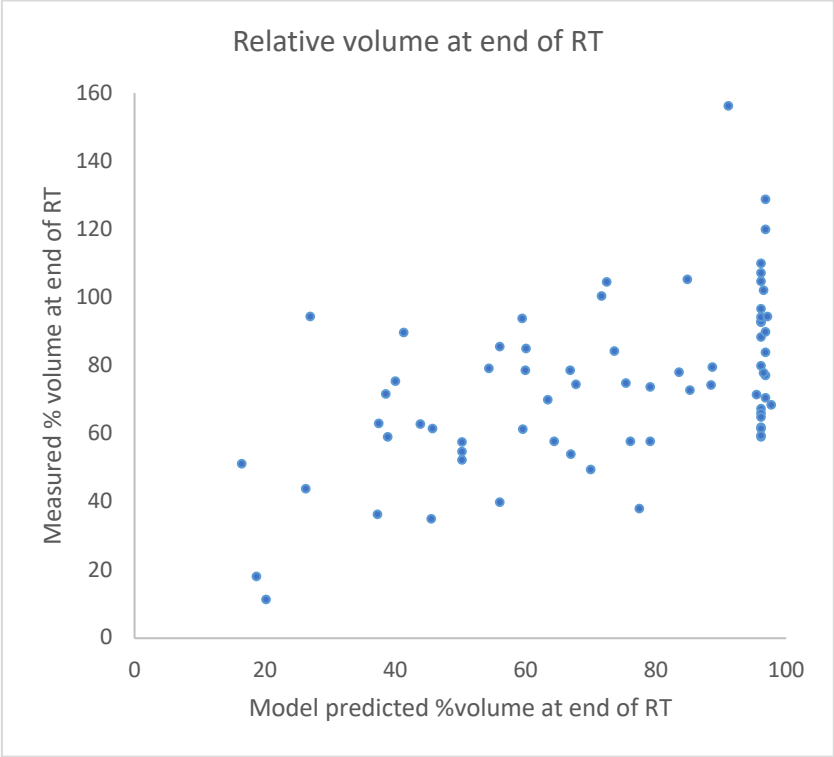


Figure 7-8 Scatter plot displaying the tumour volume at the end of RT as a percentage relative to day 1 volume demonstrating that the model predicts TVR in all patients.

X axis displays the model predicted % regression and y axis the actual percentage regression in n= 71 patients.

As the model always predicts regression, its performance predicting patients with poor TVR (≤ 10%) was evaluated as a surrogate measure of little or no regression in response to RT. Model sensitivity of 77.7%, specificity of 73.6%, positive predictive value of 73.6%, and negative predictive value of 90.7% were calculated, see Table 7-3 for details.

Table 7-3 PSI performance predicting poor TVR (≤10%) using early response dynamics (CBCT 1-4).
TP= True positive, FP= False Positive, FN= False Negative, TN= True Negative

PSI prediction	***Regression ≤10%***		
≤10%	**Yes**	**No**	**Total**
Positive	14 (TP)	14 (FP)	28
Negative	4 (FN)	39 (TN)	43
Total	18	53	71

Parameter sensitivity analysis on the predictions made using 4 CBCTs revealed the model robust to parameter variations of up to 20% with a maximum variation of 1.53% in the R^2 measured when varying α and λ by +20%, (Table 7-4, Figure 7-9).

Table 7-4 Sensitivity analysis of parameter variation on resulting R2 and PCC values.
Varying α and λ by +/- 20% from the optimal derived values

	R^2	***% Δ R^2***	***PCC***	***% Δ PCC***
α 0.03, λ 0.012	0.9083		0.95	
α, λ -20%	0.9153	0.77%	0.96	1.05%
α, λ +20%	0.8944	-1.53%	0.95	0.00%

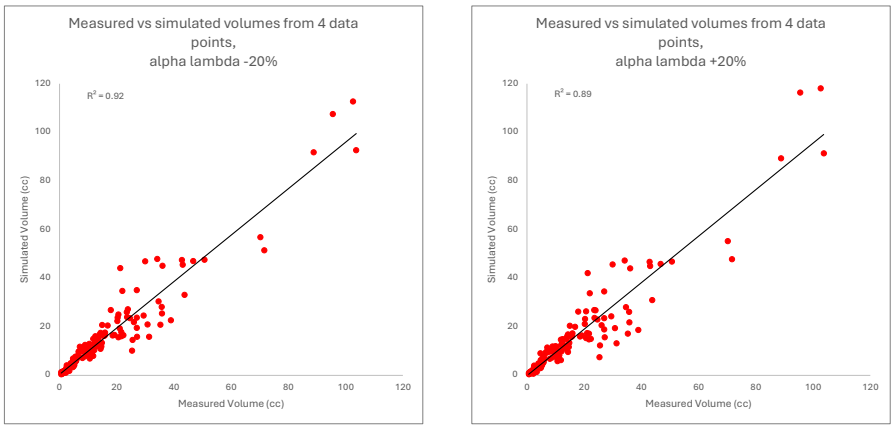


Figure 7-9 Parameter variation sensitivity analysis scatter plots in external validation data.

Scatter plots display measured vs simulated tumour volumes on the x and y axes respectively, with α and λ parameters varying by +/- 20% from the optimal values.

The range of PSI_i values derived using data from all CBCTs, 3, 4 and 5 CBCTs were compared using Box and Whisker plot and no significant variation was found between the groups (Figure 7-10, p=0.95 (ANOVA)).

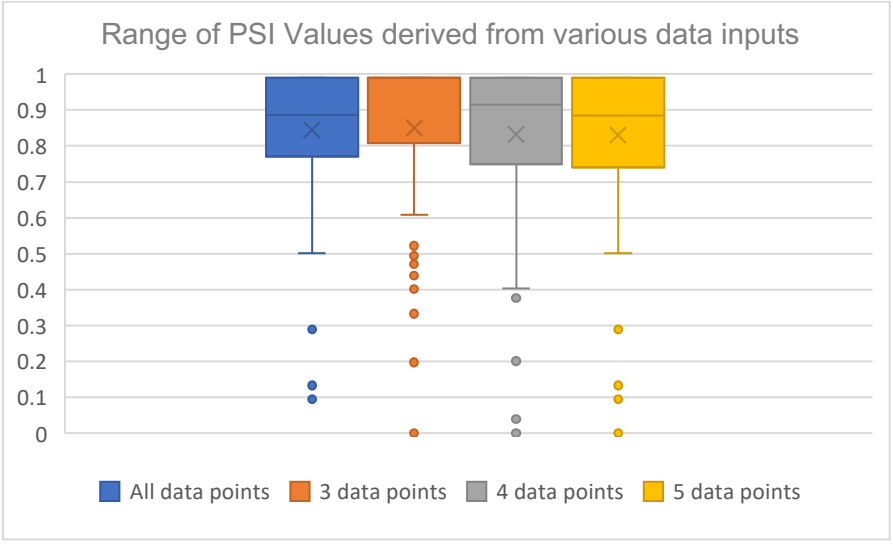


Figure 7-10 Box and Whisker plot displaying variation of PSI values associated with varying data inputs.

Values derived from CBCTs 1-3, 1-4 and 1-5 in orange, grey and yellow respectively, ANOVA p=0.95

7.3.3 Clinical Relevance of TVR in this cohort

The sampling approach of excluding patients with node positive disease resulted in a cohort who had RT alone as the definitive treatment without concurrent or sequential chemotherapy. As a result, this was a relatively homogeneous group from a treatment perspective.

7.3.3.1 Kaplan Meier Analysis

KM plots showed no significant difference for OS between patients with T1, 2 or 3 diseases, or histological subtype (Figure 7-11). Data on histology is limited with almost 50% of patients having a clinical diagnosis (n= 31) or unspecified histology (n= 3).

KM plots also demonstrated no significant difference between the various histological subtypes for local disease progression (LP), but a significant difference in LP for T stage (p= 0.006, Log Rank Test, Figure 7-12).

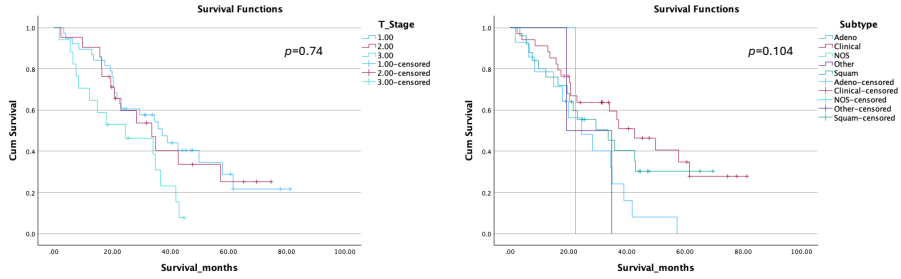


Figure 7-11 Kaplan Meier Curves for Overall Survival split by T Stage and by Histological Subtype

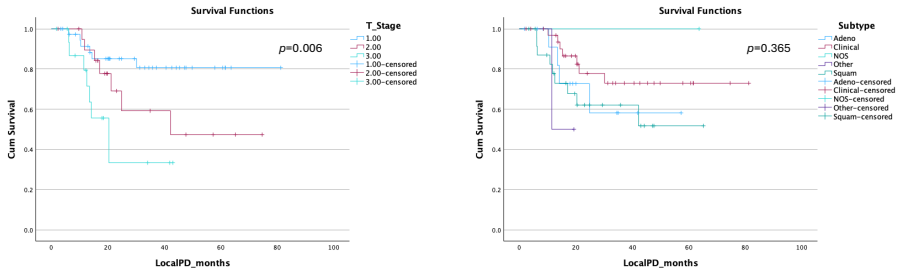


Figure 7-12 Kaplan Meier Curves for Local Progression, split by T Stage and by Histological Subtype

7.3.3.2 Univariable Cox Regression Analysis

Univariable Cox Proportional Hazards Regression did not identify the clinical factors of Age, Sex or Histology as significant predictors of OS. PS of 3 was associated with a significantly increased risk of death, but the confidence intervals were wide, suggesting a small sample size in at least one of these comparison groups (HR 8.788, 95% CI 1.559-49.549, p=0.014).

Tumour volumetric measures of V_{Day1RT} (HR 1.019, 95% CI 1.007-1.031, p=0.003), V_{EndRT} (HR 1.021, 95% CI 1.006-1.036, p=0.006), and $V_{\%REG}$ (HR 1.016, 95% CI 1.002-1.029, p=0.021), were significantly associated with OS (Table 7-5).

For every increase in V_{EndRT} the risk of death was increased by 2.1%, however, for every increase in $V_{\%REG}$, (i.e. those with more pronounced TVR) the risk of death also increased by 1.6% indicating that larger absolute volume and greater TVR in response to RT were negatively associated with OS in this population. The degree of volume change prior to treatment ($V_{\%PRE\_RT}$) was not statistically significant in this population.

Table 7-5 Univariable Cox Regression Analysis Summary Overall Survival.

V_{Day1RT} = absolute volume at day 1 of RT in cm3, V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT, $V_{\%PRE\_RT}$ = tumour growth pre-treatment expressed as a percentage of the volume. PS = Performance Status, Cat. Sig = Category significance, NOS= Not Otherwise Specified

	Hazards Ratio	***95.0% CI Lower***	***95.0% CI Upper***	***Sig.***	***Cat. Sig.***
Histology					0.203
Adeno	Reference				
Clinical	1.73	0.813	3.683	0.155	
NOS	0.747	0.379	1.473	0.4	
Other	2.077	0.271	15.936	0.482	
Squam	1.935	0.439	8.532	0.383	
Sex	0.697	0.396	1.226	0.21	
T Stage					0.15
T 1	Reference				
T 2	1.073	0.543	2.12	0.84	
T 3	1.944	0.973	3.882	0.06	
PS					0.083
PS 0	Reference				
PS 1	2.821	0.655	12.145	0.164	
PS 2	3.066	0.721	13.034	0.129	
PS 3	8.788	1.559	49.549	**0.014**	
V_{EndRT}	1.021	1.006	1.036	**0.006**	
V_{Day1RT}	1.019	1.007	1.031	**0.003**	
$V_{\%REG}$	1.016	1.002	1.029	**0.021**	
$V_{\%PRE\_RT}$	1.008	0.995	1.02	0.232	

When examining local disease progression, univariable analysis identified T Stage of T3 as associated with an increased risk of LP compared to the reference variable of T1, (HR 6.297, 95% CI 2.031-19.526, p=0.001), but with a wide 95% CI. Of the volumetric variables only V_{EndRT} and V_{Day1RT} were statistically associated with LP (HR 1.031, 95% CI 1.011-1.051, p=0.003), (HR 1.024, 95% CI 1.007-1.041, p=0.006), both indicating that a larger absolute volume at the start or end of RT, is associated with an increase in risk of LP (Table 7-6).

Table 7-6 Univariate Cox Regression Analysis Summary Local Progression

V_{Day1RT} = absolute volume at day 1 of RT in cm^3, V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT, $V_{\%PRE\_RT}$ = tumour growth pre-treatment expressed as a percentage of the volume. PS = Performance Status, Cat. Sig = Category significance, NOS= Not Otherwise Specified

	Hazards Ratio	***95.0% CI Lower***	***95.0% CI Upper***	***Sig.***	***Cat. Sig***
Histology					0.492
Adeno	Reference				
Clinical	1.001	0.3	3.339	0.998	
NOS	0.462	0.16	1.333	0.153	
Other	0	0	.	0.984	
Squam	2.249	0.273	18.538	0.451	
Sex	0.605	0.245	1.493	0.276	
PS					0.337
PS 0	Reference				
PS 1	0.275	0.061	1.233	0.092	
PS 2	0.671	0.186	2.423	0.542	
PS 3	0.708	0.072	6.922	0.767	
T Stage					**0.005**
T 1	Reference				
T 2	2.298	0.699	7.562	0.171	
T 3	6.297	2.031	19.526	**0.001**	
V_{EndRT}	1.031	1.011	1.051	**0.003**	
V_{Day1RT}	1.024	1.007	1.041	**0.006**	
$V_{\%REG}$	1.013	0.991	1.036	0.241	
$V_{\%PRE\_RT}$	1.021	1	1.043	0.053	

7.3.3.3 Multivariable Cox Regression Analysis

While LP is a clinically relevant outcome, only 19/71 patients had LP events. To ensure statistical power was maintained Multivariable Cox Proportional Hazards Regression was performed only on the clinical endpoint of OS, where 49 deaths were recorded [219].

The statistically significant variables identified from the univariable analysis were tested for multicollinearity using a correlation matrix in SPSS. Absolute volume at first and last timepoint were significantly correlated, PCC=0.957, p<0.001 (Table 7-7) and so V_{Day1RT} was excluded from the adjusted model and only V_{EndRT} included.

Table 7-7 Multicollinearity testing prior to Multivariable Cox Regression Analysis

PS= Performance status V_{Day1RT} = absolute volume at day 1 of RT in cm^3, V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT

Correlation Matrix

		PS	***V_{Day1RT}***	***V_{EndRT}***	***$V_{\%REG}$***
PS	Pearson Correlation	1	0.005	-0.024	0.086
	Sig. (2-tailed)		0.968	0.842	0.476
	N	71	71	71	71
V_{Day1RT}	Pearson Correlation	0.005	1	**.957****	-0.024
	Sig. (2-tailed)	0.968		**<.001**	0.845
	N	71	71	71	71
V_{EndRT}	Pearson Correlation	-0.024	**.957****	1	-0.208
	Sig. (2-tailed)	0.842	**<.001**		0.081
	N	71	71	71	71
$V_{\%REG}$	Pearson Correlation	0.086	-0.024	-0.208	1
	Sig. (2-tailed)	0.476	0.845	0.081	
	N	71	71	71	71

***. Correlation is significant at the 0.01 level (2-tailed).*

In the adjusted model, all variables remained significantly associated with OS. The magnitude of effects for the volume metrics were unchanged, with increasing absolute volume and increasing TVR at last timepoint

associated with an increased risk of death, HR 1.031, 95% CI 1.015-1.047, p<0.001, HR 1.02, 95% CI 1.005-1.035, p=0.007 respectively, (Table 7-8).

Table 7-8 Multivariable Cox Regression for Overall Survival.

PS= Performance Status V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT PS= Performance Status

	Hazards Ratio	***95.0% CI Lower***	***95.0% CI Upper***	***Sig.***	***Cat. Sig.***
PS					0.055
PS 0	Reference				
PS 1	4.666	1.062	20.495	**0.041**	
PS 2	4.453	1.039	19.09	**0.044**	
PS 3	11.759	2.03	68.106	**0.006**	
V_{EndRT}	1.031	1.015	1.047	**<.001**	
$V_{\%REG}$	1.02	1.005	1.035	**0.007**	

While not statistically significant, KM plot comparing those patients with tumour regression during RT, to those with no change or progression, suggesting worse OS in the regression group, p=0.166 (Figure 7-13).

Splitting the patients on the median measured TVR (25.6%) revealed no significant differences between groups for either OS or LC (Figure 7-14).

Subgroup analysis of patients with histological confirmed Squamous Cell Carcinoma (n=24) or Adenocarcinoma (n=13) yielded no statistically significant findings, however there was trend in the SCC group for improved LC with TVR greater than the median value, and in the Adenocarcinoma group this was reversed with worse LC in those with greater TVR (Figure 7-15).

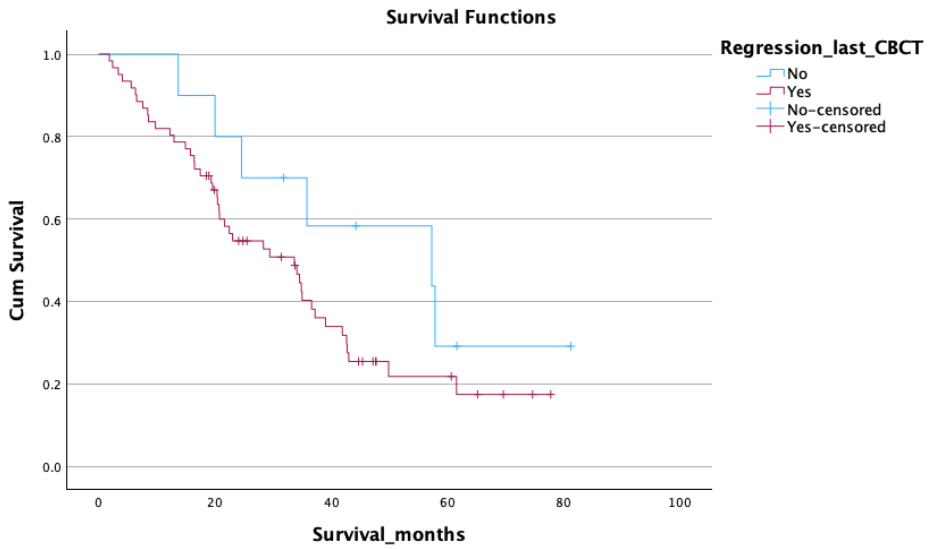


Figure 7-13 Kaplan Meier curves for Overall Survival, splitting data on TVR or no TVR at last CBCT p=0.166

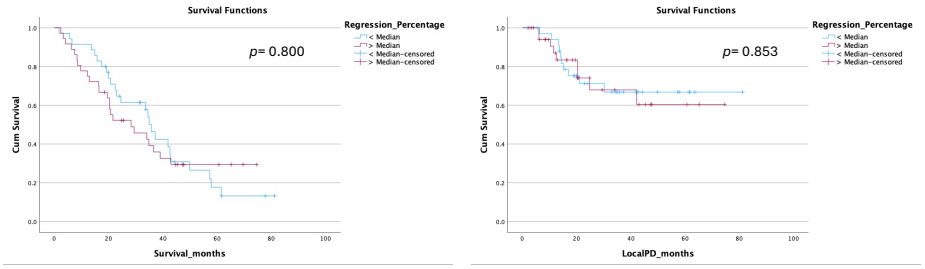


Figure 7-14 Kaplan Meier curves for Overall Survial and Local Progression, splitting the data on the median % TVR (25.6%)

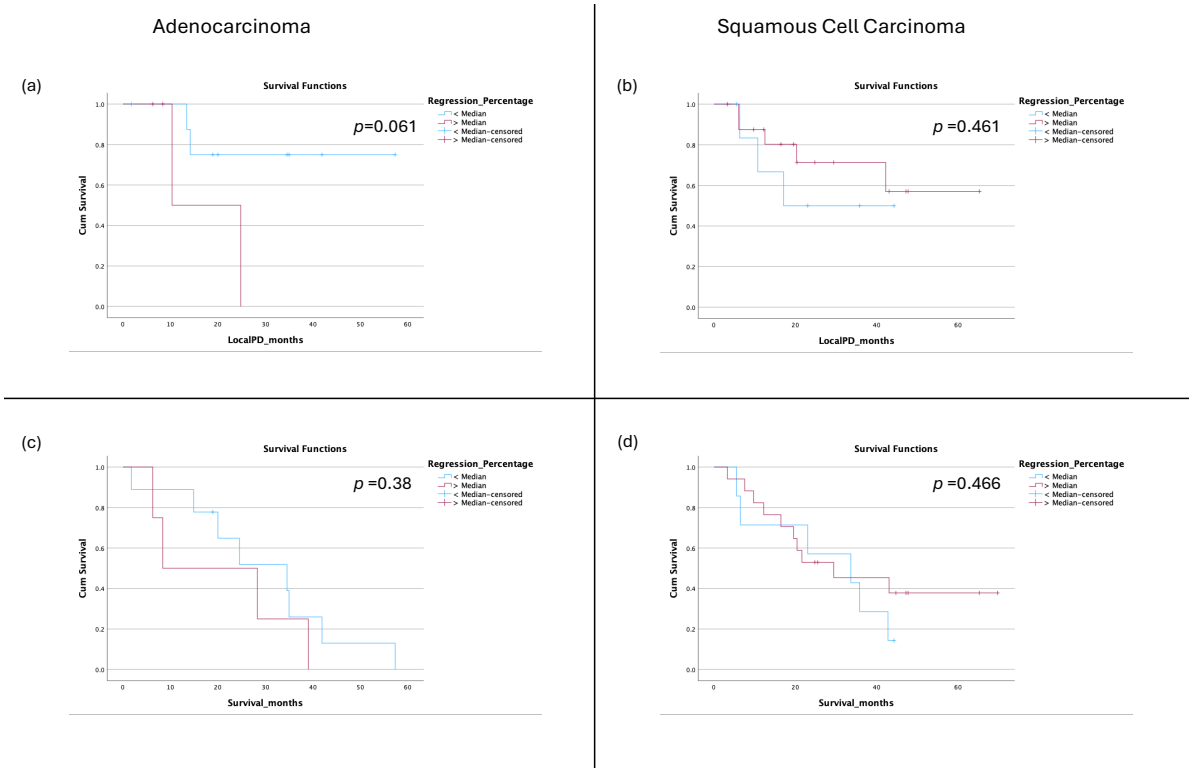


Figure 7-15 Kaplan Meier subgroup analysis of Adenocarcinoma or Squamous Cell Carcinoma subtypes, splitting the data on the median % regression (25.6%)

Panel (a) and (b) display the outcome of local progression and panel (c) and (d) display overall survival in Adenocarcinoma and Squamous Cell Carcinoma respectively

OS prediction at 3 years was assessed with ROC curves and AUC values. Multivariable Cox Regression of the volumetric variables of V_{EndRT} and $V_{\%REG}$ and, based on domain knowledge, the clinical variables of PS, Histology and T Stage were performed to generate a linear predictive score (LPS) in SPSS (Table 7-9). The LPS was utilised to create ROC curves predicting OS at 36 months.

Volumetric variables resulted in marginally higher AUC values than the clinical variables (0.703 and 0.702 respectively). The combination of clinical and volumetric variables resulted in the highest AUC value of 0.816 for 3-year OS (Figure 7-16).

Table 7-9 Multivariable Cox Regression Analysis for 3-year OS.

V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT, PS= Performance Status, NOS= Not Otherwise Specified

3-Year OS

	Hazards Ratio	95.0% CI Lower	95.0% CI Upper	Sig.	Cat. Sig.
PS					**0.041**
PS 0	Reference				
PS 1	6.055	0.748	49.013	0.091	
PS 2	6.401	0.821	49.933	0.077	
PS 3	23.125	2.312	231.271	**0.008**	
T Stage					0.3
T 1	Reference				
T 2	0.836	0.365	1.916	0.672	
T 3	0.44	0.154	1.262	0.127	
Hist.					0.654
Adeno	Reference				
Clinical	0.585	0.226	1.516	0.27	
NOS	1.432	0.159	12.878	0.748	
Other	1.672	0.286	9.783	0.569	
Squam	0.774	0.312	1.918	0.579	
V_{EndRT}	1.039	1.018	1.06	**<.001**	
$V_{\%REG}$	1.031	1.011	1.052	**0.003**	

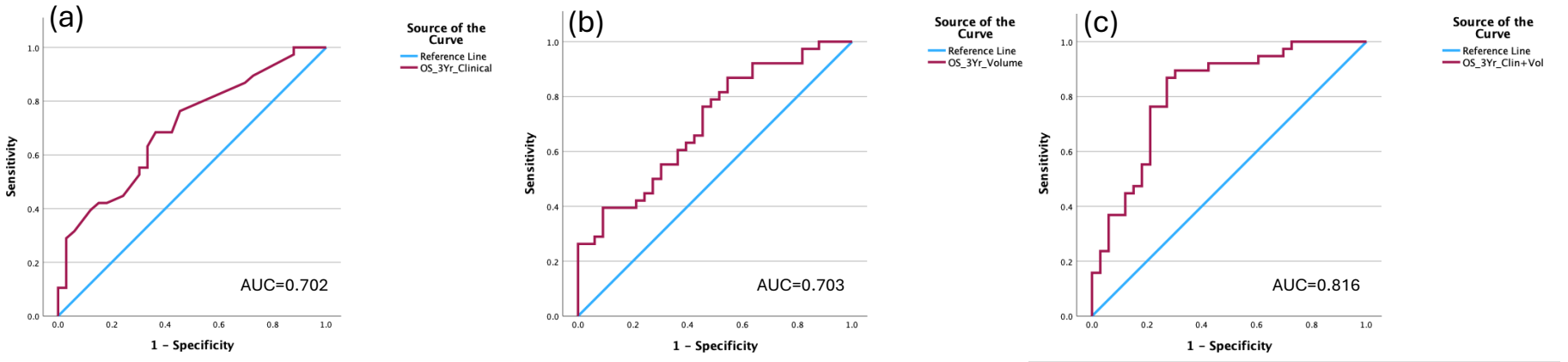


Figure 7-16 ROC curves for prediction of 3-year OS and associated AUC values.

Panel (a) displays prediction using clinical variables (Performance Status, Histology and T Stage). Panel (b) displays prediction using volumetric variables (V_{EndRT} and $V_{\%REG}$), and panel (c) displays combined clinical and volumetric variables.

7.4 Discussion

7.4.1 Validation of model performance using early tumour response dynamics

The PSI model was proposed as a mechanism to predict TVR in response to RT and prior work trained and tested this model in a population of patients treated with definitive RT for NSCLC, however external validation was lacking in these studies [79, 80, 220].

This chapter examined the PSI model performance in an independent dataset using the α and λ parameters derived from prior data. Internal validation does not ensure generalisability and therefore the model was validated in a fully independent dataset, in this TRIPOD Type 4 study [98].

Volumetric regression predicted on only the first 3 images, acquired days 1-3 of RT, demonstrated sub-optimal agreement. As there are limited tumour volume dynamics in this short timeframe, this was expected. The inclusion of volume derived from CBCT 4 (V_{CBCT4}), the first weekly image acquired on median of Day 10 RT, improved the prediction. Testing of the model in Study 3 generated R^2 and PCC values of 0.94 and of 0.97 [220] respectively, compared to 0.91 and 0.95 in this independent dataset indicating the model was able to predict TVR for the remainder of treatment using data from the first 10 days of RT. A Bland Altman plot also demonstrated good agreement, particularly in the smaller volume tumours.

Other groups have predicted tumour regression by linear regression following delivery of 1/3 of the course of RT, in 99 patients treated with 60Gy/2Gy per fraction, and reported a PCC of 0.73 when comparing the final predicted and actual volumes [68]. Here we report a substantially

higher PCC of 0.93 when comparing the measured to simulated end of treatment volume for each patient.

Patients in this dataset were treated with the moderately accelerated schedule of 20#. For those patients being treated with the common fractionation schedule of 30-33#, prediction at day 10 of RT could be more clinically beneficial, indicating in advance those that may need adaptive replanning due to tumour shrinkage. GTV shrinkage can lead to mediastinal shifts and without plan adaptation an increase in dose to critical OARs such as the heart can occur [221]. Adaptive replanning can improve the dosimetric integrity of a treatment plan [157, 170, 171].

While model performance was excellent in this cohort, all patients were treated with a single prescription of 55Gy/20# and future studies should validate this in a variety of prescriptions.

7.4.2 Clinical Relevance of TVR in this cohort

The data thus far demonstrated that the PSI model generates accurate predictions of TVR, and the next step was to ascertain the clinical relevance of predicted TVR in conjunction with the clinical demographics of the cohort.

KM plots revealed that T stage had a significant association with LC in this cohort but had no influence on OS and PS of 3 was statistically significant on Univariable Cox Regression for OS prediction.

Histological subtype was not significantly associated with either LC or OS. One limitation segregating this cohort by histology was the high proportion of patients with a clinical diagnosis, 47.9% had either a clinical or not specified diagnosis, 33.8% Squamous Cell Carcinoma and 18.3% Adenocarcinoma. The lack of histological diagnosis is not uncommon in the UK and Ireland setting. Data evaluating utilisation of SABR in early

stage disease in England, reported 28% of patients did not have a biopsy confirmed diagnosis [222]. The absence of disease classification limits interpretation of this result, as there are likely to be mixed histologies in the clinical group.

When examining the volumetric metrics of absolute volume at the start or end of RT both were significantly associated with outcomes, and this is in line with prior knowledge that absolute tumour volume is an established prognostic factor [223]. TVR ($V_{\%REG}$) was a negative prognostic indicator in this cohort, and patients with larger regression had poorer clinical outcomes. This significance persisted on multivariable analysis when the model was adjusted for PS and V_{EndRT} (Table 7-8). One limitation of the PSI model is the assumption that all tumours will respond and regress during RT as seen in Figure 7-8. To assess the model's ability to predict a clinically meaningful response to RT, we tested its prediction of patients who will have poor regression (TVR ≤ 10%) and the negative predictive value was high (90.7%). This indicates that PSI model performs well ruling out those that will not be in this category and who consequently may have a poor prognosis.

The relevance of TVR is debated in the literature, with data suggesting there is no association between GTV regression and outcomes [76], that increased regression is predictive of improved outcomes [69, 74, 105] or, similar to this study, that larger regression is a poor prognostic indicator [68, 224]. However, many of these studies investigated patients receiving CCRT unlike this cohort who had definitive RT only, and this, along with histological subtype are likely to be confounding factors [70, 71, 225].

The regression measured in this dataset was on the final CBCT, and not a diagnostic quality scan, and the timepoint of this final image varied slightly between patients (range 17-29 days). These limitations of timing and modality of contouring regression are factors which make a

meaningful comparison to other data challenging [211]. While not statistically significant, contrary trends for LC were observed in the Adenocarcinoma vs SCC groups (Figure 7-15), which mirrors the findings by Brink et al [68] and Van Timmeren et al [71] who report that different prognostic impacts were observed in those with adenocarcinoma vs non-adenocarcinoma subtypes of disease.

Having identified increased TVR as a negative association with clinical outcome, we cannot infer a causal relationship, and TVR is most likely a pseudo marker for the underlying biology, potentially aggressive disease pre-treatment. Response to RT is associated with increased proliferation rates [34], but the prognostic relevance of proliferation can vary between adenocarcinoma and squamous cell carcinomas [199]. While we identified greater reduction in tumour volume to be associated with increased risk of death, the TVR is unlikely to drive the clinical response. These data demonstrate that, in this cohort rapid TVR is a poor prognostic factor, and future studies should aim to unravel the underlying confounding factors, particularly the relevance of histology and proliferation rates. Correlation of proliferative markers by tissue staining of Ki-67 or by 18F-3'-deoxy-3'-fluoro-L-thymidine (18F-FLT) PET imaging, which reflects proliferation rates, could add biological meaning to the macroscopic changes in volume measured on CBCT [226].

ROC analysis of 3-year OS identified that the volumetric variables improved the predictive power of a model based on clinical features with AUC increasing from 0.702 to 0.816. This compares positively to prior studies with reported AUC values of 0.746 [227], 0.759 [228] and 0.62 [229] using machine learning and radiomics approaches. One recent study evaluates a deep learning model that integrates clinical and radiomic features in addition to GTV volume pre-treatment and at mid-treatment [230]. This study reports an AUC value in the training cohort of 0.869, decreasing to 0.798 in the validation dataset. The added benefit of the

model presented in this work compared to these more complex options, is its simplicity, as it is based on accessible data routinely acquired during RT.

The survival analysis discussed here followed a stepwise process to select variables for a multivariable model, and there are limitations this method. The process can incorrectly lead to the exclusion of significant variables and as a result mask interaction between variables which may be relevant [231]. The analysis presented here is exploratory, and future work should evaluate the clinical outcomes using domain knowledge to pragmatically include clinically relevant variables in addition to those identified as significant on univariable analysis.

7.5 Conclusion

To conclude, external validation of PSI model confirmed its ability to predict TVR from approximately day 10 of RT in a homogeneous cohort of n=71 patients receiving 55Gy/20# definitive RT alone.

Future work to prospectively validate the PSI model in conjunction with biological or imaging biomarkers could robustly test, not only the model performance at predicting TVR, but further elucidate the role TVR may have in predicting clinical outcomes.

CHAPTER 8 STUDY 5

DEVELOPING THE PSI MODEL IN THE PRE-TREATMENT SETTING IN NON-SMALL CELL LUNG CANCER

8.1 Introduction

In the evolving clinical landscape, prediction in advance of treatment commencing enables the potential to personalise the therapeutic approach for each patient. Predictive or prognostic models could facilitate a patient centred approach, optimising RT for that individual using data unique to their clinical status. Currently there are no widely accepted clinical implemented biomarkers that are routinely considered prior to RT [61], but several categories of these exist, including genomic biomarkers, radiomics approaches, circulating tumour biomarkers such as ctDNA and imaging biomarkers [61, 232, 233]. RT is a data driven specialty, in particular it is image intensive, which lends itself readily to advanced analysis and exploitation of these images to increase the prevalence of imaging biomarkers in our field [234]. Patient-specific computer simulations can be considered 'digital twins', which are virtual representations of a complex system, and in RO, this digital twin would facilitate *in-silico* testing of treatment options, and aid the selection of the personalised approach for each individual [235]. The PSI model can be considered in this context if predictions are made in advance of RT and alternative prescriptions modelled to guide clinical decision making. The dataset generated in Study 2 can facilitate a pre-treatment

prediction, as GTV delineation was completed on the RT planning CT and day 1 CBCT is acquired prior to RT delivery.

In this chapter the performance of the PSI model predictions based only on pre-treatment information is tested and *in-silico* modelling of alternative prescriptions will be presented to probe the potential impact of a variety of prescriptions on tumour regression during RT.

8.1.1 Aims and Objectives

1. To ascertain if the PSI model can accurately predict TVR over the course of RT using only pre-treatment data.
    a. To model tumour regression in response to RT using only pre-treatment volume measurements acquired at planning CT and day 1 of RT
    b. To establish the agreement between the modelled and measured volumes using scatter plots, R^2 and PCC.
    c. To optimise λ on a dataset that includes pre-treatment tumour growth dynamics.
    d. To test the impact of delineation uncertainty across 2 imaging modalities by robustness analysis varying the planning CT volume by +/- 5% and 10% and quantifying the variation in agreement metrics.

2. To examine the clinical application of the PSI model in the pre-treatment setting
    a. To ascertain if PSI model predicted TVR (absolute or relative) is significant in Cox Proportional Hazards Regression analysis.
    b. To model response to several clinically relevant dose fractionation schedules, and determine the variation in modelled volume at the end of RT.

8.2 Materials and Methods

8.2.1 Patient population.

The cohort of n=71 patients utilised in Study 4 were accessed. An underlying assumption of the PSI model is that tumours grow when unperturbed by RT. Therefore, those patients where no increase in volume was measured between planning CT and day 1 CBCT were excluded (n=32). A further 2 patients were excluded as their planning CT was 3D, only tumour volumes contoured on AVIP reconstructions of 4DCT were included to minimise the variation between modalities [175], resulting in n=37 patients included Table 8-1.

Table 8-1 Demographics of pre-treatment subgroup n=37.

IMRT: Intensity Modulated Radiation Therapy, VMAT: Volumetric Arc Radiation Therapy, 3DCRT: 3-Dimensional Conformal Radiation Therapy

Patient Demographics	***N (%)***
Age (median, range) years	74 (54-89)
Gender	
Female	19 (51.4%)
Male	18 (48.6%)
Performance Status	
0	3 (8.1%)
1	10 (27%)
2	21 (56.8%)
3	3 (8.1%)
T-stage	
1	16 (43.2%)
2	12 (32.4%)
3	9 (24.3%)
Histology	
Squamous Cell Carcinoma	16 (43.2%)
Adenocarcinoma	5 (13.6%)
Clinical Diagnosis	15 (40.5%)
Other/Not otherwise specified	1 (2.7%)
Radiotherapy Planning	
3DCRT	11 (29.7%)
IMRT	8 (21.7%)
VMAT	18 (48.6%)

8.2.2 Testing model prediction using only pre-treatment dynamics

Prior training and testing of the PSI model in Studies 3 and 4, included no pre-treatment data, and both α and λ parameters were derived using on treatment imaging data.

The PSI code was modified to include these data and solve only on pre-treatment volumes extracted from planning CT and day 1 CBCT.

The modified code solved for carrying capacity (K) using the following formula for a given λ:

$$K = (V_{Day1RT} * V_{PCT} * \left(e^{\lambda * \Delta t}\right) - 1)/(V_{PCT} * \left(e^{\lambda * \Delta t}\right) - V_{Day1RT})$$

Where V_{Day1RT} = Volume at day 1 of RT, V_{PCT} = Initial Volume on planning CT and Δt = time elapsed between scans. PSI is then calculated as V_{PCT}/K.

Initially, parameters of α=0.03 (derived from Study 3) and λ=0.012 (derived from the literature [80]) were tested. The input parameters were V_{PCT}, V_{Day1RT} and the time in days between these values. The model was solved on the pre-treatment data and the response to RT was simulated for the prescription dose of 2.75Gy/#. Agreement was evaluated using scatter plots to compare the simulated and measured volumes, R^2 value and PCC.

8.2.3 Re-optimisation of α and λ parameters

Following testing with the previously derived parameters, the λ parameter was reoptimized on the pre-treatment data, firstly keeping the α value of 0.03, followed by a re-optimisation of both α and λ parameters.

Optimisation was performed by testing a vector of four α and λ values simultaneously and evaluating a MSE matrix for each pair of values. This process was iteratively repeated, reducing the upper and lower bounds around the optimal pair values, to find the pair with the lowest MSE, and ensuring the results were not constrained by the bounds.

8.2.4 Clinical relevance of model simulated TVR

PSI model simulated TVR was examined to ascertain if the predicted regression was associated with clinical outcomes. This tests the hypothesis that we can predict in advance of treatment, both tumour volume

dynamics in response to RT, and that this prediction is clinically meaningful.

In this subgroup analysis KM plots were generated, and Cox Proportional Hazards modelling was performed in the same manner as the full cohort analysis outlined in Study 4. Additionally, the model simulated end of RT volume and simulated percentage TVR were included as variables and defined as:

- Simulated absolute volume at the final image RT in cm^3 (SIM_V_{EndRT})
- Simulated TVR on final image RT ($SIM\_V_{\%REG}$)

8.2.5 In-silico modelling of alternative prescriptions

Finally, to investigate the potential of the PSI model as a tool for optimising an individual patient's prescription, an *in-silico* modelling study of 6 prescriptions was performed. These dose fractionation schedules were taken from the clinical guidelines and clinical trial protocols, documented in Table 8-2.

Table 8-2 Dose fractionation schedules tested in *in-silico* prescription modelling study.

BED= biological equivalent dose, EQD2= dose in 2Gy equivalent fractions, OTT= Overall Treatment Time in days

Prescription	***Dose per #***	***OTT***	***BED***	***EQD2***	***Source***
57Gy /25#	2.28Gy	32	70Gy	58.33Gy	Ohri et. al. [236]
55Gy/20#	2.75Gy	25	70.13Gy	58.44Gy	Maguire et. al. [31]
60Gy /30#	2Gy	39	72Gy	60Gy	Postmus et. al. [30]
66Gy /33#	2Gy	44	79.2Gy	66Gy	Postmus et. al.[30]
65Gy /25#	2.6Gy	32	81.9Gy	68.25Gy	Ohri et. al. [236]
60Gy/15#	4Gy	18	84Gy	70Gy	Iyengar et. al. [237]

To calculate the Overall Treatment Time (OTT), timepoint at end of RT, for each prescription, an assumption of commencing RT on Monday was made, and end point of RT was calculated as the elapsed days, including weekends until the final fraction. Simulated tumour volume at the end of RT for each prescription was extracted, and Analysis of Variance performed to test for significant differences in the resulting simulated volumes.

8.3 Results

8.3.1 Patient population

Image acquisition timepoints for this subgroup can be seen in Figure 8-1. Planning CT was acquired on a median of 20 days prior to starting RT, (range 13-49 days).

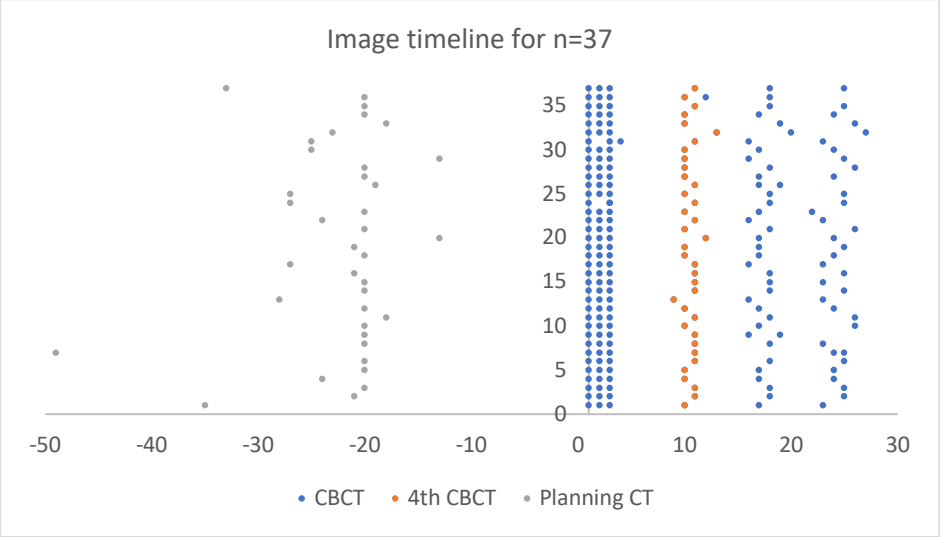


Figure 8-1 Scan acquisition timepoints for each patient in pre-treatment prediction subgroup

The start of RT delivery is marked as Day 1. Planning CT scan is displayed in grey, CBCT in blue and the 4th CBCT is highlighted as orange.

8.3.2 Testing model prediction using only pre-treatment dynamics

8.3.3 Testing using previously derived parameters

TVR predictions using the previously derived parameters overestimated regression and underestimated tumour volume at end of RT for many patients, R^2=0.87 and PCC=0.93 (Figure 8-2,Figure 8-3). In addition, individual PSI values were calculated as negative values (Figure 8-4). In solving the model using pre-treatment data, K is calculated as a function of V_{PCT} and $V_{day1RT,}$ elapsed time between measures, and growth rate (λ). The negative PSI values indicate the λ value tested was too low to account

for tumour growth rate in some patients in this new population. The tumours with the fastest growth rate in this dataset were growing faster than the model parameters would allow resulting in a negative value for K and consequently a negative PSI (V_{PCT}/K). Individual patient plots of simulated regression can be seen in Figure 8-3.

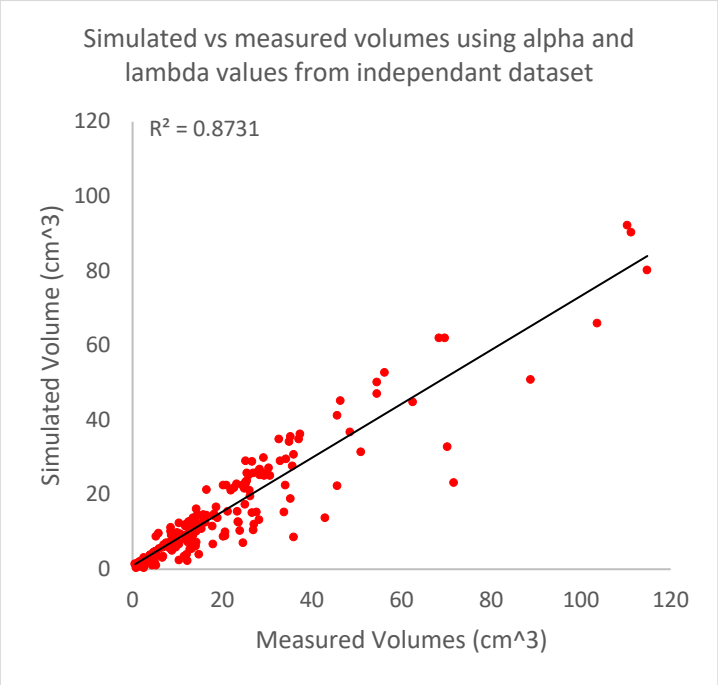


Figure 8-2 Scatter plot of volumes predicted using only pre-treatment data vs measured tumour volumes (α=0.03, λ=0.012)

X axis displays measured volume and y axis model-simulated volume (cm^3) Model parameters of α=0.03, λ=0.012 were derived from independent dataset.

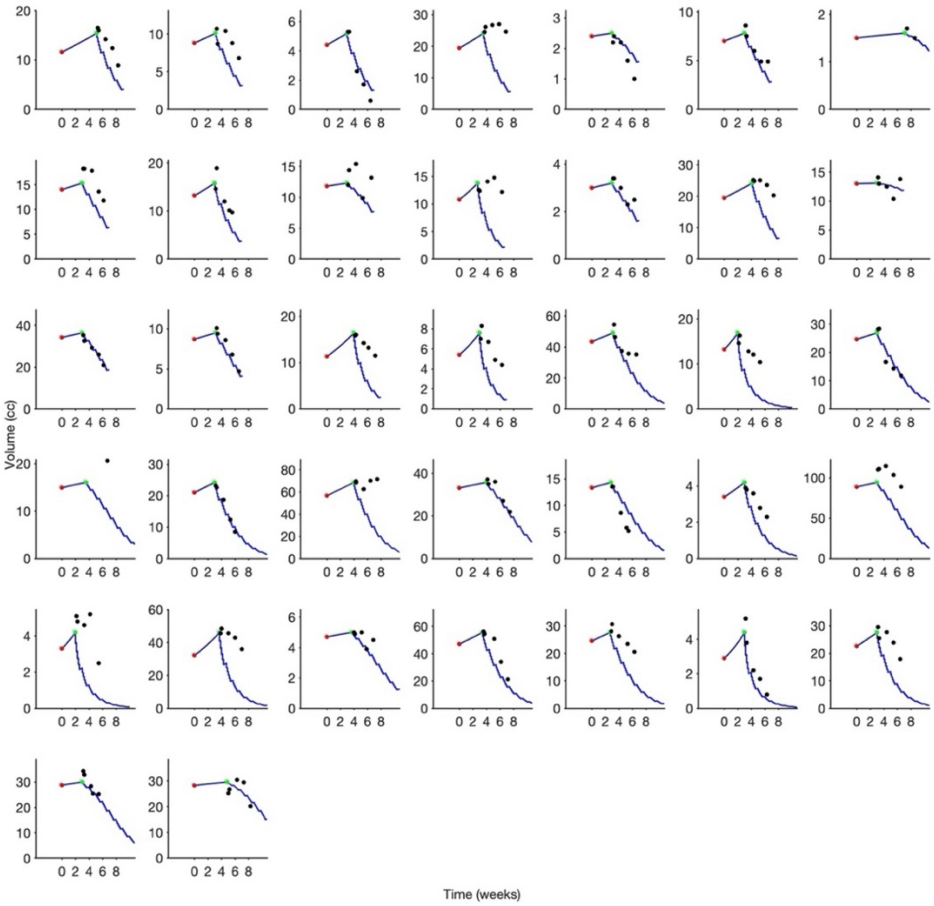


Figure 8-3 Individual plots of predicted TVR in pre-treatment prediction (α=0.03, λ=0.012)

Red dot is planning CT data, green dot is day 1 CBCT, Black dots are subsequent CBCTs. Blue curve is model simulated volumes. X axis displays time in weeks, y axis displays absolute volume in cm^3

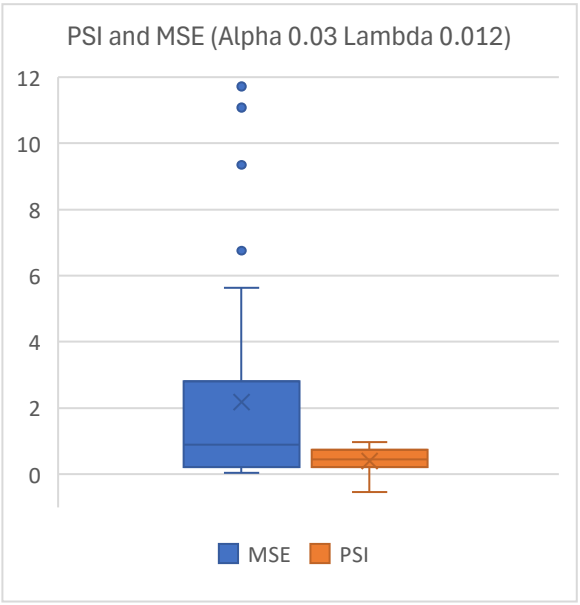
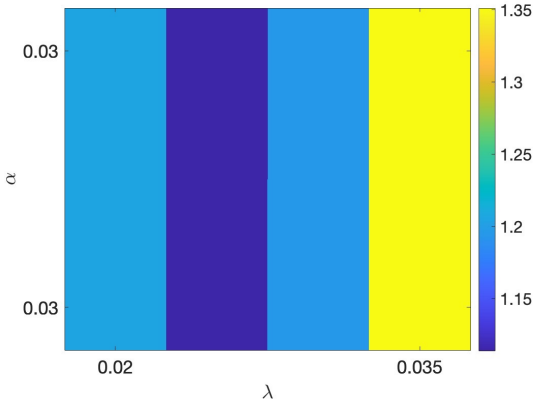


Figure 8-4 Box and whisker plot displaying PSI and MSE outputs in pre-treatment prediction (α=0.03, λ=0.012)

8.3.4 Re-optimisation of model parameters.

The α parameter of 0.03 was maintained and a vector of λ values were tested, iteratively reducing the upper and lower bounds of the vector to find the λ value that resulted in the lowest MSE. Full data are available in Appendix 10.6.2. The lowest MSE (1.11) was achieved using λ=0.025, Figure 8-5.



1.20	1.11	1.20	1.35
1.20	1.11	1.20	1.35
1.20	1.11	1.20	1.35
1.20	1.11	1.20	1.35

Figure 8-5 Heat map and associated error matrix, showing lowest MSE in λ re-optimisation (α=0.03 and λ=0.025)

Prediction of TVR using α=0.03 and λ=0.025 improved agreement between simulated and measured volumes with R^2=0.94, and Pearson Correlation Coefficient = 0.97 (Figure 8-6). Individual patient prediction plots can be seen in Figure 8-7.

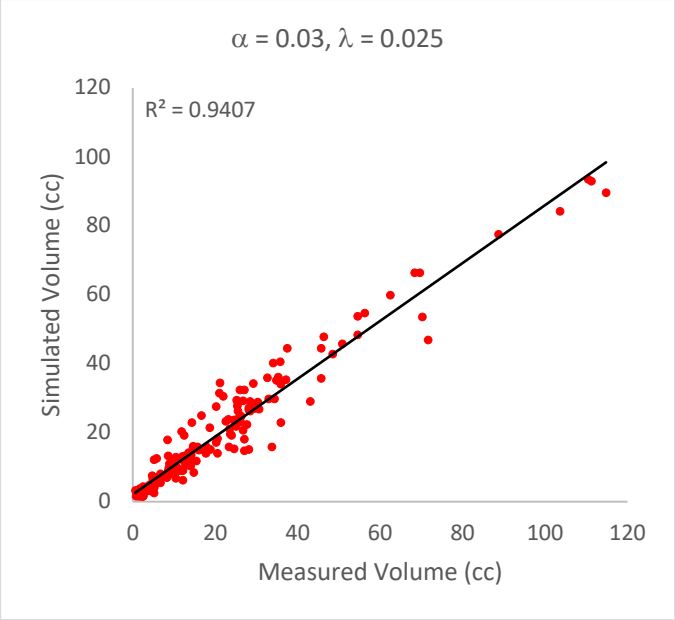


Figure 8-6 Scatter plot of volumes predicted using only pre-treatment data vs measured tumour volumes (α=0.03, λ=0.025)

X axis displays measured volume and y axis model-simulated volume (cm^3).

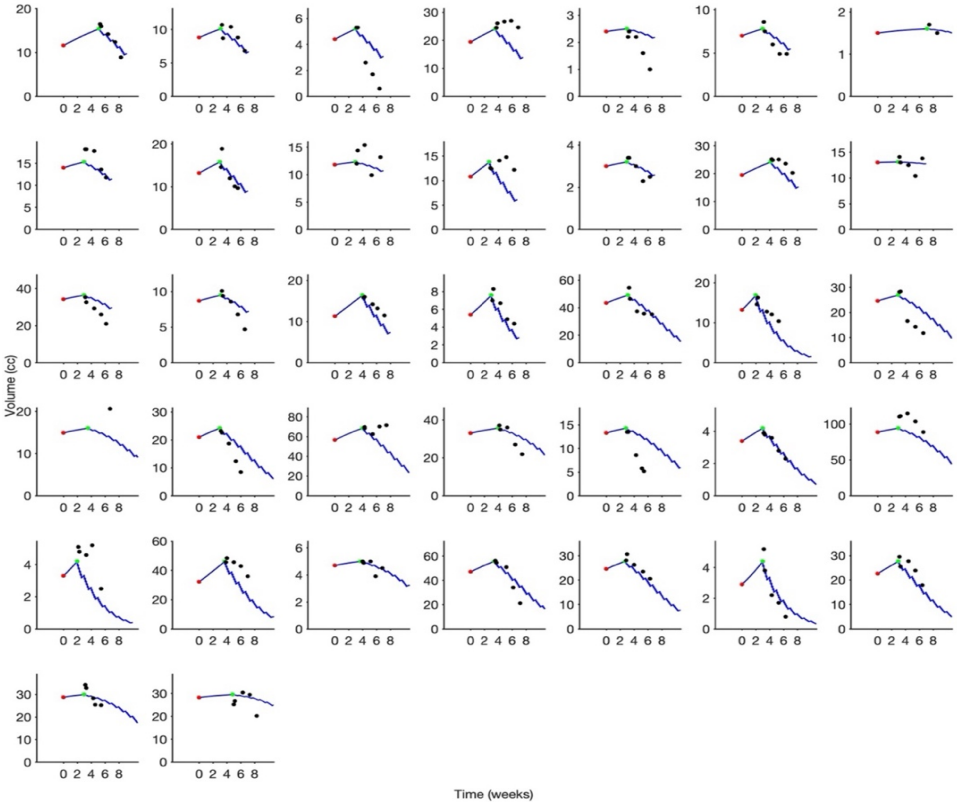
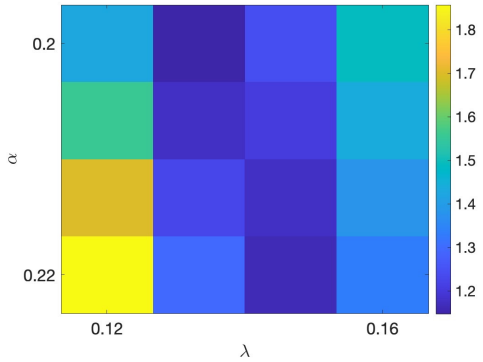


Figure 8-7 Individual plots of predicted TVR in pre-treatment prediction (α=0.03, λ=0.025)

Red dot is planning CT data, green dot is day 1 CBCT, Black dots are subsequent CBCTs. Blue curve is model simulated volumes. X axis displays time in weeks, y axis displays absolute volume in cm^3

Re-optimisation of both α and λ, using the same approach resulted in a parameter pair of α=0.21, λ=0.13, MSE=1.15 (Figure 8-8) with full data available in Appendix 10.6.3

The re-optimisation of both parameters did not improve prediction compared to re-optimising λ alone, with R^2=0.92, PCC=0.96 Figure 8-9, and individual predictions can be seen in Figure 8-10.



1.43	1.15	1.24	1.49
1.56	1.18	1.20	1.43
1.70	1.23	1.18	1.38
1.86	1.29	1.16	1.33

Figure 8-8 Heat map and associated error matrix showing for re-optimisation of both parameters displaying the lowest MSE achieved of 1.15 (α=0.21 and λ=0.13)

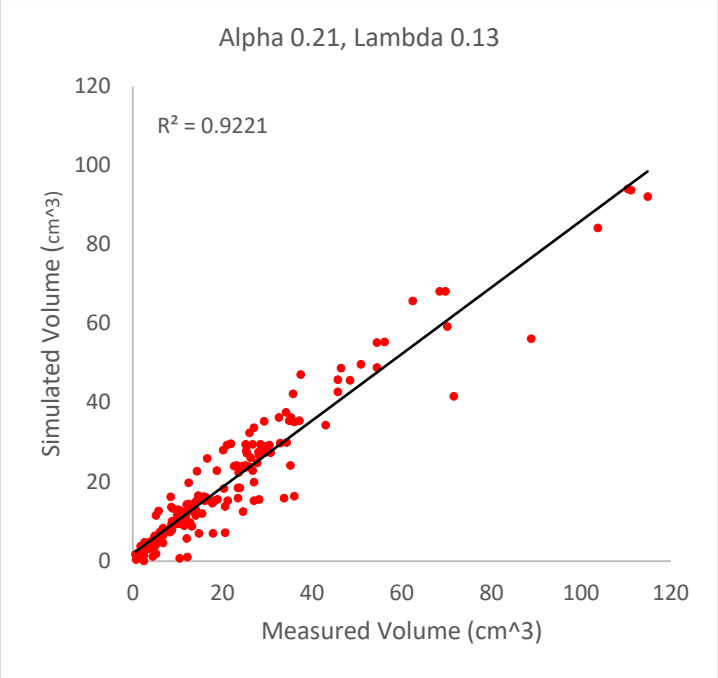


Figure 8-9 Scatter plot of volumes predicted using only pre-treatment data vs measured tumour volumes (α=0.21, λ=0.13)

X axis displays measured volume and y axis model-simulated volume (cm^3).

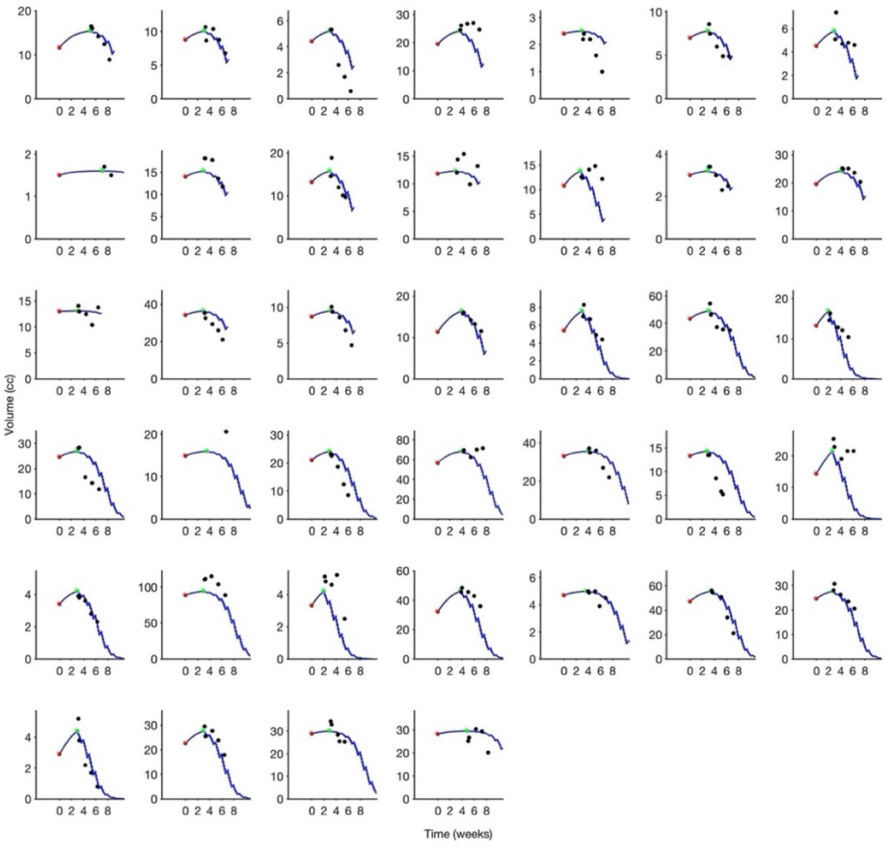


Figure 8-10 Individual plots of predicted TVR in pre-treatment prediction (α=0.21, λ=0.13)

X axis is time in weeks, y axis is absolute volume in cm3. Red dot is planning CT data, green dot is day 1 CBCT, Black dots are subsequent CBCTs. Blue curve is model simulated volumes.

Taking α=0.03 and λ=0.025 as the final parameter pair for prediction using pre-treatment data only, the association between measured and simulated volumes at the end timepoint of RT can be seen in Figure 8-11, R^2=0.86, PCC=0.93. Relative regression of each tumour compared to model-simulated relative regression is displayed in Figure 8-12.

Figure 8-11 Predicted volume vs measured volume at last image based on pre-treatment data only

X axis displays measured volume and y axis model-simulated volume (cm^3).

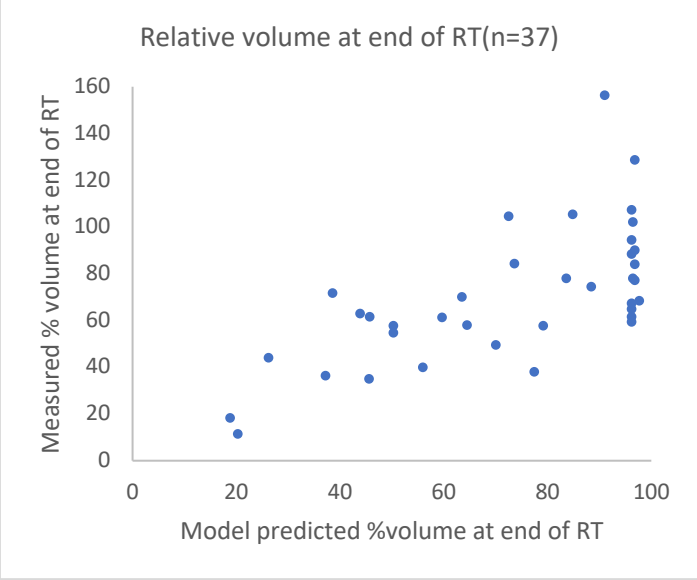


Figure 8-12 Scatter plot displaying the relative tumour volume at the end of vs modelled in pre-treatment prediction

X axis displays predicted volume (as percentage of day 1 volume) and y axis the measured volume (as percentage of day 1 volume).

8.3.5 Predictive performance of the PSI model

Performance of the PSI model predicting poor TVR (≤10%) was evaluated using only the pre-treatment data and sensitivity, specificity, positive predictive value and negative predictive value were found to be 75%, 69%, 40% and 90.9% respectively (Table 8-3).

Table 8-3 PSI performance predicting poor TVR (≤10%) using pre-treatment data only

TP= True positive, FP= False Positive, FN= False Negative, TN= True Negative

PSI prediction ≤10%	***Regression ≤10%***		
	Yes	**No**	**Total**
Positive	6 (TP)	9 (FP)	15
Negative	2 (FN)	20 (TN)	22
Total	8	29	37

8.3.6 Robustness to imaging uncertainty

As pre-treatment data were derived on 2 imaging modalities, AVIP reconstruction of 4DCT and CBCT, a robustness analysis was performed to quantify the potential impact of tumour volume uncertainty on prediction. Varying the AVIP volume for each patient by +/- 5% had limited impact on agreement metrics. Varying it by +/-10% resulted in a maximum difference in any metric of a 9.3% reduction in the R^2 value for the comparison of volumes at the end of RT (Table 8-4).

Table 8-4 Image uncertainty robustness analysis displaying the impact of varying the average intensity projection tumour volume by +/- 5-10%.

PCC= Pearson Correlation Coefficient, Δ%= percentage difference from the measured volume, R^2= Coefficient of Determination

	All Data				***Last image only***			
	PCC	**Δ%**	**R^2**	**Δ%**	**PCC**	**Δ%**	**R^2**	**Δ%**
Measured volume	0.97		0.94		0.93		0.86	
-5%	0.97	0.00	0.94	0.00	0.93	0.00	0.86	0.00
+5%	0.97	0.00	0.93	0.65	0.92	1.08	0.85	1.16
-10%	0.96	1.03	0.92	2.13	0.89	4.30	0.80	6.98
+10%	0.96	1.03	0.93	1.06	0.88	5.38	0.78	9.30

8.3.7 Clinical relevance of model simulated TVR

The clinical relevance of model simulated regression was examined in this subgroup. KM plot comparing those with TVR to those with stable or progressive disease on imaging at end of RT, demonstrated the same tendency as in the full cohort with a non-significant split in the data, indicating patients with TVR had worse OS (p=0.186) Figure 8-13. Splitting the data on TVR ≤10% was not significant (p=0.247) but again followed the same pattern (Figure 8-14).

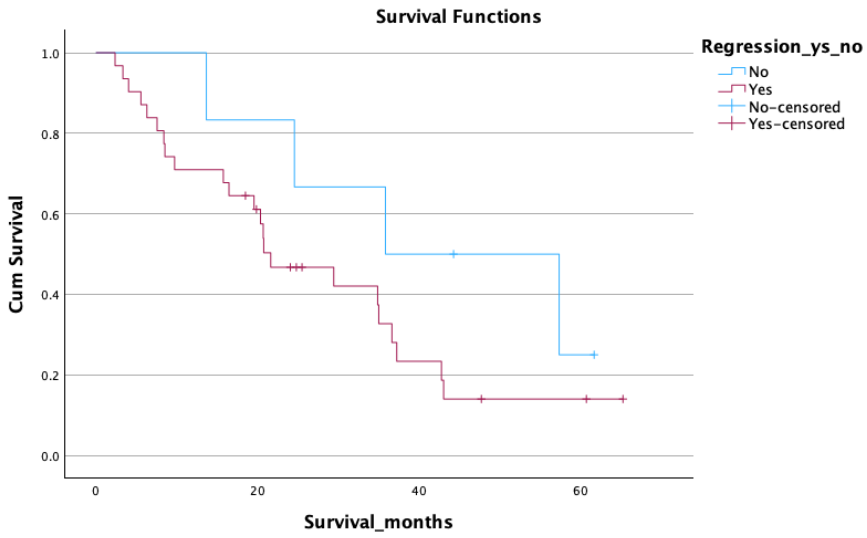


Figure 8-13 Kaplan Meier curves splitting the data on TVR vs no TVR at last CBCT in n=37 patients (p=0.186)

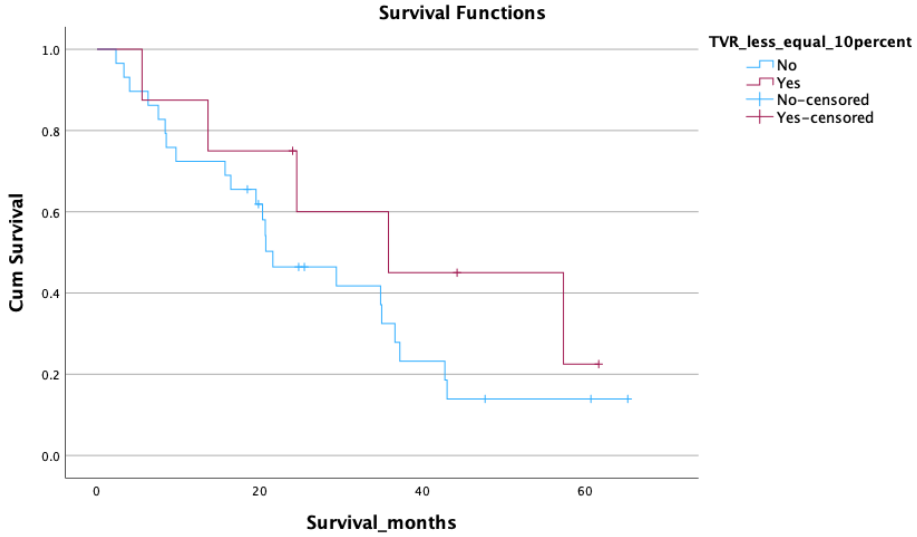


Figure 8-14 Kaplan Meier curves splitting the data on TVR ≤10% in n=37 patients (p=0.247)

Univariable Cox Regression for OS and LC was performed testing the same clinical and volumetric variables as the full cohort, but additionally the simulated volume at end of RT (SIM_V_{EndRT}) and simulated TVR ($SIM\_V_{\%REG}$) were tested. Neither simulated metric was significantly associated with OS or LC (Table 8-5, Table 8-6). In this subgroup $V_{\%PRE\_RT}$ was significantly associated with both OS and LC with increased pre-treatment tumour growth associated with increased risk of

death and disease progression (HR 1.048, 95% CI 1.016-1.082, p=0.003 and HR 1.078, 95% CI 1.023-1.135, p=0.005 respectively).

Table 8-5 Subgroup (n=37) Univariable Cox Regression Summary for Overall Survival

V_{Day1RT} = absolute volume at day 1 of RT in cm^3, V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT, $V_{\%PRE\_RT}$ =tumour growth pre-treatment expressed as a percentage of the volume, SIM_ V_{EndRT} = model simulated absolute volume at the final image RT in cm^3 SIM_$V_{\%EndRT}$ = model simulated percentage of TVR on final image RT, PS= Performance Status

	Hazards Ratio	***95.0% CI Lower***	***95.0% CI Upper***	***Sig.***	***Cat. Sig.***
Hist.					0.863
Adeno	Reference				
Clinical	1.578	0.545	4.57	0.401	
Other	1.183	0.499	2.805	0.704	
Squam	1.4	0.177	11.058	0.75	
Sex	0.975	0.667	1.427	0.898	
T Stage					0.816
T 1	Reference				
T 2	0.995	0.411	2.411	0.992	
T 3	1.325	0.51	3.445	0.563	
PS					0.446
PS 0	Reference				
PS 1	1.502	0.315	7.164	0.61	
PS 2	2.056	0.464	9.106	0.342	
PS 3	4.7	0.608	36.361	0.138	
V_{EndRT}	1.001	0.975	1.029	0.922	
V_{Day1RT}	1.003	0.98	1.026	0.81	
$V_{\%PRE\_RT}$	1.048	1.016	1.082	**0.003**	
$V_{\%REG}$	1.02	1.004	1.038	**0.017**	
$SIM\_\ V_{\%EndRT}$	0.994	0.979	1.009	0.45	
$SIM\_\ V_{EndRT}$	1.015	0.989	1.042	0.254	

Table 8-6 Subgroup (n=37) Univariate Cox Regression Summary for Local Progression

V_{Day1RT} = absolute volume at day 1 of RT in cm^3, V_{EndRT} = absolute volume at the final image RT in cm^3, $V_{\%REG}$ = percentage of TVR on final image RT, $V_{\%PRE\_RT}$ = tumour growth pre-treatment expressed as a percentage of the volume, SIM_ V_{EndRT} = model simulated absolute volume at the final image RT in cm^3 SIM_$V_{\%EndRT}$ = model simulated percentage of TVR on final image RT, PS= Performance Status

	Hazards Ratio	***95.0% CI Lower***	***95.0% CI Upper***	***Sig.***	***Cat. Sig.***
Hist					0.221
Adeno	Reference				
Clinical	0.651	0.129	3.283	0.603	
Other	0.356	0.097	1.304	0.119	
Squam	5.453	0.966	30.773	0.055	
Sex	1.092	0.546	2.186	0.804	
T Stage					**0.03**
T 1	Reference				
T 2	2.668	0.239	29.725	0.425	
T 3	12.355	1.419	107.553	**0.023**	
PS					0.405
PS 0	Reference				
PS 1	0.135	0.012	1.504	0.104	
PS 2	0.35	0.067	1.824	0.213	
PS 3	0	0	.	0.987	
V_{EndRT}	1.026	0.989	1.064	0.176	
V_{Day1RT}	1.024	0.985	1.064	0.233	
$V_{\%PRE\_RT}$	1.078	1.023	1.135	**0.005**	
$V_{\%REG}$	1.014	0.983	1.046	0.38	
SIM_ $V_{\%EndRT}$	1.005	0.975	1.037	0.742	
SIM_ V_{EndRT}	1.045	0.998	1.094	0.062	

8.3.8 In-silico modelling of alternative prescriptions

For each patient 6 dose fractionation schedules were modelled and the simulated tumour volume at the end of the course of RT was extracted. No significant variation in volume was detected between the prescriptions (ANOVA p=0.89, Figure 8-15). Change in simulated volume relative to V_{Day1RT} were median -0.39cm^3 (range -12.24-1.96cm^3) All patients followed a trend in response to a given prescription, for all patients the lowest rate of TVR was simulated in response to 60Gy/30# and the

highest TVR was predicted using 60Gy/15# Figure 8-17. The magnitude of TVR difference between these prescriptions varied per patient, and was not linearly linked to rates of pre-treatment growth (Figure 8-16, Figure 8-17)

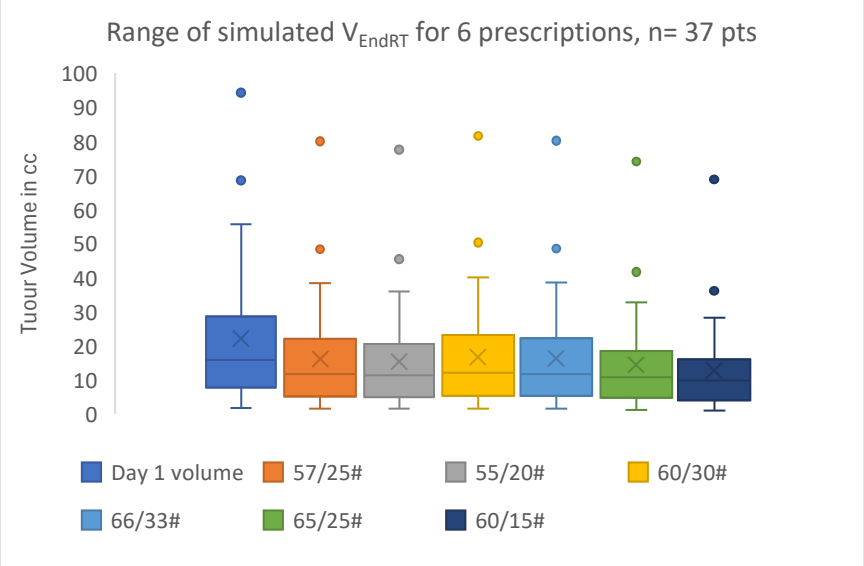


Figure 8-15 Box and Whisker plot displaying simulated volume at end of RT for a range of prescriptions,

Day 1 volume for reference (differences were not significant p=0.89 ANOVA)

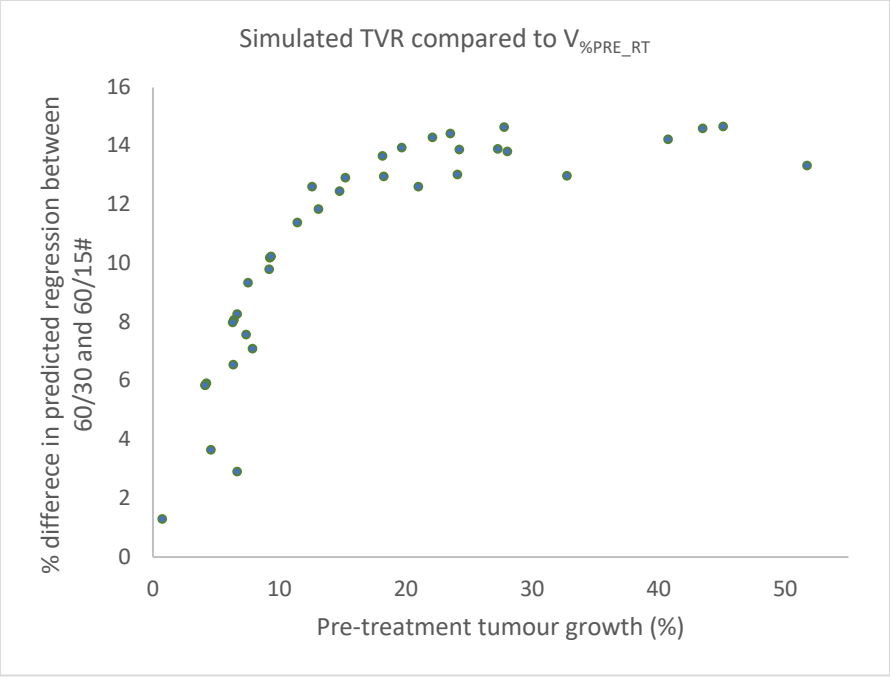


Figure 8-16 Variation in simulated TVR comparing 60Gy/30# to 60Gy/15#

X axis displays the rate of tumour growth pre-treatment, $V_{\%PRE\_RT,}$ Y axis displays the magnitude of difference between simulated TVR between the prescriptions predicting the highest and lowest TVR

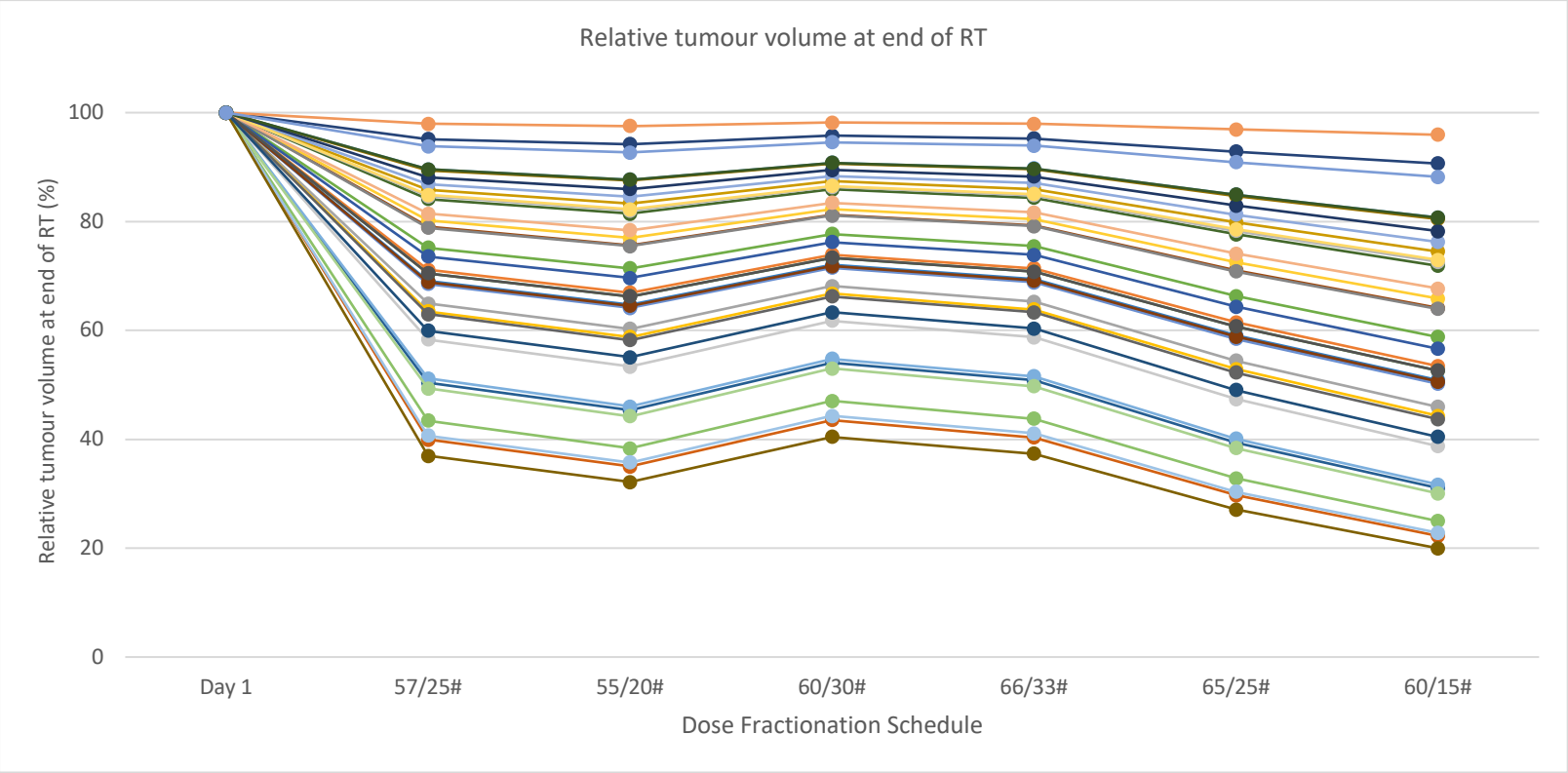


Figure 8-17 Variation in simulated volume at the end of RT for each prescription modelled

Each colour represents an individual patient. X axis displays the modelled prescriptions, Y axis note the volume in % relative to absolute volume on Day 1 of RT

8.4 Discussion

8.4.1 Testing model prediction using only pre-treatment dynamics

Tracking tumour volumes at planning CT and day 1 CBCT provided data on growth dynamics that was not considered in prior studies and enabled PSI predictions based on pre-treatment data alone.

Following the derivation of a new λ parameter, the model was solved only on the pre-treatment image data and a simulated response to RT derived. Model performance was encouraging with comparison of measured and simulated volumes at the final timepoint displaying excellent agreement (R^2=0.86, PCC=0.93, Figure 8-7, Figure 8-11).

Predictive models at a pre-treatment timepoint are more clinically valuable than predictions during RT, as they facilitate advance determination of patients that are likely to have greater TVR. Knowledge of TVR could allow appropriate selection of patients for adaptive RT studies [61, 200]. This has been comprehensively explored in the head and neck setting, where multiple approaches have been reported on identifying patients that will require re-planning during RT. These methods range from machine learning statistics to models based on anatomical features, and the inclusion of predicted TVR could augment these models and improve their predictive accuracy [238-240].

One challenge in tracking growth dynamics pre-treatment is that the data is acquired on 2 different imaging modalities for the longitudinal GTV measures, which could introduce a systematic error. This uncertainty was minimised by only including patients with a 4D planning CT and the GTV was recontoured on the AVIP reconstruction as outlined in Study 3, which has been shown to most closely align with the CBCT image [175]. Future work should measure the GTV on consistent modalities, but

longitudinal planning CT acquisition is uncommon outside the framework of a clinical trial. Repeated diagnostic imaging prior to RT is potentially useful, but these scans are not acquired for the purposes of GTV delineation and are most frequently acquired in breath hold which would not align with a CBCT derived volume [241]. There is a paucity of data on the correlation of lung tumour volumes defined across varying CT modalities. One study comparing 3DCT to CBCT reported a ratio of 0.89:1 for volumes on 3DCT:CBCT [191]. In this thesis a comparison of 4DCT AVIP to 4DCBCT AVIP found a ratio of 0.98:1 as reported in Study 1. A robustness analysis of AVIP delineation uncertainty demonstrated that variation of +/-5% had limited impact on agreement metrics of the modelled volumes. Future work is required to externally validate this prediction on pre-treat data.

8.4.2 Clinical relevance of TVR and model simulated TVR

In this subgroup the absolute measured volumes were not significantly associated with clinical outcomes, however the relative tumour growth pre-treatment ($V_{\%PRE\_RT}$) and TVR in response to RT were significantly linked to OS. $V_{\%PRE\_RT}$ had the largest effect size with 4.8% increase in the risk of death and 7.8% increase in the risk of local progression for each percentage increase in $V_{\%PRE\_RT}$. Model simulated TVR was also assessed and was not significantly associated with clinical outcomes. This may relate back to the model assumption that all tumours will regress in response to RT, and some tumours do not actually shrink during treatment, Figure 8-12. Echoing the analysis of the full cohort in Study 4, $V_{\%REG}$ was a negative predictor of OS, albeit with a small effect size, and so poor TVR (≤10%) was utilised as a cutoff to test PSI model performance as a predictive test, and in line with the full cohort there was a high negative predictive value. However, as noted in the previous chapter, future work should include all clinically relevant variables in the multivariable analysis.

8.4.3 In-silico modelling of alternative prescriptions

A proposed utility of the PSI model in the pre-treatment scenario is the selection of personalised prescription, as has been done in the head and neck setting [97]. We hypothesised that the PSI model would facilitate testing of a variety of dose fractionation schedules in NSCLC, and identify a patient specific regression optimised prescription. While variation in the simulated volume at end of RT was observed between prescriptions, the differences were not significant (ANOVA p= 0.89). At an individual level, however, the impact on tumour regression at the end of RT varied widely. Comparison of simulated TVR difference between the two prescriptions with the smallest and largest TVR, against $V_{\%PRE\_RT}$ revealed there was not a linear trend, indicating that pre-treatment growth rate alone is inadequate to select an optimal prescription for a patient (Figure 8-16). Moving from a CFRT schedule to a hypofractionated approach comes at the cost of increased acute toxicity [50, 52], and advance knowledge of the expected benefit of employing a hypofractionated prescription could support informed decision making. If future studies, with larger patient numbers, confirm the signals seen in Study 4 for patients with adenocarcinoma histology, where increased regression was linked to increased LC, modelling prescriptions could optimise volumetric regression in those patients. In the IO era, tumour burden is linked to ICI effectiveness [242] and therefore maximising TVR may be desirable for all patients proceeding to IO. Furthermore in other sites the integration of tumour volume into models have improved prediction of response to ICI [243].

As discussed in Chapter 3, several approaches to personalise RT prescriptions have been tested, or suggested as strategies [54] but lack the true personalisation aspect as the common focus was dose intensification and the research moving closer to personalised prescriptions in RT continues [244].

Again, the confounding factor in this analysis is the data were generated on 2 different modalities. Additionally, n= 32 patients were excluded from this analysis due to no detected pre-treatment growth. It is implausible that these patients had regression of their tumour prior to commencing any treatment and therefore we must consider that the uncertainty arises from the imaging modalities.

8.5 Conclusion

Prediction using pre-treatment data required recalibration of the tumour growth parameter but resulted in excellent agreement between simulated and measured tumour volumes at the end of RT. These findings are limited by the uncertainty of the robustness of GTV delineation to image modality and external validation is warranted.

CHAPTER 9 DISCUSSION

9.1 Mathematical modelling in radiation therapy

The statement that 'All models are wrong, but some are useful' is attributed to the statistician George Box in 1976. While a model is inherently wrong, as it is a simulation rather than a measurement, mathematical modelling in RT is fully integrated into many aspects of practice, particularly in radiobiology. The linear quadratic equation, α/β ratios, TCP and NTCP modelling are all part of our daily lexicon in research and clinical practice as they have demonstrated their utility.

Expansion beyond these core models is desirable, and predictive or prognostic models can supplement clinical decision-making. The PSI model evaluated here in NSCLC RT appears useful based on the data presented in this thesis. Limitations exist in its performance and further validation and training is necessary in expanded cohorts of patients to advance the transferability of the model to new patients, but this work is a building block on which the robustness of the model can be further tested. Pragmatism and clinical relevance are key when taking models from theory to reality. Models most likely to reach clinical implementation are those that are built on routinely collected, robust data, and those that address an unmet need or uncertainty [245]. The PSI model predicting TVR in response to RT does address an unmet need in the clinic and would aid decision making in relation to ART or enrolment in clinical trials [61]. Input parameters in the PSI model are tumour volume measurements with associated timepoints, and while current practice does not gather GTV data routinely, the imaging data is acquired as a standard part of daily or weekly image guidance, and this thesis has demonstrated the feasibility of routinely defining the target volume on these images.

9.2 Predicting treatment outcomes: an ongoing challenge

9.2.1 The potential of TVR as a biomarker

Forecasting clinical outcomes in NSCLC remain a challenge, and established nomograms are based on the clinical features of a patient or their disease, such as age, stage, histology and tumour grade [246]. Now, more than ever, there is a desire to predict clinical outcomes to empower clinicians and patients, and provide tailored information on potential benefits or risks of a given treatment [61]. TVR in response to RT could add additional insight, alongside the recognised clinical factors, to augment predictive models and optimise treatment for each patient. Absolute volume of the tumour at the start of RT is a known prognostic factor [211], but TVR can supplement this single timepoint measure of volume, with data on tumour dynamics over the course of RT. Evidence is emerging that mid-treatment or end of treatment TVR is a more robust prognostic indicator than pre-treatment volume alone [73, 75, 212]. Longitudinal imaging is standard in RT and CBCT images are acquired daily or weekly for geometric verification; however, their potential remains under-utilised. There is a wealth of unexplored data acquired routinely, and CBCT imaging could be a valuable untapped resource. This thesis sought to examine the feasibility of quantifying and predicting TVR in response to RT in NSCLC using these images.

Study 1 demonstrated the viability of quantifying tumour dynamics over the course of RT using a commercially available auto-segmentation algorithm. Contours generated compared favourably to the physician defined MC in the geometric agreement metrics of DSC, volume and COM shift. Dosimetric examination further identified that the AC target volumes did not result in statistically different DVH metrics compared to the manual contour over the course of RT. The approach was then piloted

on CBCT data in Study 2, and the method was reasonable and efficient in 66.6% of patients. These chapters have demonstrated that it is viable to extract longitudinal tumour volume dynamics in a selected proportion of the NSCLC population using a widely available software. This could provide a framework to exploit the untapped resource of CBCT imaging data.

Studies 3, 4 and 5 focused on prediction of TVR, evaluating the PSI model performance. Study 3 trained and tested this model in a heterogeneous population of patients, treated with a variety of dose fractionation schedules. From this work a radiation response parameter was derived, which was used to parameterise the PSI model, and resulted in excellent agreement between modelled and measured tumour volumes over the course of RT. Study 4 validated, in an external dataset, the model performance predicting TVR by day 10 of RT. In Study 5 pre-treatment prediction based only on 2 images acquired prior to RT delivery showed promising results. The PSI model predicted end of treatment tumour volume with good agreement between the final measured volumes and the model-simulated volumes (R^2=0.93) and these findings warrant validation testing in an external cohort. The clinical relevance of regression in this cohort was examined and TVR was found to be a negative prognostic factor for OS. Modelling of alternative prescriptions indicated that there was a variable rate of change in TVR per patient and this suggests the potential to personalise the RT prescription per patient.

9.2.2 Managing a fast-moving treatment landscape

Implementing personalised RT based on TVR is complicated by the rapidly evolving clinical landscape in NSCLC. In recent years the role of immunotherapy (IO) has expanded, and the impact of this has been immense. Patients receiving CCRT for Stage III NSCLC are now recommended to proceed to consolidative IO following the PACIFIC trial, which has improved 5-year OS rates to 42.9% [14, 15]. More recently the

use of neo-adjuvant IO/chemotherapy in the form of Nivolumab and chemotherapy has shown increased event-free survival and pathological complete response in patients proceeding to surgical resection [247]. This has led to a renewed interest in surgical management of patients previously deemed unresectable and referred for CCRT and patients with up to T3 N2 or T4 N1 disease are now considered potentially resectable [248]. All patients in the model validation study had N0 disease and therefore were treated with definitive RT alone, however, these patients may now be eligible for SABR, or neo-adjuvant chemoRT. These data highlight that the future work on evaluating the PSI model is warranted in patients with the bulky stage III disease, and that the cohort assessed in Studies 4 and 5 may now undergo different clinical management. Acknowledging this challenge of researching in a fast changing field should lead to improved methodologies in future studies, such as rapid-learning approaches [249].

9.2.3 Challenges associated with the PSI model

A substantial challenge with applying the PSI model to NSCLC is the limitations associated with lung tumour imaging. Lung tumours are in constant motion due to respiratory and cardiac motion and imaging a moving target is challenging. To most accurately define a lung GTV a breath hold (BH) scan can be acquired to eliminate the motion, and this strategy is regularly employed in the SABR setting. However, BH is not commonly used for more protracted schedules, and the CBCT is acquired in free-breathing (FB). This leads to a time weighted image of the tumour trajectory during the scan acquisition and the tumour boundaries are less distinct from the surrounding normal tissue, adding uncertainty to the GTV volume measured on these images [250]. This was considered in this analysis by performing the PSI model training and validation using only on-treatment image data for consistency in image modality. While uncertainty remains in the ground truth of tumour volume on CBCT measurements were carried out on a consistent modality to offset this.

Using only the on-treatment data allowed prediction of end of RT volume from a median of day 10 of treatment as shown in Study 4.

Modelling response to RT in advance of treatment would be more advantageous, and at least 2 measures of TV over 2 separate timepoints are required to predict TVR in the PSI model. The second challenge in lung imaging is the variety of image acquisition types and, outside of CBCT, repeated imaging with the same modality is rare. Diagnostic CT, PET CT and planning CT scans all have different acquisition parameters which influence the final image. In Study 5 the AVIP reconstruction of the planning CT was utilised to derive the tumour volume, to mimic the CBCT and reduce uncertainty between the 2 scans. Additionally, a sensitivity analysis varying the AVIP volume by +/- 10% was performed to ascertain the influence of this uncertainty which identified minimal impact. Predicting TVR in response to RT from only the planning CT and day 1 CBCT resulted in good agreement between the measured and simulated volumes at the end of treatment (R^2=0.86).

Future work should validate the pre-treatment prediction performance and a prospective approach with pre-planned image acquisition for the purpose of PSI modelling would be beneficial. Longitudinal BH scans acquired pre-treatment would improve delineation accuracy, and a final BH scan at the end of RT would facilitate more reliable measures of regression to truly test the model performance.

9.2.4 PSI model development

The PSI model is one of a multiplicity of modelling options that exist in this arena. The PSI model evaluated here is simplistic, and others have added layers of depth to the model, but this comes with the cost of increased data required to adequately and accurately parameterise the model.

In this PSI model, tumour response to RT is modelled as an instantaneous reduction in the GTV following RT. An extension of the PSI, where the GTV is subdivided into two compartments to represent the living and necrotic cells has been examined [251]. The adjusted PSI model predicted GTV trajectory from day 21 or day 28 of RT in patients with head and neck cancer. This is significantly later than the predictions made here, and the study reports training and internal validation on a limited cohort of 51 patients.

Another variation of the PSI model simulated the impact of RT on the environment, rather than the tumour itself, by modelling an instantaneous reduction in the carrying capacity of the environment following RT as opposed to a reduction in tumour volume [252]. This study reported prediction of LC and disease-free survival rather than examining the GTV trajectory and reported good sensitivity and specificity (75%, 83% and 68%, 85% respectively for LC and DFS). Again, this was a smaller study in head and neck RT (n=39pts) and while an external dataset was available for validation of the model parameters, clinical predictions were made using a leave one out cross validation approach due to the low number of recurrence events in the groups [252]. This thesis has demonstrated improved agreement at an earlier timepoint using the more simplified PSI model and validated this in a cohort of n=71 patients.

9.2.5 PSI for NSCLC within a model evaluation framework

The TRIPOD statement and checklists were published in 2015 and serve as a tool to ensure transparency and consistency in reporting of prediction models in clinical practice [98]. TRIPOD checklist for Studies 3, 4 and 5 are available in Appendix 10.4.1, Appendix 10.5.1 and Appendix 10.6.1.

While this thesis presented both Type 1a studies (model development) and a Type 4 Study (model validation in new data), further validation and testing can be pursued to ensure the robustness of the model

performance. A gold-standard could be a prospective clinical trial based on the PSI model in NSCLC, whereby prospective model predictions are assessed without the inherent uncertainty that exists in retrospective evaluations. This approach has been taken in the head and neck setting where the RT prescription was selected based on the PSI model outputs pre-treatment but this study did not report performance in predicting TVR during RT [97].

9.3 Clinical potential of the PSI model

The PSI model predicts TVR, as measured grossly on a CT or CBCT, which is a blunt tool in an era of precision medicine. TVR in response to RT needs to be appraised in the appropriate biological and clinical context for an individual patient.

The concurrent use of chemotherapy is a confounding factor, impacting not only the magnitude of TVR, but also the clinical relevance of this regression [71, 211]. Large or small TVR has been reported as having opposing associations in relation to OS and local relapse in patients treated with or without chemotherapy [71]. In the cohort examined in Study 4, TVR was negatively associated with overall survival (HR 1.02, 95% CI 1.005-1.035, p=0.007). Repeating this model validation in a patient group receiving concurrent CCRT would enable deeper examination of these confounding factors. As previously discussed, IO has transformed clinical outcomes for patients with NSCLC. Median OS has increased from 29.1 to 47.5 months and 42.9% of patients treated with consolidative IO were still alive at 5 years [14]. An eligibility criterion for the PACIFIC trial was no progression at the end of chemoRT. While the clinical relevance of TVR requires further investigation to elucidate its prognostic significance, maximising regression to enable more patients onto consolidative IO would be desirable. If the PSI model can predict in advance those patients with poor response, it could be employed to

initiate an adaptive strategy to optimise that regression. A hypofractionated schedule or GTV boost could be explored for these patients in a clinical trial setting. In Study 5 alternative fractionation schedules were modelled and the moderately hypofractionted prescription of 60Gy/15# simulated the maximum regression for each patient. The PSI model can identify those patients that might benefit from this prescription, given that there is a likely trade off of increased acute toxicity with increased daily dose [237]. We know that a blanket dose increase for all patients has not been successful, most likely due to increased toxicity [253] and selecting only those patients who have large predicted TVR could facilitate patient selection for personalised prescription trials [32, 61].

9.3.1 Personalised care path

In this cohort larger TVR was a negative prognostic factor. Prediction of a negative risk factor could signpost patients who need adjustments to the standard care path. This could signal shorter interval surveillance if we expect faster relapse post RT, it could indicate the need for an adjuvant systemic anti-cancer therapy if that was not originally included in the clinical plan. The RT prescription could be reconsidered for a residual GTV boost similar to the approach recently reported using accelerated CCRT followed by a SABR boost [58].

Larger TVR is potentially a surrogate for aggressive underlying biology with a rapidly proliferating tumour which is a known negative prognostic indicator [254]. Other studies have reported regression as a positive predictor and this highlights the importance of considering the full clinical picture of tumour volume dynamics, histology and all treatment data. This analysis could be complemented by a biological analysis of proliferation markers such as Ki67 or functional imaging to ascertain the correlation between modelled PSI and tumour proliferation markers. ^{18}F fluorothymidine (FLT) PET imaging facilitates a non-invasive evaluation

of tumour proliferation [255]. In NSCLC FLT-PET imaging is reproducible and longitudinal imaging has identified early changes in tracer uptake [256].

9.3.2 PSI in the era of adaptive RT

Outside of adjusting the prescription, the PSI model could still play a role in personalising the management of an individual patient. Predicted TVR could be a trigger for planned adaptive strategies in the clinic. Anatomical changes are common during RT for NSCLC [257] and ad-hoc reactive replanning is inefficient and a resource drain on a department. Having advance knowledge of TVR could facilitate planned re-scanning and replanning for selected patients and avoid unwanted breaks during RT.

The role of ART remains under investigation in lung cancer and questions on efficacy and safety remain despite the clinical desire to introduce ART in this site [258]. Study 2 identified that there was a degradation in plan quality over the course of RT and without replanning with only 17/36 on treatment scans met the PTV dose volume constraints in line with the original plan. There is a paucity of high-level evidence as to the effectiveness of ART in NSCLC. The LARTIA trial prospectively evaluated GTV shrinkage in 217 patients and identified that 50 of these required ART. The trial reported low rates of toxicity and no unexpected increase of in-field or marginal failures with a shrinking target volume [104].

There is a concern that shrinking the target volume due to regression identified on mid-treatment imaging could lead to underdosage of microscopic disease (MD). Retrospective TCP modelling suggests that the original CTV region of probable MD remains covered by the 50Gy isodose region when adapting to a shrinking GTV, but 60Gy coverage was compromised. If dose escalation was included alongside TV reduction, then the compromise to 60Gy coverage was compensated [259]. A

limitation of this study is it looked at 3DCRT plans rather than IMRT, and the dose gradient in modern modulated techniques are steep. Repeat TCP analysis of the CTV region in adaptive IMRT plans is warranted. In the SABR setting peritumour density and sharp dose fall off have been linked to clinical outcomes indicating that caution should be exercised when shrinking target volumes mid-RT [260, 261]. These considerations imply that we should be prudent when implementing ART, and one role of the PSI model could be to select patients with large predicted TVR to enrol them into a clinical trial investigating ART in NSCLC.

Most ART strategies aim to reduce the PTV to account for a shrinking GTV, but there is an argument for adapting to the mid-treatment anatomical variation in normal tissue caused by TVR. TVR can induce inter-thoracic anatomical changes such as mediastinal shifts and therefore ART to account for moving OARs is warranted to maintain plan quality [262]. In the current era where clinical outcomes are improving, OAR sparing is critical to reduce late toxicity and ensure access to consolidation IO. ART during CCRT can reduce rates and severity of adverse events [157, 171] and may improve the uptake of IO in patients where consolidative treatment is not recommended due to toxicity post CCRT. Not all patients who are eligible for IO proceed to consolidative therapy and the reasons for this are multifactorial. One US based survey reported that patients decline IO due to costs or anticipated adverse events, and physicians do not recommend it due to progression or poor performance status following CCRT [25].

9.3.2.1 PSI and selection of patients for adaptive RT

As discussed, the PSI model could identify which patients should be enrolled in an ART clinical trial, by identifying in advance those patients who will have large TVR. Large regression signals the need for geometric ART to ensure the dose to the PTV or OARs are not compromised by large anatomical shifts [262]. Studies 3 and 4 demonstrated the PSI model

can predict this regression with a good degree of accuracy. Or if large regression is a negative prognostic indicator, as we observed in Study 4 and 5, PSI predicted TVR could also flag patients for dosimetric ART studies or a SABR boost to residual disease [58]. Personalised prescriptions have been evaluated in the clinic, usually in the framework of iso-toxic escalation or boosting sub volumes of PET avid disease. The results of the PET Boost Trial and the NARLAL II trial have been somewhat underwhelming in relation to improving clinical efficacy and report mixed messages on the safety of dose escalation in these patients [55, 57, 263]. Full publication of the NARLAL II trial data are awaited, but both studies aimed to dose escalate all patients. Investigating a GTV boost, only in patients who have poor prognostic features, is an approach worthy of exploration. Rather than aiming to increase dose to all, we should better select patients for these trials and personalise the approach. Retrospective evidence suggests the clinical benefits of ART vary between NSCLC subtypes with reported survival benefits in those with squamous cell carcinoma only, but reduced toxicity measured in all subgroups [171]. The population included in Studies 4 and 5 of this thesis is representative of the clinical reality locally that we treat a relatively high proportion of patients with a clinical diagnosis and no histological confirmation of disease, which could be a confounding factor when making a clinical decision to adapt based on TVR alone.

9.3.2.2 PSI and timing of ART

Most studies suggest adapting the plan at the approximate mid-point of RT or after delivery of 15-20# [106, 168, 198, 264]. Adapting too early could lead to repeated ART at a later stage and adapting too late may negate the impact of the intervention. Using the PSI model, we can simulate the regression trajectory for an individual and identify a timepoint that the TVR warrants replanning. Currently no consensus exists as to the optimal time point or percentage regression at which to adapt, but regression greater than 30% has been suggested as an

appropriate marker [106] however less marked regression was observed in this research with a median regression of 15.8% measured in the cohort examined in Study 2 and 4. Nonetheless, as demonstrated in Studies 3 and 4 modelled regression can be individually plotted over time, allowing us to analysis the trajectory of TVR and select the appropriate timepoint for adaptation. It is likely that more complex modelling than simply TVR may be required to encompass additional data such as tumour location, nearby critical OARs, prescription and concurrent systemic therapies to guide the clinical decision to adapt.

9.3.2.3 CBCT delineation an underutilised support for ART?

ART requires enhanced image guidance, and various solutions are reported in the literature including mid-treatment repeat planning CT, 4DCT, MRI, PET or CBCT guided ART [76, 265-267]. While the image quality of CBCT is limited, it has the advantage of being the only image routinely acquired and does not introduce an added burden on either the department or the patient, in the form of additional imaging. Studies 1 and 2 demonstrated the potential to delineate the GTV on daily imaging with limited manual adjustments. This could, at a minimum, be utilised to screen patients where pre-planned rescanning was not flagged. Replanning on CBCT is technically feasible and a variety of dose calculation approaches can be taken to facilitate this but motion artefact, as is inherent in 3DCBCT, can reduce the dosimetric reliability of the plan [268]. As a result, current practice rarely relies on CBCT. Recent advances in on-treatment image acquisition have substantially improved image quality. Solutions such as Varian HyperSight, which incorporates an extended field of view, 6 second acquisition and improved reconstruction algorithms may increase the use of routine CBCT for ART. CBCT guided online ART is feasible using these high-quality images [269-272]. The move towards clinical implementation of higher quality of CBCT will necessitate increased auto contouring on these images and the delineation methods tested in this thesis may become more widely

applicable. Testing semi-automated GTV delineation in HyperSight images would be beneficial.

9.3.3 Clinical Implementation Considerations

The clinical implementation of the research presented in this thesis is indicated in several scenarios.

The auto-delineation workflow as described, could be used in the online adaptive replanning pathway. In this scenario, clinical accuracy is critical and more extensive validation would be needed to verify the validity of this approach. Under these circumstances, the RTT adjusting the contours would work within an advanced practice framework, and thorough credentialling would be essential to ensure the quality of the contours. Advanced Practice RT roles have demonstrated that this is feasible, and increases contouring efficiencies in the clinical workflow [273].

Alternatively, the tool could be used as an efficient on-treatment check of TVR, to quantify substantial regression and flag the need for a rescan and replan. In this situation, where the GTV contour is not created for the purpose of treatment delivery, it is reasonable to suggest that more RTTs, with less experience, could complete the task. The workflow presented here suggests manual edits off gross anatomy are required, not the generation of a target volume without any guidance. Additionally, the original physician defined volume is visible as a guide. This would allow for the efficient generation of an on-treatment GTV which could aid in the generation of large research datasets.

Similarly, the PSI model has multiple potential applications where the associated tolerances of performance may be varied. For example, if using the PSI predicted TVR to schedule a planned adaptive rescan, the consequence of a poor prediction is minimal. If not required a rescan could be cancelled with no impact on the patient and clinical outcomes.

If the PSI model is used to prospectively select an optimal prescription based on predicted TVR, then the threshold of an accurate prediction must be set at a higher level as the clinical consequences are greater.
These examples highlight the changing tolerances and workflows depending on the specific application of the research.

9.4 Limitations and future directions

This thesis opens exciting opportunities to develop the quantification and prediction of tumour volumes further and bring these data closer to clinical utility.

Repeated delineation of target volumes is increasingly common in both clinical practice and research settings. A limitation of Studies 1 and 2 is the single observer manually adjusting the auto-contoured GTVs. While this approach eliminates inter-observer variability, future analyses should include multiple observers and report on inter-observer variation in the resulting contours. The methods described here should be validated with multiple observers and compared to alternative options that employ deformable image registration algorithms to propagate the original GTV onto a new dataset. Additionally with advent of more advanced CBCT, bringing IGRT image quality closer to planning CT quality, we can potentially expand the accessibility of these methods to those patients who were excluded from Study 2 due to poor visibility of the GTV boundary on CBCT.

Future work on the PSI model should pursue 2 distinct avenues. Firstly, model validation in larger cohorts with more heterogeneous clinical care is required to improve the generalisability of the findings. Patients treated with CCRT, and better stratification based on histology would be desirable to truly test model performance. NSCLC is an umbrella term for a number of disease subtypes and individual histology is a confounding

factor in this thesis. In Study 3, data were acquired from prior publications and no clinical information was available except that there were mixed histologies within the groups. This limited our ability to ascertain the potential influence of confounding factors on the model prediction. In Study 4, histological data were available, but we observed a high percentage of patients with a clinical diagnosis without biopsy and therefore, while there was some stratification, it remains a confounding factor in the analysis. As a result, the clinical relevance of TVR is difficult to interpret and not generalisable to all patients. While the patients included in this thesis were not treated with IO, the introduction of IO may bring about a fundamental shift in evaluation of TVR. Pseudo-progression in patients treated with IO occurs in 2-8% of patients and IO adapted guidelines have been developed for response assessment in follow up imaging [274, 275].

Secondly the focus should move from on-treatment prediction to pre-treatment prediction. Study 5 demonstrated the feasibility of predicting TVR purely on pre-treatment data and if robustly validated in external patient cohorts, this could represent a paradigm shift in adaptive RT in NSCLC. If the model predicts that a patient will have significant tumour shrinkage during treatment, this will facilitate advance selection of patients for ART studies. These patients could benefit from improved TCP and NTCP due to replanning. Conversely if the model predicts the patient will not have strong TVR, this may flag the need for a review of the overall treatment approach to ascertain if this is a positive or negative indicator in this individual and adapt the care path appropriately.

The PSI model prediction using early response dynamics was validated in patients treated with 55Gy/20# over 4 weeks. This schedule is common in the UK oncology community but 60-66Gy/30-33# is more common in the EU/US settings. As validation was carried out using the on-treatment data from fraction 10, a prediction made at this timepoint could be more

clinically impactful in the longer schedule where there are many weeks of RT remaining. In the patients treated with 30-33# we could flag by day 10 if an adaptive plan may be required and schedule rescanning and replanning at an appropriate time point for an individual.

With the general movement towards accelerated and hypo-fractionated schedules in the clinic, further research on predicting using the pre-treatment data only is warranted. Furthermore PSI will need to evolve alongside this trend in RT practice, and the efficacy of SABR may be less dependent on tumour proliferation rates in comparison to the known association between proliferation and response to RT in CFRT schedules. Vascular changes, immunomodulation and endothelial cell damage have all been linked with SABR efficacy [276]. Testing of the underlying assumptions of the model in this more complex clinical scenario will be necessary, and recalibration of the parameters required.

9.5 Conclusion

This thesis has demonstrated the feasibility of employing a semi-automatic delineation tool to define the tumour volume on longitudinal imaging in NSCLC. This addresses the challenge of generating data on large scale datasets of daily acquired CBCT images, without the resource implications of repeated manual delineation. These data of tumour volume kinetics, can then be used to parametrise and train a mathematical model and predict TVR during RT treatment. The work presented in this thesis has validated the viability of using early response dynamics to predict end of treatment tumour volumes. Furthermore the use of pre-treatment data only to predict regression in response to RT has been presented and future research is warranted to validate these positive results in external, heterogenous cohorts.

The use of TVR as an imaging biomarker has the capacity to assist in the personalisation of RT, with potential benefits in informing adaptive strategies and selecting an optimal prescription for each patient. In the current era of personalised medicine the RT community is moving towards a patient specific approach, with the ultimate aim of improving the clinical of efficacy of RT in the treatment of NSCLC. The work presented in this thesis provides a framework for individually optimised RT in NSCLC.

CHAPTER 10 APPENDICES

10.1 Appendices Chapter 3

10.1.1 Appendix 10.1.1

Medline/ PubMed Search Strategy

Non-Small Cell Lung Cancer[Title/Abstract]
AND dose fractionation schedule[Title/Abstract])
OR optimised dose[Title/Abstract])
OR (personalized[All Fields]

AND radiotherapy dose[Title/Abstract]))
OR hypofractionated radiotherapy[Title/Abstract])
OR Hyperfractionated radiotherapy[Title/Abstract])
OR Accelerated fractionation[Title/Abstract])
OR (Isotoxic[All Fields]
AND planning[Title/Abstract]))
OR Dose escalation[Title/Abstract])
NOT Breast[Title/Abstract])
NOT Prostate[Title])

AND ("radiotherapy"[Subheading] OR "radiotherapy"[All Fields] OR "radiotherapy"[MeSH Terms])) AND ("carcinoma, non-small-cell lung"[MeSH Terms] OR ("carcinoma"[All Fields] AND "non-small-cell"[All Fields] AND "lung"[All Fields]) OR "non-small-cell lung carcinoma"[All Fields] OR ("non"[All Fields] AND "small"[All Fields] AND "cell"[All Fields] AND "lung"[All Fields] AND "cancer"[All Fields]) OR "non small cell lung cancer"[All Fields]) AND ("2000/01/01"[PDAT] : "2017/12/31"[PDAT])

Embase Search Strategy
Population
Non Small cell Lung cancer

Intervention
Personalized Medicine
Radiotherapy

Outcome
Local control

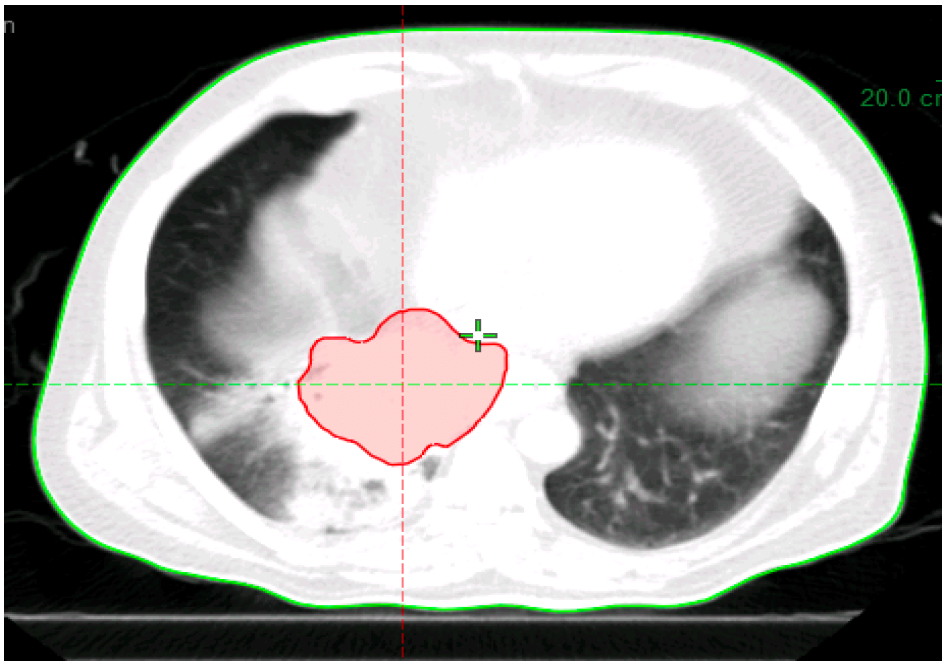
Tumour control probability
Normal tissue complication probability

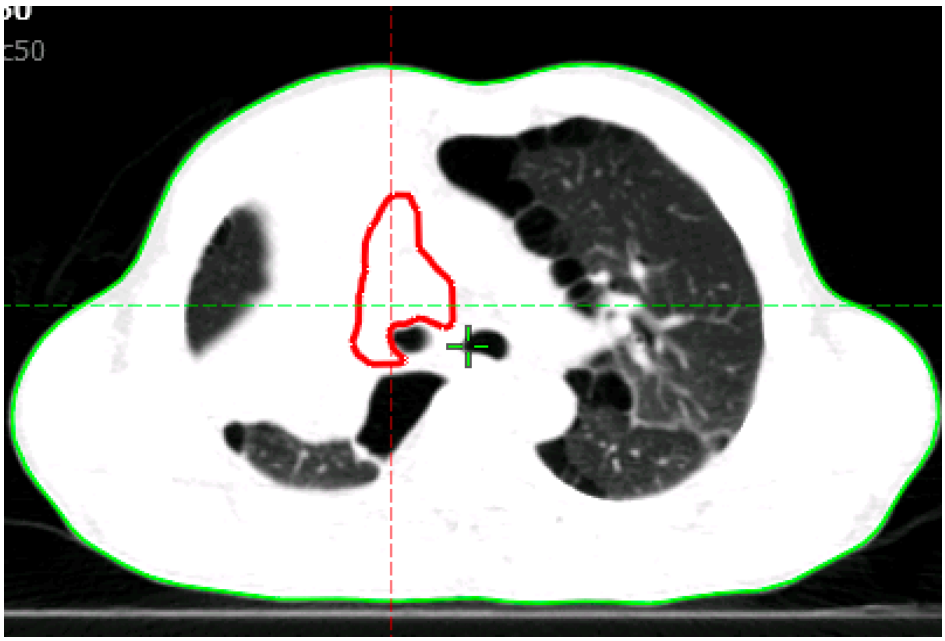
3'non small cell lung cancer'/exp OR 'bronchial non small cell cancer' OR 'bronchial non small cell carcinoma' OR 'carcinoma, non-small-cell lung' OR 'lung cancer, non small cell' OR 'lung non small cell cancer' OR 'lung non small cell carcinoma' OR 'non small cell bronchial cancer' OR 'non small cell cancer, lung' OR 'non small cell lung cancer' OR 'non small cell lung carcinoma' OR 'non small cell pulmonary cancer' OR 'non small cell pulmonary carcinoma' OR 'pulmonary non small cell cancer' OR 'pulmonary non small cell carcinoma' AND ('personalized medicine'/exp OR 'individualised medicine' OR 'individualised therapy' OR 'individualized medicine' OR 'individualized therapy' OR 'personalised medicine' OR 'personalised therapy' OR 'personalized medicine' OR 'personalized therapy' OR 'precision medicine' OR 'radiotherapy'/exp OR 'bioradiant therapy' OR 'bucky irradiation' OR 'bucky radiation' OR 'bucky radiotherapy' OR 'bucky ray' OR 'bucky ray radiation' OR 'bucky therapy' OR 'fractionated radiotherapy' OR 'hemibody irradiation' OR 'hypophysectomy, radiation' OR 'hypophysis irradiation' OR 'hypophysis radiation' OR 'irradiation therapy' OR 'irradiation treatment' OR 'irradiation, hypophysis' OR 'lymphatic irradiation' OR 'pituitary irradiation' OR 'radiation beam centration' OR 'radiation repair' OR 'radiation therapy' OR 'radiation treatment' OR 'radio therapy' OR 'radio treatment' OR 'radiohypophysectomy' OR 'radiology, therapeutic' OR 'radiotherapy' OR 'radiotherapy setup errors' OR 'radiotreatment' OR 'roentgen irradiation, therapeutic' OR 'roentgen therapy' OR 'roentgen treatment' OR 'rontgen therapy' OR 'therapeutic radiology' OR 'therapy, irradiation' OR 'therapy, radiation' OR 'therapy, roentgen' OR 'treatment, irradiation' OR 'treatment, radiation' OR 'treatment, roentgen' OR 'x radiotherapy' OR 'x ray therapy' OR 'x ray treatment' OR 'x-ray therapy') AND ('local control'/exp OR 'tumour control probability' OR 'normal tissue complication probability'/exp)

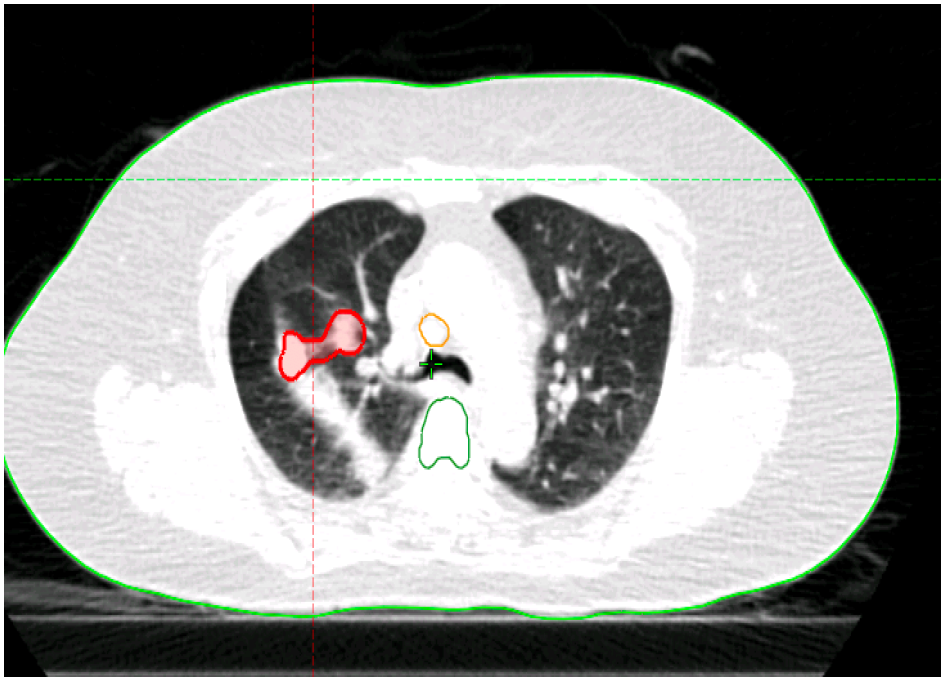
10.2 Appendices Chapter 4

10.2.1 Appendix 10.2.1

Examples of patient's excluded from analysis due to tumour within atelectasis or no associated PET-CT.

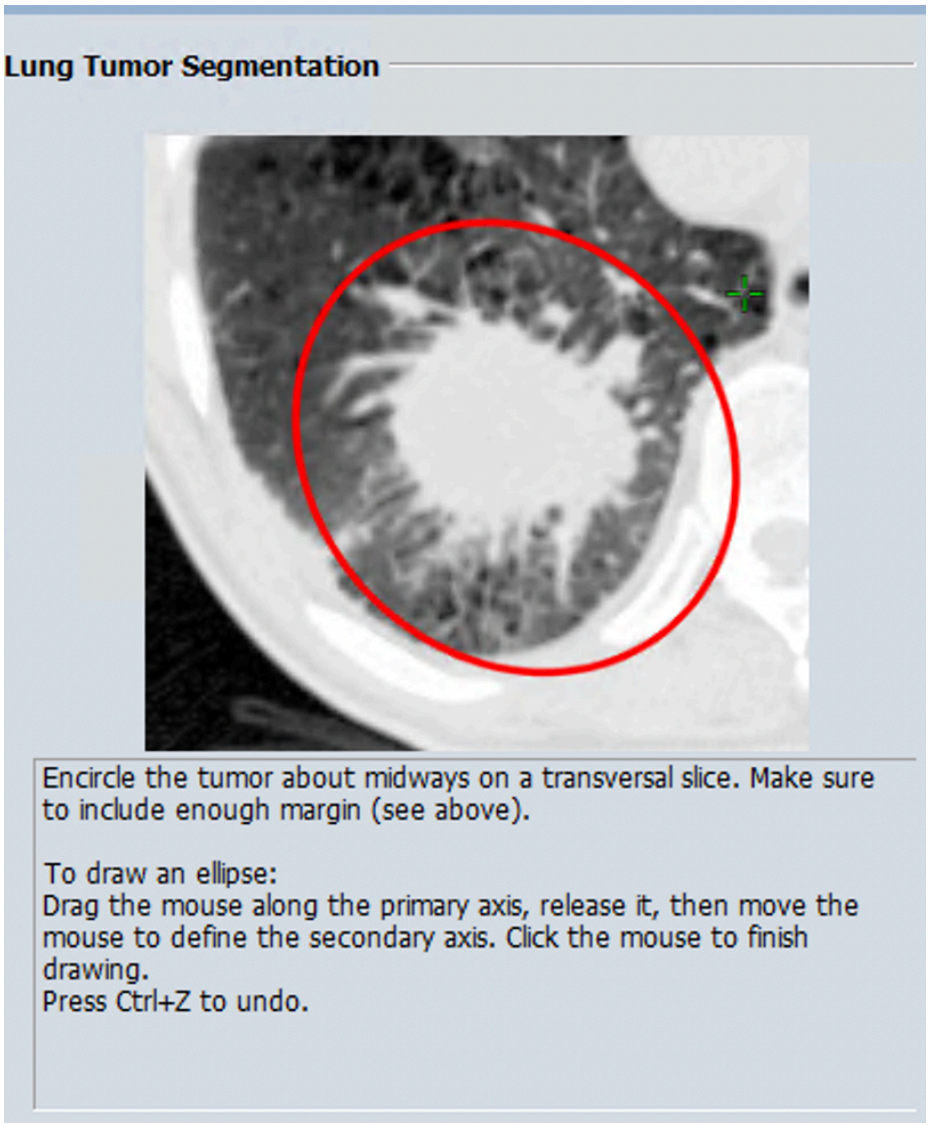






10.2.2 Appendix 10.2.2

Manufacturers instructs on Lung Tumour Segmentation



10.2.3 Appendix 10.2.3

Multilevel modelling

Observations 264

		Min Dose	
Predictors	*Estimates*	*CI*	*p*
MC	-2.82	-6.34 – 0.70	0.115
Abutting soft tissue	-11.54	-34.38 – 11.30	0.263

		Max Dose	
Predictors	*Estimates*	*CI*	*p*
MC	0.23	-0.00 – 0.46	0.054
Abutting soft tissue	1.22	-1.94 – 4.39	0.381

		Mean Dose	
Predictors	*Estimates*	*CI*	*p*
MC	-0.10	-0.33 – 0.12	0.365
Abutting soft tissue	-0.89	-3.18 – 1.39	0.375

		D 2%	
Predictors	*Estimates*	*CI*	*p*
MC	-0.03	-0.08 – 0.02	0.188
Abutting soft tissue	0.10	-2.47 – 2.67	0.928

		D 50%	
Predictors	*Estimates*	*CI*	*p*
MC	-0.06	-0.10 – -0.01	**0.009**
Abutting soft tissue	-0.48	-2.28 – 1.33	0.541

		D 95%	
Predictors	*Estimates*	*CI*	*p*
MC	-0.17	-1.55 – 1.21	0.807
Abutting soft tissue	-3.67	-10.83 – 3.50	0.257

		D 98%	
Predictors	*Estimates*	*CI*	*p*
MC	-0.42	-2.72 – 1.87	0.716
Abutting soft tissue	-5.40	-17.98 – 7.17	0.334
		V 107%	
Predictors	*Estimates*	*CI*	*p*
MC	0.14	0.03 – 0.24	**0.012**
Abutting soft tissue	0.72	-0.76 – 2.20	0.277

10.2.4 Appendix 10.2.4

Table 1 Clinical evaluation of plan acceptability

Table displays PTV plan dose metrics for each patient.

Plan 1 is the original plan and plans 2-8 are the dose metrics from the propagation of this plan onto subsequent CTs.

PTV_High (MC) denotes the physician defined contour, UA_PTV is the PTV generated from the UA-AC.

Pt ID	Structure:	Plan:	D95.0% [%]	PTV D95>90% (ideally 95%)	V107.0% [cm³]	1cc maximum <107%
107	PTV_High (MC)	1	94.5	TRUE	0.157	TRUE
107	PTV_High (MC)	2	94.9	TRUE	1.9456	FALSE
107	PTV_High (MC)	3	94	TRUE	0.47	TRUE
107	PTV_High (MC)	4	95.4	TRUE	1.351	FALSE
107	PTV_High (MC)	5	98.6	TRUE	3.4736	FALSE
107	UA_PTV	1	94.8	TRUE	0.1534	TRUE
107	UA_PTV	2	97.1	TRUE	1.96	FALSE
107	UA_PTV	3	96.3	TRUE	0.4797	TRUE
107	UA_PTV	4	98.6	TRUE	1.4917	FALSE
107	UA_PTV	5	99.5	TRUE	3.4005	FALSE
108	PTV_High (MC)	1	94.8	TRUE	0.0191	TRUE
108	PTV_High (MC)	2	64.8	FALSE	0.1241	TRUE
108	PTV_High (MC)	3	88.4	FALSE	0.0395	TRUE
108	PTV_High (MC)	4	72.2	FALSE	0.0195	TRUE
108	PTV_High (MC)	5	82.7	FALSE	0.3509	TRUE
108	UA_PTV	1	87.5	FALSE	0.0191	TRUE
108	UA_PTV	2	87.8	FALSE	0.1242	TRUE
108	UA_PTV	3	88.4	FALSE	0.0395	TRUE
108	UA_PTV	4	94.5	TRUE	0.0185	TRUE
108	UA_PTV	5	87	FALSE	0.3507	TRUE
110	PTV_High (MC)	1	91.8	TRUE	0.1347	TRUE
110	PTV_High (MC)	2	74.2	FALSE	0	TRUE
110	PTV_High (MC)	3	70.6	FALSE	0	TRUE
110	PTV_High (MC)	4	81.4	FALSE	0.0003	TRUE
110	PTV_High (MC)	5	86.7	FALSE	0	TRUE

110	UA_PTV	1	91.8	TRUE	0.1346	TRUE
110	UA_PTV	2	76.1	FALSE	0	TRUE
110	UA_PTV	3	84.5	FALSE	0	TRUE
110	UA_PTV	4	66.7	FALSE	0.0003	TRUE
110	UA_PTV	5	79.7	FALSE	0	TRUE
112	PTV_High (MC)	1	91.8	TRUE	0.2225	TRUE
112	PTV_High (MC)	2	89.7	FALSE	0.0286	TRUE
112	PTV_High (MC)	3	92.9	TRUE	0.1033	TRUE
112	PTV_High (MC)	4	91.6	TRUE	0.0008	TRUE
112	PTV_High (MC)	5	91.4	TRUE	0.0479	TRUE
112	PTV_High (MC)	6	88.3	FALSE	0.0503	TRUE
112	UA_PTV	1	81.5	FALSE	0.0077	TRUE
112	UA_PTV	2	88.4	FALSE	0	TRUE
112	UA_PTV	3	92.9	TRUE	0.0002	TRUE
112	UA_PTV	4	91.9	TRUE	0.0014	TRUE
112	UA_PTV	5	93.1	TRUE	0.025	TRUE
112	UA_PTV	6	91.8	TRUE	0.0014	TRUE
113	PTV_High (MC)	1	95	TRUE	0.5772	TRUE
113	PTV_High (MC)	2	82.1	FALSE	0.6102	TRUE
113	PTV_High (MC)	3	80.4	FALSE	0.5699	TRUE
113	PTV_High (MC)	4	73.2	FALSE	0.8353	TRUE
113	PTV_High (MC)	5	83.2	FALSE	0.3449	TRUE
113	UA_PTV	1	59.5	FALSE	0.5033	TRUE
113	UA_PTV	2	65.7	FALSE	0.455	TRUE
113	UA_PTV	3	78.3	FALSE	0.4214	TRUE
113	UA_PTV	4	26.2	FALSE	0.8409	TRUE
113	UA_PTV	5	77.2	FALSE	0.3099	TRUE
114	PTV_High (MC)	1	94.2	TRUE	0.5715	TRUE
114	PTV_High (MC)	2	80.5	FALSE	0.8284	TRUE
114	PTV_High (MC)	3	69	FALSE	0.8118	TRUE
114	PTV_High (MC)	4	70.5	FALSE	6.9277	FALSE
114	PTV_High (MC)	5	64.3	FALSE	13.3146	FALSE
114	UA_PTV	1	93.6	TRUE	0	TRUE
114	UA_PTV	2	97.4	TRUE	0.0236	TRUE
114	UA_PTV	3	96	TRUE	0.0136	TRUE
114	UA_PTV	4	98	TRUE	4.3142	FALSE
114	UA_PTV	5	97.8	TRUE	8.7472	FALSE

115	PTV_High (MC)	1	96.7	TRUE	0.1848	TRUE
115	PTV_High (MC)	2	87	FALSE	0.0504	TRUE
115	PTV_High (MC)	3	93.8	TRUE	0.121	TRUE
115	PTV_High (MC)	4	92	TRUE	0.1175	TRUE
115	PTV_High (MC)	5	95.8	TRUE	0.0431	TRUE
115	UA_PTV	1	96.4	TRUE	0.1511	TRUE
115	UA_PTV	2	93	TRUE	0.0437	TRUE
115	UA_PTV	3	93.5	TRUE	0.1168	TRUE
115	UA_PTV	4	94.5	TRUE	0.1176	TRUE
115	UA_PTV	5	97.6	TRUE	0.0144	TRUE
117	PTV_High (MC)	1	97	TRUE	0.0948	TRUE
117	PTV_High (MC)	2	96.4	TRUE	0.0235	TRUE
117	PTV_High (MC)	3	95.7	TRUE	0.0146	TRUE
117	PTV_High (MC)	4	98.8	TRUE	0.4906	TRUE
117	PTV_High (MC)	5	98.7	TRUE	0.1676	TRUE
117	PTV_High (MC)	6	99.4	TRUE	2.0797	FALSE
117	PTV_High (MC)	7	99.2	TRUE	1.2341	FALSE
117	PTV_High (MC)	8	98.5	TRUE	0.6404	TRUE
117	UA_PTV	1	94.7	TRUE	0.0946	TRUE
117	UA_PTV	2	93.9	TRUE	0.0235	TRUE
117	UA_PTV	3	96.4	TRUE	0.4145	TRUE
117	UA_PTV	4	98.6	TRUE	0.81	TRUE
117	UA_PTV	5	96.8	TRUE	0.213	TRUE
117	UA_PTV	6	98.7	TRUE	2.3625	FALSE
117	UA_PTV	7	98.8	TRUE	1.0607	FALSE
117	UA_PTV	8	97.8	TRUE	0.6392	TRUE

Table 2. Raw data of dosimetric metrics analysed in multilevel modelling

Table displays PTV plan dose metrics for each patient.

Plan 1 is the original plan and plans 2-8 are the dose metrics from the propagation of this plan onto subsequent CTs.

Tumor_c50 denotes the physician defined contour (MC), CTV_High(MC) and PTV_High(MC) are the planning volumes generated from the MC.

UA_GTV is UA-AC, UA_CTV and UA_PTV are the planning volumes generated from the UA_AC

Pt ID	Structure:	Plan:	Min Dose [%]:	Max Dose [%]:	Mean Dose [%]:	D2.0% [%]	D50.0% [%]	D95.0% [%]	D98.0% [%]	V107.0% [cm³]
107	Tumor_c50	1	95.6	107.9	100.4	104	100.4	97.7	97.1	0.0051
107	Tumor_c50	2	96.5	109.8	102.9	106.5	102.8	100	99.4	0.2177
107	Tumor_c50	3	96	107.7	101.5	104.8	101.4	99.1	98.5	0.0017
107	Tumor_c50	4	97.8	107.8	103.3	106.4	103.3	100.4	99.9	0.0235
107	Tumor_c50	5	99.3	108.8	104.6	107.7	104.8	101.8	101.2	0.4506
107	UA_GTV	1	95.6	107.5	100.5	104.2	100.5	97.8	97.2	0.0022
107	UA_GTV	2	96.5	109.8	103.1	107.1	103	100.3	99.5	0.3287
107	UA_GTV	3	95.6	107	101.5	104.9	101.5	99	98	0
107	UA_GTV	4	98.1	107.8	103.5	106.5	103.5	101.1	100.5	0.0249
107	UA_GTV	5	98.5	108.8	104.7	107.7	104.8	101.7	101	0.4725
107	CTV_High (MC)	1	94.4	108.7	100	104.4	100	97.2	96.6	0.0677
107	CTV_High (MC)	2	95.6	110.7	102.4	107.1	102.2	99.4	98.9	0.9179

107	CTV_High (MC)	3	92.2	108.6	100.9	105.3	100.8	98	97.2	0.1525
107	CTV_High (MC)	4	96.5	109.1	102.5	106.6	102.4	99.6	98.9	0.388
107	CTV_High (MC)	5	98.2	110.5	103.8	108	103.7	100.9	100.4	1.4257
107	UA_CTV	1	94.2	108.7	100	104.5	100	97.2	96.5	0.0638
107	UA_CTV	2	96.5	110.7	102.5	107.3	102.3	99.5	99	0.9613
107	UA_CTV	3	94.8	108.9	100.9	105.2	100.8	98	97.4	0.1116
107	UA_CTV	4	97	110.1	102.7	106.8	102.6	99.7	99.2	0.3893
107	UA_CTV	5	98.2	110.3	103.7	107.8	103.6	100.8	100.2	1.1095
107	PTV_High (MC)	1	77.2	109.5	99	104	99.3	94.5	93	0.157
107	PTV_High (MC)	2	68.2	111.9	101	106.8	101.4	94.9	90.7	1.9456
107	PTV_High (MC)	3	67.3	109.8	99.6	105.1	100.1	94	88.6	0.47
107	PTV_High (MC)	4	73.5	111	101.3	106.7	101.6	95.4	90.7	1.351
107	PTV_High (MC)	5	77.2	112	102.8	108.1	103	98.6	95.5	3.4736
107	UA_PTV	1	37.5	109.5	99	104	99.4	94.8	92.8	0.1534
107	UA_PTV	2	75.9	111.9	101.5	106.9	101.7	97.1	95.2	1.96
107	UA_PTV	3	74.9	109.8	100.3	105.3	100.3	96.3	94.6	0.4797
107	UA_PTV	4	87.5	111.6	102.1	107	101.9	98.6	97.4	1.4917
107	UA_PTV	5	86.4	112	103.2	108.2	103.1	99.5	98.4	3.4005
108	Tumor_c50	1	97.1	104	100.5	102.8	100.5	98.8	98.5	0

108	Tumor_c50	2	97.8	109.3	101.6	104.6	101.5	99.5	99.1	0.0166
108	Tumor_c50	3	96.3	105.7	100.9	103.6	100.9	98.6	98.1	0
108	Tumor_c50	4	97.7	107.4	102.1	105	102.1	99.8	99.3	0.0011
108	Tumor_c50	5	98.5	108.9	102.8	106.3	102.8	100.5	100.1	0.0774
108	UA_GTV	1	97.1	104.7	100.4	102.8	100.4	98.5	98.2	0
108	UA_GTV	2	98.1	106.9	101.9	104.6	101.8	99.8	99.4	0
108	UA_GTV	3	97.1	105.7	101	103.7	101.1	98.9	98.5	0
108	UA_GTV	4	97.7	107.4	101.8	104.6	101.8	99.7	99.2	0.0011
108	UA_GTV	5	98.5	108.9	102.7	105.9	102.6	100.4	100	0.0559
108	CTV_High (MC)	1	96.8	105	100	102.5	100	98.3	97.9	0
108	CTV_High (MC)	2	74.5	109.3	101.2	104.7	101.3	98.6	97.1	0.1236
108	CTV_High (MC)	3	92.4	105.7	100.5	103.3	100.4	98.4	98.1	0
108	CTV_High (MC)	4	76	108.9	100.9	104.4	101	98.6	97.2	0.0164
108	CTV_High (MC)	5	91.8	110.1	102.1	105.8	102	99.8	99.4	0.2951
108	UA_CTV	1	96.5	106	100	102.5	99.9	98.2	97.8	0
108	UA_CTV	2	96.9	106.9	101.4	104.3	101.4	99.4	99	0
108	UA_CTV	3	91.9	105.9	100.5	103.4	100.5	98.4	97.9	0
108	UA_CTV	4	97.1	107.4	101.1	104.2	101	99	98.6	0.0011
108	UA_CTV	5	95.8	110.1	102.1	105.5	102	99.9	99.5	0.1717

108	PTV_High (MC)	1	83.4	108.4	99	102.5	99.4	94.8	93.1	0.0191
108	PTV_High (MC)	2	12.6	109.3	95	104.2	100.1	64.8	38.4	0.1241
108	PTV_High (MC)	3	57.9	108.9	98.2	103.5	99.6	88.4	81.7	0.0395
108	PTV_High (MC)	4	12.6	108.9	96.6	104.2	100.2	72.2	39.5	0.0195
108	PTV_High (MC)	5	19	110.1	98.5	105.3	101.1	82.7	64.3	0.3509
108	UA_PTV	1	52.1	108.4	97.4	102.4	99	87.5	78.9	0.0191
108	UA_PTV	2	49.6	109.3	98.9	104.3	100.5	87.8	77.6	0.1242
108	UA_PTV	3	56.1	108.9	98.6	103.5	99.8	88.4	81.8	0.0395
108	UA_PTV	4	47.8	108.9	99.9	104.2	100.5	94.5	89.4	0.0185
108	UA_PTV	5	37.4	110.1	99.5	105.3	101.2	87	76.1	0.3507
110	Tumor_c50	1	96	103.9	100.2	102.5	100.2	98	97.6	0
110	Tumor_c50	2	65	104.4	100	102.9	100.1	97.4	96.8	0
110	Tumor_c50	3	63	104.5	100.1	102.9	100.1	97.7	97.1	0
110	Tumor_c50	4	96.5	105.9	101.1	104.7	100.9	98.7	98	0
110	Tumor_c50	5	94.3	105.5	100.5	104.1	100.4	98	97	0
110	UA_GTV	1	95.6	104	100	102.5	100.1	97.6	97	0
110	UA_GTV	2	92.4	103.8	99.9	102.6	99.9	97.2	96.6	0
110	UA_GTV	3	94.9	104.3	99.7	102.6	99.7	97	96.5	0
110	UA_GTV	4	96.3	105.9	101.2	104.7	101	98.4	97.7	0

110	UA_GTV	5	94.7	105.5	100.6	104.2	100.4	98.2	97.5	0
110	CTV_High (MC)	1	95.1	104.3	100	102.6	100.1	97.6	97.1	0
110	CTV_High (MC)	2	55.8	105	99	103	99.6	95.6	88.4	0
110	CTV_High (MC)	3	52.7	105	99.2	102.9	99.7	96.4	89.9	0
110	CTV_High (MC)	4	79.4	107	100.7	104.5	100.7	97.6	96.7	0
110	CTV_High (MC)	5	91.5	106.5	100.3	104.2	100.3	96.7	95.4	0
110	UA_CTV	1	94.4	105.9	99.9	102.6	100	97.4	96.8	0
110	UA_CTV	2	58	106.2	99.2	102.8	99.6	96.1	94.2	0
110	UA_CTV	3	79.1	104.3	99.2	102.5	99.4	96.2	95.5	0
110	UA_CTV	4	73.2	107	100.4	104.5	100.6	97	95.4	0
110	UA_CTV	5	87.3	106	99.8	103.9	100	96.1	94.7	0
110	PTV_High (MC)	1	51.8	111.7	98.3	102.7	99.1	91.8	87.3	0.1347
110	PTV_High (MC)	2	42.2	106.2	95.4	103	98.5	74.2	60.2	0
110	PTV_High (MC)	3	25.8	105	95.1	102.6	98.3	70.6	55.4	0
110	PTV_High (MC)	4	14.4	107.1	97.5	104.7	100	81.4	49.5	0.0003
110	PTV_High (MC)	5	27.1	106.7	97.8	104.3	99.5	86.7	71.1	0
110	UA_PTV	1	26.5	111.7	98.2	102.7	99.4	91.8	81.9	0.1346
110	UA_PTV	2	13.6	106.2	95.8	102.9	98.9	76.1	43.2	0
110	UA_PTV	3	17.5	104.8	96.5	102.6	98.7	84.5	57.8	0

110	UA_PTV	4	12.6	107.1	96.2	104.6	99.9	66.7	36.1	0.0003
110	UA_PTV	5	23.9	106.7	96.6	104	99.1	79.7	57.8	0
112	Tumor_c50	1	95.4	104.5	100.6	103	100.6	98.8	98.1	0
112	Tumor_c50	2	94.8	104.1	100.5	102.6	100.5	98.8	98.3	0
112	Tumor_c50	3	96.9	104.4	100.7	102.7	100.7	99.4	99	0
112	Tumor_c50	4	97.2	104.1	100.7	102.7	100.7	99.2	98.8	0
112	Tumor_c50	5	96.4	105.3	101.4	103.6	101.3	99.9	99.5	0
112	Tumor_c50	6	95.6	104.2	100.7	102.6	100.7	99.1	98.6	0
112	UA_GTV	1	90.1	104.5	100.2	102.8	100.5	97.2	95.3	0
112	UA_GTV	2	94.9	103.6	100.4	102.4	100.4	98.5	97.8	0
112	UA_GTV	3	97.6	104.1	100.7	102.5	100.7	99.4	98.9	0
112	UA_GTV	4	96.8	103.9	100.7	102.6	100.7	99.2	98.7	0
112	UA_GTV	5	97.7	105.3	101.4	103.5	101.3	99.9	99.6	0
112	UA_GTV	6	95.8	103.9	100.6	102.4	100.6	99.1	98.7	0
112	CTV_High (MC)	1	91	105.1	100	103.1	100.4	96.2	95	0
112	CTV_High (MC)	2	92	104.8	99.8	102.8	100.1	96.2	95.1	0
112	CTV_High (MC)	3	93.6	104.4	100.2	102.8	100.4	97.2	96.3	0
112	CTV_High (MC)	4	92.8	104.6	100.1	102.8	100.4	97.2	96.1	0
112	CTV_High (MC)	5	93	107	100.9	103.8	101	97.7	96.4	0

112	CTV_High (MC)	6	89.4	107.1	100.2	102.7	100.4	97.3	95.1	0.0001
112	UA_CTV	1	71.9	105.3	99.3	103	100.1	94	91.7	0
112	UA_CTV	2	91.1	105	99.6	102.8	100	95.6	94.4	0
112	UA_CTV	3	93.6	104.4	100.2	102.8	100.4	97.3	96.4	0
112	UA_CTV	4	93.1	104.6	100.1	102.8	100.4	97.1	96.1	0
112	UA_CTV	5	93.2	106.1	100.9	103.8	101	97.9	96.7	0
112	UA_CTV	6	92.2	104.4	100.1	102.6	100.3	97.2	95.9	0
112	PTV_High (MC)	1	75.7	112.7	98.9	103.5	99.9	91.8	89.8	0.2225
112	PTV_High (MC)	2	57.2	110.5	98.1	103.3	99.6	89.7	83.9	0.0286
112	PTV_High (MC)	3	78	111.8	99.3	103.4	100.1	92.9	91.2	0.1033
112	PTV_High (MC)	4	63.7	107.5	98.8	103.1	99.9	91.6	88.8	0.0008
112	PTV_High (MC)	5	56.9	109.3	99.3	104.2	100.7	91.4	86.9	0.0479
112	PTV_High (MC)	6	53.3	109.2	98.2	103.2	99.8	88.3	80.7	0.0503
112	UA_PTV	1	13.1	108.4	96.2	103.3	99.6	81.5	51.9	0.0077
112	UA_PTV	2	60	106.2	97.9	103.1	99.5	88.4	81.9	0
112	UA_PTV	3	71.3	107.4	99.2	103.3	100.1	92.9	91	0.0002
112	UA_PTV	4	66.3	108	98.9	103.2	99.9	91.9	89.3	0.0014
112	UA_PTV	5	72.4	108.4	99.9	104.2	100.8	93.1	91.2	0.025
112	UA_PTV	6	74.6	107.3	99	103	99.9	91.8	89.3	0.0014

113	Tumor_c50	1	94.9	107.5	99.7	103.1	99.6	97.4	97	0.0015
113	Tumor_c50	2	89.7	108.2	100	104	99.8	97.4	96.8	0.0245
113	Tumor_c50	3	95.3	107.5	100.2	104	100.1	97.8	97.3	0.0023
113	Tumor_c50	4	85.3	110.7	98.9	104	98.6	96.4	95.9	0.0907
113	Tumor_c50	5	94.1	105.7	99.1	102.4	98.9	96.9	96.5	0
113	UA_GTV	1	90.7	107.5	99.9	103.8	99.7	97.3	96.9	0.0029
113	UA_GTV	2	87.8	108.5	100.1	104.3	100	97.4	96.9	0.0325
113	UA_GTV	3	95	107.9	100.2	103.9	100.1	97.7	97.2	0.0083
113	UA_GTV	4	23	110.6	98.1	104.8	98.7	95.3	75.7	0.2369
113	UA_GTV	5	94.1	107.3	99.3	103.9	99.1	96.9	96.5	0.0027
113	CTV_High (MC)	1	94.8	107.5	100	103.8	99.9	97.3	96.8	0.0124
113	CTV_High (MC)	2	80.2	108.2	99.9	104.1	100	96.7	95.5	0.0507
113	CTV_High (MC)	3	87.6	112.8	100.3	104.2	100.3	97.4	96.8	0.2586
113	CTV_High (MC)	4	76	110.7	99	104.3	99	95.8	93.6	0.2927
113	CTV_High (MC)	5	83.1	108	99.5	103.9	99.3	96.6	95.9	0.0376
113	UA_CTV	1	58.9	113.3	99.8	104	99.9	96.8	93.8	0.4115
113	UA_CTV	2	37.8	112.2	99.9	104.7	100.1	96.9	95.6	0.4528
113	UA_CTV	3	89	112.8	100.4	104.4	100.3	97.5	96.9	0.3528
113	UA_CTV	4	9.2	112.1	95.6	104.6	98.8	65.9	36.2	0.6686

113	UA_CTV	5	80.9	111.9	99.5	104.4	99.4	96.6	96	0.2695
113	PTV_High (MC)	1	56	113.3	99.3	104	99.7	95	87.4	0.5772
113	PTV_High (MC)	2	35.3	112.2	97.1	104	99.2	82.1	74.7	0.6102
113	PTV_High (MC)	3	23.5	112.8	97.2	104.1	99.6	80.4	67.1	0.5699
113	PTV_High (MC)	4	18.3	112.1	95.3	104	98.4	73.2	60.1	0.8353
113	PTV_High (MC)	5	33.7	111.9	97.2	104	98.8	83.2	72.4	0.3449
113	UA_PTV	1	15.4	113.3	94.4	103.7	99.2	59.5	45.5	0.5033
113	UA_PTV	2	9.5	112.2	95.1	104	99.3	65.7	38.7	0.455
113	UA_PTV	3	23.7	112.8	97.2	104.1	99.7	78.3	57.8	0.4214
113	UA_PTV	4	4.9	112.1	88.6	103.8	97.9	26.2	9	0.8409
113	UA_PTV	5	23.4	111.9	96.6	103.9	98.9	77.2	59.6	0.3099
114	Tumor_c50	1	94.9	105.3	100	103	100	97.6	97.1	0
114	Tumor_c50	2	92.3	109.7	101.7	105.6	101.6	98.9	98.4	0.3168
114	Tumor_c50	3	62.7	107.4	100.9	105.1	101	97.7	96.4	0.0063
114	Tumor_c50	4	80.3	110	102.4	107.3	102.3	98.9	98.2	3.4299
114	Tumor_c50	5	54	111	102.7	108	102.8	98.9	97.3	6.4424
114	UA_GTV	1	94.9	105.8	100.1	103.2	100.1	97.8	97.3	0
114	UA_GTV	2	96.5	107.5	101.9	105.5	101.8	99.3	98.8	0.0036
114	UA_GTV	3	95.4	107.3	101.4	105.4	101.3	98.8	98.3	0.0041

114	UA_GTV	4	97.7	110	103.1	107.6	102.9	99.9	99.3	2.2411
114	UA_GTV	5	97.6	111	103.4	108.4	103.1	100.2	99.7	3.8131
114	CTV_High (MC)	1	88.9	108.8	100	103.2	100	97.5	96.9	0.0832
114	CTV_High (MC)	2	44.2	109.7	100.9	105.3	101.1	97.9	96.4	0.6542
114	CTV_High (MC)	3	48.2	110.9	100	105	100.5	94.8	87.8	0.3212
114	CTV_High (MC)	4	32.9	111.7	101.1	107	101.6	96.9	88.4	4.7616
114	CTV_High (MC)	5	49.3	111	100.9	107.7	102	90.6	78.1	8.8998
114	UA_CTV	1	92.9	106.2	100.1	103.2	100	97.7	97.2	0
114	UA_CTV	2	95.9	107.7	101.4	105.2	101.2	98.5	98	0.0233
114	UA_CTV	3	94.7	107.4	100.9	105.1	100.7	97.8	97	0.0085
114	UA_CTV	4	95.4	110	102.4	107.2	102.3	99.1	98.4	3.2996
114	UA_CTV	5	97	111	102.7	107.9	102.4	99.1	98.6	5.6056
114	PTV_High (MC)	1	31.5	110.4	98.7	103.5	99.6	94.2	84.5	0.5715
114	PTV_High (MC)	2	9.5	110.3	97.7	105	100.3	80.5	48.6	0.8284
114	PTV_High (MC)	3	12	110.9	95.5	104.9	99.4	69	50	0.8118
114	PTV_High (MC)	4	7.4	111.7	96.6	106.7	100.4	70.5	35	6.9277
114	PTV_High (MC)	5	28.3	111.2	96.1	107.4	100.3	64.3	54.7	13.3146
114	UA_PTV	1	58.2	106.3	99.1	103.2	99.6	93.6	88.5	0
114	UA_PTV	2	79.6	107.7	100.7	104.9	100.7	97.4	96.3	0.0236

114	UA_PTV	3	76.2	107.6	100	104.8	100.1	96	94.2	0.0136
114	UA_PTV	4	88.2	110	101.7	106.9	101.5	98	97.2	4.3142
114	UA_PTV	5	89.9	111	101.8	107.6	101.5	97.8	96.9	8.7472
115	Tumor_c50	1	97	104.3	99.9	102.4	99.9	98.1	97.8	0
115	Tumor_c50	2	95.5	103.2	98.6	101.3	98.5	96.6	96.4	0
115	Tumor_c50	3	96.3	103.7	99.3	101.6	99.2	97.6	97.4	0
115	Tumor_c50	4	96.8	103.7	99.5	102.2	99.4	97.8	97.6	0
115	Tumor_c50	5	97	102.5	99.4	101.3	99.3	98.2	97.9	0
115	UA_GTV	1	97	104.3	99.9	102.4	99.8	98.1	97.8	0
115	UA_GTV	2	95.8	103.2	98.7	101.5	98.5	96.9	96.5	0
115	UA_GTV	3	96.7	103.7	99.4	101.8	99.3	97.7	97.5	0
115	UA_GTV	4	97.1	103.7	99.6	102.2	99.5	98	97.7	0
115	UA_GTV	5	97.2	102.9	99.3	101.4	99.2	98.2	98	0
115	CTV_High (MC)	1	96.4	104.3	100	102.4	100.1	97.9	97.5	0
115	CTV_High (MC)	2	94.3	103.2	98.9	101.3	98.9	96.7	96.3	0
115	CTV_High (MC)	3	96.2	103.7	99.7	102.2	99.7	97.6	97.2	0
115	CTV_High (MC)	4	96.3	103.9	99.9	102.5	99.9	97.8	97.5	0
115	CTV_High (MC)	5	96.3	103.9	99.6	102.3	99.6	97.8	97.4	0
115	UA_CTV	1	96.4	104.3	99.9	102.3	100	97.8	97.5	0

115	UA_CTV	2	95.4	103.2	98.7	101.3	98.8	96.6	96.2	0
115	UA_CTV	3	96.2	103.7	99.6	102	99.7	97.6	97.2	0
115	UA_CTV	4	96.3	103.9	99.8	102.4	99.8	97.8	97.4	0
115	UA_CTV	5	96.3	103.6	99.6	101.9	99.5	97.9	97.6	0
115	PTV_High (MC)	1	82.7	112.9	100	103.4	100.1	96.7	95.1	0.1848
115	PTV_High (MC)	2	65.7	109.6	97.2	101.9	98.5	87	81.8	0.0504
115	PTV_High (MC)	3	72.8	110.1	99.2	102.9	99.7	93.8	89.8	0.121
115	PTV_High (MC)	4	67.2	113.1	98.9	102.7	99.6	92	88	0.1175
115	PTV_High (MC)	5	74.7	110.7	99.6	103.2	99.9	95.8	91.9	0.0431
115	UA_PTV	1	83.5	112.9	99.9	103.3	100.2	96.4	94.4	0.1511
115	UA_PTV	2	79.9	109.6	98.3	101.8	98.8	93	89.3	0.0437
115	UA_PTV	3	78.2	110.1	99.2	102.9	99.7	93.5	90.5	0.1168
115	UA_PTV	4	75	113.1	99.4	102.8	99.7	94.5	91.5	0.1176
115	UA_PTV	5	91.8	109.1	100	103.1	100	97.6	96.9	0.0144
117	Tumor_c50	1	96.5	104.1	100.2	102.3	100.2	98.7	98.3	0
117	Tumor_c50	2	96.9	104.4	100.4	102.6	100.4	98.6	98.2	0
117	Tumor_c50	3	98.7	106.5	102.1	104.5	101.9	100	99.6	0
117	Tumor_c50	4	98.5	106.2	102.7	104.7	102.7	100.7	100.2	0
117	Tumor_c50	5	99.4	105.7	102.5	104.5	102.5	100.8	100.5	0

117	Tumor_c50	6	99.8	107	103.3	105.4	103.2	101.4	101	0
117	Tumor_c50	7	99.8	106.3	103	105	103	101.2	100.7	0
117	Tumor_c50	8	100	106.3	102.9	105	102.8	101	100.6	0
117	UA_GTV	1	96.7	104.1	100.2	102.3	100.2	98.6	98.2	0
117	UA_GTV	2	96.9	104.4	100.4	102.6	100.4	98.6	98.2	0
117	UA_GTV	3	98.7	106.6	102.3	104.6	102.2	100.3	99.9	0
117	UA_GTV	4	98.7	106.5	102.8	105	102.9	100.9	100.4	0
117	UA_GTV	5	99.4	105.1	102.6	104.5	102.5	100.9	100.6	0
117	UA_GTV	6	100.2	106.2	103.4	105.2	103.4	101.7	101.3	0
117	UA_GTV	7	99.9	105.9	103	105	102.9	101.2	100.7	0
117	UA_GTV	8	100.2	105.6	102.9	105	102.8	101.4	100.9	0
117	CTV_High (MC)	1	95.6	104.1	100	102.2	100	98.2	97.9	0
117	CTV_High (MC)	2	96.1	104.4	100	102.3	99.9	98.1	97.8	0
117	CTV_High (MC)	3	96.2	106.8	101.3	104.4	101.3	98.7	98.2	0
117	CTV_High (MC)	4	98.1	107.5	102.3	105.2	102.2	99.9	99.5	0.0093
117	CTV_High (MC)	5	98.3	107.1	102.2	104.7	102.2	100.2	99.7	0.0002
117	CTV_High (MC)	6	98.8	108.8	103.1	106.4	103	100.5	100	0.1051
117	CTV_High (MC)	7	98.2	107.8	102.7	105.5	102.7	100.4	99.9	0.0129
117	CTV_High (MC)	8	98.1	107.6	102.7	105.3	102.6	100.5	100.1	0.0102

117	UA_CTV	1	95.6	104.1	100.1	102.3	100	98.2	97.9	0
117	UA_CTV	2	96.1	104.4	100	102.3	100	98.1	97.7	0
117	UA_CTV	3	97	106.8	101.6	104.7	101.5	99	98.5	0
117	UA_CTV	4	98.1	107.5	102.3	105.2	102.2	100	99.5	0.0066
117	UA_CTV	5	98.3	107.1	102.1	104.6	102.1	100.1	99.7	0.0003
117	UA_CTV	6	98.3	108.8	103	106.2	102.9	100.6	100.1	0.0675
117	UA_CTV	7	98.2	107.1	102.7	105.4	102.7	100.3	99.8	0.0001
117	UA_CTV	8	98.1	107.6	102.7	105.3	102.6	100.5	100	0.0092
117	PTV_High (MC)	1	81.9	112.3	100	103.2	100.1	97	95.7	0.0948
117	PTV_High (MC)	2	84.3	109.4	99.8	103.5	99.9	96.4	94.7	0.0235
117	PTV_High (MC)	3	71.8	108.1	100.3	104.9	100.6	95.7	93.2	0.0146
117	PTV_High (MC)	4	93.5	108.9	101.8	106.3	101.6	98.8	98	0.4906
117	PTV_High (MC)	5	84.2	110.9	101.7	105.7	101.8	98.7	97.1	0.1676
117	PTV_High (MC)	6	91	110.1	102.6	107.7	102.5	99.4	98.5	2.0797
117	PTV_High (MC)	7	84.1	109.6	102.4	107.2	102.3	99.2	97.9	1.2341
117	PTV_High (MC)	8	70.4	108.9	101.9	106.7	102.1	98.5	91.8	0.6404
117	UA_PTV	1	75.8	112.3	99.6	103.2	100	94.7	92	0.0946
117	UA_PTV	2	75.6	109.4	99.3	103.5	99.8	93.9	91	0.0235
117	UA_PTV	3	80.7	109.7	100.8	106.1	100.9	96.4	94	0.4145

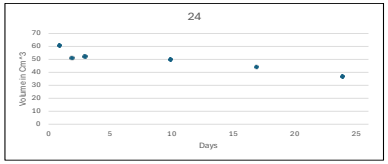
117	UA_PTV	4	91.1	110.6	101.8	106.7	101.7	98.6	97.8	0.81
117	UA_PTV	5	48.6	110.9	101.1	105.8	101.6	96.8	91.5	0.213
117	UA_PTV	6	83.8	110.3	102.5	107.9	102.4	98.7	96.3	2.3625
117	UA_PTV	7	80.6	109.6	102.2	107.2	102.2	98.8	95.8	1.0607
117	UA_PTV	8	70.2	108.9	101.8	106.7	102.1	97.8	91.5	0.6392

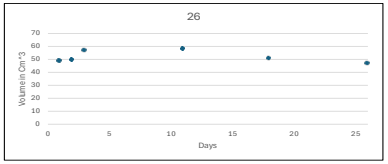
10.3 Appendices Chapter 5

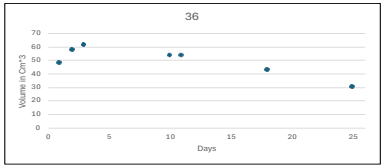
10.3.1 Appendix 10.3.1

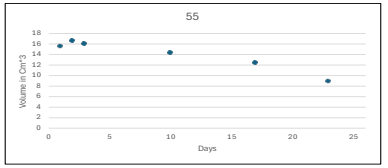
Individual plots of absolute tumour volume changes over the course of RT

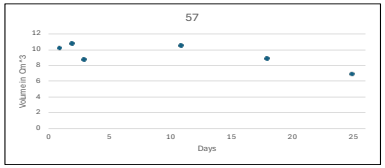
X axis displays the time in days with Day 1 being the start of RT. Y axis displays the absolute tumour volume in cm^3. Plot titles are pseudonymised patient IDs

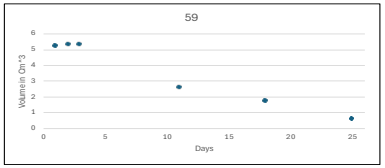

24
Volume in Cm^3
Days

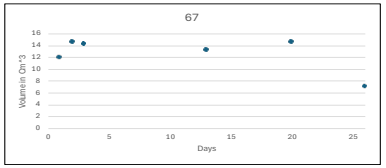

26
Volume in Cm^3
Days

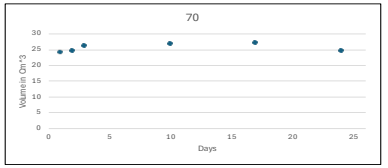

36
Volume in Cm^3
Days

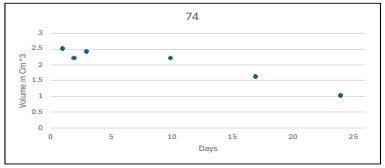

55
Volume in Cm^3
Days

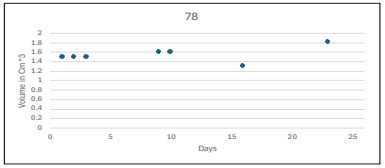

57
Volume in Cm^3
Days

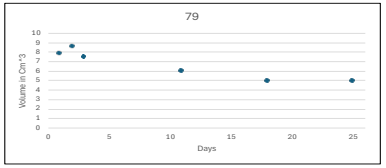

59
Volume in Cm^3
Days

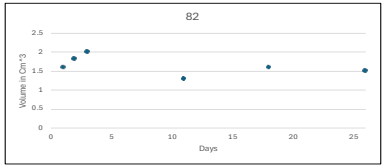

67
Volume in Cm^3
Days

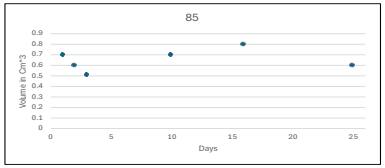

70
Volume in Cm^3
Days

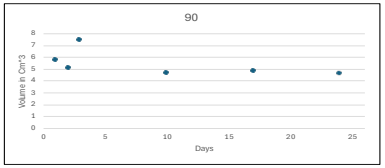

74
Volume in Cm^3
Days

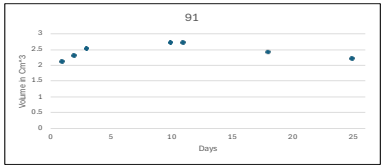

78
Volume in Cm^3
Days

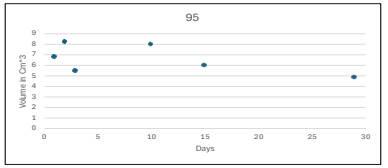

79
Volume in Cm^3
Days

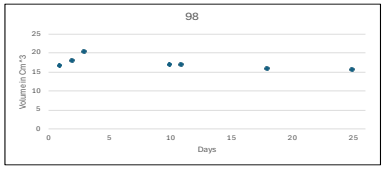

82
Volume in Cm^3
Days

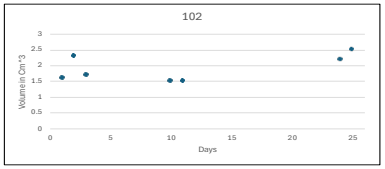

85
Volume in Cm^3
Days

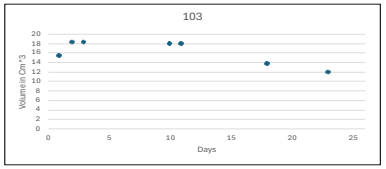

90
Volume in Cm^3
Days

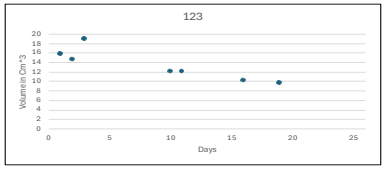

91
Volume in Cm^3
Days

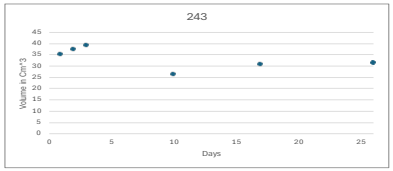

95
Volume in Cm^3
Days

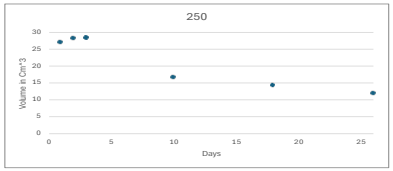

98
Volume in Cm^3
Days

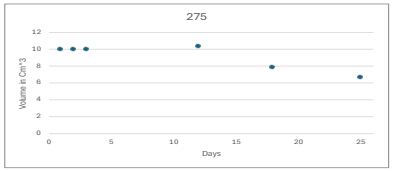

102
Volume in Cm^3
Days

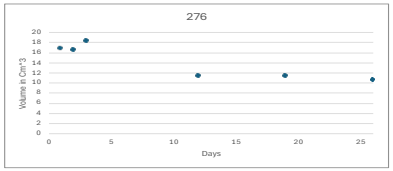

103
Volume in Cm^3
Days

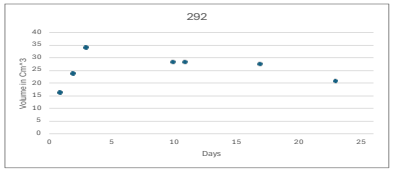

123
Volume in Cm^3
Days

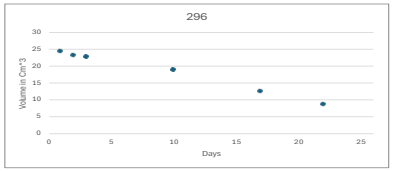

243
Volume in Cm^3
Days

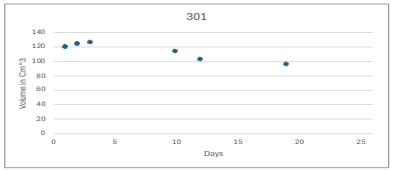

250
Volume in Cm^3
Days

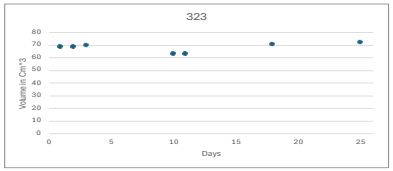

275
Volume in Cm^3
Days

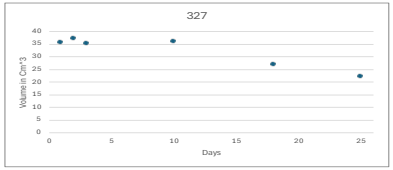

276
Volume in Cm^3
Days

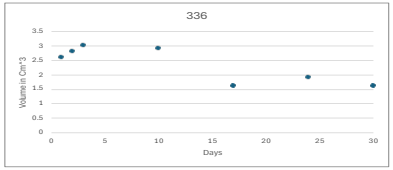

292
Volume in Cm^3
Days

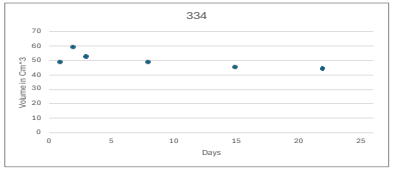

296
Volume in Cm^3
Days

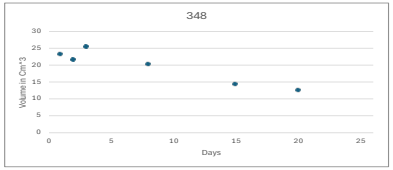

301
Volume in Cm^3
Days

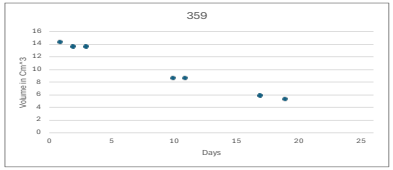

323
Volume in Cm^3
Days

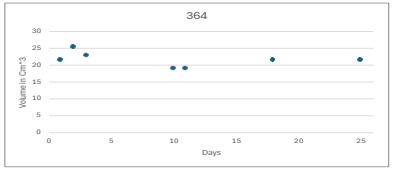

327
Volume in Cm^3
Days

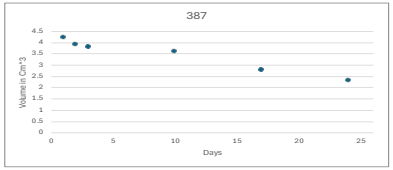

336
Volume in Cm^3
Days

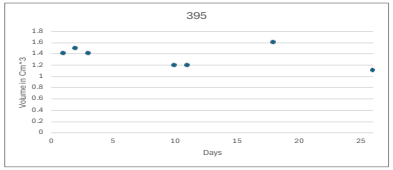

334
Volume in Cm^3
Days

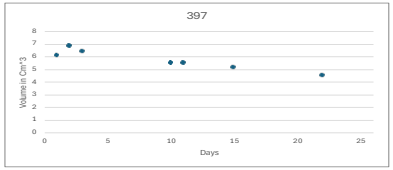

348
Volume in Cm^3
Days

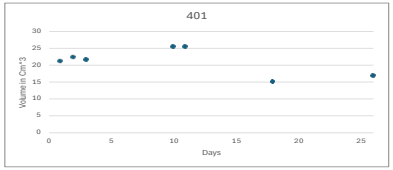

359
Volume in Cm^3
Days

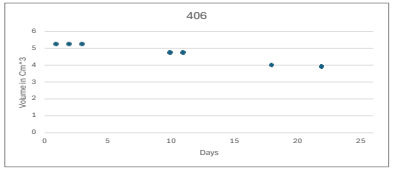

364
Volume in Cm^3
Days

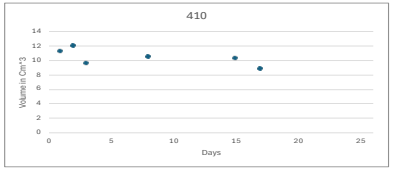

387
Volume in Cm^3
Days

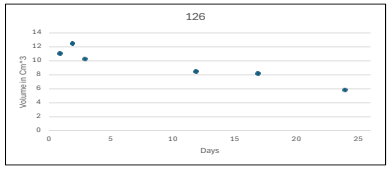

395
Volume in Cm^3
Days

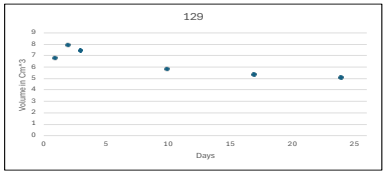

397
Volume in Cm^3
Days

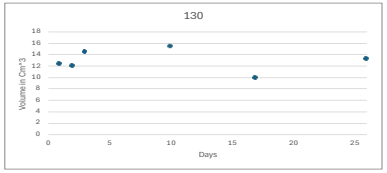

401
Volume in Cm^3
Days

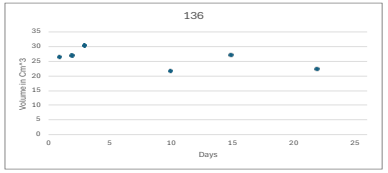

406
Volume in Cm^3
Days

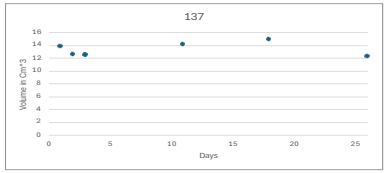

410
Volume in Cm^3
Days

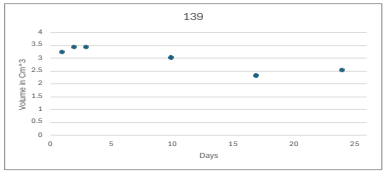

126
Volume in Cm^3
Days

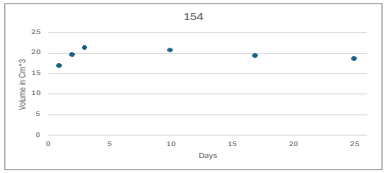

129
Volume in Cm^3
Days

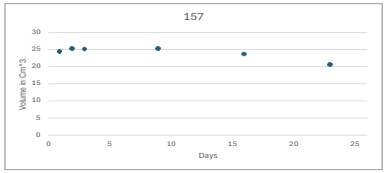

130
Volume in Cm^3
Days

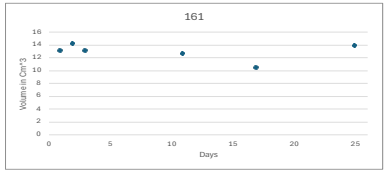

136
Volume in Cm^3
Days

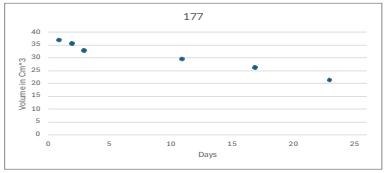

137
Volume in Cm^3
Days

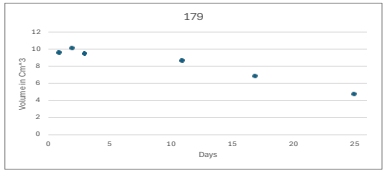

139
Volume in Cm^3
Days

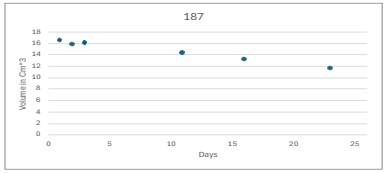

154
Volume in Cm^3
Days

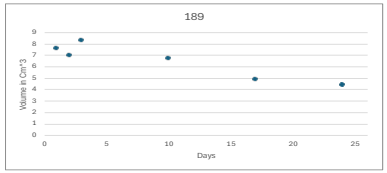

157
Volume in Cm^3
Days

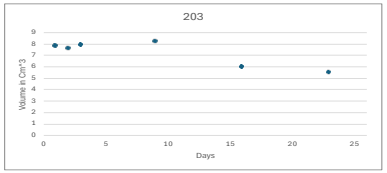

161
Volume in Cm^3
Days

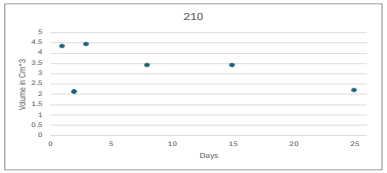

177
Volume in Cm^3
Days

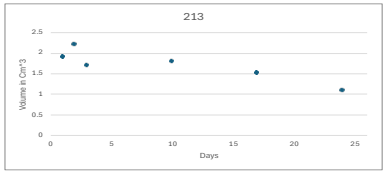

179
Volume in Cm^3
Days

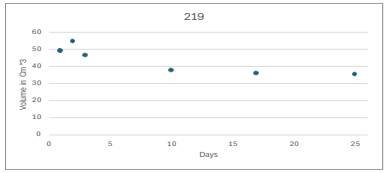

187
Volume in Cm^3
Days

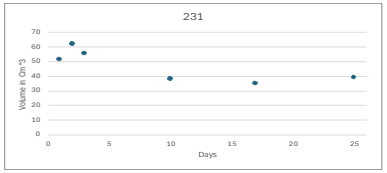

189
Volume in Cm^3
Days

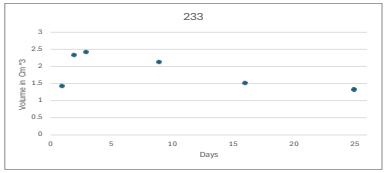

203
Volume in Cm^3
Days

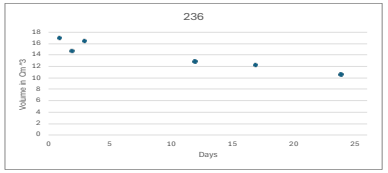

210
Volume in Cm^3
Days

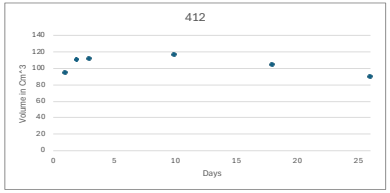

213
Volume in Cm^3
Days

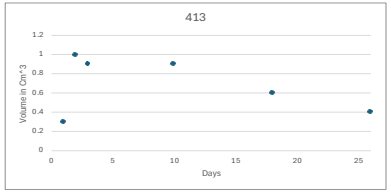

219
Volume in Cm^3
Days

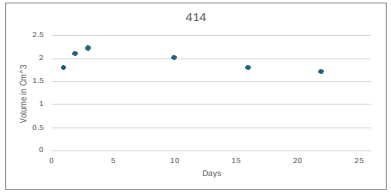

231
Volume in Cm^3
Days

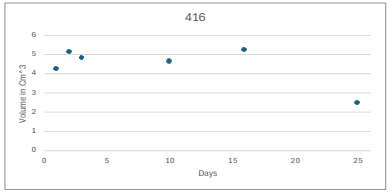

233
Volume in Cm^3
Days

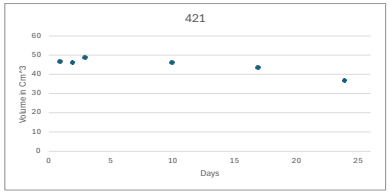

236
Volume in Cm^3
Days

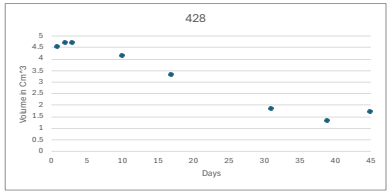

412
Volume in Cm^3
Days

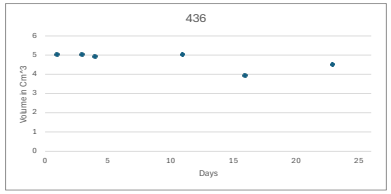

413
Volume in Cm^3
Days

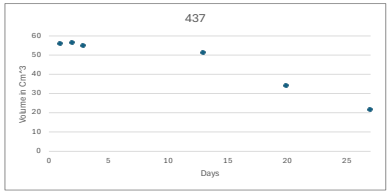

414
Volume in Cm^3
Days

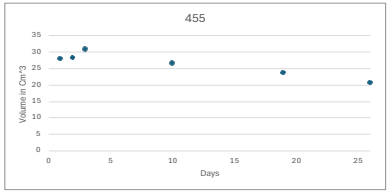

416
Volume in Cm^3
Days

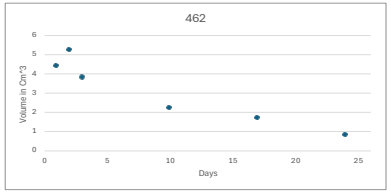

421
Volume in Cm^3
Days

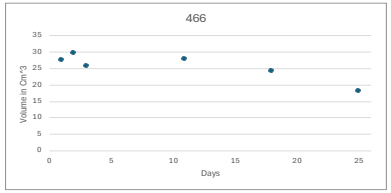

428
Volume in Cm^3
Days

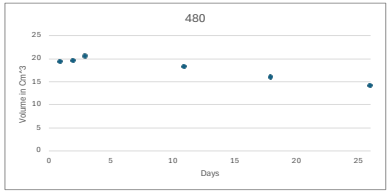

436
Volume in Cm^3
Days

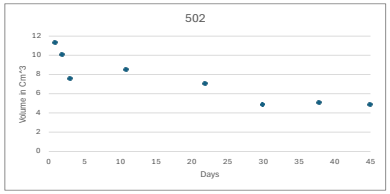

437
Volume in Cm^3
Days

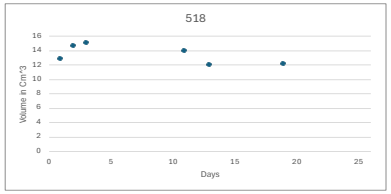

455
Volume in Cm^3
Days

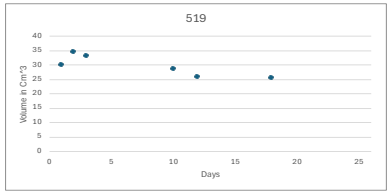

462
Volume in Cm^3
Days

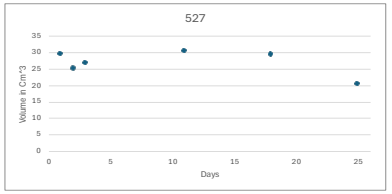

466
Volume in Cm^3
Days


480
Volume in Cm^3
Days


502
Volume in Cm^3
Days


518
Volume in Cm^3
Days


519
Volume in Cm^3
Days


527
Volume in Cm^3
Days

10.4 Appendices Chapter 6

10.4.1 Appendix 10.4.1

TRIPOD Checklist: Prediction Model Development

Section/Topic	n	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	137
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	n/a
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	137-138
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	139-140
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	141
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Where available 141
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	141
	5b	Describe eligibility criteria for participants.	141
	5c	Give details of treatments received, if relevant.	141
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	144
	6b	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	163
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	Explain how the study size was arrived at.	n/a
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	n/a
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	142-144
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	142-143

	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	144
Risk groups	11	Provide details on how risk groups were created, if done.	n/a
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	145-150
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	141
Model development	14a	Specify the number of participants and outcome events in each analysis.	n/a
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	n/a
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	n/a
	15b	Explain how to the use the prediction model.	152-153
Model performance	16	Report performance measures (with CIs) for the prediction model.	152
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	169-170
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	160-170
Implications	20	Discuss the potential clinical use of the model and implications for future research.	170
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	n/a
Funding	22	Give the source of funding and the role of the funders for the present study.	50

10.5 Appendices Chapter 7

10.5.1 Appendix 10.5.1

TRIPOD Checklist: Prediction Model Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	171
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	n/a
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	171-172
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	172-173
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	174
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	174
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	174
	5b	Describe eligibility criteria for participants.	174
	5c	Give details of treatments received, if relevant.	174-175
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	176
	6b	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	178-179
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	Explain how the study size was arrived at.	n/a
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	n/a
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	178-179

	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	175-176
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	n/a
Risk groups	11	Provide details on how risk groups were created, if done.	n/a
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	174-175
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	178
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	174-175, 178
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	n/a as unavailable in training data set
Model performance	16	Report performance measures (with CIs) for the prediction model.	179-185
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	n/a
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	198-199
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	196
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	200
Implications	20	Discuss the potential clinical use of the model and implications for future research.	196-200
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	n/a
Funding	22	Give the source of funding and the role of the funders for the present study.	50

10.6 Appendices Chapter 8

10.6.1 Appendix 10.6.1

TRIPOD Checklist: Prediction Model Development

Section/Topic	n	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	201
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	n/a
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	201
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	202-203
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	202
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	In study 4- 174
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	174
	5b	Describe eligibility criteria for participants.	174
	5c	Give details of treatments received, if relevant.	174-175
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	205
	6b	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	205-206
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	Explain how the study size was arrived at.	n/a
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	n/a
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	205-206
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	205-206

	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	205-206
Risk groups	11	Provide details on how risk groups were created, if done.	n/a
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	208
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	208
Model development	14a	Specify the number of participants and outcome events in each analysis.	n/a
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	n/a
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	n/a
	15b	Explain how to the use the prediction model.	215
Model performance	16	Report performance measures (with CIs) for the prediction model.	209-217
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	223-224
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	223-226
Implications	20	Discuss the potential clinical use of the model and implications for future research.	223-226
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	n/a
Funding	22	Give the source of funding and the role of the funders for the present study.	50

10.6.2 Appendix 10.6.2

Solving for Lambda in Validation pre-treatment dataset Alpha = 0.03

First run

num_bins = 4;

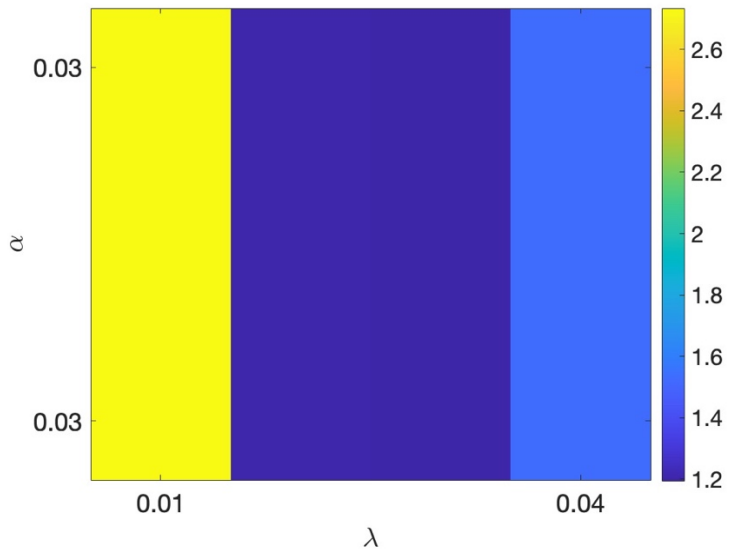
lambda_LB = 0.01; lambda_UB = 0.04;

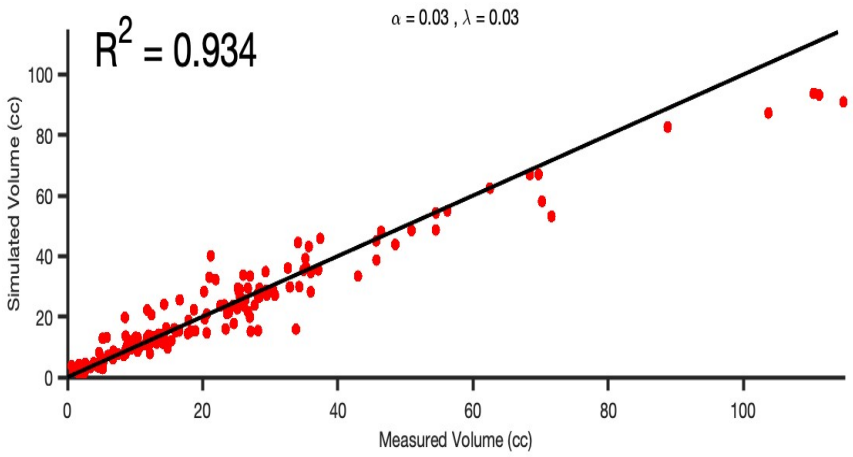
alpha_LB = 0.03; alpha_UB = 0.03;

Lambda vec 0.0100000000000000 0.0200000000000000 0.0300000000000000 0.0400000000000000

Error matrix

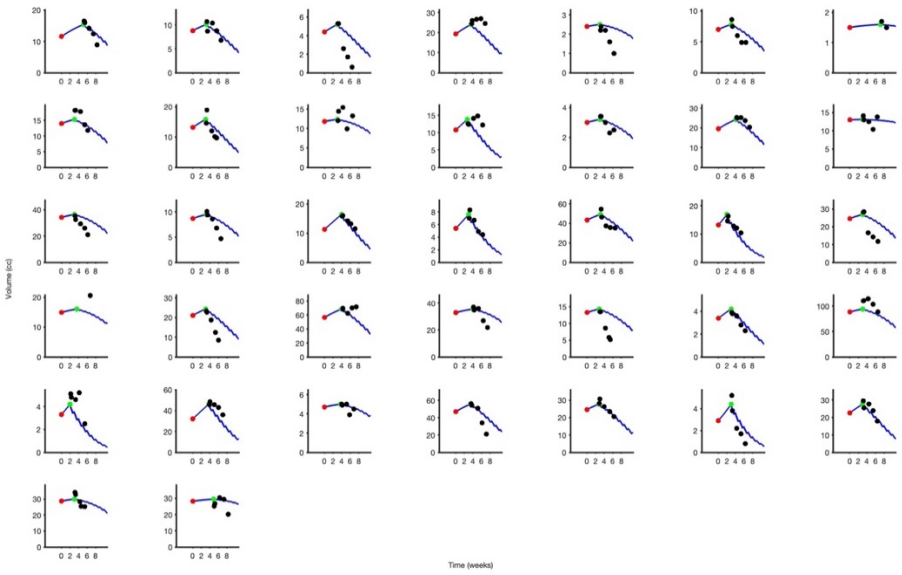
2.73394	1.20430	**1.19533**	1.52702
2.73394	1.20430	**1.19533**	1.52702
2.73394	1.20430	**1.19533**	1.52702
2.73394	1.20430	**1.19533**	1.52702





Prolif. Sat. Inde

PSI



Second run

num_bins = 4;

lambda_LB = 0.02; lambda_UB = 0.035;

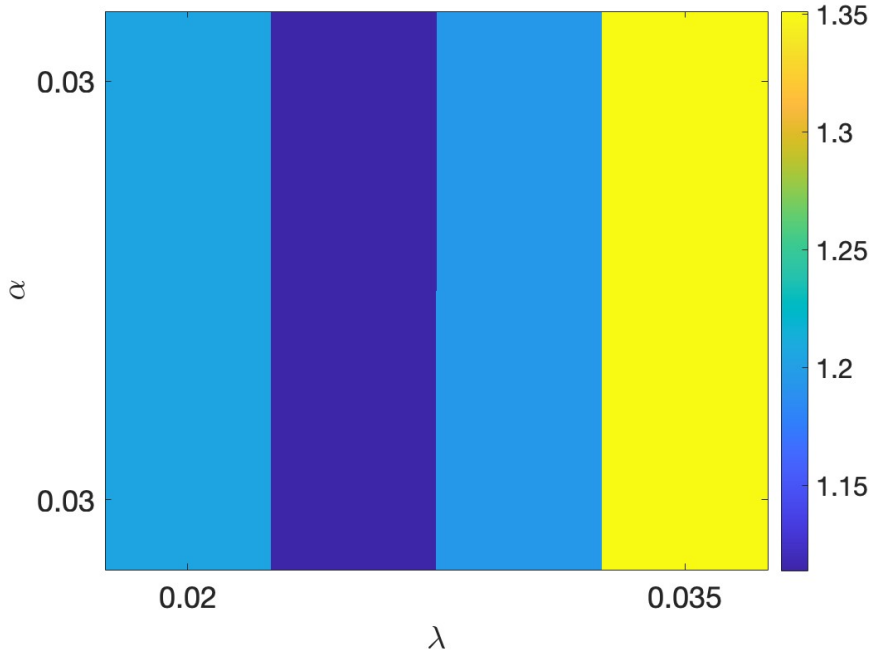
alpha_LB = 0.03; alpha_UB = 0.03;

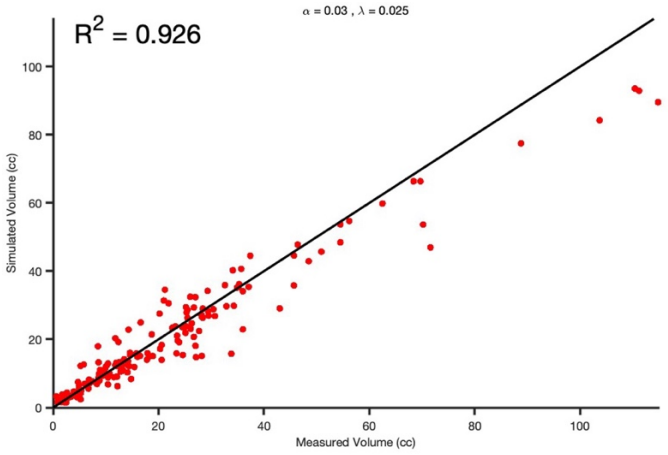
lambda vec 0.0200000000000000 0.0250000000000000

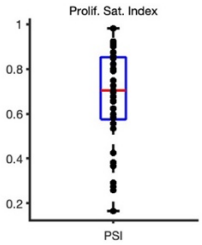
0.0300000000000000 0.0350000000000000

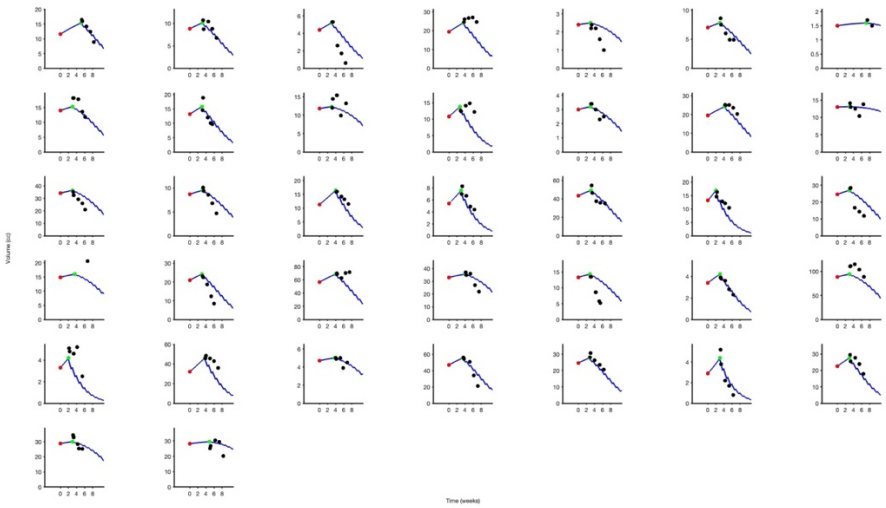
Error Matrix

1.20430	**1.11395**	1.19533	1.35136
1.20430	**1.11395**	1.19533	1.35136
1.20430	**1.11395**	1.19533	1.35136
1.20430	**1.11395**	1.19533	1.35136




α = 0.03 , λ = 0.025
R2 = 0.926
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
PSI
1
0.8
0.6
0.4
0.2



Third run

num_bins = 4;

lambda_LB = 0.02; lambda_UB = 0.03;

alpha_LB = 0.03; alpha_UB = 0.03;

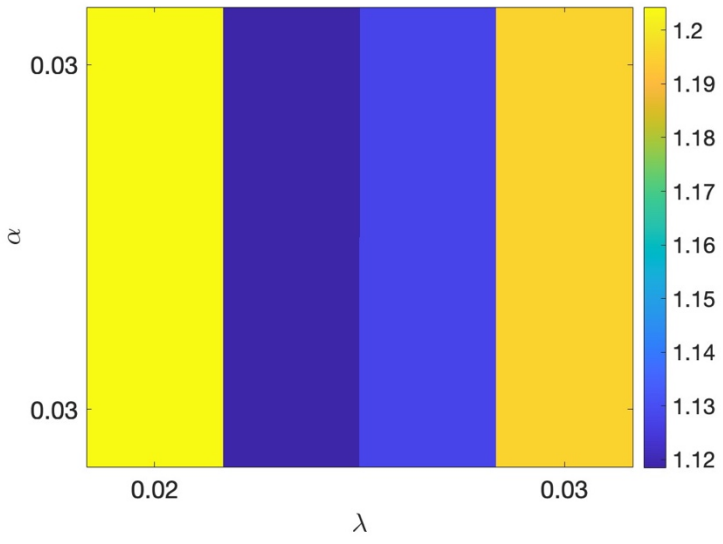
Lambda Vec

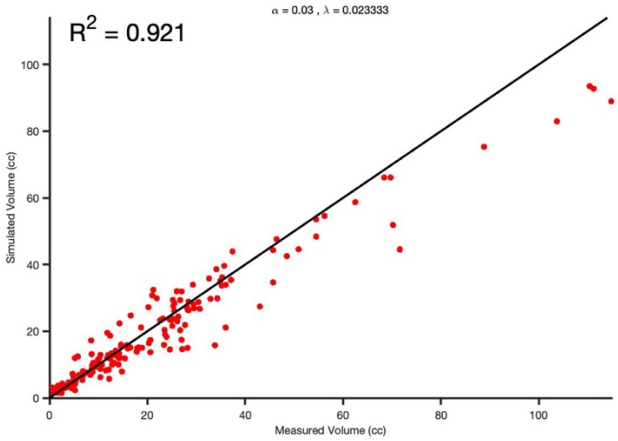
0.0200000000000000 0.0233333333333333 0.0266666666666667 0.0300000000000000

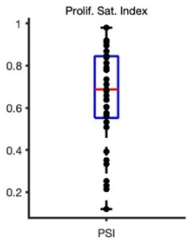
Error Matrix

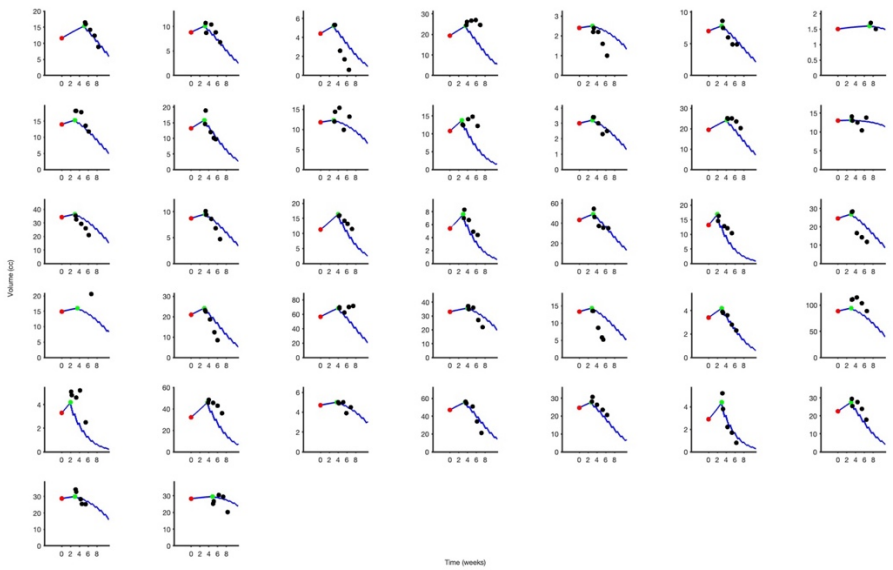
1.204296	**1.118534**	1.127891	1.195329
1.204296	**1.118534**	1.127891	1.195329
1.204296	**1.118534**	1.127891	1.195329
1.204296	**1.118534**	1.127891	1.195329

Very slightly higher error than previous value, alpha = 0.025




α = 0.03 , λ = 0.023333
R2 = 0.921
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
PSI
1
0.8
0.6
0.4
0.2


Volume (cc)
Time (weeks)

Fourth run

num_bins = 4;

lambda_LB = 0.022; lambda_UB = 0.027;

alpha_LB = 0.03; alpha_UB = 0.03;

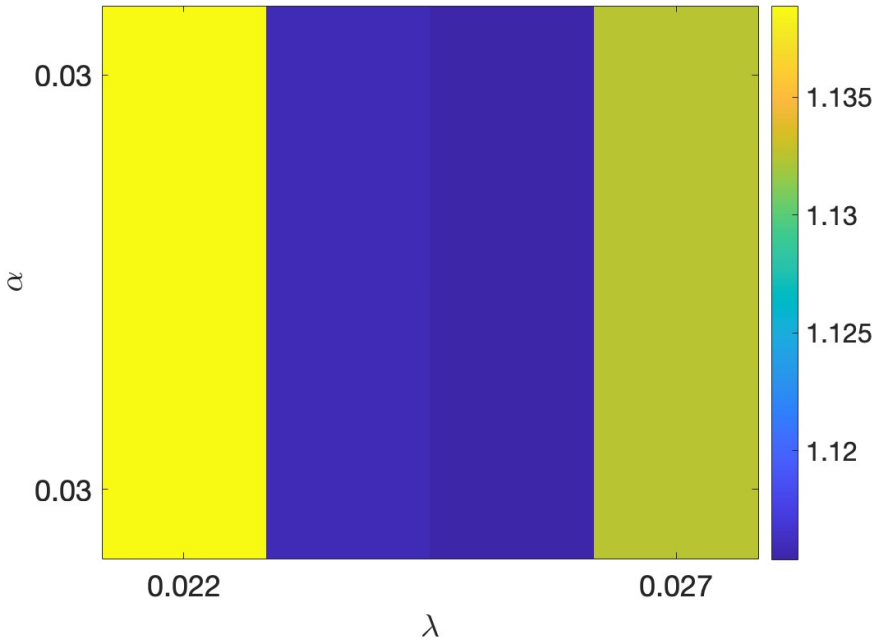
lambda vec 0 0.0220000000000000 0.0236666666666667 0.0253333333333333 0.0270000000000000

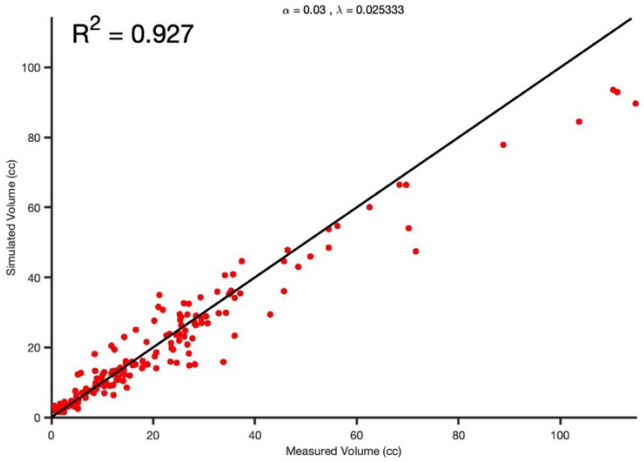
Error matrix

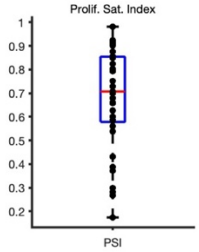
1.13887	1.11591	**1.11540**	1.13252
1.13887	1.11591	**1.11540**	1.13252
1.13887	1.11591	**1.11540**	1.13252
1.13887	1.11591	**1.11540**	1.13252

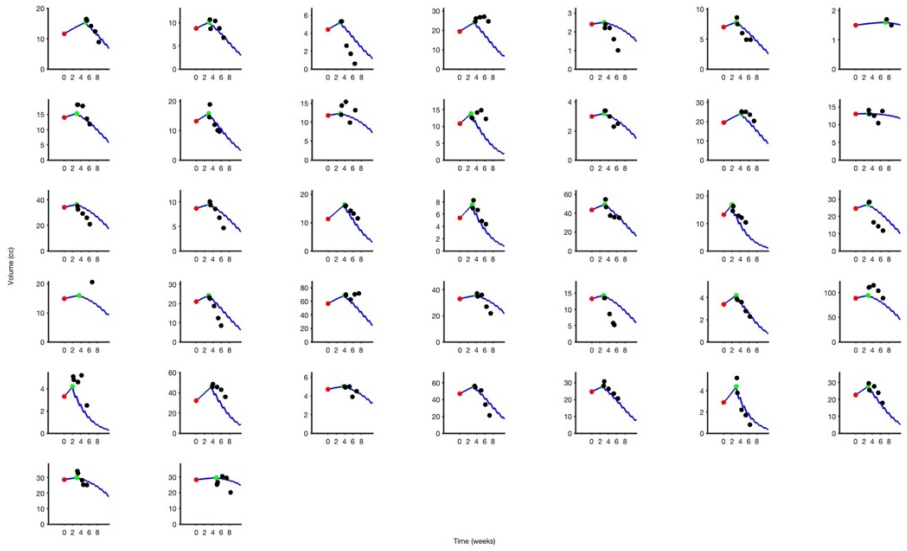
MSE very slightly higher.

Lambda 0.025 optimal




α = 0.03 , λ = 0.025333
R2 = 0.927
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
PSI
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2


Volume (cc)
Time (weeks)

10.6.3 Appendix 10.6.3

Solving for Alpha and Lambda in Validation pre-treatment dataset

First run

num_bins = 4;

lambda_LB = 0.02; lambda_UB = 0.2;

alpha_LB = 0.02; alpha_UB = 0.2;

alpha vec

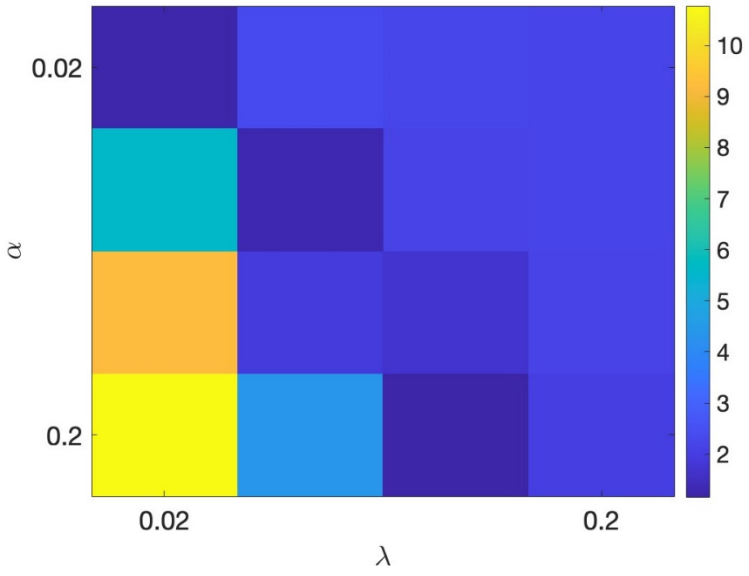
0.0200000000000000 0.0800000000000000 0.140000000000000 0.200000000000000

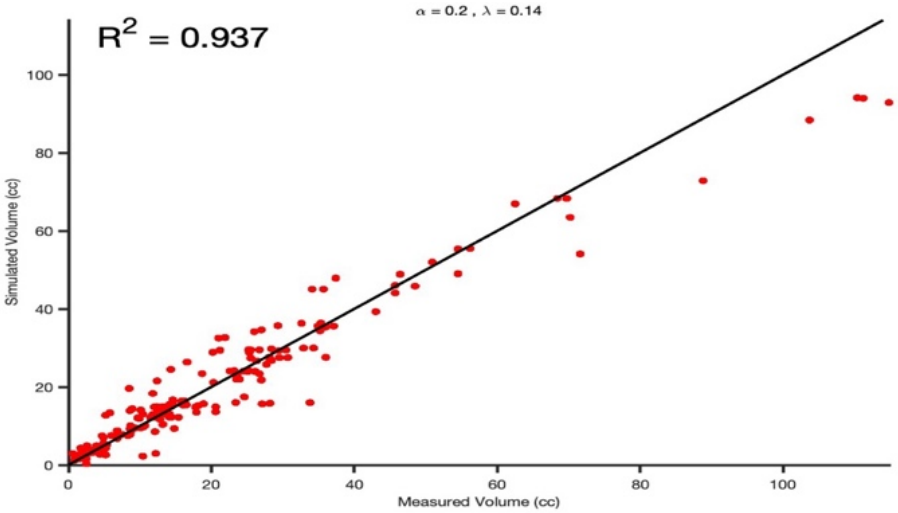
Lambda vec

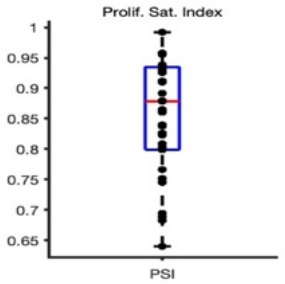
0.0200000000000000 0.0800000000000000 0.140000000000000 0.200000000000000

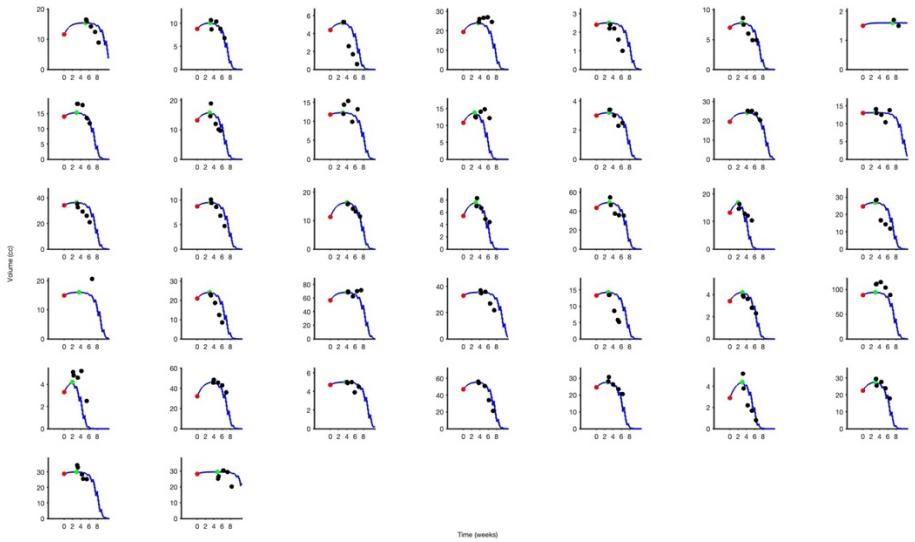
Error matrix

1.19849	2.33057	2.24768	2.20527
5.60706	1.30970	2.13736	2.19589
9.20938	1.90075	1.69339	2.15260
10.77348	4.38727	**1.16066**	2.02228




α = 0.2 , λ = 0.14
R2 = 0.937
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
PSI


Volume (cc)
Time (weeks)

Second run

num_bins = 4;

lambda_LB = 0.1; lambda_UB = 0.18;

alpha_LB = 0.018; alpha_UB = 0.22;

alpha vec

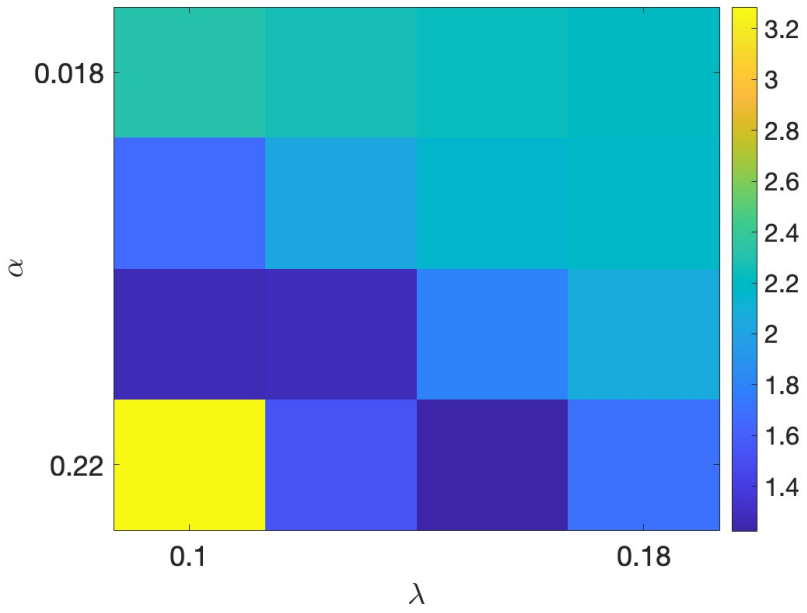
0.0180000000000000 0.0853333333333333 0.152666666666667 0.220000000000000

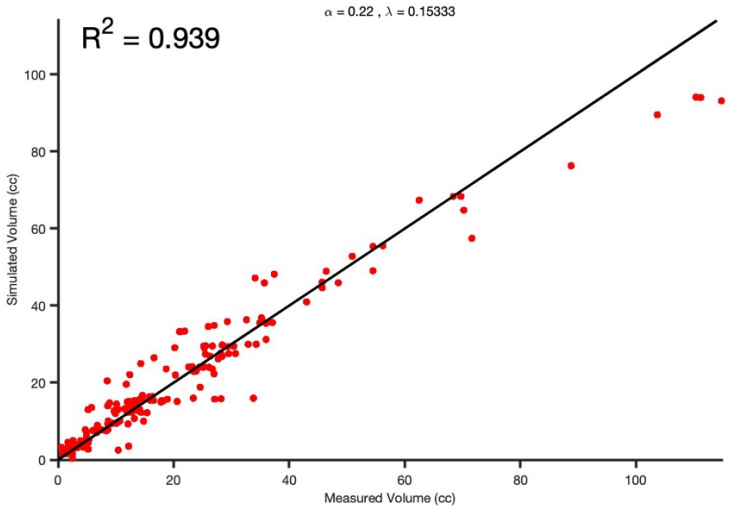
Lambda vec

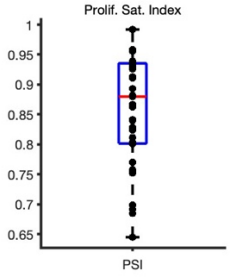
0.100000000000000 0.126666666666667 0.153333333333333 0.180000000000000

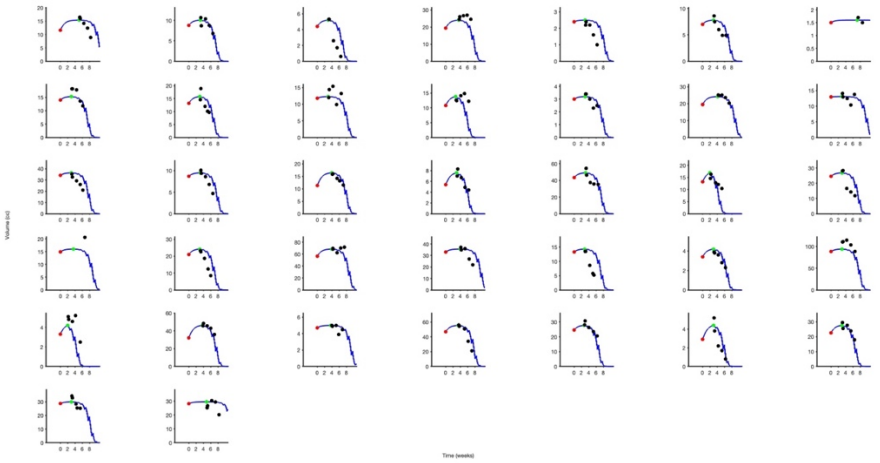
Error matrix

2.31298	2.26661	2.23460	2.21481
1.66672	2.03469	2.15799	2.18936
1.27031	1.28405	1.78931	2.05485
3.28529	1.52413	**1.22498**	1.69957




α = 0.22 , λ = 0.15333
R2 = 0.939
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
PSI



Third run

num_bins = 4;

lambda_LB = 0.14; lambda_UB = 0.16;

alpha_LB = 0.21; alpha_UB = 0.3;

alpha vec

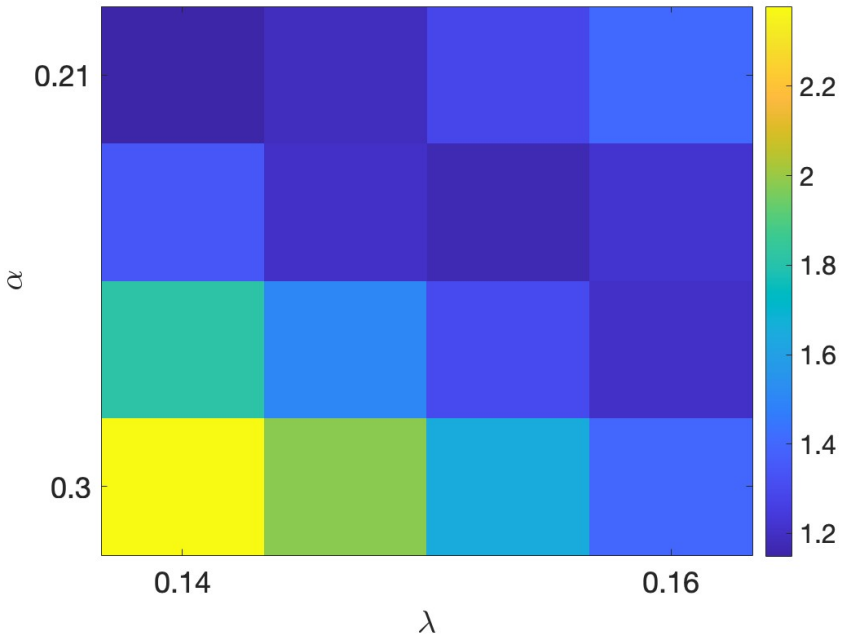
0.210000000000000 0.240000000000000 0.270000000000000 0.300000000000000

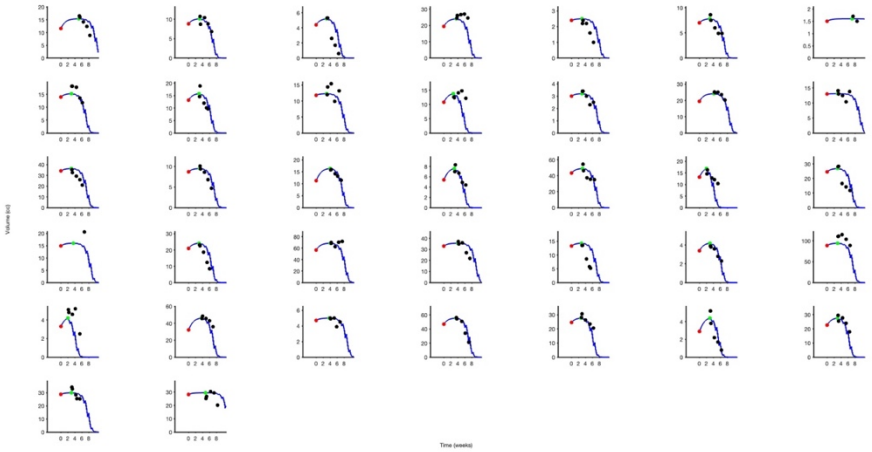
Lambda vec

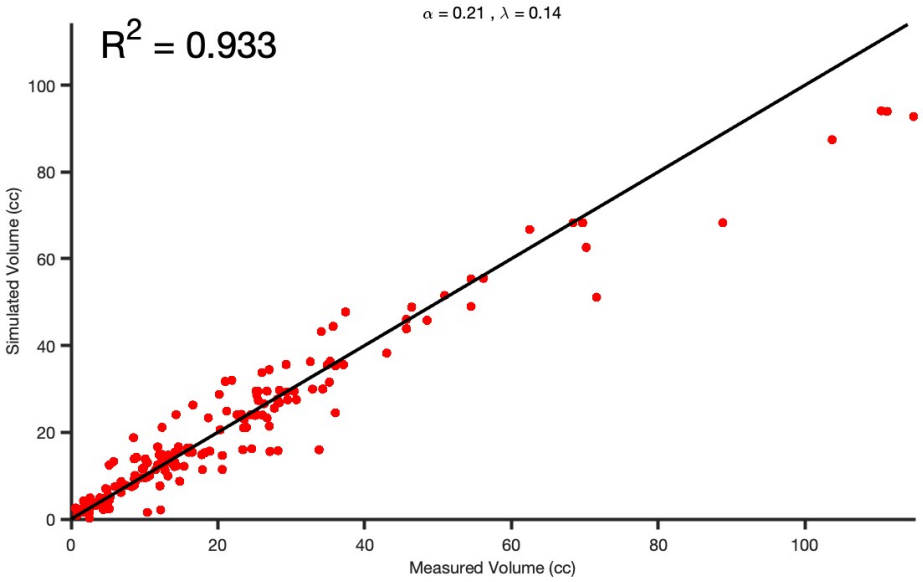
0.140000000000000 0.146666666666667 0.153333333333333 0.160000000000000

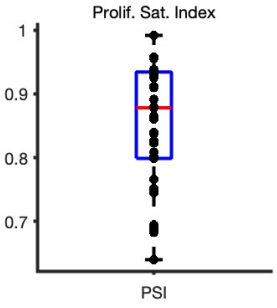
Error matrix

1.1499	1.1887	1.2845	1.4077
1.3465	1.2070	1.1727	1.2212
1.8139	1.5090	1.3024	1.2036
2.3784	1.9828	1.6479	1.3988






α = 0.21 , λ = 0.14
R2 = 0.933
Simulated Volume (cc)
Measured Volume (cc)
0
20
40
60
80
100


Prolif. Sat. Index
1
0.9
0.8
0.7
PSI

Fourth run

num_bins = 4;

lambda_LB = 0.12; lambda_UB = 0.16;

alpha_LB = 0.2; alpha_UB = 0.22;

alpha vec

0.200000000000000 0.206666666666667 0.213333333333333 0.220000000000000

Lambda vec

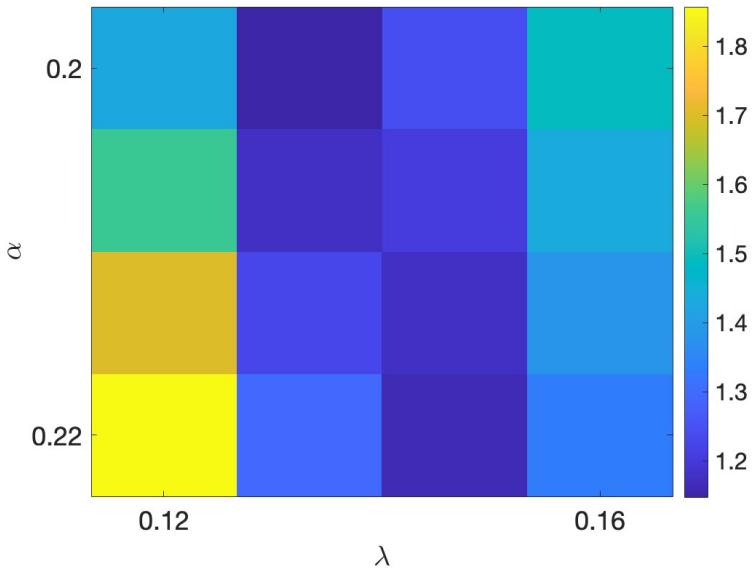
0.120000000000000 0.133333333333333 0.146666666666667 0.160000000000000

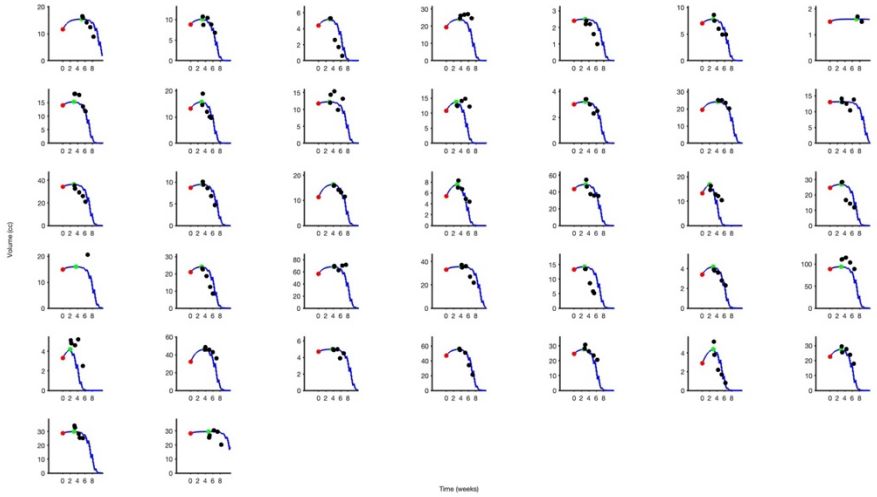
Minimal difference in MSE or values

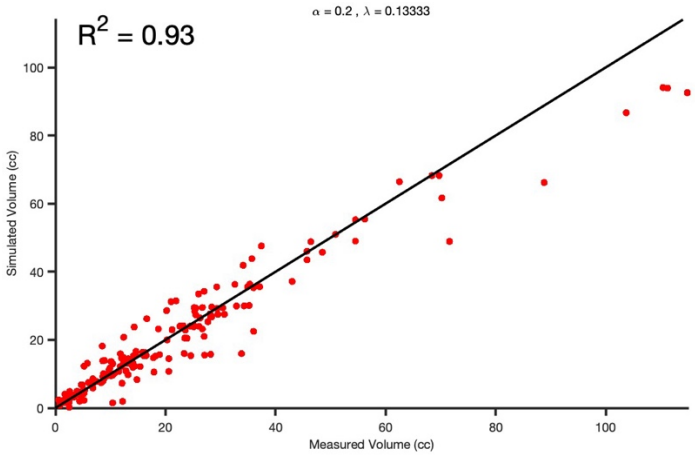
Alpha = 0.21, Lambda = 0.13

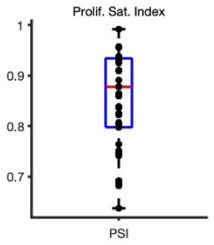
Error Matrix

1.42882	1.14744	1.24192	1.48996
1.55770	1.17686	1.20367	1.43451
1.70171	1.22624	1.17668	1.38178
1.85715	1.29466	1.16252	1.33289









Minimal difference in MSE or values

Final parameters

Alpha = 0.21, Lambda = 0.13

10.7 Appendix 10.7

Data Sharing Agreement

This deed is dated 7/6/2024

DEED OF AGREEMENT FOR DATA SHARING BETWEEN

The Belfast Health and Social Care Trust
and
The Provost, Fellows, Foundation Scholars, and the other members of Boar, of the College of the holy and undivided Trinity of Queen Elizabeth near Dublin

<table>
<tr><td>1.</td><td colspan="2">Parties to this Data Sharing Agreement</td></tr>
<tr><td></td><td>The Belfast Health and Social
Care Trust

Moira Kearney
Interim Director for Cancer Services

moira.kearney@belfasttrust.hscni.net

Non – Clinical Support Building,
Royal Hospital,
Belfast BT12 6BA
028 950 48858</td><td>The Provost, Fellows, Foundation Scholars, and the other members of Board, of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin

Sarah Barrett
Assistant Professor

Trinity Centre for Health Sciences,
St James
Dublin 8
+353 1 896 3248
Barrets7@tcd.ie</td></tr>
<tr><td>2.</td><td colspan="2">Introduction

The aim of this project is to investigate if a simple mathematical model can predict tumour volume changes during treatment delivery.

Advance knowledge of how a tumour may respond to radiation therapy could allow us to adapt the treatment plan, ensuring dose to critical normal tissues, such as the heart, are minimised with the aim of reducing side effects.

Furthermore, this prediction may facilitate an optimised prescription for each patient, improving local control while maintaining acceptable toxicity.

The model has been tested in data acquired from prior publications but must be externally validated. This data would evaluate if the model can accurately predict the tumour response to radiation therapy in a clinical population. Furthermore, it would allow us to investigate the clinical relevance of this prediction by determining if it can ascertain what patients may need an adapted radiotherapy plan, and examining if the</td></tr>
</table>

	predicted response is linked to patient outcomes such as time-to-recurrence, overall survival and cardiopulmonary toxicity.
3.	Purpose This Data Sharing Agreement has been set up for the purpose of sharing information on a collaborate research project titled 'Reducing Incidental Cardiopulmonary Doses in Lung Cancer by Harnessing Routine Imaging for Adaptive Planning'
4.	Details of data to be shared and data flows Upon approval of this application, the data will be transferred electronically to Trinity College Dublin via an encrypted email. The data will include: volume measurements of the tumour from routine radiotherapy scans during a course of treatment. We further require the following pseudonymised clinical data for these patients: • Diagnosis: o Age at diagnosis o Histological subtype o TNM staging • Treatment: o Fractionation schedule and dose o Dates of treatment o Radiotherapy dose summary metrics • Follow-up: o Time-to-recurrence o Cardiopulmonary toxicity o Survival time No patient details are transferred and pseudo-anonymised data will be encrypted. The Belfast Health & Social Care Trust will retain an ID key for de-anonymising patients in the future if required.
5.	Data Flow Map Only the Belfast Trust CI / primary care team will hold the identification keys to the database that will be kept in the

	Belfast Health & Social Care Trust and will not be shared. Only unidentified pseudo-anonymized data will be transferred to Trinity College Dublin. Data will be stored on servers located behind the Trinity College Dublin firewall and physically located in Trinity College Dublin. Access to the university network is via password. Servers are administered by the research group and access is controlled by individual passwords. Data will be stored securely will access granted to the local study team only. Patient data will be only be transferred and analysed in a coded form Individual patients will not be identifiable from the presented or published material
6.	Information use The parties will ensure the information shared under this Data Sharing Agreement will only be used for the specific purpose set out in Section 3. There will be no onward sharing of personal data beyond meeting the purposes explained in this Data Sharing Agreement. The necessary Data Protection and UK GDPR principles, as set out below, will be adhered to. - Lawfulness, fairness and transparency - Purpose limitation - Data minimization - Accuracy - Storage limitation - Integrity and confidentiality (security) - Accountability
7.	Legal Basis for Data Sharing It is acknowledged that the information to be shared as part of this agreement services will be Special Category (sensitive) data, as defined in Data Protection legislation. For the purposes of this agreement, the following lawful basis for processing is relied upon by the respective parties

	For Trinity College Dublin: • Necessary for the performance of a task carried out in the public interest – Art. 6 (1)(e), GDPR • Necessary for scientific research in the public interest – Art. (9)(2)(j), GDPR – in accordance with Article 89 (1) The Lawful basis for data processing is it is necessary for reasons of public interest in the area of public health.
8.	Relationship and Responsibilities of parties Separate Controllers
9.	Requests for information (Subject Access Requests and Freedom of Information Requests) Patients will not be informed about the sharing of their pseudo-anonymised data, in line with the original governance approvals. No identity key will be shared with the partner institution so their dataset can be considered to be anonymous. The data does not involve tissue or imaging, is retrospective in nature, and robust data security measures are in place at the partner institution. Patients were not required to be informed of their data being retrospectively and anonymously used for the broader project. Parties acknowledge the legal rights of Data Subjects and individual members of the public, where appropriate, to access information under the General Data Protection Regulation (subject access requests), the Freedom of Information Act (FOI requests) and associated legislation. The parties each agree to provide such assistance as is reasonably required to enable the other party to comply with requests submitted under this legislation within the time limits imposed. The point of contact (service lead) for each party is responsible for maintaining a record of individual requests for information, the decisions made and any information that was exchanged. Such records will be retained and stored as per respective information retention and disposal policies.

10.	Security and Storage of Data The parties undertake to have in place appropriate technical and organisational security measures to: prevent: unauthorised or unlawful processing of the Shared Personal Data; and the accidental loss or destruction of, or damage to, the Data ensure a level of security appropriate to: the harm that might result from such unauthorised or unlawful processing or accidental loss, destruction or damage; and the nature of the Shared Personal Data to be protected. To ensure confidentiality, all relevant staff involved in discussions across the organisations will adopt steps that can reduce the risks to personal data The sharing of recorded / written information will be by secure means at all times. If transferred electronically, this will be through password protected document and encrypted email only. Only authorised staff that would normally have access to this data as part of their duties will transfer / receive this information Measures will be in place to ensure that personal data, both manual and electronic, is stored and transported by secure means at all times and only authorised staff will have access to this data as part of their duties. It is the responsibility of each party to ensure that its staff members are appropriately trained to handle and process the shared Personal Data in accordance with appropriate technical and organisational security measures and Data Protection Legislation.
11.	Retention and disposal All data held is for the specific purpose stated in section 3 of this Data Sharing Agreement. Data must not be held for any longer than is necessary to carry out the task(s) agreed in Section 3. All parties must retain and dispose of information in accordance with their organisational policy.

12.	Security incidents or data breaches Security of information should be managed in line with each organisation's protocols and standards to eliminate any data breaches. Any security breach incident involving personal data should be reported at the earliest opportunity through the organisations incident reporting system and to the relevant Information Governance Lead. Where appropriate the parties will inform the point of contact in the other organisation immediately. Parties to this Data Sharing Agreement recognise the need to report certain types of personal data breach to the ICO within 72 hours of becoming aware of the breach, where feasible; and if the breach is likely to result in a high risk of adversely affecting individuals' rights and freedoms, also inform those individuals without undue delay. The relevant Information Asset Owner (IAO) or appropriate official will instigate an investigation in line with the established security reporting / incident reporting protocol / policy. Each party must be fully engaged in the resolution of an incident by assisting in the investigation being carried out by the responsible party.
13.	Review/Termination of Data Sharing Agreement If any significant change takes place which means this Data Sharing Agreement becomes an unreliable reference point, this Data Sharing Agreement will be updated as needed and a new version circulated to replace it. Either signatory to this Data Sharing Agreement can request an extraordinary review at any time. This Data Sharing Agreement will be reviewed three (3) years after the date stated at the beginning of it (or earlier if considered necessary by either party) and or when there is a change in the nature or content of data being shared. The responsible officer (service lead) in each organisation will carry out the reviews.
14.	Declaration The Parties to this Data Sharing Agreement will comply with Data Protection Legislation; the Information Commissioner's Data Sharing Code of Practice; Freedom of Information legislation; and any associated guidelines issued by relevant authorities.

	The parties agree that: The information shared is for a specified, explicit and legitimate purpose It is adequate, relevant and limited to the stated purpose It will be processed fairly and lawfully and used only for the stated purpose It will be processed and stored in a manner that ensures appropriate security It will be held no longer than is necessary for the stated purpose It will be disposed of fully and in such a way that it is not possible to reconstitute it All measures will be taken to ensure identifiable data is not disclosed to third parties without appropriate prior authority of the other party/parties. Where appropriate, the parties to this Data Sharing Agreement will be informed of the identifiable data being deleted / destroyed Any loss, theft or corruption of the shared data will be immediately reported to the authorised officer of the owning organisation and we will assist fully in any investigation. Any serious breaches, data loss, theft or corruption will be reported to the ICO within 72 hours of the breach first being discovered.
15	Indemnity Each party shall indemnify the other against all liabilities, costs, expenses, damages and losses suffered or incurred by the indemnified party arising out of or in connection with the breach of this Data Sharing Agreement or Data Protection Legislation by the indemnifying party or any third party processor acting on its behalf in connection with this Data Sharing Agreement, or any of their employees or agents, provided that the indemnified party gives to the indemnifier prompt notice of such claim, full information about the circumstances giving rise to it, reasonable co-operation and assistance in dealing with the claim and sole authority to manage, defend and/or settle it. The provisions of this section 15 shall survive the termination or expiry of this Data Sharing Agreement.
16	Governing Law and Jurisdiction This Data Sharing Agreement and any dispute or claim (including non-contractual disputes or claims) arising out of or in connection with it or its subject matter or formation shall be

	governed by and construed in accordance with the law of Northern Ireland. Each party irrevocably agrees that the courts of Northern Ireland shall have exclusive jurisdiction to settle any dispute or claim (including non-contractual disputes or claims), arising out of or in connection with this Data Sharing Agreement or its subject matter or formation.
	This document has been executed as a deed and is delivered and takes effect on the date stated at the beginning of it.

Executed as deed by **Belfast Health and Social Care Trust** acting by Moira Kearney a director, in the presence of:

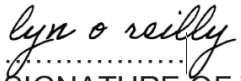
…… ……
[SIGNATURE OF WITNESS]

Debbie Wightman, Co – Director, Cancer and Specialist Medicine
…… 09.07.24……..
[NAME, ADDRESS [AND OCCUPATION] OF WITNESS] Date

…… 09/07/2024…….
[SIGNATURE OF DIRECTOR] Date

Director Interim Director Cancer & Specialist Services

Executed as deed by **The Provost, Fellows, Foundation Scholars, and the other members of Board, of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin** acting through Dr. Gordon Elliott, Senior Patents and Licensing Manager, in the presence of:



……
SIGNATURE OF WITNESS

…… ….06/06/24…
Lyn O'Reilly, Paralegal Officer Date

……
SIGNATURE

….07 JUN 2024…
Date

Gordon Elliot
Senior Patients and Licensing Manager

10.8 Appendix 10.8

Data Protection Risk Assessment (DPRA) approval dates

Project Name: Reducing Incidental Cardiopulmonary Doses in Lung Cancer by Harnessing Routine Imaging for Adaptive Planning	**Date: 4/3/24**
Project Owner: Sarah Barrett	**Site: Trinity Centre for Health Sciences**
Email address: barrets7@tcd.ie	**Phone Number: 3248**

Risk Assessment Circulation

Name	Date	Reviewed/Consulted
Project Manager Details	[Insert Date]	*Reviewed/Consulted*
Other Details	[Insert Date]	*Reviewed/Consulted*
Imelda O'Keeffe	15.03.2024	*Reviewed*
DPO Details Evelyn Fox, RDPO	21.11.24	*Reviewed to close out file on register*

CHAPTER 11

REFERENCES

1. Bray, F., et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: A Cancer Journal for Clinicians, 2024. **74**(3): p. 229-263.
2. *National Cancer Registry of Ireland, Lung Cancer Factsheet.* 2022 [cited 2024 23-7-2024]; Available from: https://www.ncri.ie/sites/ncri/files/factsheets/Dashboardv3.3_C34.pdf.
3. Malhotra, J., et al., *Risk factors for lung cancer worldwide.* European Respiratory Journal, 2016. **48**(3): p. 889.
4. LoPiccolo, J., et al., *Lung cancer in patients who have never smoked — an emerging disease.* 2024. **21**(2): p. 121-146.
5. Florez, N., et al., *Lung Cancer in Women: The Past, Present, and Future.* Clinical Lung Cancer, 2024. **25**(1): p. 1-8.
6. Ganti, A.K., et al., *Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non–Small Cell Lung Cancer in the US.* JAMA Oncology, 2021. **7**(12): p. 1824-1832.
7. McAleese, J., et al., *Differential Relapse Patterns for Non-small Cell Lung Cancer Subtypes Adenocarcinoma and Squamous Cell Carcinoma: Implications for Radiation Oncology*, in *Clin Oncol (R Coll Radiol)*. 2019, © 2019 The Royal College of Radiologists. Published by Elsevier Ltd: England. p. 711-719.

8. Keogh, R.J., et al., *Genomic Landscape of NSCLC in the Republic of Ireland.* 2024. **5**(2): p. 100627.
9. Remon, J., J.C. Soria, and S. Peters, *Early and locally advanced non-small-cell lung cancer: an update of the ESMO Clinical Practice Guidelines focusing on diagnosis, staging, systemic and local therapy.* Annals of Oncology, 2021. **32**(12): p. 1637-1642.
10. Hendriks, L.E., et al., *Non-oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up[☆].* Annals of Oncology, 2023. **34**(4): p. 358-376.
11. Hendriks, L.E., et al., *Oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up[☆].* Annals of Oncology, 2023. **34**(4): p. 339-357.
12. Chih-Hsin Yang et al., *Osimertinib after Chemoradiotherapy in Stage III EGFR-Mutated NSCLC.* New England Journal of Medicine. **0**(0).
13. Chan, B.A. and B.G.M. Hughes, *Targeted therapy for non-small cell lung cancer: current standards and the promise of the future.* Translational Lung Cancer Research, 2014. **4**(1): p. 36-54.
14. Spigel, D.R., et al., *Five-Year Survival Outcomes From the PACIFIC Trial: Durvalumab After Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer.* Journal of Clinical Oncology, 2022. **40**(12): p. 1301-1311.

15. Antonia, S.J., et al., *Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer.* New England Journal of Medicine, 2017. **377**(20): p. 1919-1929.
16. Peters, S., et al., *Position of a panel of international lung cancer experts on the approval decision for use of durvalumab in stage III non-small-cell lung cancer (NSCLC) by the Committee for Medicinal Products for Human Use (CHMP).* Annals of Oncology, 2019. **30**(2): p. 161-165.
17. Steinfort, D.P., et al., *Systematic endoscopic staging of mediastinum to guide radiotherapy planning in patients with locally advanced non-small-cell lung cancer (SEISMIC): an international, multicentre, single-arm, clinical trial.* The Lancet Respiratory Medicine, 2024. **12**(6): p. 467-475.
18. Detterbeck, F.C., *The eighth edition TNM stage classification for lung cancer: What does it mean on main street?* The Journal of Thoracic and Cardiovascular Surgery, 2018. **155**(1): p. 356-359.
19. Detterbeck, F.C., et al., *The Proposed Ninth Edition TNM Classification of Lung Cancer.* CHEST, 2024. **166**(4): p. 882-895.
20. Ball, D., et al., *Stereotactic ablative radiotherapy versus standard radiotherapy in stage 1 non-small-cell lung cancer (TROG 09.02 CHISEL): a phase 3, open-label, randomised controlled trial.* The Lancet Oncology, 2019. **20**(4): p. 494-503.
21. Dingemans, A.M., et al., *OA06.05 Consensual Definition of Stage III NSCLC Resectability: EORTC-Lung Cancer Group Initiative with Other Scientific Societies.* Journal of Thoracic Oncology, 2023. **18**(11): p. S57-S58.

22. Lababede, O. and M.A. Meziane, *The Eighth Edition of TNM Staging of Lung Cancer: Reference Chart and Diagrams.* Oncologist, 2018. **23**(7): p. 844-848.
23. Allemani, C., et al., *Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2).* The Lancet. **385**(9972): p. 977-1010.
24. Quaresma, M., M.P. Coleman, and B. Rachet, *40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study.* The Lancet, 2015. **385**(9974): p. 1206-1218.
25. Cotarla, I., et al., *Treatment Decision Drivers in Stage III Non-Small-Cell Lung Cancer: Outcomes of a Web-Based Survey of Oncologists in the United States.* JCO Oncol Pract, 2020. **16**(10): p. e1232-e1242.
26. Bradley, J.D., et al., *PACIFIC-2: Phase 3 study of concurrent durvalumab and platinum-based chemoradiotherapy in patients with unresectable, stage III NSCLC.* Journal of Clinical Oncology, 2019. **37**(15_suppl): p. TPS8573-TPS8573.
27. J.D. Bradley, et al., *Durvalumab in combination with chemoradiotherapy for patients with unresectable stage III NSCLC: Final results from PACIFIC-2.* Annals of Oncology, 2024. **9**: p. 1-12.
28. Lu, S., et al., *Osimertinib after Chemoradiotherapy in Stage III EGFR-Mutated NSCLC.* New England Journal of Medicine, 2024.
29. Machtay, M., et al., *Defining Local-Regional Control and Its Importance in Locally Advanced Non-small Cell Lung*

Carcinoma. Journal of Thoracic Oncology, 2012. **7**(4): p. 716-722.

30. Postmus, P.E., et al., *Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†.* Annals of Oncology, 2017. **28**(suppl_4): p. iv1-iv21.
31. Maguire, J., et al., *SOCCAR: A randomised phase II trial comparing sequential versus concurrent chemotherapy and radical hypofractionated radiotherapy in patients with inoperable stage III Non-Small Cell Lung Cancer and good performance status*, in *Eur J Cancer*. 2014, © 2014. Published by Elsevier Ltd.: England. p. 2939-49.
32. Cooke, S.A., et al., *18F-FDG-PET guided vs whole tumour radiotherapy dose escalation in patients with locally advanced non-small cell lung cancer (PET-Boost): Results from a randomised clinical trial.* Radiotherapy and Oncology, 2023. **181**: p. 109492.
33. Palcic, B., J.W. Brosing, and L.D. Skarsgard, *Survival measurements at low doses: oxygen enhancement ratio.* British journal of cancer, 1982. **46**(6): p. 980-984.
34. Pawlik, T.M. and K. Keyomarsi, *Role of cell cycle in mediating sensitivity to radiotherapy.* International Journal of Radiation Oncology*Biology*Physics, 2004. **59**(4): p. 928-942.
35. Marcu, L.G., *Altered fractionation in radiotherapy: From radiobiological rationale to therapeutic gain.* Cancer Treatment Reviews, 2010. **36**(8): p. 606-614.
36. Allemani, C., et al., *Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18*

cancers from 322 population-based registries in 71 countries. The Lancet, 2018. **391**(10125): p. 1023-1075.

37. Kong, F.-M., et al., *Radiation dose effect in locally advanced non-small cell lung cancer.* Journal of Thoracic Disease, 2014. **6**(4): p. 336-347.

38. Bradley, J.D., et al., *Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study.* The Lancet Oncology, 2015. **16**(2): p. 187-199.

39. Bradley, J. and C. Hu, *Learning From Trials on Radiation Dose in Non-Small Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2016. **96**(4): p. 748-750.

40. McMillan, M.T., et al., *Radiation Treatment Time and Overall Survival in Locally Advanced Non-small Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2017. **98**(5): p. 1142-1152.

41. Fowler Jack F., C. and Rick, *Non-small cell lung tumors repopulate rapidly during radiation therapy.* International Journal of Radiation Oncology*Biology*Physics, 2000. **46**(2): p. 516-517.

42. Speirs, C.K., et al., *Heart Dose Is an Independent Dosimetric Predictor of Overall Survival in Locally Advanced Non–Small Cell Lung Cancer.* Journal of Thoracic Oncology, 2017. **12**(2): p. 293-301.

43. McWilliam, A., et al., *Demystifying the Results of RTOG 0617: Identification of Dose Sensitive Cardiac Subregions*

Associated With Overall Survival. J Thorac Oncol, 2023. **18**(5): p. 599-607.

44. McWilliam, A., et al., *Radiation dose to heart base linked with poorer survival in lung cancer patients.* Eur J Cancer, 2017. **85**: p. 106-113.
45. Mauguen, A., et al., *Hyperfractionated or accelerated radiotherapy in lung cancer: an individual patient data meta-analysis.* J Clin Oncol, 2012. **30**(22): p. 2788-97.
46. Saunders, M., et al., *Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer: mature data from the randomised multicentre trial.* Radiotherapy and Oncology, 1999. **52**(2): p. 137-148.
47. Hatton, M.Q., et al., *Continuous hyperfractionated accelerated radiotherapy - Escalated dose (CHART-ED): A phase I study.* Radiother Oncol, 2016. **118**(3): p. 471-7.
48. Yom, S.S., *Accelerated repopulation as a cause of radiation treatment failure in non-small cell lung cancer: review of current data and future clinical strategies.* Semin Radiat Oncol, 2015. **25**(2): p. 93-9.
49. Kaster, T.S., et al., *Radical-intent hypofractionated radiotherapy for locally advanced non-small-cell lung cancer: a systematic review of the literature.* Clin Lung Cancer, 2015. **16**(2): p. 71-9.
50. Cannon, D.M., et al., *Dose-limiting toxicity after hypofractionated dose-escalated radiotherapy in non-small-cell lung cancer.* J Clin Oncol, 2013. **31**(34): p. 4343-8.
51. Zwitter, M., et al., *Gemcitabine, cisplatin, and hyperfractionated accelerated radiotherapy for locally*

advanced non-small cell lung cancer. J Thorac Oncol, 2006. **1**(7): p. 662-6.

52. Dieleman, E.M.T., et al., *Concurrent Daily Cisplatin and High-Dose Radiation Therapy in Patients With Stage III Non-Small Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2018. **102**(3): p. 543-551.

53. Kong, F.-M.S., et al., *NRG-RTOG 1106/ACRIN 6697: A phase IIR trial of standard versus adaptive (mid-treatment PET-based) chemoradiotherapy for stage III NSCLC—Results and comparison to NRG-RTOG 0617 (non-personalized RT dose escalation).* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 8548-8548.

54. Barrett, S., G.G. Hanna, and L. Marignol, *An overview on personalisation of radiotherapy prescriptions in locally advanced non-small cell lung cancer: Are we there yet?* Radiotherapy and Oncology, 2018. **128**(3): p. 520-533.

55. van Diessen, J., et al., *The acute and late toxicity results of a randomized phase II dose-escalation trial in non-small cell lung cancer (PET-boost trial).* Radiotherapy and Oncology, 2019. **131**: p. 166-173.

56. Møller, D.S., et al., *Heterogeneous FDG-guided dose-escalation for locally advanced NSCLC (the NARLAL2 trial): Design and early dosimetric results of a randomized, multi-centre phase-III study.* Radiotherapy and Oncology, 2017. **124**(2): p. 311-317.

57. Schytte, T., et al., *3531: The NARLAL II trial: First results on loco-regional control in stage 3 NSCLC.* Radiotherapy and Oncology, 2024. **194**: p. S117-S119.

58. Wu, T.C., et al., *Accelerated Hypofractionated Chemoradiation Followed by Stereotactic Ablative Radiotherapy Boost for Locally Advanced, Unresectable Non–Small Cell Lung Cancer: A Nonrandomized Controlled Trial.* JAMA Oncology, 2024. **10**(3): p. 352-359.

59. Ohri, N., et al., *Positron Emission Tomography–Adjusted Intensity Modulated Radiation Therapy for Locally Advanced Non-Small Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2018. **102**(4): p. 709-715.

60. Ohri, N., et al., *Pre-treatment FDG-PET predicts the site of in-field progression following concurrent chemoradiotherapy for stage III non-small cell lung cancer.* 2015. **87**(1): p. 23-27.

61. Horne, A., et al., *Treatment Response Biomarkers: Working Toward Personalized Radiotherapy for Lung Cancer.* Journal of Thoracic Oncology, 2024. **19**(8): p. 1164-1185.

62. Lambin, P., et al., *Radiomics: Extracting more information from medical images using advanced feature analysis.* European Journal of Cancer, 2012. **48**(4): p. 441-446.

63. van Timmeren, J.E., et al., *Radiomics in medical imaging—"how-to" guide and critical reflection.* Insights into Imaging, 2020. **11**(1): p. 91.

64. Li, Q., et al., *Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy.* Medical Physics, 2017. **44**(8): p. 4341-4349.

65. Cui, S., R.K. Ten Haken, and I. El Naqa, *Integrating Multiomics Information in Deep Learning Architectures for Joint Actuarial Outcome Prediction in Non-Small Cell Lung*

Cancer Patients After Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics, 2021. **110**(3): p. 893-904.

66. Fornacon-Wood, I., et al., *Radiomics as a personalized medicine tool in lung cancer: Separating the hope from the hype.* Lung Cancer, 2020. **146**: p. 197-208.
67. van Timmeren, J.E., et al., *Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images.* Radiotherapy and Oncology, 2017. **123**(3): p. 363-369.
68. Brink, C., et al., *Locoregional Control of Non-Small Cell Lung Cancer in Relation to Automated Early Assessment of Tumor Regression on Cone Beam Computed Tomography.* International Journal of Radiation Oncology*Biology*Physics, 2014. **89**(4): p. 916-923.
69. Wald, P., et al., *Prognostic Value of Primary Tumor Volume Changes on kV-CBCT during Definitive Chemoradiotherapy for Stage III Non–Small Cell Lung Cancer.* Journal of Thoracic Oncology, 2017. **12**(12): p. 1779-1787.
70. Kanzaki, H., et al., *Impact of early tumor reduction on outcome differs by histological subtype in stage III non-small-cell lung cancer treated with definitive radiotherapy*, in *Int J Clin Oncol*. 2016: Japan. p. 853-861.
71. van Timmeren, J.E., et al., *Tumor regression during radiotherapy for non-small cell lung cancer patients using cone-beam computed tomography images.* Strahlentherapie und Onkologie, 2019.
72. Käsmann, L., et al., *Predictive and prognostic value of tumor volume and its changes during radical radiotherapy*

*of			stage III non-small cell lung cancer.* Strahlentherapie und Onkologie, 2018. **194**(2): p. 79-90.

73. Ostheimer, C., et al., *Prognostic impact of gross tumor volume during radical radiochemotherapy of locally advanced non-small cell lung cancer—results from the NCT03055715 multicenter cohort study of the Young DEGRO Trial Group.* Strahlentherapie und Onkologie, 2021. **197**(5): p. 385-395.
74. In Lee, H., et al., *Predictive value of primary tumor volume change during concurrent chemoradiotherapy in patients with unresectable stage III non-small cell lung cancer.* Radiotherapy and Oncology, 2024. **198**.
75. Trommer, S., et al., *Tumor volume change at radiation boost planning to estimate the response to chemoradiotherapy in stage III unresectable NSCLC (TORCH): a multicenter retrospective observational study.* Strahlenther Onkol, 2025.
76. Kwint, M., et al., *The prognostic value of volumetric changes of the primary tumor measured on Cone Beam-CT during radiotherapy for concurrent chemoradiation in NSCLC patients.* Radiotherapy and Oncology, 2020. **146**: p. 44-51.
77. Hormuth, D.A., A.M. Jarrett, and T.E. Yankeelov, *Forecasting tumor and vasculature response dynamics to radiation therapy via image based mathematical modeling.* Radiation Oncology, 2020. **15**(1): p. 4.
78. Amugongo, L.M., et al., *Early prediction of tumour-response to radiotherapy in NSCLC patients.* Phys Med Biol, 2021. **66**(22).
79. Prokopiou, S., et al., *A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation.* Radiation Oncology (London, England), 2015. **10**: p. 159.

80. Sunassee, E.D., et al., *Proliferation saturation index in an adaptive Bayesian approach to predict patient-specific radiotherapy responses.* International Journal of Radiation Biology, 2019: p. 1-6.

81. Zahid, M.U., et al., *Proliferation Saturation Index to Characterize Response to RT and Evaluate Altered Fractionation in Head and Neck Cancer.* Appl Radiat Oncol, 2021. **10**: p. 32-39.

82. Zahid, M., et al., *Proliferation Saturation Index to Prospectively Predict Patient-Specific Responses to Radiation in Oropharyngeal Cancer.* International Journal of Radiation Oncology, Biology, Physics, 2019. **105**(1): p. E792.

83. Kutuva, A.R., et al., *Mathematical modeling of radiotherapy: impact of model selection on estimating minimum radiation dose for tumor control.* Frontiers in Oncology, 2023. **13**.

84. Zahid, M.U., et al., *Dynamics-Adapted Radiotherapy Dose (DARD) for Head and Neck Cancer Radiotherapy Dose Personalization.* Journal of Personalized Medicine, 2021. **11**(11): p. 1124.

85. Norton, L. and R. Simon, *Growth curve of an experimental solid tumor following radiotherapy.* J Natl Cancer Inst, 1977. **58**(6): p. 1735-41.

86. Enderling, H., et al., *Mathematical modelling of radiotherapy strategies for early breast cancer.* J Theor Biol, 2006. **241**(1): p. 158-71.

87. Gao, X., et al., *Acute and fractionated irradiation differentially modulate glioma stem cell division kinetics.* Cancer Res, 2013. **73**(5): p. 1481-90.

88. Jeong, J., et al., *Modeling the Cellular Response of Lung Cancer to Radiation Therapy for a Broad Range of Fractionation Schedules.* Clin Cancer Res, 2017. **23**(18): p. 5469-5479.

89. Leder, K., et al., *Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules.* Cell, 2014. **156**(3): p. 603-616.

90. McClatchy, D.M., et al., *Modeling Resistance and Recurrence Patterns of Combined Targeted-Chemoradiotherapy Predicts Benefit of Shorter Induction Period.* Cancer Res, 2020. **80**(22): p. 5121-5133.

91. Scott, J.G., et al., *Spatial Metrics of Tumour Vascular Organisation Predict Radiation Efficacy in a Computational Model.* PLoS Comput Biol, 2016. **12**(1): p. e1004712.

92. Sung, W., et al., *Mathematical Modeling to Simulate the Effect of Adding Radiation Therapy to Immunotherapy and Application to Hepatocellular Carcinoma.* Int J Radiat Oncol Biol Phys, 2022. **112**(4): p. 1055-1062.

93. Wilson, N., et al., *Modelling Radiation Cancer Treatment with a Death-Rate Term in Ordinary and Fractional Differential Equations.* Bull Math Biol, 2023. **85**(6): p. 47.

94. Brady, R. and H.A.-O. Enderling, *Mathematical Models of Cancer: When to Predict Novel Therapies, and When Not to.* (1522-9602 (Electronic)).

95. Harshe, I., H.A.-O. Enderling, and R.A.-O. Brady-Nicholls, *Predicting Patient-Specific Tumor Dynamics: How Many Measurements Are Necessary? LID - 10.3390/cancers15051368 [doi] LID - 1368.* (2072-6694 (Print)).

96. Zahid, M.U., et al., *In Silico Trial to Estimate Personalized RT Dose in Head and Neck Cancer.* International Journal of Radiation Oncology, Biology, Physics, 2021. **111**(3): p. e146-e147.

97. Caudell, J., et al., *OC-0106 Prospective Trial of Personalized Fractionation in Low-risk HPV Positive Oropharyngeal Cancerroph.* Radiotherapy and Oncology, 2023. **182**: p. S67-S68.

98. Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement.* 2015. **13**(1): p. 1.

99. Louie, A.V., et al., *Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era.* Radiotherapy and Oncology, 2010. **95**(2): p. 166-171.

100. Mercieca, S., J.S.A. Belderbos, and M. van Herk, *Challenges in the target volume definition of lung cancer radiotherapy.* Translational Lung Cancer Research, 2020. **10**(4): p. 1983-1998.

101. Cole, A.J., et al., *Motion management for radical radiotherapy in non-small cell lung cancer.* Clin Oncol (R Coll Radiol), 2014. **26**(2): p. 67-80.

102. Konert, T., et al., *PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014.* Radiotherapy and Oncology, 2015. **116**(1): p. 27-34.

103. Nestle, U., et al., *ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-*

small cell lung cancer. Radiotherapy and Oncology, 2018. **127**(1): p. 1-5.

104. Ramella, S., et al., *Local Control and Toxicity of Adaptive Radiotherapy Using Weekly CT Imaging: Results from the LARTIA Trial in Stage III NSCLC.* Journal of Thoracic Oncology, 2017. **12**(7): p. 1122-1130.
105. Jabbour, S.K., et al., *Reduction in Tumor Volume by Cone Beam Computed Tomography Predicts Overall Survival in Non-Small Cell Lung Cancer Treated With Chemoradiation Therapy.* International journal of radiation oncology, biology, physics, 2015. **92**(3): p. 627-633.
106. Woodford, C., et al., *Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images*, in *Int J Radiat Oncol Biol Phys*. 2007: United States. p. 1316-22.
107. Oliveira, C., et al., *Advanced practice roles of therapeutic radiographers/radiation therapists: A systematic literature review.* Radiography, 2022. **28**(3): p. 605-619.
108. Adair Smith, G., et al., *Evaluation of therapeutic radiographer contouring for magnetic resonance image guided online adaptive prostate radiotherapy.* Radiotherapy and Oncology, 2023. **180**.
109. van Pelt, V.W.J., et al., *Evaluation of delineating the target volume by radiation therapists in breast cancer patients.* (2405-6324 (Electronic)).
110. Batumalai, V., et al., *Interobserver Variability in Clinical Target Volume Delineation in Tangential Breast Irradiation: a Comparison between Radiation Oncologists and Radiation Therapists.* Clinical Oncology, 2011. **23**(2): p. 108-113.

111. Black, S.L., et al., *155 Advanced Practice Clinical technologist contouring GTV volumes for radical lung radiotherapy.* Lung Cancer, 2024. **190**: p. 107716.

112. Lustberg, T., et al., *Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer.* Radiotherapy and Oncology, 2018. **126**(2): p. 312-317.

113. Clark, K., et al., *The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository.* Journal of Digital Imaging, 2013. **26**(6): p. 1045-1057.

114. Hugo, G.D., et al., *A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer.* Medical Physics, 2017. **44**(2): p. 762-771.

115. Hugo, G.D., Weiss, Elisabeth, Sleeman, William C., Balik, Salim, Keall, Paul J., Lu, Jun, & Williamson, Jeffrey F., *Data from 4D Lung Imaging of NSCLC Patients.* 2016.

116. Clark, K., et al., *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.* J Digit Imaging, 2013. **26**(6): p. 1045-57.

117. Walls, G.M., et al., *Association between statin therapy dose intensity and radiation cardiotoxicity in non-small cell lung cancer: Results from the NI-HEART study.* Radiotherapy and Oncology, 2023. **186**.

118. Tariq, I., et al., *Modelling and Bayesian adaptive prediction of individual patients' tumour volume change during radiotherapy.* Physics in Medicine & Biology, 2016. **61**(5): p. 2145.

119. Chang, J.Y., et al., *Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a*

pooled analysis of two randomised trials. The Lancet Oncology. **16**(6): p. 630-637.

120. Murray, P., K. Franks, and G.G. Hanna, *A systematic review of outcomes following stereotactic ablative radiotherapy in the treatment of early-stage primary lung cancer.* The British Journal of Radiology, 2017. **90**(1071): p. 20160732.

121. Ramroth, J., et al., *Dose and Fractionation in Radiation Therapy of Curative Intent for Non-Small Cell Lung Cancer: Meta-Analysis of Randomized Trials.* International Journal of Radiation Oncology*Biology*Physics, 2016. **96**(4): p. 736-747.

122. Belderbos, J., et al., *Radiotherapy dose and fractionation for stage III NSCLC.* The Lancet Oncology, 2015. **16**(4): p. e156-e157.

123. Faivre-Finn, C., *Dose escalation in lung cancer: have we gone full circle?* The Lancet Oncology, 2015. **16**(2): p. 125-127.

124. Christodoulou, M., et al., *New radiotherapy approaches in locally advanced non-small cell lung cancer.* European Journal of Cancer, 2014. **50**(3): p. 525-534.

125. Lambin, P., et al., *Decision support systems for personalized and participative radiation oncology.* Advanced Drug Delivery Reviews, 2017. **109**: p. 131-153.

126. Haslett, K., et al., *Protocol for the isotoxic intensity modulated radiotherapy (IMRT) in stage III non-small cell lung cancer (NSCLC): a feasibility study.* BMJ Open, 2016. **6**(4).

127. Warren, M., et al., *An isotoxic planning comparison study for stage II-III non-small cell lung cancer: is intensity-*

modulated radiotherapy the answer? Clin Oncol (R Coll Radiol), 2014. **26**(8): p. 461-7.

128. Guckenberger, M., A. Kavanagh, and M. Partridge, *Combining advanced radiotherapy technologies to maximize safety and tumor control probability in stage III non-small cell lung cancer.* Strahlenther Onkol, 2012. **188**(10): p. 894-900.

129. Even, A.J., et al., *PET-based dose painting in non-small cell lung cancer: Comparing uniform dose escalation with boosting hypoxic and metabolically active sub-volumes.* Radiother Oncol, 2015. **116**(2): p. 281-6.

130. Hoffmann, A.L., et al., *Individualized dose prescription for hypofractionation in advanced non-small-cell lung cancer radiotherapy: an in silico trial.* Int J Radiat Oncol Biol Phys, 2012. **83**(5): p. 1596-602.

131. van Elmpt, W., et al., *The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer.* Radiother Oncol, 2012. **104**(1): p. 67-71.

132. van Baardwijk, A., et al., *Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer.* J Clin Oncol, 2010. **28**(8): p. 1380-6.

133. van Baardwijk, A., et al., *Mature results of a phase II trial on individualised accelerated radiotherapy based on normal tissue constraints in concurrent chemo-radiation for stage III non-small cell lung cancer.* European Journal of Cancer, 2012. **48**(15): p. 2339-2346.

134. Landau, D.B., et al., *IDEAL-CRT: A Phase 1/2 Trial of Isotoxic Dose-Escalated Radiation Therapy and Concurrent Chemotherapy in Patients With Stage II/III Non-Small Cell*

Lung Cancer. International Journal of Radiation Oncology*Biology*Physics, 2016. **95**(5): p. 1367-1377.

135. Wanet, M., et al., *An individualized radiation dose escalation trial in non-small cell lung cancer based on FDG-PET imaging.* Strahlenther Onkol, 2017.

136. Petit, S.F. and W. van Elmpt, *Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients.* Radiother Oncol, 2015. **117**(3): p. 453-8.

137. Kim, M., R.D. Stewart, and M.H. Phillips, *A feasibility study: Selection of a personalized radiotherapy fractionation schedule using spatiotemporal optimization.* Med Phys, 2015. **42**(11): p. 6671-8.

138. Zhang, P., et al., *Predictive treatment management: incorporating a predictive tumor response model into robust prospective treatment planning for non-small cell lung cancer.* Int J Radiat Oncol Biol Phys, 2014. **88**(2): p. 446-52.

139. Zhang, P., et al., *A geometric atlas to predict lung tumor shrinkage for radiotherapy treatment planning.* Phys Med Biol, 2017. **62**(3): p. 702-714.

140. Lindblom, E., A. Dasu, and I. Toma-Dasu, *Optimal fractionation in radiotherapy for non-small cell lung cancer - A modelling approach.* Acta Oncologica, 2015. **54**(9): p. 1592-1598.

141. Chvetsov, A.V., et al., *Theoretical effectiveness of cell survival in fractionated radiotherapy with hypoxia-targeted dose escalation.* Med Phys, 2017.

142. Uzan, J. and A.E. Nahum, *Radiobiologically guided optimisation of the prescription dose and fractionation scheme in radiotherapy using BioSuite.* British Journal of Radiology, 2012. **85**(1017): p. 1279-1286.

143. Keller, H., et al., *A novel dose-volume metric for optimizing therapeutic ratio through fractionation: retrospective analysis of lung cancer treatments.* Med Phys, 2013. **40**(8): p. 084101.

144. Marks, L.B., et al., *Use of Normal Tissue Complication Probability Models in the Clinic.* International Journal of Radiation Oncology • Biology • Physics. **76**(3): p. S10-S19.

145. Fleming, C., et al., *Normal tissue considerations and dose–volume constraints in the moderately hypofractionated treatment of non-small cell lung cancer.* Radiotherapy and Oncology, 2016. **119**(3): p. 423-431.

146. Speirs, C.K., et al., *Heart Dose Is an Independent Dosimetric Predictor of Overall Survival in Locally Advanced Non–Small Cell Lung Cancer.* Journal of Thoracic Oncology. **12**(2): p. 293-301.

147. Della Gala, G., et al., *Fully automated VMAT treatment planning for advanced-stage NSCLC patients.* Strahlenther Onkol, 2017. **193**(5): p. 402-409.

148. Lewin, T., et al., *Proliferation Saturation Index Predicts Oropharyngeal Squamous Cell Cancer Gross Tumor Volume Reduction to Prospectively Identify Patients for Adaptive Radiation Therapy.* International Journal of Radiation Oncology • Biology • Physics. **94**(4): p. 903.

149. Poleszczuk, J., et al., *Predicting Patient-Specific Radiotherapy Protocols Based on Mathematical Model*

Choice for Proliferation Saturation Index. Bull Math Biol, 2017.

150. Hudson, A., et al., *Is heterogeneity in stage 3 non-small cell lung cancer obscuring the potential benefits of dose-escalated concurrent chemo-radiotherapy in clinical trials?* Lung Cancer, 2018. **118**: p. 139-147.
151. Njeh, C.F., *Tumor delineation: The weakest link in the search for accuracy in radiotherapy.* Journal of medical physics, 2008. **33**(4): p. 136-140.
152. van Dam, I.E., et al., *Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool.* Radiotherapy and Oncology, 2010. **96**(1): p. 67-72.
153. Daisne, J.-F. and A. Blumhofer, *Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation.* Radiation Oncology, 2013. **8**(1): p. 154.
154. Walls, G.M., et al., *Validation of an established deep learning auto-segmentation tool for cardiac substructures in 4D radiotherapy planning scans.* Physics and Imaging in Radiation Oncology, 2022. **23**: p. 118-126.
155. Wong Yuzhen, N. and S. Barrett, *A review of automatic lung tumour segmentation in the era of 4DCT.* Reports of Practical Oncology & Radiotherapy, 2019. **24**(2): p. 208-220.
156. Velazquez, E.R., et al., *Volumetric CT-based segmentation of NSCLC using 3D-Slicer.* Scientific reports, 2013. **3**: p. 3529-3529.

157. Møller, D.S., et al., *Adaptive radiotherapy for advanced lung cancer ensures target coverage and decreases lung dose.* Radiotherapy and Oncology, 2016. **121**(1): p. 32-38.

158. Apolle, R., et al., *Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer.* Acta Oncologica, 2019: p. 1-8.

159. Hanna, G.G., A.R. Hounsell, and J.M. O'Sullivan, *Geometrical Analysis of Radiotherapy Target Volume Delineation: a Systematic Review of Reported Comparison Methods.* Clinical Oncology, 2010. **22**(7): p. 515-525.

160. Gaede, S., et al., *An evaluation of an automated 4D-CT contour propagation tool to define an internal gross tumour volume for lung cancer radiotherapy.* Radiotherapy and Oncology, 2011. **101**(2): p. 322-328.

161. Martin Bland, J. and D. Altman, *Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement.* The Lancet, 1986. **327**(8476): p. 307-310.

162. Giavarina, D., *Understanding Bland Altman analysis.* Biochemia medica, 2015. **25**(2): p. 141-151.

163. Goldstein, H., *Multilevel mixed linear model analysis using iterative generalized least squares.* Biometrika, 1986. **73**(1): p. 43-56.

164. Eldesoky, A.R., et al., *Internal and external validation of an ESTRO delineation guideline – dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer.* Radiotherapy and Oncology, 2016. **121**(3): p. 424-430.

165. Wright, J.L., et al., *Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO*

Consensus Paper. Practical Radiation Oncology, 2019. **9**(2): p. 65-72.

166. Reinke, A., et al., *Understanding metric-related pitfalls in image analysis validation.* Nature Methods, 2024. **21**(2): p. 182-194.

167. Tsang, Y., et al., *Assessment of contour variability in target volumes and organs at risk in lung cancer radiotherapy.* Technical Innovations & Patient Support in Radiation Oncology, 2019. **10**: p. 8-12.

168. Berkovic, P., et al., *Adaptive radiotherapy for locally advanced non-small cell lung cancer, can we predict when and for whom?* Acta Oncol, 2015. **54**(9): p. 1438-44.

169. Britton, K.R., et al., *Consequences of Anatomic Changes and Respiratory Motion on Radiation Dose Distributions in Conformal Radiotherapy for Locally Advanced Non–Small-Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2009. **73**(1): p. 94-102.

170. Møller, D.S., et al., *Survival benefits for non-small cell lung cancer patients treated with adaptive radiotherapy.* Radiotherapy and Oncology, 2022. **168**: p. 234-240.

171. Tvilum, M., et al., *Different benefits of adaptive radiotherapy for different histologies of NSCLC.* Acta Oncol, 2023. **62**(11): p. 1426-1432.

172. Shepherd, M., et al., *Pathway for radiation therapists online advanced adapter training and credentialing.* Technical Innovations & Patient Support in Radiation Oncology, 2021. **20**: p. 54-60.

173. Willigenburg, T., et al., *Evaluation of daily online contour adaptation by radiation therapists for prostate cancer*

treatment on an MRI-guided linear accelerator. Clinical and Translational Radiation Oncology, 2021. **27**: p. 50-56.

174. Richmond, N.D., et al., *Positioning accuracy for lung stereotactic body radiotherapy patients determined by on-treatment cone-beam CT imaging.* The British Journal of Radiology, 2012. **85**(1014): p. 819-823.
175. Oechsner, M., et al., *Registration uncertainties between 3D cone beam computed tomography and different reference CT datasets in lung stereotactic body radiation therapy.* Radiation Oncology, 2016. **11**(1): p. 142.
176. Dumas, M., et al., *Comparison of the internal target volume (ITV) contours and dose calculations on 4DCT, average CBCT, and 4DCBCT imaging for lung stereotactic body radiation therapy (SBRT).* Journal of applied clinical medical physics, 2020. **21**(11): p. 288-294.
177. Mohamed, A.A., et al., *Clinical validation of a semi-automated segmentation algorithm for target volume definition on planning CT and CBCT in stereotactic body radiotherapy (SBRT) for peripheral lung lesions.* Journal of Medical Radiation Sciences, 2023. **70**(S2): p. 37-47.
178. Ebadi, N., et al., *CBCT-guided adaptive radiotherapy using self-supervised sequential domain adaptation with uncertainty estimation.* Medical Image Analysis, 2023. **86**: p. 102800.
179. He, Y., et al., *Geometric and dosimetric accuracy of deformable image registration between average-intensity images for 4DCT-based adaptive radiotherapy for non-small cell lung cancer.* Journal of Applied Clinical Medical Physics, 2021. **22**(8): p. 156-167.

180. Huesa-Berral, C., et al., *Detailed dosimetric evaluation of inter-fraction and respiratory motion in lung stereotactic body radiation therapy based on daily 4D cone beam CT images.* Physics in Medicine & Biology, 2023. **68**(1): p. 015005.

181. Nenoff, L., et al., *Review and recommendations on deformable image registration uncertainties for radiotherapy applications.* Physics in Medicine & Biology, 2023. **68**(24): p. 24TR01.

182. Barrett, S., et al., *Geometric and Dosimetric Evaluation of a Commercially Available Auto-segmentation Tool for Gross Tumour Volume Delineation in Locally Advanced Non-small Cell Lung Cancer: a Feasibility Study.* Clin Oncol (R Coll Radiol), 2021. **33**(3): p. 155-162.

183. Lindberg, J., et al., *A national perspective about the current work situation at modern radiotherapy departments.* Clin Transl Radiat Oncol, 2020. **24**: p. 127-134.

184. Mackay, K., et al., *A Review of the Metrics Used to Assess Auto-Contouring Systems in Radiotherapy.* Clinical Oncology, 2023. **35**(6): p. 354-369.

185. Josipovic, M., et al., *Deep inspiration breath-hold radiotherapy for lung cancer: impact on image quality and registration uncertainty in cone beam CT image guidance.* Br J Radiol, 2016. **89**(1068): p. 20160544.

186. Josipovic, M., et al., *Deep inspiration breath hold in locally advanced lung cancer radiotherapy: validation of intrafractional geometric uncertainties in the INHALE trial.* Br J Radiol, 2019. **92**(1104): p. 20190569.

187. Moran, K., C. Poole, and S. Barrett, *Evaluating deep learning auto-contouring for lung radiation therapy: A*

review of accuracy, variability, efficiency and dose, in target volumes and organs at risk. Physics and Imaging in Radiation Oncology, 2025.

188. Rasing, M.J.A., et al., *Online adaptive MR-guided radiotherapy: Conformity of contour adaptation for prostate cancer, rectal cancer and lymph node oligometastases among radiation therapists and radiation oncologists.* Technical Innovations & Patient Support in Radiation Oncology, 2022. **23**: p. 33-40.
189. Goudschaal, K., et al., *Clinical implementation of RTT-only CBCT-guided online adaptive focal radiotherapy for bladder cancer.* Clinical and Translational Radiation Oncology, 2025. **50**: p. 100884.
190. van Pelt, V.W.J., et al., *Evaluation of delineating the target volume by radiation therapists in breast cancer patients.* Technical Innovations & Patient Support in Radiation Oncology, 2021. **17**: p. 78-81.
191. Michienzi, A., et al., *Cone-beam computed tomography for lung cancer–validation with CT and monitoring tumour response during chemo-radiation therapy.* Journal of medical imaging and radiation oncology, 2017. **61**(2): p. 263-270.
192. Das, I.J., et al., *Intra- and inter-physician variability in target volume delineation in radiation therapy.* Journal of Radiation Research, 2021. **62**(6): p. 1083-1089.
193. Hou, Y., et al., *Inter-scan and inter-observer tumour volume delineation variability on cone beam computed tomography in patients treated with stereotactic body radiation therapy for early-stage non-small cell lung cancer.* Journal of Medical Imaging and Radiation Oncology, 2017. **61**(1): p. 93-98.

194. Mao, W., et al., *Evaluation of Auto-Contouring and Dose Distributions for Online Adaptive Radiation Therapy of Patients With Locally Advanced Lung Cancers.* Practical Radiation Oncology, 2022. **12**(4): p. e329-e338.

195. Moran, K., C. Poole, and S. Barrett, *Evaluating deep learning auto-contouring for lung radiation therapy: A review of accuracy, variability, efficiency and dose, in target volumes and organs at risk.* Physics and Imaging in Radiation Oncology.

196. Hosny, A., et al., *Clinical validation of deep learning algorithms for radiotherapy targeting of non-small-cell lung cancer: an observational study.* The Lancet Digital Health, 2022. **4**(9): p. e657-e666.

197. Wang, S., et al., *Automated Lung Cancer Segmentation Using a PET and CT Dual-Modality Deep Learning Neural Network.* International Journal of Radiation Oncology*Biology*Physics, 2023. **115**(2): p. 529-539.

198. Lim, G., et al., *Tumor Regression and Positional Changes in Non-small Cell Lung Cancer During Radical Radiotherapy.* Journal of Thoracic Oncology, 2011. **6**(3): p. 531-536.

199. Warth, A., et al., *Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role.* British Journal of Cancer, 2014. **111**(6): p. 1222-1229.

200. Piperdi, H., et al., *Adaptive Radiation Therapy in the Treatment of Lung Cancer: An Overview of the Current State of the Field.* Front Oncol, 2021. **11**: p. 770382.

201. Hoppen, L., et al., *Dosimetric benefits of adaptive radiation therapy for patients with stage III non-small cell lung cancer.* 2023. **18**(1): p. 34.

202. Liting Shi and Yi Rong and Megan Daly and Brandon Dyer and Stanley Benedict and Jianfeng Qiu and Tokihiro, Y., *Cone-beam computed tomography-based delta-radiomics for early response assessment in radiotherapy for locally advanced lung cancer.* Physics in Medicine & Biology, 2020. **65**(1): p. 015009.

203. Bertholet, J., et al., *Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part II: Offline and online plan adaption for interfractional changes.* Radiother Oncol, 2020. **153**: p. 88-96.

204. Davey, A., et al., *Automated gross tumor volume contour generation for large-scale analysis of early-stage lung cancer patients planned with 4D-CT.* Med Phys, 2021. **48**(2): p. 724-732.

205. Knap, M.M., et al., *Daily cone-beam computed tomography used to determine tumour shrinkage and localisation in lung cancer patients.* Acta Oncologica, 2010. **49**(7): p. 1077-1084.

206. Burnham, K.P. and D.R. Anderson, *Multimodel Inference: Understanding AIC and BIC in Model Selection.* Sociological Methods & Research, 2004. **33**(2): p. 261-304.

207. Park, L.J., et al., *Application of genetic algorithms to parameter estimation of bioprocesses.* (0140-0118 (Print)).

208. Akaike, H., *Information Theory and an Extension of the Maximum Likelihood Principle*, in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa, Editors. 1998, Springer New York: New York, NY. p. 199-213.

209. Sonke, J.J. and J. Belderbos, *Adaptive radiotherapy for lung cancer.* Semin Radiat Oncol, 2010. **20**(2): p. 94-106.

210. Hattu, D., et al., *A traffic light protocol workflow for image-guided adaptive radiotherapy in lung cancer patients.* 2022. **175**: p. 152-158.

211. Kasmann, L., et al., *Predictive and prognostic value of tumor volume and its changes during radical radiotherapy of stage III non-small cell lung cancer : A systematic review.* Strahlenther Onkol, 2018. **194**(2): p. 79-90.

212. Lee, et al.,., *Predictive value of primary tumor volume change during concurrent chemoradiotherapy in patients with unresectable stage III non-small cell lung cancer.* Radiotherapy and Oncology, 2024. **198**.

213. Gronau, Q.F. and E.-J. Wagenmakers, *Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection.* Computational Brain & Behavior, 2019. **2**(1): p. 1-11.

214. Coates, J. and I. El Naqa, *Outcome modeling techniques for prostate cancer radiotherapy: Data, models, and validation.* Physica Medica, 2016. **32**(3): p. 512-520.

215. Cui, S., et al., *Artificial Intelligence for Outcome Modeling in Radiotherapy.* Seminars in Radiation Oncology, 2022. **32**(4): p. 351-364.

216. Ramspek, C.L., et al., *External validation of prognostic models: what, why, how, when and where?* Clinical Kidney Journal, 2020. **14**(1): p. 49-58.

217. El Naqa, I., *Prospective clinical deployment of machine learning in radiation oncology.* Nature Reviews Clinical Oncology, 2021. **18**(10): p. 605-606.

218. Loong, T.W., *Understanding sensitivity and specificity with the right side of the brain.* BMJ, 2003. **327**(7417): p. 716-9.

219. Peduzzi, P., et al., *A simulation study of the number of events per variable in logistic regression analysis.* Journal of clinical epidemiology, 1996. **49**(12): p. 1373-1379.

220. Barrett S., Z.M.U., Enderling H., Marignol L., *Predicting Individual Tumour Response Dynamics in Locally Advanced Non-Small Cell Lung Cancer Radiation Therapy; a Mathematical Modelling Study.* International Journal of Radiation Oncology • Biology • Physics, 2025, **121** (4): 1077-1087

221. Appel, S., et al., *Image-guidance triggered adaptive replanning of radiation therapy for locally advanced lung cancer: an evaluation of cases requiring plan adaptation.* British Journal of Radiology, 2019. **93**(1105).

222. Damhuis, R.A.M., et al., *Age-related treatment patterns for stage I NSCLC in three European countries.* Journal of Geriatric Oncology, 2021. **12**(8): p. 1214-1219.

223. Oberije, C., et al., *A Validated Prediction Model for Overall Survival From Stage III Non-Small Cell Lung Cancer: Toward Survival Prediction for Individual Patients.* International Journal of Radiation Oncology*Biology*Physics, 2015. **92**(4): p. 935-944.

224. Koo, T.R., et al., *The effect of tumor volume and its change on survival in stage III non-small cell lung cancer treated with definitive concurrent chemoradiotherapy.* Radiation Oncology, 2014. **9**(1): p. 283.

225. Elsayad, K., et al., *Is tumor volume reduction during radiotherapy prognostic relevant in patients with stage III non-small cell lung cancer?* Journal of Cancer Research and Clinical Oncology, 2018. **144**(6): p. 1165-1171.

226. Everitt, S., et al., *Imaging Cellular Proliferation During Chemo-Radiotherapy: A Pilot Study of Serial 18F-FLT Positron Emission Tomography/Computed Tomography Imaging for Non–Small-Cell Lung Cancer.* International Journal of Radiation Oncology*Biology*Physics, 2009. **75**(4): p. 1098-1104.

227. Yang, F., et al., *CT-based radiomics signatures can predict the tumor response of non-small cell lung cancer patients treated with first-line chemotherapy and targeted therapy.* European Radiology, 2022. **32**(3): p. 1538-1547.

228. Hindocha, S., et al., *A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models.* eBioMedicine, 2022. **77**.

229. Jochems, A., et al., *Developing and Validating a Survival Prediction Model for NSCLC Patients Through Distributed Learning Across 3 Countries.* International Journal of Radiation Oncology*Biology*Physics, 2017. **99**(2): p. 344-352.

230. Chen, X., et al., *Establishing a deep learning model that integrates pre- and mid-treatment computed tomography to predict treatment response for non-small cell lung cancer.* International Journal of Radiation Oncology *Biology* Physics, 2025.

231. Smith, G., *Step away from stepwise.* Journal of Big Data, 2018. **5**(1): p. 32.

232. Scott, J.G., et al., *Personalizing Radiotherapy Prescription Dose Using Genomic Markers of Radiosensitivity and*

Normal Tissue Toxicity in NSCLC. J Thorac Oncol, 2021. **16**(3): p. 428-438.

233. Walls, G.M., B. Pellini, and A.A. Chaudhuri, *Circulating tumor DNA kinetics: A future tool for radiation therapy personalization in lung cancer?* The Journal of Liquid Biopsy, 2024. **6**: p. 100160.

234. Osman, A.F.I., *Radiation Oncology in the Era of Big Data and Machine Learning for Precision Medicine*, in *Artificial Intelligence - Applications in Medicine and Biology*, M.A. Aceves-Fernández, Editor. 2019, IntechOpen: Rijeka.

235. Deasy, J.O., *Data Science Opportunities To Improve Radiotherapy Planning and Clinical Decision Making.* Seminars in Radiation Oncology, 2024. **34**(4): p. 379-394.

236. Ohri, N., et al., *Positron Emission Tomography-Adjusted Intensity Modulated Radiation Therapy for Locally Advanced Non-Small Cell Lung Cancer.* Int J Radiat Oncol Biol Phys, 2018. **102**(4): p. 709-715.

237. Iyengar, P., et al., *Accelerated Hypofractionated Image-Guided vs Conventional Radiotherapy for Patients With Stage II/III Non–Small Cell Lung Cancer and Poor Performance Status: A Randomized Clinical Trial.* JAMA Oncology, 2021. **7**(10): p. 1497-1505.

238. Brouwer, C.L., et al., *Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help?* Radiotherapy and Oncology, 2015. **115**(3): p. 285-294.

239. Iliadou, V., et al., *Early Prediction of Planning Adaptation Requirement Indication Due to Volumetric Alterations in*

Head and Neck Cancer Radiotherapy: A Machine Learning Approach. Cancers (Basel), 2022. **14**(15).

240. Brown, E., et al., *Predicting the need for adaptive radiotherapy in head and neck cancer.* Radiother Oncol, 2015. **116**(1): p. 57-63.
241. Bhalla, A.S., et al., *Imaging protocols for CT chest: A recommendation.* Indian J Radiol Imaging, 2019. **29**(3): p. 236-246.
242. Lynch, C., S.P. Pitroda, and R.R. Weichselbaum, *Radiotherapy, immunity, and immune checkpoint inhibitors.* Lancet Oncol, 2024. **25**(8): p. e352-e362.
243. Lim, S.Y., et al., *Size matters: integrating tumour volume and immune activation signatures predicts immunotherapy response.* Molecular Cancer, 2024. **23**(1): p. 228.
244. Citrin, D. and Z.S. Morris, *Advancing Towards Personalized Prescription of Radiotherapy Dose.* Seminars in Radiation Oncology, 2023. **33**(3): p. 219-220.
245. Craddock, M., et al., *Evaluation of Prognostic and Predictive Models in the Oncology Clinic.* Clinical Oncology, 2022. **34**(2): p. 102-113.
246. Zhang, J., et al., *A nomogram to predict overall survival of patients with early stage non-small cell lung cancer.* J Thorac Dis, 2019. **11**(12): p. 5407-5416.
247. Forde, P.M., et al., *Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer.* N Engl J Med, 2022. **386**(21): p. 1973-1985.
248. Houda, I., et al., *An international and multidisciplinary EORTC survey on resectability of stage III non-small cell lung cancer.* Lung Cancer, 2025. **199**: p. 108061.

249. Claire, P., *Embracing rapid learning in radiotherapy: feasible and acceptable with stakeholder corroboration.* BMJ Oncology, 2024. **3**(1): p. e000327.

250. Barrett, S., *On-Treatment Verification for Lung Stereotactic Ablative Radiation Therapy.* Journal of Medical Imaging and Radiation Sciences, 2017. **48**(4): p. 343-345.

251. Browning, A.P., et al., *Predicting Radiotherapy Patient Outcomes with Real-Time Clinical Data Using Mathematical Modelling.* Bulletin of Mathematical Biology, 2024. **86**(2): p. 19.

252. Zahid, M.U., et al., *Forecasting Individual Patient Response to Radiation Therapy in Head and Neck Cancer With a Dynamic Carrying Capacity Model.* International Journal of Radiation Oncology*Biology*Physics, 2021. **111**(3): p. 693-704.

253. Bradley, J.D., et al., *Long-Term Results of NRG Oncology RTOG 0617: Standard- Versus High-Dose Chemoradiotherapy With or Without Cetuximab for Unresectable Stage III Non-Small-Cell Lung Cancer.* J Clin Oncol, 2020. **38**(7): p. 706-714.

254. Wei, D.-m., et al., *Augmented expression of Ki-67 is correlated with clinicopathological characteristics and prognosis for lung cancer patients: an up-dated systematic review and meta-analysis with 108 studies and 14,732 patients.* Respiratory Research, 2018. **19**(1): p. 150.

255. Bollineni, V.R., et al., *A systematic review on [18F]FLT-PET uptake as a measure of treatment response in cancer patients.* European Journal of Cancer, 2016. **55**: p. 81-97.

256. Trigonis, I., et al., *Early reduction in tumour [18F]fluorothymidine (FLT) uptake in patients with non-*

small cell lung cancer (NSCLC) treated with radiotherapy alone. Eur J Nucl Med Mol Imaging, 2014. **41**(4): p. 682-93.

257. Kwint, M., et al., *Intra thoracic anatomical changes in lung cancer patients during the course of radiotherapy.* Radiother Oncol, 2014. **113**(3): p. 392-7.

258. Harris, J.P., et al., *Adapt or Perish: Adaptive RT for NSCLC.* International Journal of Radiation Oncology*Biology*Physics, 2024. **119**(4): p. 1047-1051.

259. Guckenberger, M., et al., *Adaptive radiotherapy for locally advanced non-small-cell lung cancer does not underdose the microscopic disease and has the potential to increase tumor control.* Int J Radiat Oncol Biol Phys, 2011. **81**(4): p. e275-82.

260. Davey, A., et al., *Predicting cancer relapse following lung stereotactic radiotherapy: an external validation study using real-world evidence.* Frontiers in Oncology, 2023. **13**.

261. Davey, A., et al., *OC-0096: Peritumoural density as a biomarker of distant failure in NSCLC patients treated with SABR.* Radiotherapy and Oncology, 2020. **152**: p. S44-S45.

262. Appel, S., et al., *Image-guidance triggered adaptive replanning of radiation therapy for locally advanced lung cancer: an evaluation of cases requiring plan adaptation.* Br J Radiol, 2020. **93**(1105): p. 20190743.

263. Schytte, T., et al., *Overall survival following heterogeneous FDG-guided dose-escalation for locally advanced NSCLC in the international phase III NARLAL2 trial.* 2024, American Society of Clinical Oncology.

264. Lee, Y.H., et al., *Tumour volume changes assessed with high-quality KVCT in lung cancer patients undergoing concurrent*

chemoradiotherapy. Br J Radiol, 2015. **88**(1052): p. 20150156.

265. Moller, D.S., et al., *A planning study of radiotherapy dose escalation of PET-active tumour volumes in non-small cell lung cancer patients.* Acta Oncol, 2011. **50**(6): p. 883-8.

266. Fox, J., et al., *Quantification of tumor volume changes during radiotherapy for non-small-cell lung cancer.* Int J Radiat Oncol Biol Phys, 2009. **74**(2): p. 341-8.

267. Henke, L.E., et al., *Stereotactic MR-Guided Online Adaptive Radiation Therapy (SMART) for Ultracentral Thorax Malignancies: Results of a Phase 1 Trial.* Advances in Radiation Oncology, 2019. **4**(1): p. 201-209.

268. Giacometti, V., A.R. Hounsell, and C.K. McGarry, *A review of dose calculation approaches with cone beam CT in photon and proton therapy.* Physica Medica, 2020. **76**: p. 243-276.

269. Duan, J., et al., *Enhancing Precision in Radiation Therapy for Locally Advanced Lung Cancer: A Case Study of Cone-Beam Computed Tomography (CBCT)-Based Online Adaptive Techniques and the Promise of HyperSight™ Iterative CBCT.* Cureus, 2024. **16**(8): p. e66943.

270. MacDonald, R.L., et al., *The feasibility of CT simulation-free adaptive radiation therapy.* Journal of Applied Clinical Medical Physics, 2024. **25**(9): p. e14438.

271. Robar, J.L., et al., *Novel Technology Allowing Cone Beam Computed Tomography in 6 Seconds: A Patient Study of Comparative Image Quality.* Practical Radiation Oncology, 2024. **14**(3): p. 277-286.

272. Duan, J., et al., *Assessing Dosimetric Benefits of CBCT-Guided Online Adaptive Radiation Treatment Frequencies for Lung Cancer.* Advances in Radiation Oncology, 2025.

273. Beckert, R., et al., *The impact of an Advanced Practice Radiation Therapist contouring for a CBCT-based adaptive radiotherapy program.* Technical Innovations & Patient Support in Radiation Oncology, 2024. **30**: p. 100242.

274. Seymour, L., et al., *iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics.* Lancet Oncol, 2017. **18**(3): p. e143-e152.

275. Ferrara, R., et al., *Pseudoprogression in Non–Small Cell Lung Cancer upon Immunotherapy: Few Drops in the Ocean?* Journal of Thoracic Oncology, 2019. **14**(3): p. 328-331.

276. Brown, J.M., D.J. Carlson, and D.J. Brenner, *The Tumor Radiobiology of SRS and SBRT: Are More Than the 5 Rs Involved?* International Journal of Radiation Oncology, Biology, Physics, 2014. **88**(2): p. 254-262.

PUBLICATIONS

Radiotherapy and Oncology 128 (2018) 520–533

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Lung cancer

An overview on personalisation of radiotherapy prescriptions in locally advanced non-small cell lung cancer: Are we there yet?

Sarah Barrett [a,*], Gerard G. Hanna [b], Laure Marignol [a]

[a] Applied Radiation Therapy Trinity, Discipline of Radiation Therapy, Trinity College Dublin, Dublin, Ireland; and [b] Centre for Cancer Research and Cell Biology, Queen's University Belfast, United Kingdom

ARTICLE INFO

Article history:
Received 6 March 2018
Received in revised form 21 May 2018
Accepted 28 May 2018
Available online 13 June 2018

Keywords:
Non-small-cell lung cancer
Personalised prescriptions
Radiotherapy
Personalised medicine
Dose fractionation

ABSTRACT

Standard of care radiotherapy in LA-NSCLC is 60–66 Gy in 30–33 fractions. However outcomes for these patients are poor with 5-year survival in the range of 10–20%. Randomised controlled trials have shown that dose escalation in a linear fashion does not improve outcomes for all patients, thus there is a need to tailor the prescription to the individual patient. This review assesses the strategies published to personalise the radiation therapy dose prescription in LA-NSCLC. A systematic and scoping search of the literature was performed to identify studies that met the inclusion criteria. 19 relevant studies were identified ranging from prospective clinical trials to mathematically modelled concept studies. Heterogeneity existed between all clinical studies. Nine heterogeneous publications proposed methodology to adapt the dose prescription to the individual patient. A number of encouraging strategies have been identified but fall short of the evidence level required to influence clinical practice.

Non-Small Cell Lung Cancer (NSCLC) is the most common form of lung cancer and with poor treatment outcomes and an age-standardised 5-year survival reported to be 10–20% world-wide [1]. Early stage disease treatment for medically inoperable patients has been revolutionised with the advent of stereotactic ablative body radiotherapy (SABR), which has become a standard of care for patients with medically inoperable NSCLC and which has comparable outcomes to surgery for Stage I disease [2,3]. Outcomes in the locally advanced (LA) setting remain poor with little improvement in overall survival (OS) over the past 40 years [4]. Radiotherapy (RT) is a mainstay of treatment for LA-NSCLC and is delivered either as radiotherapy alone or in conjunction with chemotherapy either concurrently or sequentially. An established dose fractionation of 60–66 Gy in 30–33 fractions remains the RT standard of care [5] for patients with LA-NSCLC.

The benefit of RT dose escalation has been explored and a dose response relationship has been reported in the setting of curative intent RT for NSCLC [6]. Theoretically, dose escalation is key to improving local control (LC), progression free survival (PFS) and OS. The randomised control trial, RTOG 0617, sought to compare outcomes following standard dose RT of 60 Gy versus a higher dose of 74 Gy in LA-NSCLC [7]. Dose escalation was found to be associated with worse outcomes and poorer overall survival. A more recent meta-analysis investigating RT dose response relationships in NSCLC reported survival benefits for escalation in the RT alone setting and no survival benefit in the concurrent chemoradiotherapy (CRT) setting [8].

It has been hypothesised that any potential benefit of dose escalation has not translated clinically in the findings of the RTOG 0617 due, at least in part, to the increased toxicity induced from both the RT and chemotherapy regimens as well as potential variation in RT delivery and quality assurance [9,10]. The overall extension of treatment duration in a conventional 2 Gy per day fractionation schedule, may be a contributing factor to the poorer outcomes given that longer treatment times may facilitate tumour repopulation. As a result adapting the dose fractionation schedule, without extending overall treatment duration, has been suggested as an alternative approach [11].

While the evidence is somewhat conflicting, it seems clear that universal dose escalation in a linear fashion is not the solution for this patient cohort. It thus follows that adopting a personalised, patient specific approach to modifying the radiotherapy prescription may permit safe RT dose escalation [11].

While recent advances in RT technology have been vast, the evolution is now moving away from increasing the geometric capabilities of the equipment and towards personalising and individualising treatment parameters to a patient's specific needs and

* Corresponding author at: Applied Radiation Therapy Trinity, Discipline of Radiation Therapy, Trinity College Dublin, Dublin, Ireland.
E-mail addresses: barrets7@tcd.ie (S. Barrett), g.hanna@qub.ac.uk (G.G. Hanna), marignol@tcd.ie (L. Marignol).

https://doi.org/10.1016/j.radonc.2018.05.029
0167-8140/

requirements [12]. The poor OS associated with the treatment of LA-NSCLC with RT calls for an urgent need to transform outcomes for patients. Personalisation of radiotherapy dose prescription and safe dose escalation may be one strategy to address this.

Isotoxic treatment planning has been suggested as a method of personalising radiotherapy for each patient based on their individual treatment plan [13]. In isotoxic RT planning, the prescribed dose is increased incrementally until an organ at risk (OAR) maximum tolerated dose is reached. As a result, each patient will have a personalised prescription. However, this approach of dose escalation is not based on any patient or disease specific features.

Several methodologies have been reported in the literature to modify LA-NSCLC prescriptions, which may prove more successful than RTOG 0617 in improving outcomes for these patients. This review aims to assess both the on-going clinical work being carried out in developing modified dose fractionation schedules, as well as evaluate emerging novel proposals to personalise LA-NSCLC radiotherapy prescriptions.

Search strategy methodology

A systematic search of the literature was carried out in PubMed and Embase databases using the key words 'Non-small cell lung cancer' 'Personalised Radiotherapy' 'Dose Optimization' and associated Mesh Terms. See Supplementary Material for complete search strategy. A scoping search of the literature was also performed.

Inclusion criteria were defined as personalised radiotherapy studies relating to NSCLC (including a pre-treatment change in dose per fraction, escalation, de-escalation based on Tumour Control Probability (TCP) or Normal Tissue Complication Probability (NTCP)), modelled personalised radiotherapy studies and studies published after January 2000. Exclusion criteria were defined as any other site (including small cell lung cancer), adaptive studies based on mid-treatment response rather than pre-treatment personalisation, change to chemotherapy prescription, no change to RT prescription, stereotactic radiotherapy studies, protons/carbon ion studies, genetic or biomarker based studies and non-EBRT escalation studies. Studies that reported prescription modification for stratified groups of patients were excluded and only those that individually adjusted the prescription on a per patient basis were included.

Due to heterogeneity between the study types and a limited sample size secondary analysis of the data was not feasible. As a result, the findings were evaluated in a qualitative manner. 19 relevant publications were identified (Fig. 1) [14–32].

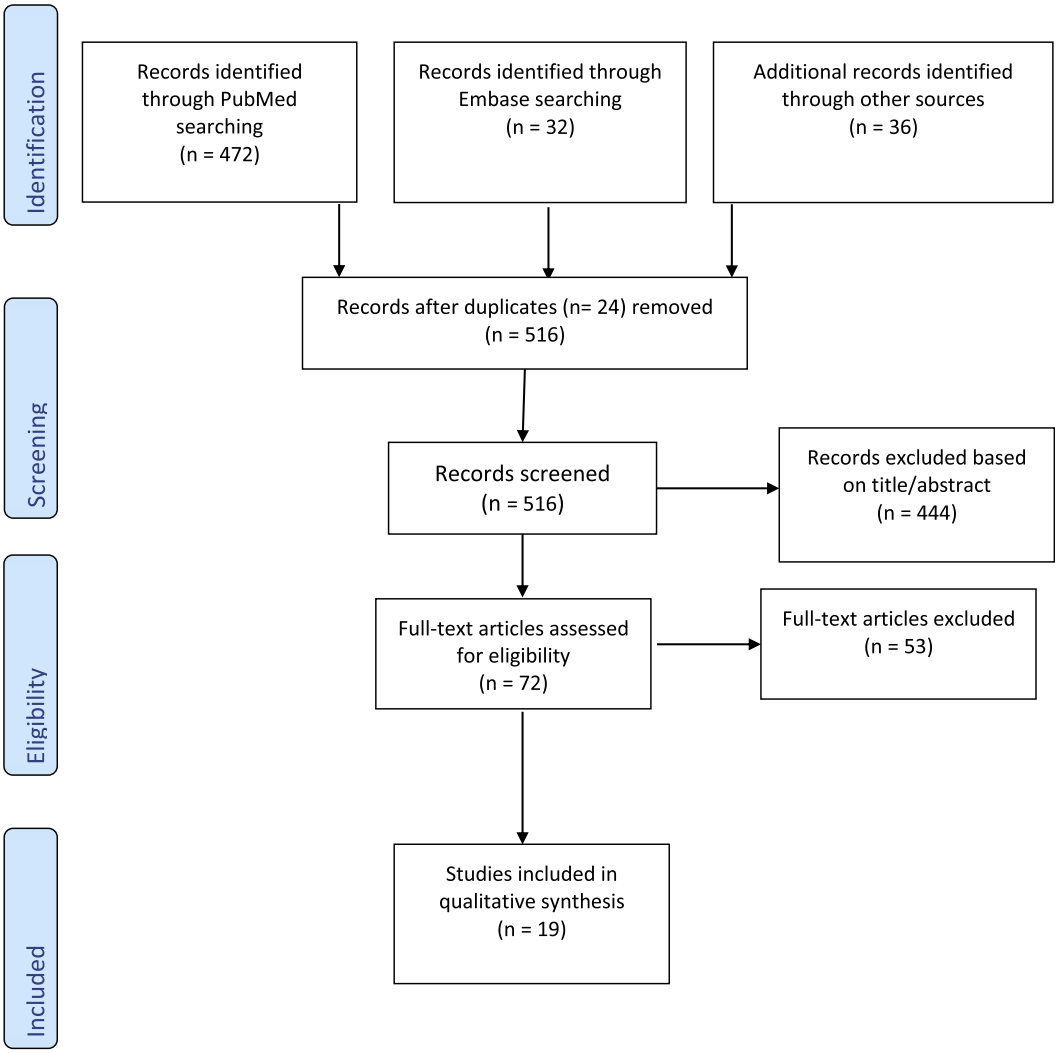


Fig. 1. Flow chart of the literature search process.

Results

Feasibility of dose modification in LA-NSCLC – in-silico studies

Five retrospective treatment planning studies [14–18] were identified which evaluated the feasibility of adapting the NSCLC RT prescription dose, based essentially on isotoxic planning and established OAR dose volume constraints. Findings from these four studies are summarised in Table 1.

Two studies [14,15] compared the effectiveness of RT techniques for the purpose of dose modification by evaluating 3DCRT, IMRT, simultaneous integrated boost (SIB) and adaptive RT (ART) techniques. The remaining three [16–18] focused solely on the dose adaptation using their current institutional planning techniques. All five studies concluded that a form of prescription personalisation via uniform dose escalation or SIB is achievable while adhering to accepted dose volume constraints on surrounding normal tissue. Mean Lung Dose (MLD) was the most frequently reported dose limiting parameter.

While all five planning studies [14–18] reported that dose modification was feasible in most patients with NSCLC Stage Ib-III, this was a small cohort of patients, $n = 101$ in total. Additionally, a variety of technologies (3DCRT, IMRT, VMAT and helical IMRT), techniques (ART, SIB), and dose modification approaches (1.8 Gy twice daily, 2.5 Gy-5.4 Gy daily) were reported making these findings difficult to interpret in a clinically meaningful way. While the treatment prescription was personalised to the patient based on their normal anatomy and target volumes, the overarching aim was to escalate the dose for all patients rather than ascertain differences between those that may or may not benefit from a modified dose prescription. RTOG 0617 has demonstrated that this may not be beneficial to all patients [7].

Clinical evidence for personalised radiotherapy prescriptions

Following the establishment of feasibility, prospective clinical trials were implemented to evaluate the use of personalised prescriptions in a patient population. Five prospective clinical trials [19–23] were identified that utilised isotoxic treatment planning techniques to individually prescribe RT for LA-NSCLC. Findings from these five trials are summarised in Table 2.

There was sizeable heterogeneity between these trials. The study by van Baardwijk et al. (2010) and Wanet et al. were prospective single arm studies [19,23] while the remaining three studies were Phase I and Phase II clinical trials [20–22]. The patient population inclusion criteria varied. Two studies [19,21] included only patients receiving RT, whereas patients receiving concurrent CRT were included in the remaining three studies [20,22,23]. Disease stages I-IV were included over the five studies. There was heterogeneity reported in the RT techniques utilised. Three studies [19,20,22] report the use of 3DCRT techniques and EPID imaging in contrast with the more advanced techniques of IMRT and tomotherapy reported in two trials [21,23]. RT prescription was modified for each patient based on their individual plan features, mainly normal tissue dose volume constraints or FDG PET uptake values (see Table 2 for details). The resultant prescriptions ranged between 1.5–1.8 Gy twice daily and 2.28–4.8 Gy daily, with treatment duration ranging from 25 to 40 days.

Interpreting these trials in a clinically meaningful way is difficult due to the vast heterogeneity observed between patient and disease characteristics as well as the technical approaches taken. One of the main sources of clinical variation between these studies is the parameters used to determine the personalised dose. Standard normal tissue dose volume constraints were used as dose limiting parameters in two studies [19,20] which had adopted a twice-daily accelerated fractionation schedule. One study [21] reported the mean fraction-normalised lung dose divided by the normalised prescription dose as the dose limiting constraint. One study reported DVCs in EQD2 doses [23]. The remaining trial [22] consisted of 2 trial arms, one with an escalating oesophageal dose limiting constraint. In the other treatment arm dose limiting parameters were modified based on the oesophageal toxicity outcomes from the first cohort. As a result, any direct comparative analysis between these studies is difficult and potentially unreliable.

The clinical outcomes reported in these trials were generally positive, with three of the five trials reporting an increased rate of OS with increased dose, which resulted from personalising the prescription. One study [21] found no significant difference in OS between the 57 Gy and 63.25 Gy cohorts. One study did not report OS in relation to dose modification [23]. Varying levels of toxicity were reported, ranging from mild and transient to Grade 5 fatal toxicities, depending on target location and dose delivered.

The findings presented suggest that dose escalation is not only clinically feasible, but also potentially desirable in terms of improved outcomes (Table 2). An increase in OS was associated with an increase in dose in three of the five studies. However, significant toxicity was reported in all studies with Grade 5 adverse events also reported in four of the five studies. Cannon et al. [21] reported that long term follow-up revealed unacceptable late toxicity and recommended a maximum tolerated hypofractionated dose of 63.25 Gy in 25 fractions. This toxicity may well be due to the lack of established evidence-based dose volume constraints (DVCs) in the moderately hypofractionated setting. QUANTEC guidelines caution the user on the limitations of applying standard DVCs to modified fractionation schedules and highlighted the need for further work in this area [33]. A review of DVCs utilised in moderately hypofractionated RT for NSCLC highlights the lack of consensus in the clinical setting [34]. Additionally, it has recently been shown that heart dose is a key factor in predicting OS, however, the optimal DVC is yet to be determined [35].

One potential method of reducing this toxicity could be improving the dose conformity to the target using more modulated techniques. In these trials, a number of 3DCRT planning approaches were reported. It is possible that the use of isotoxic planning techniques which require the increased conformity that comes with IMRT or VMAT, could translate to reduced clinical toxicity and the published outcomes of the isotoxic IMRT trial are eagerly awaited [13]. However, two studies that reported Grade 5 haemoptysis utilised modulated techniques, but these studies observed this toxicity in patients with centrally located tumours [21,23].

All trials summarised here modified the RT prescription to increase the dose delivered per fraction or per day, rather than extending the prescription in a linear 2 Gy per fraction manner. In spite of the negative outcome of RTOG 0617 [9,10], these results are encouraging and warrant further investigation. However, these studies do not consider RT dose personalisation based on the molecular or genomic characteristics of the tumour thus there is potential for a further personalised medicine approach to the treatment of NSCLC with RT.

Future directions of personalised radiotherapy prescriptions

Nine studies [24–32] were identified that outlined novel concepts which may prove to be a future basis for personalised RT prescriptions, findings are summarised in Table 3. These studies can be broadly categorised depending on the patient specific feature they are utilising as a basis for personalisation.

Spatial relationship between target and OARs

Exploiting the patient specific spatial relationship or distance between target volumes and OARs in order to individually modify

Table 1
Feasibility of dose modification in LA-NSCLC – *in silico* studies.

Author, year (ref)	Study type	Participants *n* = ()	Participants stage	Target definition	Target volumes	Dose escalation limiting parameter	Maximum allowed dose	RT technique	Dose achieved	Dose achieved in biologically equivalent 2 Gy fractions (EQD2)	Study conclusions
Warren, 2014 [14]	Retrospective planning study (3DCRT vs. IMRT) for isotoxic planning	*n* = 20	II–III	Union of GTV delineated on all phases of a 4DCT	GTV and involved nodes	Spinal cord PRV < 50 Gy to 1 cm^3, Lungs (excluding GTV) MLD < 20 Gy, Brachial plexus PRV < 66 Gy to 1 cm^3, Proximal bronchial tree and great vessels PRV < 75.1 Gy to 1 cm^3	79.2 Gy/ 1.8 Gy per fraction twice daily	3DCRT vs. Inverse-planned CRT vs. IMRT (step and shoot)	70.2 Gy/ 39 fractions, twice daily RT.	69.0 Gy	Use of IMRT allows for higher isotoxic dose escalation than 3DCRT
Guckenberger, 2012 [15]	Retrospective planning study (Only Isotoxic dose escalation section included)	*n* = 13n = 1 Radiotherapy, *n* = 12 Concurrent chemoRT	III	Planning CT and FDG PET	GTV and involved nodes (no elective nodal irradiation)	MLD level achieved with a prior 3DCRT plan	Not stated	IMRT ± SIB ± ART	Average for *n* = 7 planned with IMRT + SIB+ ART 78.8 Gy ± 2.7 Gy SD	Not reported	Isotoxic SIB adaptive replanning is a promising strategy to safely escalate dose
Even, 2015 [16]	Feasibility study comparing various boost volumes	*n* = 10	IIb-IIIb	Mid ventilation phase of 4D FDG PET (Boost volumes SUV50% max and HX4 TBR > 1.4 or whole tumour)	GTV and involved nodes	Lungs Dmean < 20 Gy (corrected to EQD2); spinal cord D0.1 < 51 Gy (EQD2 < 52 Gy); oesophagus V36 < 80%; brachial plexus D0.1 < 66 Gy (EQD2 < 66 Gy); whole heart Dmean < 46 Gy (EQD2 Dmean < 46 Gy); planning organ-at-risk volume mediastinal structures (OAR + 5-mm margin) D0.1 < 76 Gy (EQD2 < 94 Gy), (D0.1 the dose delivered to 0.1% of the OAR)	129.6 Gy/ 24 fractions to PTV boost volume	VMAT SIB to FDG and HX4 volumes	Average prescribed doses to the boost volumes were 87.1 ± 10.1 Gy (PTV Prim), 107.3 ± 20.6 Gy (PTV FDG) and 117.6 ± 15.2 Gy (PTV HX4).	Not reported	Statistically higher prescriptions achieved for boost volume plans than primary alone. No significant difference in OAR DVHs
Hoffmann, 2012 [17]	Retrospective planning study (increasing dose per fraction ± the number of fractions, up to 33, based on normalised total doses)	*n* = 38	III	FDG PET CT	Not reported	NTDmean $\leq$ 19 Gy, NTDmax, (Dmax in 2-Gy fractions using EQD2) oesophagus NTDmax 80 Gy, NTDmean 34 Gy; spinal cord NTDmax 50 Gy; brachial plexus NTDmax 66 Gy; heart NTDmean 26 Gy	82.5 Gy/33 fractions	Step and Shoot IMRT	Median total tumour dose of 74.2 Gy (Range:67.5–79.9 Gy) in 33 fractions	64.9 Gy (Range 57.0–71.9 Gy) TED in 2-Gy fractions corrected for overall treatment time. *note greater TED was achieved using 15 fraction schedule (67.0 Gy (57.0–75.4 Gy)	Individualised prescription enabled dose escalation and theoretical therapeutic gain in 79% of cases
Van Elmpt, 2012 [18]	Planning analysis of Phase II Trial Randomised to Arm A boost whole GTV, Arm B, boost	*n* = 20, prescription modification successful in *n* = 15	I-IIIb	Mid ventilation planning CT dataset in conjunction with co-	GTV only for dose escalation (nodal volumes	MLD 20 Gy	129.6 Gy/ 24 fractions once daily (max 5.4	Step and Shoot IMRT	Average prescribed dose Arm A: 78.4 ± 7.6 Gy (72.2–102.7 Gy) for 24 fractions Arm B: 87.2 ± 13.1	For whole population: average 77 Gy in 24 fractions were achieved (EQD2 of 85 Gy, α/β 10 Gy) for the entire	Dose-escalation by boosting the primary tumour or the high FDG uptake region within the primary

(*continued on next page*)

Table 1 (*continued*)

Author, year (ref)	Study type	Participants *n* = ( )	Participants stage	Target definition	Target volumes	Dose escalation limiting parameter	Maximum allowed dose	RT technique	Dose achieved	Dose achieved in biologically equivalent 2 Gy fractions (EQD2)	Study conclusions
	> 50% SUVmax region only			registered FDG PET CT	given non-modified dose of 66 Gy/24#		Gy per fraction)		Gy (76.8–129.6 Gy) for 24 fractions	primary tumour, or for the boost region up to 87 Gy in 24 fractions (EQD2 of 99 Gy, α/β 10 Gy)	tumour as feasible in 75% of the patients analysed

IGRT: Image Guided Radiotherapy.
3D CRT: 3-Dimensional Conformal Radiation Therapy.
IMRT: Intensity Modulated Radiation Therapy.
GTV: Gross Tumour Volume.
4DCT: 4 Dimensional Computed Tomography.
PRV: Planning Organ at Risk Volume.
MLD: Mean Lung Dose.
Gy: Gray.
RT: Radiation Therapy.
EQD2: Equivalent Dose in 2 Gy per fraction.
FDG PET: Fluorodeoxyglucose Positron Emission Tomography.
SIB: Simultaneous Integrated Boost.
ART: Adaptive Radiation Therapy.
HX4: 3-[18F]fluoro-2-(4-((2-nitro-1H-imidazol-1-yl)methyl)-1H-1,2,3-triazol-1-yl)propan-1-ol (Hypoxia tracer).
TBR: Tumour-to-Background Ratio.
Dmean: Mean dose.
OAR: Organ at Risk.
PTV: Planning Target Volume.
VMAT: Volumetric Modulated Arc Therapy.
DVH: Dose Volume Histogram.
NTDmean: mean normalised total lung dose: MLD, adjusted for the dose per fraction.
NTDmax: maximum normalised total dose.
TED: Tumour Effective Dose.
D_V: The absorbed dose in Gray that covers a specified percent volume V.
V_D: The percentage volume that receives a dose of at least D in Gray.

Table 2
Clinical evidence for personalised radiotherapy prescriptions.

Author, Year (ref)	Trial Type	Participants *n* = ( )	Participants stage	Target Definition Dataset	Target Volumes	Dose Escalation Limiting Parameter	Maximum Allowed Dose	RT Technique	IGRT technique	Dose Achieved	Dose Achieved in biologically equivalent 2 Gy fractions (EQD2,T)	Outcomes (LC/OS)	Outcomes (Toxicity)
van Baardwijk, 2010 [19]	Prospective, single arm study	*n* = 116, no concurrent chemoRT	I & II if inoperable, III	Fused PET-CT with planning CT	GTV and involved nodes (no elective nodal irradiation)	MLD dependant on lung function between 10 and 19 Gy, Cord Dmax 54 Gy ± 0.5 Gy, Great Vessels or Bronchi Dmax 70.2 Gy, Brachial Plexus Dmax 66 Gy	79.2 Gy, 1.8 Gy per fraction, bidaily treatment, 8 h apart	3D CRT	EPID	Median prescribed TTD = 64.8 ± 11.4 Gy with OTT 25 ± 5.8 days (range, 2 to 50 days)	66.0 ± 7.1 Gy (range, 51.9 to 73.1 Gy)	OS was higher with a higher EQD 2,T. Median OS was 21.0 months (95% CI, 15.8 to 26.2 months)	Acute toxicity was mild and transient (4.8% experienced grade 3 dysphasia) Late toxicity: Grade 3 and Grade 4 was reported in 10% of patients, mainly those with dyspnoea prior to treatment
van Baardwijk, 2012 [20]	Phase II	*n* = 137, Concurrent chemo RT	Stage III (Stage II *n* = 1)	Fused PET-CT with Mid ventilation planning CT dataset	GTV and involved nodes (no elective nodal irradiation)	MLD 19.0 ± 1.0 Gy, Spinal Cord Dmax 54.0 ± 0.5 Gy, and Brachial Plexus Dmax 66 Gy, Oesophagus Dmax 75 Gy	69 Gy (due to proximity to central structures) 1.5 Gy fractions twice daily up to 45 Gy with an interfraction interval of at least 8-h, followed by once daily fractions of 2 Gy based on the ESPATU phase III trial scheme	3D CRT	EPID	Median prescribed dose 65.0 ± 6.0 Gy (51–69 Gy) OTT of 35 ± 5.7 days (18–48 days)	53.9 ± 3.9 Gy (range 43.1–63.1 Gy)	OS was better with a WHO-PS of 0 (p = 0.009), smaller nodal volume (GTV-2; p = 0.015) and a higher EQD2,T (p = 0.037). Median OS 25.0 months (95%CI 19.8–30.3 months)	Acute Grade 3 dysphagia 25.5%, Late Grade 3 dysphagia 4.6%. Acute Grade 3/4 dyspnoea 2.9% and 0.7% respectively. Late grade 3 dyspnoea 2.3%, 0.8% (n = 1) grade 5 pneumonitis
Cannon, 2013 [21]	Phase I	*n* = 79, *n* = 75 evaluable per protocol patients (no concurrent chemo)	All stages eligible Stage I/II = 9%, Stage IIIA = 27%, stage IIIB = 44%, Stage IV = 13%, Recurrent patients = 8%	CT ± PET	GTV and involved nodes (no elective nodal irradiation)	rNTDmean. At a given dose level, patients with higher rNTDmean (and therefore higher bin number) were predicted to be at higher risk for radiation pneumonitis	Dose per fraction escalation to 57 Gy at 2.28 Gy/ fraction, 63.25 Gy at 2.53 Gy/ fraction, 69.25 Gy at 2.77 Gy/fraction, 75 Gy at 3.0 Gy/ fraction, 80.5 Gy at 3.22 Gy/fraction, and 85.5 Gy at 3.42 Gy/fraction (all delivered in 25 Fractions, once per day, 5 days per week)	IMRT/Helical tomotherapy	Not reported	Median delivered dose was 57 Gy (range, 22.8 to 85.5 Gy) in 25 fractions	60, 70, 80, 90, 100, and 110 Gy, respectively	Median overall survival rate was 16.0 months, no significant difference in local control in patients treated to 69.25 Gy and higher compared with those treated at the 57 and 63.25 Gy dose levels	No incident of Grade 3 acute or late oesophageal toxicity, no incidents of grade 3 or greater radiation pneumonitis. 6 incidents of grade 4–5 toxicity that were possibly related to RT (fatal haemoptysis and lung abscess) found to relate to bronchial tree high dose regions and central tumours

(*continued on next page*)

Table 2 (*continued*)

Author, Year (ref)	Trial Type	Participants *n* = ( )	Participants stage	Target Definition Dataset	Target Volumes	Dose Escalation Limiting Parameter	Maximum Allowed Dose	RT Technique	IGRT technique	Dose Achieved	Dose Achieved in biologically equivalent 2 Gy fractions (EQD2,T)	Outcomes (LC/OS)	Outcomes (Toxicity)
Landau, 2016 [22]	Phase I/II	*n* = 84 with concurrent chemo (*n* = 82 evaluable patients	Stage II & III	3DCT or 4DCT	Reported as GTV (no distinction made as to nodal volumes)	Split into 2 groups 1: an escalating Oesophageal dose constraint (progressively raised from 65 Gy to 68 Gy and then 71 Gy to 1 cm^3 of Oesophagus). Group 2: Lung and other normal tissue dose constraints (MLD EQD2mean 18.2 Gy, Heart D100% <45 Gy, D67% <53 Gy, D33% <60 Gy, Cord D0.1 cc $\leq$ 47 Gy, Brachial Plexus D30%$\leq$ 60 Gy, D0.1 cc $\leq$ 65 Gy, Oesophagus initially 63 Gy but increasing to 65 Gy then 68 Gy as trial progressed	73 Gy/30 fractions, once daily over 40 days	3DCRT (*n* = 81) and VMAT (*n* = 3)	Not reported	Overall mean of 67.7 Gy, with means of 68.9 Gy and 66.8 Gy for groups 1 and 2, respectively	not reported	Median OS was 36.9 months (95% CI 31.7–42.1 months) he risk of death was lower for the higher tumour dose subgroup (HR 0.53, 95% CI 0.28–1.02, P = .06)	Grade 2 to 5 RTPN was seen in 30.5% of patients (*n* = 3 Grade 3, *n* = 0 Grade 4 or 5) Grade 2 to 5 oesophagitis rate overall was 82.9% (*n* = 5 Grade 3, *n* = 0 Grade 4, *n* = 1 Grade 5) Oesophageal dose was limited to 68 Gy following Grade 5 toxicity in 71 Gy cohort

Table 2 (*continued*)

Author, Year (ref)	Trial Type	Participants *n* = ()	Participants stage	Target Definition Dataset	Target Volumes	Dose Escalation Limiting Parameter	Maximum Allowed Dose	RT Technique	IGRT technique	Dose Achieved	Dose Achieved in biologically equivalent 2 Gy fractions (EQD2,T)	Outcomes (LC/OS)	Outcomes (Toxicity)
Wanet, 2017 [23]	Prospective, single arm study	*N* = 13	Stage II–III	Contrast enhanced 3DCT and 4D FDG PET to create PET ITV for boost	GTV and involved nodes, individualised boost prescribed to PTV_{PET} only (62.5 Gy/25# prescribed to remainder of PTV)	Reported in EQD2: Lung- GTV, MLD dependent on FEV_1 (<10–19 Gy,) V20 Gy < 30%, V13 Gy < 40%, V10 Gy < 45%, V5 Gy < 65%. Oesophagus V40 Gy < 50%, V50 Gy < 30%, V70 Gy < 20%. Cord D2% <50 Gy. Brachial plexus D2% <60 Gy. Large Vessels D2% <94 Gy. Bronchi D2% <94 Gy. Whole Heart Dmean < 46 Gy, V40 Gy < 50%	4.8 Gy/25# to PTV_{PET}	Helical IMRT or IMAT	Daily online correction using MV-CT or CBCT	PTV_{PET} reached 82.1 ± 17.9 Gy for all patients, with values of 89.2 ± 24.0 and 75.0 ± 3.2 Gy for peripheral and central tumors respectively	90.8 Gy, 100.8 Gy and 81.3 Gy for all patients, peripheral and central tumours respectively	OS at 36 months was 52.8%, Overall PFS at 36 months was 34.6%, Local PFS at 36 months was 52.8% (from supplementary material)	Acute toxicity: grade 3 toxicities included oesophagitis (23.1%), dyspnoea (23.1%), fatigue (30.8%), dermatitis (15.4%), anorexia (15.4%) and nausea (7.7%). Late toxicities: 5 patients presented grade 1 pulmonary fibrosis (38.5%), 4 patients developed grade 2 pneumonitis (30.8%). 2 patients developed grade 5 hemoptysis

RT: Radiation Therapy.
PET-CT: Positron emission tomography-CT.
GTV: Gross Tumour Volume.
ITV: Internal Target Volume.
MLD: Mean Lung Dose.
Dmax: Maximum point dose.
Gy: Gray.
3DCRT: 3-Dimensional Conformal Radiation Therapy.
EPID: Electronic Portal Imaging Device.
TTD: Total Tumour Dose.
OTT: Overall Treatment Time.
OS: Overall Survival.
PFS: Progression Free Survival.
EQD2: Equivalent Dose in 2 Gy per fraction.
WHO PS: World Health Organisation Performance Status.
rNTDmean: mean fraction-normalised lung dose divided by the normalised prescription dose.
IMRT: Intensity Modulated Radiation Therapy.
IMAT: Intensity Modulated Arc Therapy.
3DCT: 3 Dimensional Computed Tomography.
4DCT: 4 Dimensional Computed Tomography.
VMAT: Volumetric Modulated Arc Therapy.
HR: Hazard Ratio.
RTPN: Radiation Therapy Pneumonitis.

the dose prescription was the focus of two publications [24,25]. Petit et al. [24] utilised retrospective planning data to model a patient specific maximum achievable dose, based on PTV to OAR distance. The model performed well in predicting the max dose in 79% of patients ($n = 89$).

A similar concept utilised a spatio-temporal model [25]. This study considered the distance between target and OARs but also integrated factors, such as tissue α/β ratios and the linear quadratic model of response, to select an optimal prescription for an individual. This incorporated both the physical features of the patient's anatomy but also tumour doubling and lag times to predict the optimal fractionation schedule. It concluded that the biologically effective dose (BED) could be increased. However, in some patients, there was no significant difference in the tumour BED achieved with the model-determined optimal fractionation and the standard of care approach. The use of this model may result in a more personalised approach for patients, as it was successful in distinguishing between those who may or may not benefit from a dose prescription modification, albeit only in relation to BED. This may lead to improved stratification of patients and potentially reduce the toxicity observed in the clinical trials.

Targeting sub-volumes of disease

These studies focused on identifying a sub volume within the tumour that may represent a radio-resistant portion of disease. Once identified, a boost was prescribed to this volume only. One approach attempted to generate a mathematical model that would predict visible tumour response to radiation therapy during treatment and consequently also predicted a residual disease volume that would remain at the end of RT delivery. This model was generated from a database of treated patients was the focus of both a feasibility study, [26] and a follow up analysis [27]. Disease response to treatment was predicted prior to treatment delivery and an individualised simultaneous integrated boost (SIB) to this 'residual' volume would then be prescribed at the outset. Dose escalation to the residual volume was determined to be feasible [26] but the cohort of patients was small ($n = 5$). The follow up study $n = 10$ [27] found that use of this predictive approach facilitated mean dose escalation of 10.4 Gy, to the actual residual tumour volume. One limitation of these studies, similar to the planning studies and clinical trials, is that they assume that dose escalation is the most desirable strategy. Modification of the prescription is optimised to the patient's anatomy and the highest achievable dose, rather than predictive disease features.

Hypoxia targeting

Tumour hypoxic regions are known to be an area of radioresistance and as such they make an attractive target for dose modification studies. A radiobiology focused personalised model was proposed [28] which recommended a dose fractionation schedule that was based on a reoxygenation model, revised to also account for repopulation and repair. Changes in tumour hypoxia over the duration of treatment delivery was considered in this model, which lead to a heterogeneous fractionation approach, with the dose per fraction increasing after short breaks, such as weekends. This mathematically modelled study identified that a theoretical benefit in tumour control probability (TCP) was only observed in well-oxygenated tumours. In a hypoxic setting, the heterogeneous dose schedule was not comparable to the extreme hypofractionation as is commonly seen in the SABR protocols. At present, extreme hypofractionation does not seem a feasible approach for patients with LA-NSCLC, but this method of hypoxia targeting may still be clinically applicable in this cohort of patients. This methodology of modelling tumour hypoxia would require routine hypoxic imaging, perhaps in a longitudinal setting, in order to accurately determine tumour oxygenation status, which is not widely clinically available.

A recent study [29] modelled individualised dose escalation based on either hypoxic sub-volume targeting or whole tumour dose escalation. This study evaluated the uniform vs. non-uniform dose escalation in mathematically modelled population exhibiting varying ranges of hypoxia. It was reported that uniform dose escalation was more beneficial at the early stages of treatment, followed by a hypoxia-targeted dose escalation at a later stage. It was postulated that this is due to the hypoxic regions changing rapidly in the early stages of RT delivery. Again, routine hypoxia imaging would be required to implement such a technique.

TCP/NTCP models

Two studies [30,31] incorporated radiobiological modelling into the treatment planning process in order to select the most appropriate dose fractionation schedule for the individual case. The basis for the selection was TCP and NTCP models [30] and a novel 'Bifurcation Number' [31]. Both studies reported the proposed methodologies to be beneficial in selecting an optimal prescription.

BioSuite, a novel software program, was developed and was the subject of a feasibility study [30], which incorporates a number of TCP and NTCP models into the RT prescription process. The method proposed dictates that an optimal treatment plan is created and exported to BioSuite. Based on the patient specific anatomy and the individual's treatment plan, BioSuite identifies the optimal dose fractionation considering TCP and NTCP models to maximise the therapeutic window for the patient. The concept was used in the treatment of two patients with NSCLC and utilised lung TD_{50} and oesophageal max dose as the dose limiting constraints. The software calculates a desirable dose fractionation schedule suited to the individual. The individualised approach is clear here, however it is reliant on an optimal dosimetric plan being produced to determine what can be achieved. As such is potentially limited by the skill of the individual dosimetrist. This uncertainty may be mitigated by the introduction of automatic planning techniques, which have been shown to produce high quality treatment plans in patients with late stage NSCLC [36].

The novel concept of a Bifurcation Number has been developed by Keller et al. [31], and is defined as a dose ratio between the tumour and one OAR. The Bifurcation Number was then compared to the α/β ratios between normal tissue and tumour, to determine the optimal fractionation schedule for an individual. This methodology begins to take into account patient specific features as well as known radiobiological principles. Furthermore, this option indicates the optimal dose modification strategy, which is in keeping with the desire to tailor treatment to the patient. However, as only one OAR is considered in these calculations, this may be a limiting factor in the clinical application of this concept.

Modelled proliferation rate

A patient specific Proliferation Saturation Index (PSI) was the focus of one study [32], which is generated from tumour volume changes over time and the host organ's tumour carrying capacity. The carrying capacity is an ecological concept that defines the maximum ecosystem (tumour) that a host environment can sustain (patient) and is often used in mathematical modelling of cancer dynamics. It was hypothesised that a high PSI indicates a tumour volume is close to the host's maximum carrying capacity and hence it would have a low proportion of cells rapidly proliferating. This in turn would indicate a poor sensitivity to RT and therefore a need for a modified prescription. The concept was tested on a mathematically modelled population ($n = 500$) and four

Table 3
Future directions of personalised radiotherapy prescriptions.

Author, year (ref)	Proposed parameter for individualisation	Brief overview of concept	Participants n = ()	Participants stage	Target definition	Target volumes	Dose escalation limiting parameter	Maximum allowed dose	RT technique	Reported conclusions
Petit, 2015 [24]	Patient anatomy and prior plans (OAR to PTV distance)	Optimal achievable DVHs were predicted based on OAR distance to PTV, based on this max achievable dose escalation is predicted	Training cohort n = 50, Validation Cohort n = 39	II–III	Not reported	GTV and involved nodes	OAR dose: MLD excluding GTV < 20 Gy; Plexus Brachialis Dmax < 70 Gy; Heart Dmean < 46 Gy; Dmax to the Heart, mediastinal envelop and Oesophagus < 76 Gy; D0.1 cc Spinal Cord < 54 Gy	69 Gy/2Gy per fraction, once daily	VMAT	The method accurately predicted the mean lung dose within -0.11 ± 1.9 Gy (mean ± SD) and the maximum achievable dose escalation to the PTV, within one 2 Gy fraction for 79% of the patients
Kim, 2015 [25]	Spatiotemporal optimization	Model to predict optimal dose based on spatial relationship of target to OARs and incorporating Td and Tk	n = 16	n/a	n/a	n/a	Heart Dmean $\leq$ 45 Gy, Lungs D mean $\leq$ 20 Gy, Cord Dmax $\leq$ 45 Gy, Oesophagus Dmax $\leq$ 63 Gy, and unspecified tissues D05 $\leq$ 60 Gy	60 Gy/30 fractions EUD	IMRT	Suggest that personalising a fractionation schedule could potentially deliver a larger BED and EUD to tumours without exceeding the normal tissue BED DVCs with a large Td. For tumours with a small Td, there is no significant difference in tumour BED between an optimal fractionation and a standard fractionation. The results also showed that an optimal $N*$ (number of fractions) does not depend on Tk as long as $N*$ is larger than Tk
Zhang, 2014 [26]	Predicted residual tumour based on prior patients treatment response (feasibility Study)	The predictive atlas should identify the location of residual disease following treatment, this volume is then prescribed an escalated personalised SIB	n = 5	Locally advanced	CBCT and planning CT	Tumour only	In house dosimetric OAR constraints	56–70 Gy	IMRT	PTP incorporating a tumour regression model from the start, is technically feasible. PTP represents a new approach to increase tumour dose without increasing toxicities
Zhang, 2017 [27]	Predicted residual tumour based on prior patients treatment response	The predictive atlas should identify the location of residual disease following treatment, this volume is then prescribed an escalated personalised SIB	n = 10	Locally advanced	RO and Medical physicist delineated on Planning CT and last fraction CBCT	GTV Tumour	Local standard DVCs (Lung – GTV Mean dose < 20 Gy, Cord Dmax < 50 Gy)	No max	IMRT + SIB	The error in predicting the tumour shrinkage rate averaged 6% ± 13% (mean ± standard deviation) compared to the true shrinkage rate. PTP maintains the same OAR DVHs, but yielded an increased PTVresd D95 and PTVresd Dmean to the actual residual tumour of 6.5 ± 1.8 Gy and 10.4 ± 2.9 Gy, respectively
Lindblom, 2015 [28]	Radiobiology modelling on *in silico* population	Inclusion of repopulation and repair into a model that considers reoxygenation. A heterogeneous fractionation schedule was considered (increasing dose per fraction after short breaks)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Heterogeneous fractionation has only been found to be effective in the case of an overall well-oxygenated tumour. For the hypoxic tumour, a schedule employing 2- and 4 Gy-fractions could not compete with the SBRT-like schedules in terms of achieving high levels of TCP

(*continued on next page*)

Table 3 (*continued*)

Author, year (ref)	Proposed parameter for individualisation	Brief overview of concept	Participants *n* = ()	Participants stage	Target definition	Target volumes	Dose escalation limiting parameter	Maximum allowed dose	RT technique	Reported conclusions
Chvetsov, 2017 [29]	Hypoxic areas as seen on functional imaging (uniform vs. non-uniform dose distributions) *in silico* trial	modelling of uniform vs. non uniform dose escalation using mathematical formulae accounting for hypoxia and re-oxygenation	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Uniform dose escalation is more efficient at the beginning of treatment when the density of clonogens is nearly uniform and the hypoxic tumour is subject to change quickly. Hypoxia-targeted dose escalation will be more effective when applied later, after the density of clonogens becomes non-uniform
Uzan, 2012 [30]	BioSuite software (utilising a number of TCP/NTCP models)	creating a standard plan, exporting to BioSuite and modelling TCP/NTCP based on the patient specific anatomy	*n* = 2 (NSCLC)	I & III	Not reported	GTV or GTV and involved nodes	NTCP parameters Lung TD50 = 29.2 Gy, Oesophagus Dmax = 72 Gy	Not reported	3DCRT	Dose Response curves generated by BioSuite could indicate underutilised therapeutic windows. Radiobiologically guided prescription dose and fraction number customisation of external beam radiotherapy can be easily performed using BioSuite. The isotoxic dose and fractionation or 2D optimisation effectively convert all the various effects on TCP and NTCP into a single dose response curve
Keller, 2013 [31]	Bifurcation Number(B) 'a generalised dose ratio between the normal tissue and the tumour'	Calculate the bifurcation number to determine optimal fractionation 'If B is smaller than the ratio of alpha/beta ratios between normal tissue and tumour, then a single fraction is optimal; otherwise the optimal treatment is an infinite number of doses. These fractionation schedules correspond clinically to hypo- and standard/hyperfractionation'	*n* = 46	Not stated (Conventional, hypofractionated and SBRT all included)	Not reported	GTV and involved nodes	n/a	n/a	n/a	The bifurcation number may assist in selection of an appropriate fractionation protocol once the dose distribution has been optimised

Table 3 (*continued*)

Author, year (ref)	Proposed parameter for individualisation	Brief overview of concept	Participants $n = ()$	Participants stage	Target definition	Target volumes	Dose escalation limiting parameter	Maximum allowed dose	RT technique	Reported conclusions
Prokopiou, 2015 [32]	Proliferation Saturation Index (PSI)	A PSI is generated based on tumour volume changes between diagnosis and CT simulation, predicting the rate of growth and proliferation, therefore the radiosensitvity of the tumour. This may indicate the need for alternative fractionation schedules	$n = 500$ in a modelled population (matlab), $n = 4$ (patient data from the literature)	n/a	n/a	GTV tumour (modelled and reported)	n/a	n/a	n/a	Use of the PSI during the initial planning stage could facilitate the simulation of a variety of fractionation schedules and the treatment protocol can dynamically be adapted based on this

OAR: Organ at Risk.
PTV: Planning Target Volume.
DVH: Dose Volume Histogram.
GTV: Gross Tumour Volume.
MLD: Mean Lung Dose.
VMAT: Volumetric Modulated Arc Therapy.
Gy: Gray.
Td: Tumour doubling time.
Tk: Tumour lag time.
Dmean: Mean Dose.
EUD: Equivalent Uniform Dose.
IMRT: Intensity Modulated Radiation Therapy.
BED: Biologically Equivalent Dose.
DVC: Dose Volume Constraint.
SIB: Simultaneous Integrated Boost.
CBCT: Cone Beam Computed Tomography.
CT: Computed Tomography.
PTP: Predictive treatment planning.
RT: Radiation Therapy.
Dmax: Maximum point dose.
SBRT: Stereotactic Body Radiation Therapy.
TCP: Tumour Control Probability.
NTCP: Normal Tissue Complication Probability.
NSCLC: Non-Small Cell Lung Cancer.
TD50: dose that would result in 50% probability of developing complications.
3DCRT: 3-Dimensional Conformal Radiation Therapy.
2D: 2 Dimensional.
PSI: Proliferation Saturation Index.

clinical NSCLC cases. It was reported that the PSI correlated with tumour volume reduction during RT when using standard fractionation schedules. While promising, this study was limited by the lack of patient data. Further work on this concept has been encouraging in predicting treatment early response in patients with oropharyngeal carcinoma [37] however, clinical evaluation is necessary to ascertain the potential benefit of this concept in alternative fractionation schedules. The most recent work published on this concept highlights the dependence of the model of tumour growth law utilised for the calculations [38]. One advantage of the PSI is that it is calculated from routinely acquired CT datasets that are used for diagnostic and planning purposes. As a result, this may have little impact on departmental resources or the patient pathway through treatment.

Discussion

In the literature reviewed, a variety of methodologies have been proposed and investigated to personalise RT prescriptions in patients with LA-NSCLC, however, most are not yet being tested in the clinical setting. One clear reason for this lack in clinical progression is the lack of robust *in silico* evidence. Many of the proposed models have been tested on mathematically generated populations in an academic setting. Further work is required to translate these potential models into clinical RT treatments.

The retrospective planning studies reviewed have shown that dose modification is feasible and hence implementation of this concept is clinically possible. Some clinical evidence has been gained from a small number of trials looking at the alternative dose fractionation schedules, though these are largely focused on dose escalation, mainly exploiting the spatial relationship between targets and OARs. To an extent, this approach is personalised to the individual's gross anatomy. However, to truly personalise the treatment we should also target disease features that are identifiable at a cellular level.

Furthermore, a true person-centred approach is lacking in this research, with the majority of studies and proposals aiming to simply escalate dose without a focus on the patient characteristics. A recent publication [39] highlights the vast heterogeneity within this population and the myriad of factors that are not yet considered in our treatment approach. This may well provide a starting point for improved patient stratification and thus personalisation. While our efforts should not move entirely away from personalising the RT plan, we cannot continue to do this without also giving equal consideration to patient related factors.

One potential barrier of implementing personalised radiotherapy prescriptions is the resource burden of such approaches. Clinical departments do not have unlimited resources available to deliver an entirely personalised approach to every patient at present. Those concepts and theories which have a minimal impact on workflow and resources, are most likely to be the first introduced to clinical practice, if shown to be beneficial.

Conclusions

Personalised medicine is fast becoming a basic requisite in the modern era of cancer treatment. The treatment of LA-NSCLC with RT is an area of on-going investigation in the pursuit of a personalised approach. Much of the data identified in this review was retrospective planning studies and this limits the strength of any conclusions. The current clinical evidence, although promising in terms of improving outcomes, is lacking sufficient patient numbers or focus on personalisation to lead to a change in clinical practice and results from ongoing trials are eagerly awaited. Several encouraging strategies have been identified but do not yet provide the level of evidence required to inform clinical practice and a renewed effort to personalise RT dose prescription in the treatment of NSCLC is needed.

Conflict of interest statement

The authors have no conflict of interest to declare.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2018.05.029.

References

[1] Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang X-S, et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25676887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 385: 977–1010.
[2] Chang JY, Senan S, Paul MA, Mehran RJ, Louie AV, Balter P, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. Lancet Oncol 16: 630–7.
[3] Murray P, Franks K, Hanna GG. A systematic review of outcomes following stereotactic ablative radiotherapy in the treatment of early-stage primary lung cancer. The Br J Radiol 2017;90:20160732.
[4] Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. Lancet 2015;385:1206–18.
[5] Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017;28:iv1–iv21.
[6] Kong F-M, Zhao J, Wang J, Faivre-Finn C. Radiation dose effect in locally advanced non-small cell lung cancer. J Thoracic Dis 2014;6:336–47.
[7] Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. Lancet Oncol 2015;16:187–99.
[8] Ramroth J, Cutter DJ, Darby SC, Higgins GS, McGale P, Partridge M, et al. Dose and fractionation in radiation therapy of curative intent for non-small cell lung cancer: meta-analysis of randomized trials. Int J Radiat Oncol Biol Phys 2016;96:736–47.
[9] Belderbos J, Walraven I, van Diessen J, Verheij M, de Ruysscher D. Radiotherapy dose and fractionation for stage III NSCLC. Lancet Oncol 2015;16:e156–7.
[10] Faivre-Finn C. Dose escalation in lung cancer: have we gone full circle? Lancet Oncol 2015;16:125–7.
[11] Christodoulou M, Bayman N, McCloskey P, Rowbottom C, Faivre-Finn C. New radiotherapy approaches in locally advanced non-small cell lung cancer. Eur J Cancer 2014;50:525–34.
[12] Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. Adv Drug Deliv Rev 2017;109:131–53.
[13] Haslett K, Franks K, Hanna GG, Harden S, Hatton M, Harrow S, et al. Protocol for the isotoxic intensity modulated radiotherapy (IMRT) in stage III non-small cell lung cancer (NSCLC): a feasibility study. BMJ Open 2016;6.
[14] Warren M, Webster G, Ryder D, Rowbottom C, Faivre-Finn C. An isotoxic planning comparison study for stage II-III non-small cell lung cancer: is intensity-modulated radiotherapy the answer? Clin Oncol (Royal College of Radiologists (Great Britain)) 2014;26:461–7.
[15] Guckenberger M, Kavanagh A, Partridge M. Combining advanced radiotherapy technologies to maximize safety and tumor control probability in stage III non-small cell lung cancer. Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft 2012;188:894–900.
[16] Even AJ, van der Stoep J, Zegers CM, Reymen B, Troost EG, Lambin P, et al. PET-based dose painting in non-small cell lung cancer: Comparing uniform dose escalation with boosting hypoxic and metabolically active sub-volumes. Radiother Oncol 2015;116:281–6.
[17] Hoffmann AL, Troost EG, Huizenga H, Kaanders JH, Bussink J. Individualized dose prescription for hypofractionation in advanced non-small-cell lung cancer radiotherapy: an in silico trial. Int J Radiat Oncol Biol Phys 2012;83:1596–602.

[18] van Elmpt W, De Ruysscher D, van der Salm A, Lakeman A, van der Stoep J, Emans D, et al. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. Radiother Oncol 2012;104:67–71.
[19] van Baardwijk A, Wanders S, Boersma L, Borger J, Ollers M, Dingemans AM, et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. J Clin Oncol 2010;28:1380–6.
[20] van Baardwijk A, Reymen B, Wanders S, Borger J, Öllers M, Dingemans A-MC, et al. Mature results of a phase II trial on individualised accelerated radiotherapy based on normal tissue constraints in concurrent chemo-radiation for stage III non-small cell lung cancer. Eur J Cancer 2012;48:2339–46.
[21] Cannon DM, Mehta MP, Adkison JB, Khuntia D, Traynor AM, Tome WA, et al. Dose-limiting toxicity after hypofractionated dose-escalated radiotherapy in non-small-cell lung cancer. J Clin Oncol 2013;31:4343–8.
[22] Landau DB, Hughes L, Baker A, Bates AT, Bayne MC, Counsell N, et al. IDEAL-CRT: A phase 1/2 trial of isotoxic dose-escalated radiation therapy and concurrent chemotherapy in patients with stage II/III non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2016;95:1367–77.
[23] Wanet M, Delor A, Hanin FX, Ghaye B, Van Maanen A, Remouchamps V, et al. An individualized radiation dose escalation trial in non-small cell lung cancer based on FDG-PET imaging. Strahlenther Onkol 2017.
[24] Petit SF, van Elmpt W. Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients. Radiother Oncol 2015;117:453–8.
[25] Kim M, Stewart RD, Phillips MH. A feasibility study: Selection of a personalized radiotherapy fractionation schedule using spatiotemporal optimization. Med Phys 2015;42:6671–8.
[26] Zhang P, Yorke E, Hu YC, Mageras G, Rimner A, Deasy JO. Predictive treatment management: incorporating a predictive tumor response model into robust prospective treatment planning for non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2014;88:446–52.
[27] Zhang P, Rimner A, Yorke E, Hu YC, Kuo L, Apte A, et al. A geometric atlas to predict lung tumor shrinkage for radiotherapy treatment planning. Phys Med Biol 2017;62:702–14.
[28] Lindblom E, Dasu A, Toma-Dasu I. Optimal fractionation in radiotherapy for non-small cell lung cancer - A modelling approach. Acta Oncol 2015;54:1592–8.
[29] Chvetsov AV, Rajendran JG, Zeng J, Patel SA, Bowen SR, Kim EY. Theoretical effectiveness of cell survival in fractionated radiotherapy with hypoxia-targeted dose escalation. Med Phys 2017.
[30] Uzan J, Nahum AE. Radiobiologically guided optimisation of the prescription dose and fractionation scheme in radiotherapy using BioSuite. Br J Radiol 2012;85:1279–86.
[31] Keller H, Hope A, Meier G, Davison M. A novel dose-volume metric for optimizing therapeutic ratio through fractionation: retrospective analysis of lung cancer treatments. Med Phys 2013;40:084101.
[32] Prokopiou S, Moros EG, Poleszczuk J, Caudell J, Torres-Roca JF, Latifi K, et al. A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation. Radiat Oncol (London, England) 2015;10:159.
[33] Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al., Use of normal tissue complication probability models in the clinic. Int J Radiat Oncol Biol Phys. 76: S10–S9.
[34] Fleming C, Cagney DN, O'Keeffe S, Brennan SM, Armstrong JG, McClean B. Normal tissue considerations and dose–volume constraints in the moderately hypofractionated treatment of non-small cell lung cancer. Radiother Oncol 2016;119:423–31.
[35] Speirs CK, DeWees TA, Rehman S, Molotievschi A, Velez MA, Mullen D, et al. Heart dose is an independent dosimetric predictor of overall survival in locally advanced non–small cell lung cancer. J Thoracic Oncol 12: 293–301.
[36] Della Gala G, Dirkx MLP, Hoekstra N, Fransen D, Lanconelli N, van de Pol M, et al. Fully automated VMAT treatment planning for advanced-stage NSCLC patients. Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft 2017;193:402–9.
[37] Lewin T, Kim J, Latifi K, Poleszczuk J, Bull J, Byrne H, et al. Proliferation saturation index predicts oropharyngeal squamous cell cancer gross tumor volume reduction to prospectively identify patients for adaptive radiation therapy. Int J Radiat Oncol Biol Phys 94: 903.
[38] Poleszczuk J, Walker R, Moros EG, Latifi K, Caudell JJ, Enderling H. Predicting patient-specific radiotherapy protocols based on mathematical model choice for proliferation saturation index. Bull Math Biol 2017.
[39] Hudson A, Chan C, Woolf D, McWilliam A, Hiley C, O'Connor J, et al. Is heterogeneity in stage 3 non-small cell lung cancer obscuring the potential benefits of dose-escalated concurrent chemo-radiotherapy in clinical trials? Lung Cancer 2018;118:139–47.

Clinical Oncology 33 (2021) 155–162

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

Geometric and Dosimetric Evaluation of a Commercially Available Auto-segmentation Tool for Gross Tumour Volume Delineation in Locally Advanced Non-small Cell Lung Cancer: a Feasibility Study

S. Barrett [*], A.J. Simpkin [†], G.M. Walls [‡], M. Leech [*], L. Marignol [*]

[*] *Applied Radiation Therapy Trinity, Discipline of Radiation Therapy, Trinity College Dublin, Dublin, Ireland*
[†] *School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland*
[‡] *Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, UK*

Abstract

Aims: To quantify the reliability of a commercially available auto-segmentation tool in locally advanced non-small cell lung cancer using serial four-dimensional computed tomography (4DCT) scans during conventionally fractionated radiotherapy.

Materials and methods: Eight patients with serial 4DCT scans ($n = 44$) acquired over the course of radiotherapy were assessed. Each 4DCT had a physician-defined primary tumour manual contour (MC). An auto-contour (AC) and a user-adjusted auto-contour (UA-AC) were created for each scan. Geometric agreement of the AC and the UA-AC to the MC was assessed using the dice similarity coefficient (DSC), the centre of mass (COM) shift from the MC and the structure volume difference from the MC. Bland Altman analysis was carried out to assess agreement between contouring methods. Dosimetric reliability was assessed by comparison of planning target volume dose coverage on the MC and UA-AC. The time trend analysis of the geometric accuracy measures from the initial planning scan through to the final scan for each patient was evaluated using a Wilcoxon signed ranks test to assess the reliability of the UA-AC over the duration of radiotherapy.

Results: User adjustment significantly improved all geometric comparison metrics over the AC alone. Improved agreement was observed in smaller tumours not abutting normal soft tissue and median values for geometric comparisons to the MC for DSC, tumour volume difference and COM offset were 0.80 (range 0.49–0.89), 0.8 cm^3 (range 0.0–5.9 cm^3) and 0.16 cm (range 0.09–0.69 cm), respectively. There were no significant differences in dose metrics measured from the MC and the UA-AC after Bonferroni correction. Variation in geometric agreement between the MC and the UA-AC were observed over the course of radiotherapy with both DSC ($P = 0.035$) and COM shift from the MC (ns) worsening. The median tumour volume difference from the MC improved at the later time point.

Conclusions: These findings suggest that the UA-AC can produce geometrically and dosimetrically acceptable contours for appropriately selected patients with non-small cell lung cancer. Larger studies are required to confirm the findings.

Key words: auto-contouring; delineation; non-small cell lung cancer; target volume definition

Introduction

Accurate definition of a gross tumour volume (GTV) on a computed tomography (CT) scan is a key step in the radiotherapy process. However, it remains problematic and prone to inter- and intra-observer error [1,2]. In lung cancer radiotherapy, consensus guidelines on contouring lung GTV recommend the use of a four-dimensional CT (4DCT), intravenous contrast and multimodality imaging in the form of a recent fluorodeoxyglucose positron emission tomography (FDG PET) scan [3]. In clinical practice, this approach is taken by the oncologist when defining the target volume for treatment planning purposes. However, this must be repeated on each of the 10 breathing cycle-associated phased scans, which is time and resource intensive [4]. Auto-segmentation tools have been suggested as a method of improving both the efficiency and the accuracy of lung GTVs [4,5]. Auto-segmentation algorithms can automatically define the GTV, creating true auto-contours (ACs) and these are often subsequently manually adjusted to create a semi-automatic user-adjusted auto-contour (UA-AC) [6,7].

In addition to the contouring burden of 4DCT datasets, clinical advances have led to repeated imaging over the

Author for correspondence: S. Barrett, Applied Radiation Therapy Trinity, Discipline of Radiation Therapy Trinity College Dublin, Trinity Centre for Health Sciences, Dublin 8, Ireland. Tel: +35-318-963-248.
E-mail address: barrets7@tcd.ie (S. Barrett).

https://doi.org/10.1016/j.clon.2020.07.019

course of radiotherapy treatment delivery, increasing the imaging data available for each patient. Analysis of GTV changes on repeated imaging during the course of radiotherapy may identify changes to the target volume that require plan adaptation. Although adaptive radiotherapy based on changing target volumes can ensure accurate dose delivery [8], a recent study showed that physician inter-observer variability was increased during the course of radiotherapy [9]. To facilitate rapid and repeated definition of target volumes on both planning CTs and subsequent imaging, a readily available, automated approach would be desirable.

Many of the auto-segmentation tools evaluated in the literature are locally available systems that were not widely available to clinical departments [5]. Furthermore, many available auto-segmentation approaches utilise deformable registration to propagate a physician-defined volume on to a subsequent image, meaning clinician input remains central to the process.

Here we investigated the feasibility of using a commercially available auto-contouring system (Smart Segmentation® Varian Medical Systems) in the delineation of lung GTVs without the requirement of propagation of a physician-defined manual contour (MC). We quantified the impact of user-adjustment on geometric accuracy, assessed the dosimetric reliability of ACs compared with physician-defined contours and carried out a time-trend analysis of auto-segmentation performance over the course of radiotherapy.

To complete this study we accessed a unique dataset of non-small cell lung cancer (NSCLC) patients with both a pre-treatment 4DCT and weekly research 4DCT scans acquired over the course of radiotherapy delivery. This intensive imaging schedule allowed a rare opportunity for a longitudinal analysis of AC on high-quality images at multiple time points during radiotherapy.

Materials and Methods

Dataset

A research dataset of 20 locally advanced NSCLC patients with 4DCT scans uniquely acquired longitudinally during their radiotherapy treatment delivery was identified at The Cancer Imaging Archive [10,11] and freely accessed. This atypical dataset described by Hugo *et al.* [12], consisted of a series of 4DCT scans acquired weekly and 4D cone beam CT scans acquired daily over the course of conventionally fractionated radiotherapy treatment delivery. Physicians delineated the primary GTV, nodal target volumes and organs at risk on each phase of the 4DCT datasets. The primary GTV was assessed in this study and nodal target volumes were not included.

Of the 20 patients, seven were excluded as only one 4DCT with contoured target volumes was available. The remaining 13 patients were evaluated for inclusion in the study.

Five patients were excluded due to the primary tumours being located within a region of atelectasis and a lack of associated FDG PET scan for discernment of tumour (see supplementary material for examples of excluded patients).

Eight patient datasets were included, with a total of 44 physician-delineated 4DCTs. The GTV location was reviewed and categorised as either abutting normal soft tissue in the thorax or not. See Table 1 for details of included patients. All patients were planned to receive 62–70.2 Gy in 31–39 once-daily fractions.

Contour Methods

Manual contours

Each 4DCT scan had an accompanying radiotherapy structure set, which included a physician-delineated MC, created on each phase of the 4D dataset. The CT 50% phase dataset was chosen for evaluation, as this is a stable expiration phase. The delineation of the MC was supervised by a single experienced radiation oncologist [12] and the MC was deemed to be the gold standard contour for the purpose of evaluation.

Auto-contours

For each 4DCT, an AC was generated on the same 50% phase. The structure set was opened with the corresponding CT series; a lung window was utilised to locate the approximate mid-volume slice. Automatic delineation was carried out using the Smart Segmentation function on Varian Eclipse™ V15.5 by following the manufacturer's instructions.

User-adjusted auto-contours

The AC was duplicated and manually adjusted by an experienced radiation therapist to exclude gross extensions into adjacent normal tissues only, resulting in a UA-AC for each CT50%. This process was audited by a clinical oncologist with significant experience in treating lung cancer. See Figure 1 for examples of contours generated and adjustments made.

Geometric Comparison to Manual Contour

To quantify the geometric agreement between the MC and the automatically generated contours, three metrics were utilised: the dice similarity coefficient (DSC), a shift in structure centre of mass (COM) and the volume difference in cm^3 in line with published recommendations [13].

The DSC measures the spatial overlap between two contours and is the volume of the union normalised by the mean of the two volumes. A DSC of 1 indicates perfect overlap and a DSC of 0 indicates no overlap. DSC values greater than 0.7 are generally considered to show good agreement [14]. The DSC was measured in Radialogica fullReport ™ V 1.4.5.12438.

COM coordinates in x, y and z directions for each contour were extracted from Varian Eclipse™ V15.5 statistics. They

Table 1
Demographics of included cases

Patient ID	T N M	Stage	Tumour location	Total dose (Gy)	Number of fractions	Dose per fraction (Gy)	Tumour volume (cm^3)	Tumour abutting normal tissue
107	2 2 0	IIIA	LUL	63	35	1.8	18	No
108	2 2 0	IIIA	RUL	70.2	39	1.8	12	No
110	2 3 0	IIIB	RLL	62	31	2	55	No
112	3 2 0	IIIA	RUL	63	35	1.8	31	Yes
113	1 1 0	IIA	LLL	66	33	2	78	Yes
114	2 2 0	IIIA	RLL	66	33	2	179	Yes
115	1 3 0	IIIB	RUL	66.6	37	1.8	7	No
117	3 2 0	IIIA	RUL	66	33	2	10	No

TNM, tumour nodes metastases; LUL, left upper lobe; RUL, right upper lobe; LLL, left lower lobe; RLL, right lower lobe.

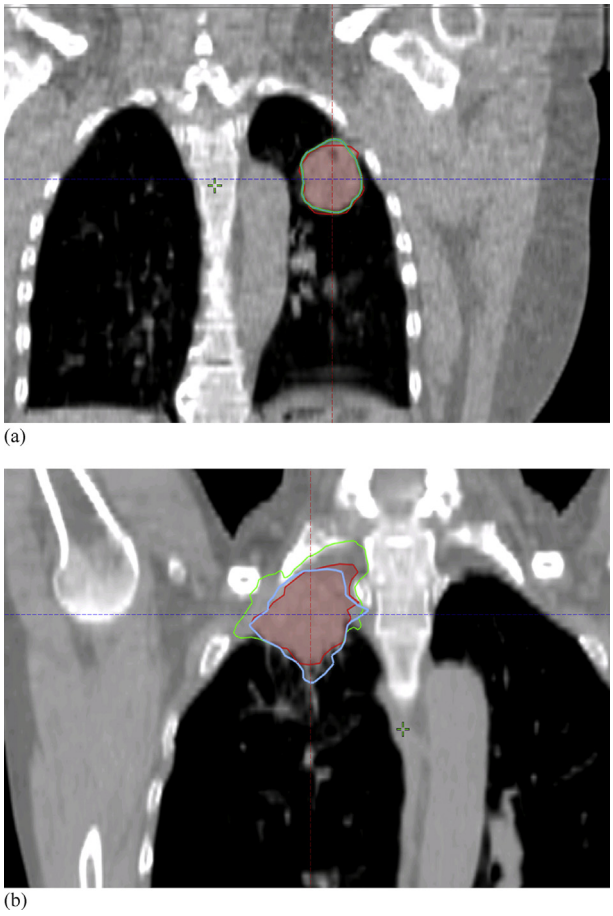


Fig 1. Examples of contours created: manual contour (MC) in red, auto-contour (AC) in green, user-adjusted auto-contour (UA-AC) in blue. (a) No user adjustment was required, (b) increased user adjustment was required after AC as tumour abuts normal tissue (chest wall).

were converted into an overall COM shift from the MC using the Euclidian distance formula. The volume difference between the AC and the UA-AC to the MC was recorded.

Geometric Agreement Analysis

To evaluate differences between the AC and UA-AC agreement with the MC, a ranked Wilcoxon test was carried out. A *P*-value of less than 0.05 was assumed to be statistically significant.

Bland-Altman plots [15,16], a graphical method to compare two measurement techniques where the differences between each method are plotted against the averages of both methods, were created to evaluate the agreement between contouring methods in determining target volume.

The geometric agreement analysis was repeated excluding three patients from this analysis, as they were considered potentially poor candidates for AC due to tumour proximity to normal soft tissue.

See Table 2 for a summary of contour methods and comparison metrics used in this study.

Dosimetric Evaluation

On the earliest scan for all eight patients from the MC, a clinical target volume (CTV) and a planning target volume (PTV) were created using an expansion of GTV + 5 mm for the CTV and CTV + 0.9 cm superior/inferior and 0.7 cm laterally to create the PTV [17]. A clinically acceptable treatment plan in line with the protocol outlined by Haslett *et al.* [17] was devised using a volumetric-modulated arc therapy (VMAT) technique. On all subsequent CT scans, a MC and UA-AC CTV and PTV were created and the original plan was copied on to each subsequent scan and recalculated using the fixed monitor units. Renormalisation was not carried out.

For each scan, the dosimetric indices of minimum dose, maximum dose mean dose D2%, D98%, D50%, D95% and V107% on MC GTV, CTV and PTV versus UA-AC GTV, CTV and PTV were measured and compared using a multilevel model to account for the clustering of the responses (dose, volume) within patients, structures and plan [18]. After

Table 2
Summary of the contour methods and comparison metrics used in this study

Contours generated		
Acronym	Name	Brief description
MC	Manual contour	Physician-defined manual contour, generated under the supervision of a single experienced radiation oncologist
AC	Auto-contour	Generated using the commercially available software with no manual corrections
UA-AC	User-adjusted auto-contour	Adjusted auto-contour to correct gross errors where contour extends into normal tissue
Agreement measures		
Acronym	Name	Brief description
DSC	Dice similarity coefficient	Geometric measure of the overlap between two structures
COM	Centre of mass	Centre of mass coordinates difference from gold standard manual contour to evaluate positional differences between contours
Vol	Volume	Absolute volume difference between auto-contours and the gold standard manual contour

taking this clustering into account, our main variable of interest is whether the contour was MC or UA-AC. Whether the tumour was deemed to be abutting normal soft tissue or not was also controlled for in each model. As there were eight responses (minimum dose, maximum dose, mean dose, D2, D50, D95, D98 and V107) Bonferroni correction was applied to account for multiple testing, such that statistical significance is taken to be differences between clinician and automated delineation with an associated *P*-value below $0.05/8 = 0.00625$.

In addition, the planning requirement of PTV D95% $>$ 90% and V107% $<$ 1 cm^3 was reviewed on all subsequent scans to assess compliance with the original plan for both the MC and the UA-AC.

Time Trend Analysis

Differences in the geometric accuracy measures from the first scan to the last of each patient were evaluated using a Wilcoxon signed rank test to assess the reliability of the UA-AC over the duration of radiotherapy.

Results

Geometric Comparison of Software-aided and Manual Contours

The geometric comparison metrics of the raw and user-adjusted contours to the MC are displayed in Figure 2.

UA-AC significantly improved geometric agreement to the MC compared with AC alone (Table 3). Clinician audit of the user adjustments showed complete agreement for 87.5% of the cases and satisfactory agreement for derivation of the UA-AC in >90% cases sampled.

As the UA-AC were shown to be geometrically superior to the AC alone, the remainder of the analysis focused on UA-AC metrics compared with MC.

The Bland Altman analysis identified improved agreement between the UA-AC and the MC when the tumour was smaller (Figure 3). However, the larger tumours also abutted adjacent normal tissues.

The impact of the anatomy surrounding the tumour on AC accuracy was evaluated next. The exclusion of the three patients whose tumour was abutting normal tissue ($n = 28$ 4DCTs included), further improved all of the geometric

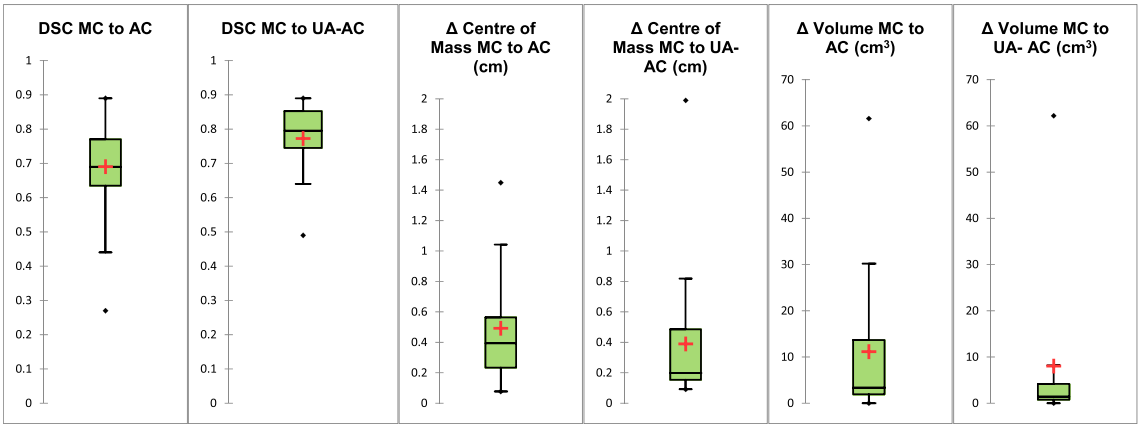


Fig 2. Box plots of geometric metrics measured on 44 datasets comparing the auto-contour (AC) and the user-adjusted auto-contour (UA-AC) to the manual contour (MC), showing an improvement in all geometric measures when user-adjustment was incorporated. All improvements were statistically significant ($P < 0.05$; Wilcoxon signed rank test). DSC, dice similarity coefficient; Δ, delta. Red cross = mean.

Table 3
Geometric comparison of both auto-contour (AC) and user-adjusted auto contour (UA-AC) with manual contour (MC)

Metric	AC congruence with MC		UA-AC congruence with MC		P^*
	Median (range)	Mean (SD)	Median (range)	Mean (SD)	
Volume-based metrics					
DSC	0.69 (0.27–0.89)	0.69 (0.122)	0.8 (0.49–0.89)	0.77 (0.099)	<0.0001
Volume difference	3.35 (0–61.6)	11.15 (15.48)	1.4 (0–62.2)	8.05 (16.498)	0.005
Distance-based metric					
COM offset	0.39 (0.08–1.45)	0.49 (0.377)	0.20 (0.09–1.99)	0.39 (0.409)	0.0003

DSC, dice similarity coefficient; COM, centre of mass; SD, standard deviation.
Median, means and COM offset are provided in centimetres; volume variance is provided in centimetres squared.
* Generated from Wilcoxon signed rank test.

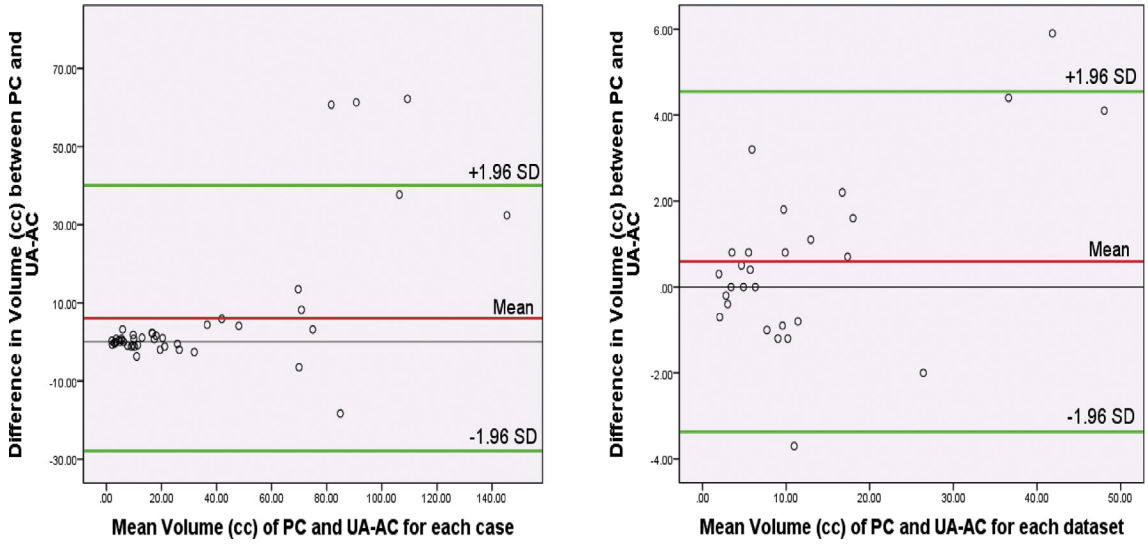


Fig 3. Bland–Altman comparison of manual contour (MC) and user-adjusted auto-contour (UA-AC) for all patients (left). Bland–Altman comparison excluding three large tumours abutting normal tissue (right), showing improved agreement.

agreement metrics measured (Table 4). Improved agreement between the MC and the UA-AC was also observed on Bland Altman analysis when these three cases were excluded (Figure 3).

Dosimetric Comparison of Software-aided and Manual Contours

There were no statistically significant differences in dose metrics measured from the MC volumes and the UA-AC volumes, after Bonferroni correction. There were two nominally significant findings in D50% (mean difference −0.06%, 95% confidence interval for difference [−0.1, −0.01]; $P = 0.009$) and V107% (mean difference 0.14 cm^3, 95% confidence interval for difference [0.03, 0.24]; $P = 0.012$). All data are available in the supplementary material.

All eight original plans met planning constraints of PTV D95% > 90% and V107% < 1 cm^3. Subsequent scans ($n = 36$) were assessed to determine if these metrics were breached in either the MC PTV or the UA-AC PTV. The PTV D95% > 90% dose constraint was still met in 17/36 re-plans using the MC and in 21/36 re-plans using the UA-AC. V107% < 1 cm^3 planning constraint was met in 29/36 re-plans using both the MC and the UA-AC.

Time Trend Analysis of Software-aided Contours

Between five and eight 4DCT scan sets were available for each case. Scan dates ranged from 26 days pre-radiotherapy to day 47 of radiotherapy. The median duration between the first and the last scan was 47.5 days (range 27–65 days).

The median DSC decreased from 0.82 (range 0.76–0.89) to 0.72 (range 0.49–0.86) ($P = 0.035$) between the first and the last scan.

The median COM shift from the MC at the first scan was 0.22 cm (range 0.09–0.71 cm) and at the last scan was 0.23 cm (range 0.09–1.99 cm). The mean COM shift increased from 0.31 to 0.52 cm ($P = 0.058$).

The median volume difference between the MC and the UA-AC at the first scan was 3.15 cm^3 (range 0.8–32.4 cm^3) and at the last scan was 1.0 cm^3 (range 0.7–60.7 cm^3). The

Table 4
Geometric comparison of user-adjusted auto-contour (UA-AC) with manual contour (MC), excluding three patients with tumour volumes abutting normal soft tissue

Metric	UA-AC congruence with MC excluding three cases	
	Median (range)	Mean (SD)
Volume-based metrics		
DSC	0.8 (0.49–0.89)	0.78 (0.091)
Volume difference	0.85 (0–5.9)	1.45 (1.5)
Distance-based metric		
COM offset	0.16 (0.09–0.69)	0.21 (0.14)

DSC, dice similarity coefficient; COM, centre of mass; SD, standard deviation.
Median, means and COM offset are provided in centimetres; volume variance is provided in centimetres squared.

mean volume difference increased from 6.6 to 9.9 cm^3 ($P = 0.866$).

Discussion

This pilot study investigated the feasibility of using a commercially available auto-contouring software to create lung tumour target volumes by evaluating the geometric and dosimetric reliability of the contours generated over the duration of radiotherapy. Although the number of cases was small, this rare image intensive dataset uniquely supported an in-depth longitudinal evaluation of the contours generated.

The study identified that UA-AC significantly improved geometric agreement in all metrics, in line with previous work on this topic evaluating non-commercially available segmentation and deformable registration [5,7,14,19]. Recent consensus guidelines highlight the ongoing requirement for manual review and adjustment of AC structures [20]. Similarly, in organ at risk delineation, UA-AC was reported to be a viable strategy to produce clinically acceptable contours [7]. Following user-adjustment, this study reported good agreement between contours compared with the MC, with mean and median DSC >0.7 (0.77 and 0.8, respectively), which is generally accepted as good agreement between the gold standard and the assessed contour [14,19].

The median volume difference between UA-AC and the MC was 1.4 cm^3 (range 0–62.2). Although the median volume difference was low, large individual differences were observed between the UA-AC and the MC, up to 62.2 cm^3. We hypothesised that large variation may occur when tumours were opposed to normal structures within the thorax, such as the mediastinum or chest wall. The analysis was therefore repeated excluding three cases where the tumour abutted normal tissue, which led to a considerable reduction in the median (0.85) and range (0–5.9 cm^3) of differences. This highlights that some tumours may be more amenable to AC than others and appropriate case selection may be a key step in implementing this technology in routine clinical practice.

Related to discernment of tumour and normal tissues, previous publications have highlighted the difficulty in generating target volumes in nodal regions [14] or in regions of atelectasis, where defining the tumour boundary with CT alone is difficult [2]. Although this study excluded tumours located within atelectasis, other studies have enacted a restriction on cases based on the percentage of tumour surrounded by lung parenchyma [21]. One limitation of this study is that it did not include node-positive cases, for which target delineation is known to be more difficult [14]. However, the auto-segmentation tool investigated is designed for primary tumour volume delineation only, so it was not appropriate to assess its performance on nodal volumes.

Exclusion of the three cases in which the tumour was abutting normal soft tissue resulted in improved geometric agreement, with a median DSC of 0.8, a median COM shift of 0.16 cm and a median volume difference of 0.84 cm^3. The mean DSC of 0.78 is comparable with the value of 0.84 reported recently evaluating inter-observer variability in CTV delineation [22]. This, in conjunction with the high level of agreement in COM and structure volume, also shows that appropriate case selection increases geometric accuracy of adjusted AC.

Clinical acceptability of UA-AC-based radiotherapy planning was considered in this study. A VMAT treatment plan was generated for each patient and then used to evaluate the dosimetric differences between the dose metrics measured from the physician-defined MC and target volumes generated using the UA-AC method.

UA-AC volumes performed as well as the MC in relation to target coverage over the course of treatment in this population, with no statistically significant differences between the measured dose metrics on either structure over the course of treatment. However, over the course of radiotherapy, the plan quality decreased and dose coverage of the PTV was compromised in both the MC and the UA-AC target volumes.

A qualitative review of the plans revealed that there was significant tumour regression, with magnitudes similar to those reported in previous studies [23,24] during radiotherapy, which may be responsible for the poor target coverage as radiotherapy progresses. This may indicate a need to introduce adaptive radiotherapy for locally advanced NSCLC radiotherapy to ensure planned dose coverage is delivered, which has been reported previously [8]. The LARTIA trial [25] reported on local control and toxicity for patients undergoing adaptive radiotherapy and reported that about 25% of patients required repeated planning due to tumour regression during treatment.

Repeated planning requires repeated contouring and so the UA-AC method could be a useful tool in the adaptive radiotherapy setting for saving time. UA-AC could be used to perform an efficient dose calculation check and trigger an adaptive plan for those with a breach of dose coverage metrics. The high level of oncologist agreement of the user-adjusted contours indicates the applicability of the software across trained staff groups, i.e. radiographers, dosimetrists and oncologists.

Other factors beyond tumour volume regression, such as patient contour changes or set-up uncertainties, may have contributed to the observed variation in plan dosimetry [26]. This study aimed to test if the dosimetry reported on the MC volumes or the UA-AC volumes were statistically different, but causes of dosimetric changes were not investigated. Future work should determine if the variation observed is a direct result of tumour volume regression and should include nodal volumes as well as the primary tumour.

Although adaptive radiotherapy may improve target coverage in this patient population, it requires repeated target delineation over the course of radiotherapy. Recently it has been shown that inter-observer variability in target delineation increases during treatment adaptive radiotherapy [9]. This analysis identified that although two of three geometric congruence metrics declined between the first and last time points, only DSC decline reached statistical significance. Both this study and Apolle *et al.*'s [9] analysis on MC inter-observer variability were carried out on a limited dataset (six observers delineating on five cases). Although less agreement was observed between the MC and the UA-AC over the course of radiotherapy, Apolle *et al.* [9] showed that MC reliability decreases during radiotherapy, and so further research is required in relation to both MC and AC.

Although the findings from this study are promising, we recognise that our analysis was limited to a single MC comparison ('gold standard') and comparison with another physician's contours may have yielded slightly different results. The relatively small sample size limits the applicability of our findings. However, the methodology was robust in that the analysis evaluated geometric agreement, dosimetric reliability and uniquely assessed trends with serial 4DCT scanning.

Conclusions

This feasibility study suggests that UA-AC can produce clinically and geometrically accurate GTV contours for NSCLC when appropriate cases are selected. Our findings suggest that adaptive strategies are merited in NSCLC radiotherapy and that automated techniques could augment adaptive radiotherapy workflows. Further work to validate the findings on larger datasets and incorporation into prospective studies of adaptive radiotherapy is warranted before clinical implementation.

Conflicts of Interest

S. Barrett and L. Marignol disclose a research grant from Varian Medical Systems, (grant number 011871), which partially funded this work. Varian Medical Systems had no input in this study design, data analysis or manuscript preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clon.2020.07.019.

References

[1] Njeh CF. Tumor delineation: the weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008;33(4):136–140.
[2] Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, *et al.* Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother Oncol* 2010;95(2):166–171.
[3] Nestle U, De Ruysscher D, Ricardi U, Geets X, Belderbos J, Pöttgen C, *et al.* ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer. *Radiother Oncol* 2018;127(1):1–5.
[4] van Dam IE, van Sörnsen de Koste JR, Hanna GG, Muirhead R, Slotman BJ, Senan S. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiother Oncol* 2010;96(1):67–72.
[5] Wong Yuzhen N, Barrett S. A review of automatic lung tumour segmentation in the era of 4DCT. *Rep Pract Oncol Radiother* 2019;24(2):208–220.
[6] Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, *et al.* Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep* 2013;3:3529.
[7] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, *et al.* Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol* 2018;126(2):312–317.
[8] Møller DS, Holt MI, Alber M, Tvilum M, Khalil AA, Knap MM, *et al.* Adaptive radiotherapy for advanced lung cancer ensures target coverage and decreases lung dose. *Radiother Oncol* 2016;121(1):32–38.
[9] Apolle R, Appold S, Bijl HP, Blanchard P, Bussink J, Faivre-Finn C, *et al.* Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. *Acta Oncol* 2019;58(10):1378–1385.
[10] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, *et al.* The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Dig Imag* 2013;26(6):1045–1057.
[11] Hugo, Geoffrey D, Weiss, Elisabeth, Sleeman, William C, *et al. Data from 4D lung imaging of NSCLC patients*. The Cancer Imaging Archive; 2016. https://doi.org/10.7937/K9/TCIA.2016.ELN8YGLE.
[12] Hugo GD, Weiss E, Sleeman WC, Balik S, Keall PJ, Lu J, *et al.* A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer. *Med Phys* 2017;44(2):762–771.
[13] Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. *Clin Oncol* 2010;22(7):515–525.
[14] Gaede S, Olsthoorn J, Louie AV, Palma D, Yu E, Yaremko B, *et al.* An evaluation of an automated 4D-CT contour propagation tool to define an internal gross tumour volume for lung cancer radiotherapy. *Radiother Oncol* 2011;101(2):322–328.

[15] Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;327(8476):307–310.
[16] Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015;25(2):141–151.
[17] Haslett K, Franks K, Hanna GG, Harden S, Hatton M, Harrow S, *et al*. Protocol for the isotoxic intensity modulated radiotherapy (IMRT) in stage III non-small cell lung cancer (NSCLC): a feasibility study. *BMJ Open* 2016;6(4):e010457. https://doi.org/10.1136/bmjopen-2015-010457.
[18] Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 1986;73(1):43–56.
[19] Eldesoky AR, Yates ES, Nyeng TB, Thomsen MS, Nielsen HM, Poortmans P, *et al*. Internal and external validation of an ESTRO delineation guideline – dependent automated segmentation tool for loco-regional radiation therapy of early breast cancer. *Radiother Oncol* 2016;121(3):424–430.
[20] Wright JL, Yom SS, Awan MJ, Dawes S, Fischer-Valuck B, Kudner R, *et al*. Standardizing normal tissue contouring for radiation therapy treatment planning: an ASTRO consensus paper. *Pract Radiat Oncol* 2019;9(2):65–72.
[21] Brink C, Bernchou U, Bertelsen A, Hansen O, Schytte T, Bentzen SM. Locoregional control of non-small cell lung cancer in relation to automated early assessment of tumor regression on cone beam computed tomography. *Int J Radiat Oncol Biol Phys* 2014;89(4):916–923.
[22] Tsang Y, Hoskin P, Spezi E, Landau D, Lester J, Miles E, *et al*. Assessment of contour variability in target volumes and organs at risk in lung cancer radiotherapy. *Techn Innov Patient Supp Radiat Oncol* 2019;10:8–12.
[23] Berkovic P, Paelinck L, Lievens Y, Gulyban A, Goddeeris B, Derie C, *et al*. Adaptive radiotherapy for locally advanced non-small cell lung cancer, can we predict when and for whom? *Acta Oncol* 2015;54(9):1438–1444.
[24] Britton KR, Starkschall G, Liu H, Chang JY, Bilton S, Ezhil M, *et al*. Consequences of anatomic changes and respiratory motion on radiation dose distributions in conformal radiotherapy for locally advanced non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2009;73(1):94–102.
[25] Ramella S, Fiore M, Silipigni S, Zappa MC, Jaus M, Alberti AM, *et al*. Local control and toxicity of adaptive radiotherapy using weekly CT imaging: results from the LARTIA trial in stage III NSCLC. *J Thorac Oncol* 2017;12(7):1122–1130.
[26] Richmond ND, Pilling KE, Peedell C, Shakespeare D, Walker CP. Positioning accuracy for lung stereotactic body radiotherapy patients determined by on-treatment cone-beam CT imaging. *Br J Radiol* 2012;85(1014):819–823.

INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Predicting Individual Tumor Response Dynamics in Locally Advanced Non-Small Cell Lung Cancer Radiation Therapy: A Mathematical Modelling Study

Sarah Barrett, MSc,* Mohammad U. Zahid, PhD,† Heiko Enderling, PhD,†,‡ and Laure Marignol, PhD*

*Applied Radiation Therapy Trinity, Trinity St. James's Cancer Institute, Discipline of Radiation Therapy, Trinity College, Dublin, Ireland; †Department of Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; and ‡Institute for Data Science in Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas

Received Mar 21, 2024; Accepted for publication Oct 28, 2024

Purpose: To predict individual tumor responses to radiation therapy (RT) in non-small cell lung cancer.
Materials and Methods: The proliferation saturation index (PSI) model, which models tumor dynamics in response to RT as an instantaneous reduction in tumor volume, was fit to n = 162 patients with 4 distinct dose fractionation schedules (30-32 fractions × 2 Gy, 23-24 fractions × 2.75 Gy, 32-42 fractions × 1.8 Gy, and 30 fractions × 1.5 Gy Bidaily, followed by 5-12 fractions × 2 Gy daily). Following initial training, the predictive power of the model was tested using only the first 3 tumor volume measurements as measured on daily imaging. The remainder of tumor volume regression during RT was simulated using the PSI model. Comparisons of the measured to the simulated volumes were made using scatter plots, coefficient of determination (R^2), and Pearson correlation coefficient values.
Results: The PSI model predicted tumor volume regression during RT with a high degree of accuracy. Comparison of the measured versus predicted volumes resulted in R^2 values of 0.968, 0.954, 0.968, and 0.937, and Pearson correlation coefficient values of 0.984, 0.977, 0.984, and 0.968 in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively.
Conclusions: The proliferation saturation model can predict, with a high degree of accuracy, non-small cell lung cancer tumor volume regression in response to RT in 4 distinct dose fractionation schedules.

Introduction

Gross tumor volume (GTV) changes during radiation therapy (RT) for non-small cell lung cancer (NSCLC) are common, with up to one-third of all patients having significant tumor shrinkage, but the ability to predict which tumors may shrink remains elusive.[1] The rate of GTV regression for individuals varies and has been reported in this population at a rate of 1.5% decrease per fraction, resulting in a mean regression of 43.1% (±23.1 SD) from the start to end of RT.[2]

Tumor volume regression in response to RT is also an emerging imaging biomarker that may be predictive of

Corresponding author: Sarah Barrett, MSc; E-mail: barrets7@tcd.ie

Disclosures: A research grant from Varian Medical Systems, (grant number 011871), partially funded this work to S.B. and L.M. Varian Medical Systems had no input in this study design, data analysis, or manuscript preparation. Supported, in part, by NIH/NCI award U01CA244100 to H.E.

Author Responsible for Statistical Analysis: Sarah Barrett, MSc; Email: barrets7@tcd.ie

Data Sharing Statement: Research data are not available at this time.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2024.10.038.

0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2024.10.038

clinical outcomes such as local control (LC) or overall survival (OS),[3-5] and research is ongoing in this area.[4,6] Current research focuses on retrospective analyses of tumor volume dynamics over time and how these relate to clinical outcomes. In this context, it would be desirable to predict the tumor regression at an earlier time point, and if the response is predicted to be suboptimal, dose-adapt RT to improve that response.

Herein, we evaluated the performance of a mathematical model to predict tumor regression over the course of RT delivery. The model, originally proposed by Prokopiou et al[7], utilizes longitudinal tumor volume measurements to calculate a patient-specific proliferation saturation index (PSI) and uses this index to simulate and predict response to RT.[8-11] The PSI model allows us to simulate tumor response to RT using an initial dose fractionation schedule, or test response to alternative prescriptions if a modeled response is suboptimal. We can then theoretically optimize the RT prescription for each patient as demonstrated recently for total dose personalization for head and neck cancer.[12]

The PSI model is a patient-specific, single-variable response model that simulates tumor volume regression over the course of RT delivery.[7] The model is based on the ecological concept that tumors exist in host environments with limited resources, and therefore each environment has a specific carrying capacity *(K)*. This carrying capacity is patient-specific because it is relevant only to that tumor and its microenvironment.[8] PSI is defined as the volume of the tumor at the start of RT divided by the *K*. It follows that tumor volumes close to their *K* with a high PSI would have a small proportion of proliferating cells, caused by the reduced availability of resources. These high PSI tumors are anticipated to show a shallow response to RT because radiation sensitivity is related to proliferation.[13] Conversely, tumors with a low PSI would have ample access to resources, be highly proliferative, and therefore respond quickly to irradiation and show greater regression.

The PSI model can dynamically evaluate the change in tumor volume over a given time for specific RT schedules. The model parameters are calibrated using tumor volume measurements from complete courses of RT. Using this information, the PSI model can determine the population growth rate, λ(unit 1/d), and a radiation sensitivity value, α (1/Gy). The only free parameter for each patient is PSI. For each subsequent patient, the model is solved on tumor volume data taken early during RT using the population level growth and radiation sensitivity rates to determine the individual PSI and predict the individual tumor volume regression in response to the remainder of the treatment course.

While a change in tumor volume is a vast oversimplification of the underlying tumor growth mechanisms and radiobiological response, this simplicity is also the appeal of the PSI model because there is only a single patient-specific variable. While a plethora of more complex mathematical models have been developed to simulate radiation response,[14-22] it is clinically infeasible to calibrate multiple interdependent parameters and make patient-specific predictions with limited uncertainty from a few clinical measurements.[23,24] The appeal of the simple PSI model is that it can predict tumor response to RT early in treatment, which could be a method of triggering plan adaptation. It enables the simulation of responses to a variety of prescriptions and could inform the selection of the optimal prescription for individual patients.

The PSI model has been simulated using a small retrospective cohort of patients with NSCLC[7,8] and has been explored more thoroughly in the head and neck RT setting where it is shown to accurately predict tumor volume dynamics during RT.[9,10,12] A prospective clinical trial has demonstrated promising results in using PSI to optimize prescription and clinical outcomes.[25]

Herein, we demonstrate the utility of the PSI modeling approach across a range of dose fractionation schedules in NSCLC. As outlined in the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement,[26] this work is a Type 1a study, describing the development of a prediction model. This study tests the robustness of the model's performance across 4 conventionally fractionated dose prescriptions commonly used in the treatment of locally advanced NSCLC (LA-NSCLC). It determines the model's performance at fitting data in these prescription groups, with 1 or 2 free parameters, and it will assess whether it accurately predicts tumor response to RT in these patients. Finally, we ascertained the model sensitivity to α and λ parameter variations.

Materials and Methods

The mathematical model

In the PSI model,[7-9] tumor growth is modeled as logistic growth using the differential equation: $\frac{dV}{dt} = \lambda V\left(1 - \frac{V}{K}\right)$,[9] where λ (1/d) is the intrinsic growth rate of the tumor, V is the gross tumor volume (cc), and *K* is defined as tumor carrying capacity (cc). PSI is defined as the ratio of tumor volume at the start of RT (V_0) to the *K* $PSI = \frac{V_0}{K}$. PSI is, therefore, a unitless value between 0 and 1 and is representative of the proportion of the tumor volume not proliferating.

Response to RT is modeled following the Norton-Simon hypothesis that RT response is proportional to the tumor growth rate,[27] using the linear quadratic model $SF(d) = e^{-\alpha d - \beta d^2}$, where *SF(d)* is the surviving fraction of cells following a given dose of radiation, d (Gy), and α (1/Gy) and β (1/Gy^2) are radiation sensitivity parameters. Finally, the change in volume following a dose of RT is modeled as an instantaneous reduction, $V_{post\ RT} = V_{pre\ RT} - V_{pre\ RT}\,\gamma\left(1 - \frac{V_{pre\ RT}}{K}\right)$, where γ is the volumetric death term per RT dose and is linked to the LQ model $\gamma = 1 - SF(d)$. For NSCLC, an α/β ratio of 10 Gy is recognized, and so the model only estimates the α parameter. Note that in this case, α is the volumetric radiosensitivity

and thus not comparable with *in vitro* estimates of radiosensitivity values.

Custom scripts in MATLAB (Mathworks) were generated to simulate tumor volume response with inputs of dose fractionation schedule, λ, α, and PSI. A variety of optimization scripts were written using the particle swarm optimization (PSO) function within the MATLAB Global Optimization Toolbox. The optimal parameter values are found by minimizing the mean square error (MSE) across all the data, therefore finding the values for λ, α and PSI, which result in the best fit to the data. Data required to determine these values are longitudinal tumor volume measurements, the associated time points for each measure in days since the start of RT, and the dose prescription. These scripts can determine individual values of λ, α and PSI (given a set α/β ratio), or alternatively, set values for λ and α can be input to only allow PSI as a patient-specific individual parameter. Finally, tumor regression is modeled beyond the input data to simulate tumor response to RT from an early time point.

Patient population

A de-identified data set containing tumor volume measurements for 141 patients with NSCLC was abstracted from a prior publication.[5] This dataset is described fully previously,[5] but briefly, contained tumor volume measurements acquired longitudinally during a course of radical RT or concurrent chemotherapy RT (ChemoRT). Patients within the data set were heterogeneous in relation to stage, histology, and use of chemotherapy; however, all patients were treated with an RT dose > 45 Gy. Tumor volume was measured using a free-breathing cone beam computed tomography (CBCT) scan. Patients were treated with several distinct dose fractionation schedules. For the purposes of this study, patients from this data set were those treated with 3 prescriptions: 23 to 24 fractions × 2.75 Gy daily (n = 28), 32 to 42 fractions × 1.8 Gy daily (n = 27), and 30 fractions × 1.5 Gy Bidaily followed by 5 to 12 fractions × 2 Gy daily (n = 84), resulting in n = 139 patients.

An additional data set was abstracted from a second publication.[28] This data set consisted of 25 patients with LA-NSCLC, treated with daily RT on the Hi-Art helical tomotherapy unit at the London Regional Cancer Program, Ontario, Canada, from 2005 to 2007 and is most fully described in Woodford et al[29] in 2007. Of the 25 patients, 2 were excluded because they were not treated with a radical dose prescription (n = 23). These data were previously included in a PSI evaluation study[8]; however, in this work, the PSI model algorithm has been updated to PSO from a genetic algorithm, which may alter the model performance. A prescription dose of 30 to 32 fractions × 2 Gy per fraction was used for these patients, and all were treated with neoadjuvant chemotherapy.[28] GTV of the primary tumor only was measured on daily megavoltage CT (MVCT) images. Each data set included tumor volume measurements and the associated time point it was acquired (daily or weekly imaging) from day 1 of RT.

Model parameters

The PSI model has 3 modifiable parameters, a tumor intrinsic growth rate (λ), a radiation sensitivity value (α), and the patient-specific PSI. As the data for this study contained no pretreatment information to calculate unperturbed tumor growth, the published λ = 0.012 1/d was utilized for all patients,[8] α and PSI were derived as per the methods below.

PSI model data fitting with 2 individual parameters

For each prescription group, an optimal fit of the data was derived using MATLAB PSO. The model was allowed to optimize parameter values for α and PSI for each patient (α_i, and PSI_i), by minimizing the residual MSE between measured and simulated tumor volumes over time. Lower bounds of 0 and upper bounds of 0.99 were set for both parameters. Scatter plots of measured versus simulated tumor volumes were generated for each prescription group. The coefficient of determination (R^2) and MSE were calculated as a measure of how well-simulated values matched measured values.

R^2 was calculated using the formula, $1 - \frac{RSS}{TSS}$, where RSS = residual sum squared (measured volume − simulated volume)^2 and TSS = total sum of squares (sum of measured volumes − mean of measured volumes)^2.

MSE for each patient, *i*, was calculated as the normalized square error at each observation, *j*, between the measured volume, $V_{m,j}$, and the simulated volume, $V_{s,j}$, averaged over the number of observations, n_m:

$$MSE_i = \frac{1}{n_m} \sum_{j=1}^{n_m} \frac{(V_{m,j} - V_{s,j})^2}{V_{m,j}},$$

and MSE per group (MSE_g) was calculated as the average of MSE_i of all patients in the group:

$$MSE_g = \frac{1}{n_{patients}} \sum_{i=1}^{n_{patients}} MSE_i.$$

Akaike information criterion (AIC) was calculated as the measure of model quality using the formula $(2 \times k) + n \times \ln(SSE_i/n)$, where k is the number of free parameters, n is the number of data points, and $SSE_i = MSE_i \times n_m$ is the sum squared error for each patient, *i*. The AIC value per group is calculated as the average AIC value of all patients in the group. When comparing models using AIC values, it is assumed that the model with the lowest AIC (AIC_{min}) results in the least information lost from the true data. A change in AIC (ΔAIC) values from $AIC_{min} \leq 2$ is substantial evidence that there is no loss in model quality; ΔAIC values ranging between 4 and 7 have

considerably less support than the AIC_{min} model, and models with ΔAIC values > 10 have essentially no support.[30]

The range of α_i and PSI_i values generated for each patient were visualized with Box and Whisker plots. Mean and standard deviations were calculated for each group. The range of α_i, values derived between prescription groups were compared using single-factor analysis of variance and Box and Whisker plots to estimate the biological volume reduction effect of each dose fractionation schedule.

PSI model data fitting with a single patient-specific parameter

An optimal α for each prescription group (α_g), was determined by testing a range of α values in each group. The value that generated the best fit to the measured data, assessed by the highest R^2 achieved, was selected. The model then derived the optimal fit to the data using α_g and was allowed the freedom to generate only an individual PSI parameter (PSI_i). Following data fitting with α_g and PSI_i, scatter plots were generated and R^2, MSE_g, and AIC in each prescription group were calculated. Finally, data fitting with a population α (α_p) was tested using the mean of the group α values. Data were compared with prior values generated using α_i and PSI_i.

Prediction of tumor response to RT

To test the predictive power of the model, only the initial 3 tumor volume measurements for each patient were inputted into the model, along with the population α value, and the tumor regression during the remainder of RT was simulated. The simulated volumes were compared with the measured volumes, excluding the initial 3 measures, using scatter plots, R^2, and Pearson correlation coefficient. A sensitivity analysis of the model parameters was carried out to quantify the model robustness by varying the input parameters by ± 20% and comparing the resulting R^2 and Pearson correlation coefficients to those generated with the optimal parameters. The individual PSI values derived using only 3 data points were compared with those derived using all data using Box and Whisker plots and paired *t* tests.

Data analysis

Model simulations and data analyses were carried out in MATLAB (Mathworks), Microsoft Excel, and SPSS.

Results

Description of patient population and available tumor volume measurements

A total of 162 patients with 1474 tumor volume measurements were included in this study.

Patients in the 2 Gy per fraction group had a median of 30 tumor volume measurements (range, 15-33) measured on MVCT. For the remaining 3 prescription groups, a median of 6 tumor volume measurements was available for each patient over the course of treatment, ranging from 5 to 7, with 2 patients having 7 CBCTs and 24 having 5 CBCTs. All patients had CBCT on day 1 of RT. See Figure 1 for an

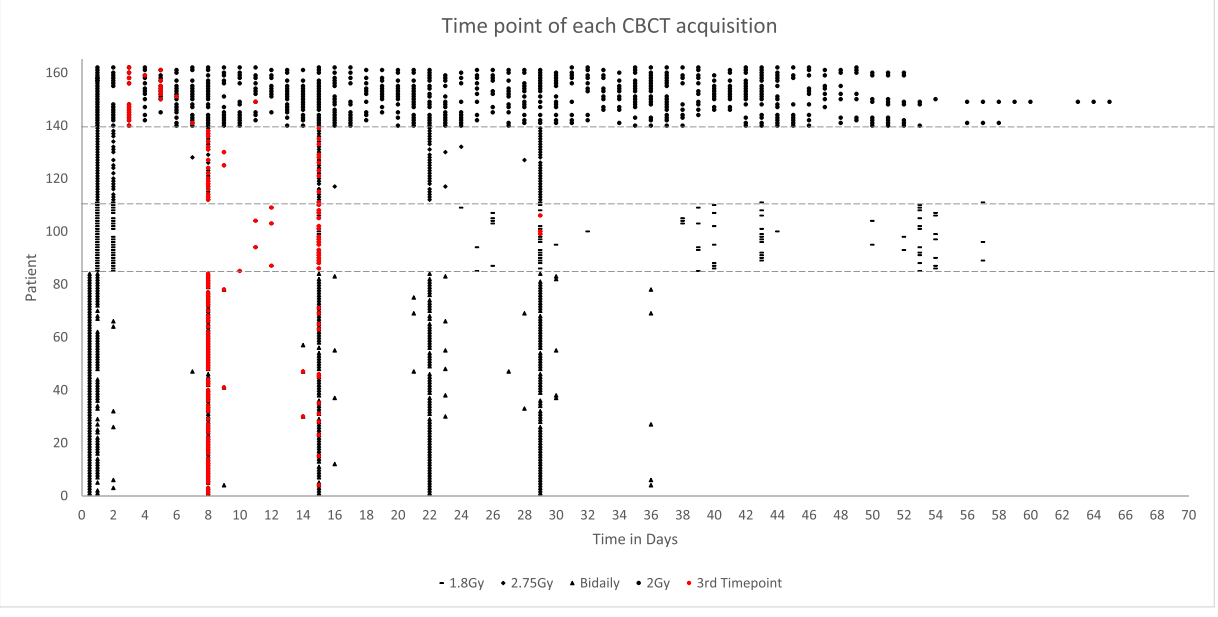


Fig. 1. CBCT acquisition time points. For each patient, the time point of each tumor volume measurement in days since RT started is displayed in black. The time point of the third measurement for each patient is highlighted in red. Patients in the Bidaily group had more than 1 image on day 1 and so these are coded as days 0.5 and 1.
Abbreviations: CBCT = cone beam computed tomography; RT = radiation therapy.

overview of scan acquisition time points. Median (range) tumor volumes on day 1 for each group were 115.30 (3.2-830.8) cc, 32.5 (0.6-198.3) cc, 29.6 (1.3-327.8) cc, and 43.4 (0.7-318.1) cc for the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively (Fig. E1). Volumes in the 2 Gy data set were significantly larger than the other 3 ($P < .0001$) but no significant difference was seen between the remaining 3 groups ($P = .667$). Demographic and clinical characteristics data for each patient were unavailable.

PSI model fits data with individual patient parameters

We first tested that the PSI model was able to fit the tumor volume kinetics data with individual α and PSI parameters. The PSI model with α_i and PSI_i fits the data in each prescription group with high accuracy. Scatter plots for each group show excellent agreement between measured and simulated tumor volumes (Fig. 2), as indicated by R^2 values of 0.993, 0.995, 0.997, and 0.956 and MSE_g values of 0.625, 0.142, 0.211, and 2.071 in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively. The range and standard deviation for derived PSI_i values were larger than α_i in all 4 prescription groups (Fig. 2). The range of α_i values derived for each prescription group (Fig. E2) was compared using single-factor analysis of variance and were not found to be significantly different, $P = .0845$. Mean ($\pm$SD) α_i values were 0.028 ($\pm$ 0.018), 0.035 ($\pm$ 0.022), 0.024 ($\pm$ 0.021), and 0.035 ($\pm$ 0.024) in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively.

PSI model fits the data with a single patient-specific parameter

We next examined whether the data fitting performance of the PSI model was significantly impacted when we used a prescription-specific α parameter or a population α parameter.

The α parameter for each prescription group (α_g) was estimated at 0.02, 0.03, 0.022, and 0.048 in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively. Using the relevant α_g and allowing individual PSI (PSI_i), excellent fit of the data was achieved as shown in scatter plots (Fig. 3). R^2 values were found to be 0.99, 0.991, 0.995, and 0.992, and MSE_g values were 1.202, 0.231, 0.289, and 0.314, respectively. An improvement in data fit was observed in the Bidaily group with an increase in R^2 from 0.956 to 0.992 and a decrease in MSE_g from 2.071 to 0.314.

Mean AIC values were calculated and compared with original data fitting with α_i and PSI_i to examine whether there was a loss in model quality moving from 2 to 1 free parameter. In the 1.8 Gy and 2.75 Gy groups, marginal changes in mean AIC values were observed, with a ΔAIC value increase of + 0.37 and a ΔAIC value decrease of -0.08 measured in the single-parameter models, respectively. A moderate increase in mean AIC value (ΔAIC value, +6.12) was measured in the 2 Gy group, indicating a moderate loss of data fit quality when reducing to one free parameter in this group. A substantial decrease in mean AIC value (ΔAIC value, -9.06) was observed in the Bidaily group, indicating that the single-parameter model resulted in a truer fit to the data in patients with this prescription.

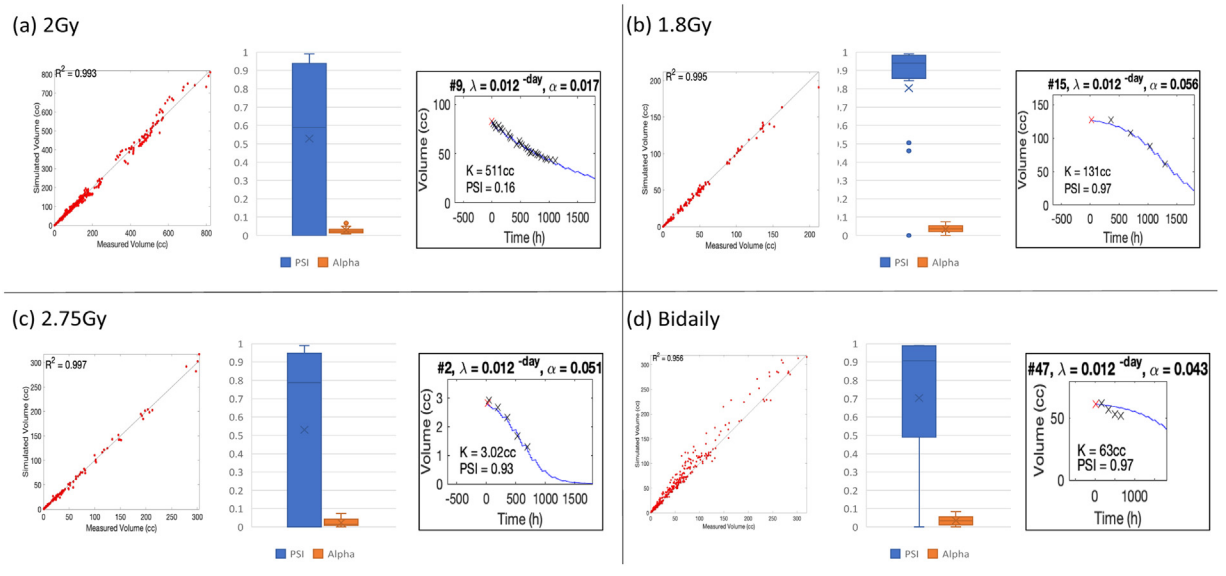


Fig. 2. PSI model data fitting and parameter output values. (a-d) Scatter plots of data fitting using α_i and PSI_i across the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily prescription groups, respectively. Box and Whisker plots display the values derived for PSI_i and α_i in each group; additionally included is a sample of individual patients fit in each prescription group with measured volumes marked as Xs and model simulated volumes displayed as blue curves.
Abbreviation: PSI = proliferation saturation index.

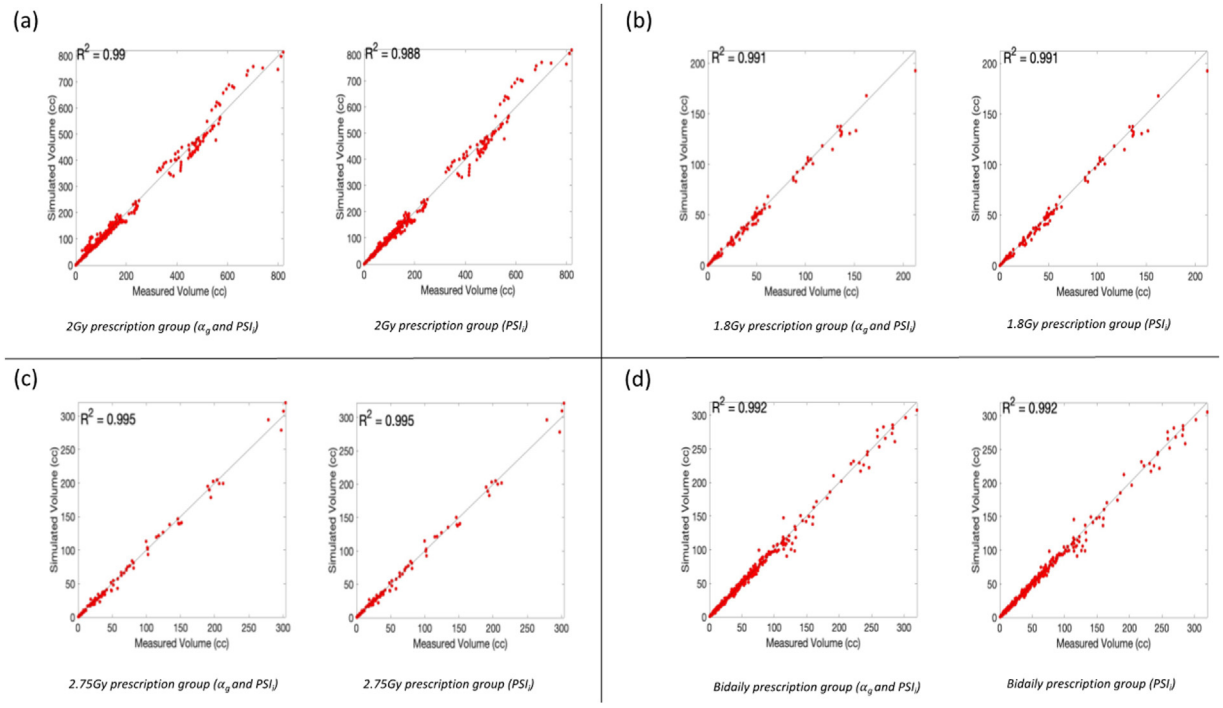


Fig. 3. Data fitting using group and population alpha (α) values. (a-d) Scatter plots of data fitting performance in the single-variable model using prescription group α (α_g, PSI_i) and population α (α_p, PSI_i) across the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily prescription groups, respectively.
Abbreviations: PSI = proliferation saturation index; PSI_i = individual proliferation saturation index.

Finally, a population α parameter, (α_p) was tested using the mean of the group α values, $\alpha_p = 0.03$.

Again we found an excellent fit of the data with R^2 values of 0.988, 0.991, 0.995, and 0.992 and MSE_g values of 0.878, 0.23, 0.293, and 0.324, in the 2 Gy, 1.8 Gy, 2.75 Gy and Bidaily groups, respectively (Fig. 3). Minimal changes in AIC values were noted moving from (α_g) to (α_p) with the same trends as those reported above. In the 1.8 Gy and 2.75 Gy groups marginal changes in mean AIC values were observed, the ΔAIC values +0.37 and −0.25, respectively, compared with the 2-parameter model. In the 2 Gy group, the ΔAIC value was +6.12, and in the Bidaily group again the ΔAIC value decreased substantially by −9.11.

PSI model accurately predicts tumor volume regression during RT

Finally, we tested the ability of the PSI model to predict tumor volume kinetics based on the initial 3 measurements of volume acquired during RT using population λ and α. PSI model-predicted tumor volume regression during RT with a high degree of accuracy, as seen in the scatter plots demonstrating measured versus predicted volumes (Fig. 4), with values for R^2, 0.968, 0.954, 0.968, and 0.937 and Pearson correlation coefficient values, 0.984, 0.977, 0.984, and 0.968 in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively. Individual patient plots of model prediction versus measured data are available in Figures E3 to E6.

Patient-specific PSI values derived with only 3 data points were compared with those generated with the complete data set. Those derived from the Bidaily group were significantly different with mean PSI in the complete data set = 0.631 and mean PSI derived from 3 data points = 0.889 ($P < .00001$) (Fig. E7).

The model was tested for sensitivity to parameter changes by varying the α and λ parameters by ± 20%. No substantial change in performance was observed, with the largest percentage difference in either R^2 or Pearson Correlation Coefficients being 2.15%. See Table E1 for all values.

Discussion

The PSI model is a mathematical tool that simulates and predicts volumetric tumor response dynamics during RT. This study examined the PSI model performance fitting data in 4 distinct prescription groups, derived a population radiation sensitivity parameter (α_p), and established the performance of the model predicting tumor response to RT in each group.

We showed that the PSI model can fit longitudinal data from patients treated for LA-NSCLC across a variety of dose prescriptions to a high degree of accuracy (R^2 range, 0.956-0.997, Fig. 2), using a population-wide tumor growth rate parameter (λ_p) derived from the literature. This is an improvement on the previously reported results in both NSCLC[8] and oropharyngeal cancer.[9] Sunassee et al[8]

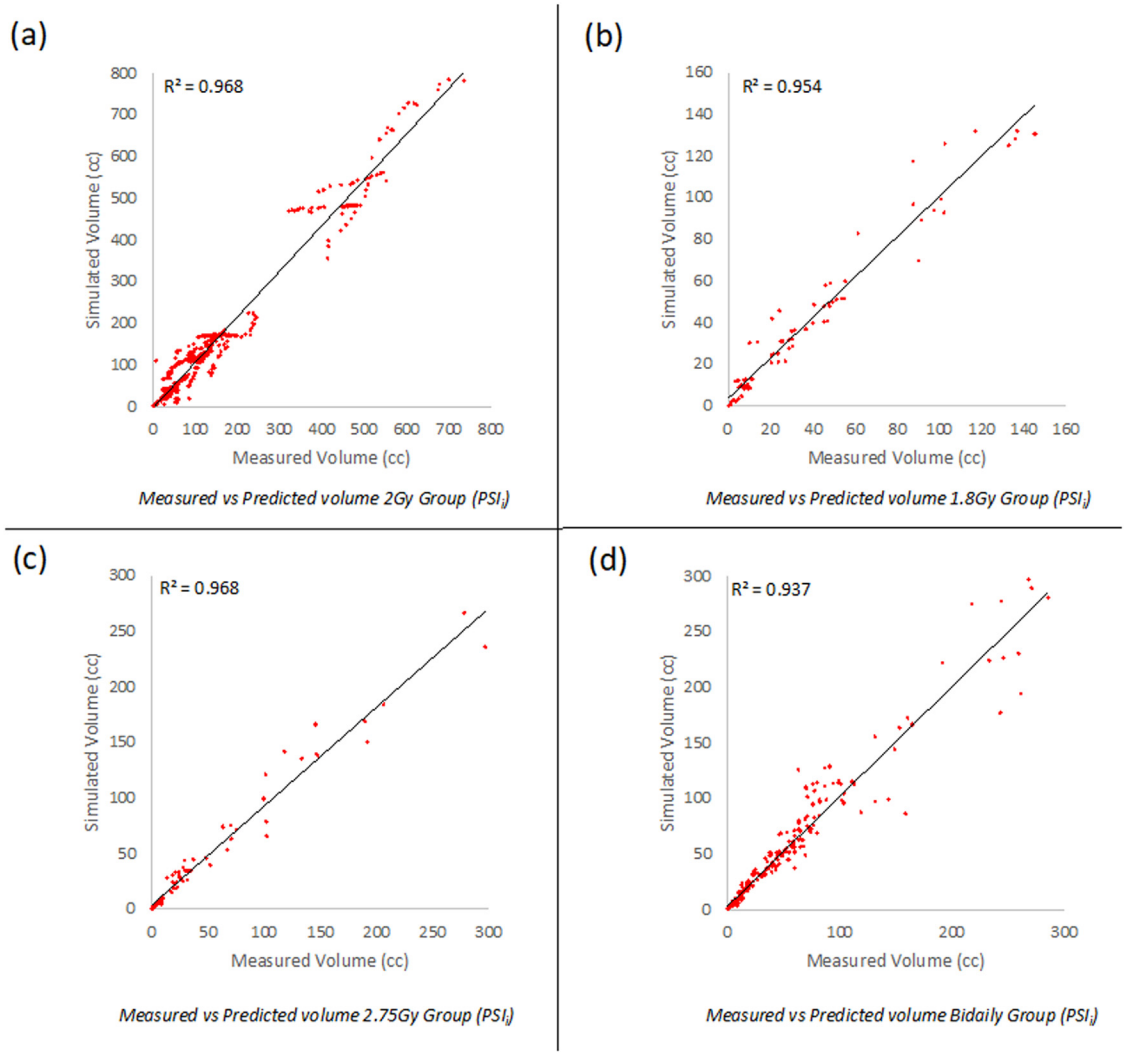


Fig. 4. Data prediction scatter plots show measured versus predicted tumor volumes in each prescription group single-variable model (α_p and PSI_i), excluding the first 3 tumor volume measurements. (a-d) Values in the 2 Gy, 1.8 Gy, 2.75 Gy, and Bidaily groups, respectively.
Abbreviations: PSI = proliferation saturation index; PSI_i = individual proliferation saturation index.

previously examined the PSI model in the patient group treated with 2 Gy/d, also included in this study. Here we found an improvement in the models' performance with a previous R^2 value of 0.91 reported compared with an R^2 value of 0.99. This prior work employed an earlier iteration of the PSI model, which did not use PSO to fit the data. PSO is a computational method that aims to iteratively solve a problem by having a number of potential solutions that move around the search space to find an optimal solution. The prior work employed a genetic algorithm to derive a combination of parameters that best fits patient data[8,31] and this may be susceptible to converging to a local minimum instead of the global minimum when solving for optimal parameters. Improvement was also seen in the model's predictive performance, with this study reaching an excellent agreement between measured and simulated volumes in this group, with an R^2 value of 0.97 compared with an R^2 value of 0.84 in the previous work,[8] which may be explained by both the change in algorithm and the omission of 2 patients, who may have been incorrectly identified as having radical RT. It is notable also that the prior work predicted regression from week 2 of RT, but this analysis used only the first 3 tumor measures, which for the majority of patients (n = 19/23) were acquired by day 5 of RT, indicating this improved performance was at an earlier time point in treatment.

We allowed the model freedom to create an individual α and PSI for each patient to maximize data fit. The range of values derived is shown in Figure 2. PSI was more variable between patients than α, with PSI values ranging from 0 to 0.99 and α ranging from 0 to 0.08. While this was previously investigated in the 2 Gy per fraction group,[8] it was not tested in a variety of dose fractionation schedules. These findings support the hypothesis that PSI is more variable and patient-specific than the α parameter across all prescription groups.

We next investigated an optimal α value for each prescription group by testing a range of α values and establishing which value resulted in the best fit to data, and the difference in individual α values did not reach significance in this small sample (P = .0845, Fig. E2). Finally, this study confirms that a population α_p value, regardless of dose per fraction, has minimal impact on data fit compared with α_g, with negligible variation in the R^2 (Fig. 3).

We hypothesized that there could be a loss in model quality when reducing from 2 to 1 free parameters. Model quality was assessed using the AIC formula. AIC is a mathematical method to evaluate how well a model fits the data,[32] and in this case, we calculated AIC for both single-parameter models (α_g/α_p and PSI_i) and the model with 2 free parameters (α_i and PSI_i).

Interestingly, an AIC_{min} value was observed in the single-parameter model (α_p, PSI_i) in 2 groups (2.75 Gy and Bidaily), indicating that a single-parameter model is preferable. Δ AIC value was marginal in the 2.75 Gy group (0.25) but was larger in the Bidaily group (9.11), indicating much less support for the 2-parameter model in this dose fractionation schedule. On review of the scatter plots and individual data fit, there was also an observable improvement in the performance of the single-parameter model in the Bidaily group (Fig. 5). This unanticipated improvement in the model was robust to adjustments of individual PSO parameters and therefore could not be attributed to poor optimization within the model. We anticipate that future examination of a larger data set will elucidate these results. This prescription was the most complex, with Bidaily

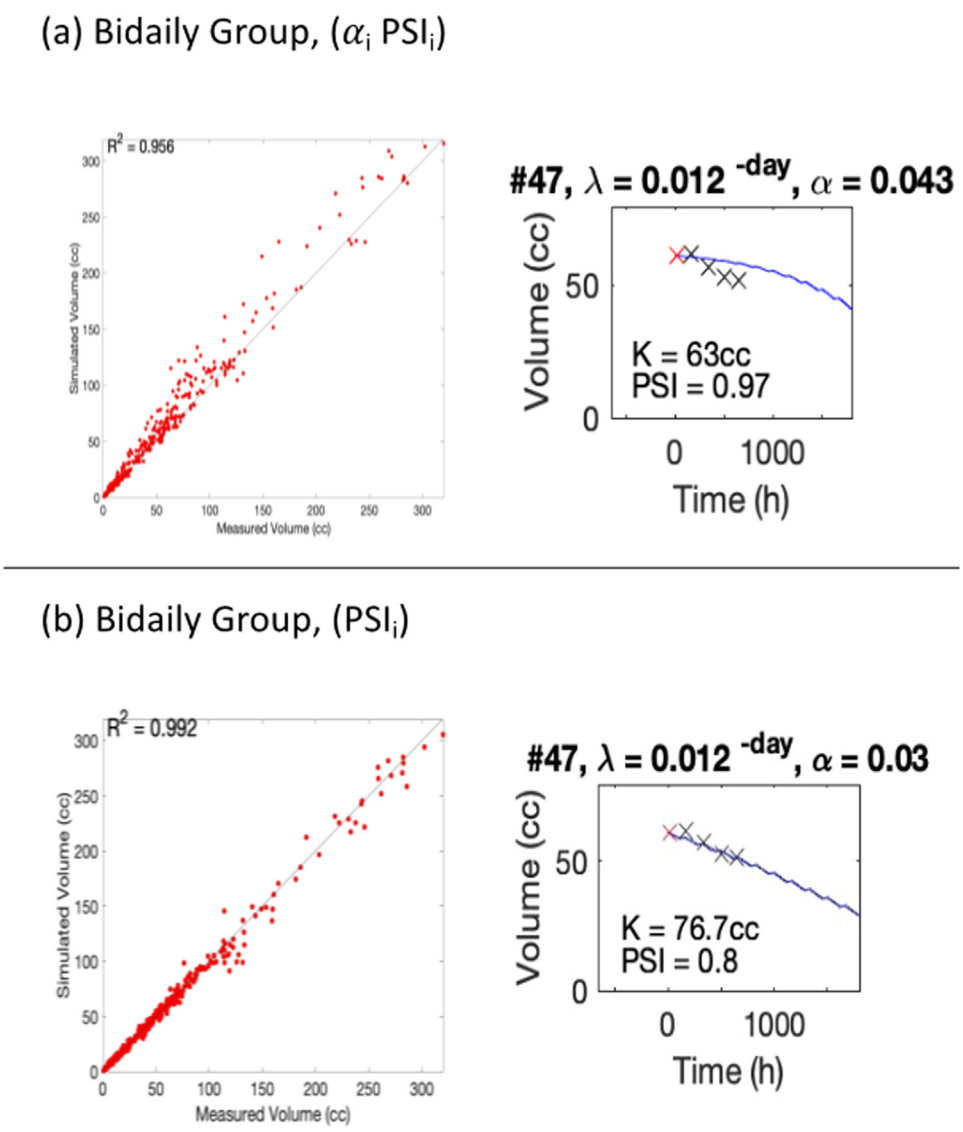


Fig. 5. Bidaily prescription group variation in fit between single- and 2-parameter models. (a) Scatter plot of data fitting using α_i and PSI_i in the Bidaily prescription group and a sample of individual patient fit with measured volumes marked as X and model simulated volume displayed as blue curve. (b) The same data and individual patient using α_p and PSI_i.
Abbreviations: PSI = proliferation saturation index; PSI_i = individual proliferation saturation index.

followed by daily fractions, and this finding warrants further investigation in future work.

There was no significant loss in model performance, either in AIC or in R^2 values observed when using α_p compared with using α_i. This suggests that we can reasonably use the single-variable model in differing prescriptions. Again, these results are in line with the prior head and neck and NSCLC papers that evaluated the PSI model,[7-9] which also found reasonable data fitting allowing only PSI as a patient-specific parameter within the model. However, this study was the first to evaluate the model in a number of dose fractionation schedules for NSCLC.

Finally, the ability of the model to predict tumor response based on early tumor volume dynamics was tested. The first 3 tumor volume measurements were used to predict the remainder of RT, and these measures were acquired at varying time points in the 4 groups. In the 2 Gy group, the majority of patients had a third volume by day 5 of RT. In the 1.8 Gy and 2.75 Gy groups, a third measure was acquired, in the majority of cases, by days 8 to 15. In the Bidaily group, the majority of third tumor volume measures were acquired by day 8. Therefore, the prediction of tumor regression took place from days 5 to 15 in most cases (Fig. 1). The performance of the model stratified using image acquisition time points was not assessed in this study because we aimed to test the prediction at the earliest possible time point and therefore used the first 3 measures for all patients. Future work in a homogeneous population could examine the performance improvement when adding more data to the prediction. We found that the single-variable PSI model-predicted tumor volume regression across the entire population with a high degree of accuracy (Fig. 4), with R^2 values ranging from 0.937 to 0.969 and all Pearson correlation coefficient values > 0.968. The findings demonstrate that the model can predict RT response in varying dose fractionation schedules, a prerequisite if we wish to use the PSI model to simulate the response to alternative prescriptions. Prior work by Sunassee et al[8] only evaluated the model performance in patients receiving 2 Gy/fraction, which may not be representative of the current approaches in lung cancer RT.

While we see from these values that the model performed well across the population, the scatter plots reveal some patients where a poor correlation was observed between measured and predicted tumor volume. This is most notable in the 2 Gy group (Fig. 4a). Two patients had predicted volumes that changed very little over the course of RT, but the actual tumor volume reduced. On closer review of these 2 cases, one showed a small volume growth in the first 3 measures (477.7-492.9 cc), and the second had only a very slight volume decrease (570.7-564.5 cc). In both cases, the model predicted little to no tumor volume regression during RT. These patients demonstrated that while on a population level the model predicted to a high degree of accuracy, in individual scenarios where tumor volume does not decrease there may be difficulty in predicting the remainder of treatment. Of note, both cases occurred in the 2 Gy data group, which had daily MVCT images, meaning the model was using data from just the first 3 to 5 days to simulate RT response for the remainder of treatment.

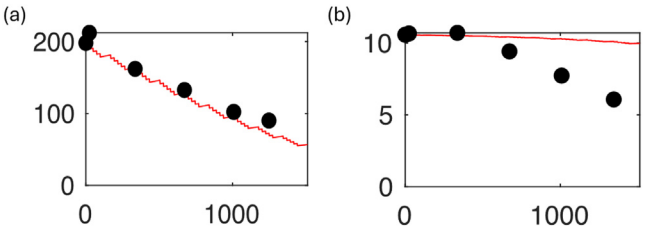


Fig. 6. Individual patient plots of measured versus predicted volume. Black dots are tumor volume in cc versus time in hours over the course of RT. The red curve is model-predicted tumor regression over the course of RT based on the first 3 tumor measures. (a) A patient where the modeled and actual volume agreed. (b) A patient where the model failed to predict the tumor regression.
Abbreviations: RT = radiation therapy.

Plots showing residual absolute volume difference between measured and predicted volumes are available in Figures E8 to E11 and also demonstrate very good agreement for the majority of patients, with less agreement observed for several patients in the 2 Gy group.

Reviewing the individual fitting plots for all patients further enforces that the model may underperform in a subset of individual cases (Figs. E3-E6). Examples can be seen in Figure 6, where the model performed well and poorly. In the case where the model failed to predict the tumor regression, the first 3 data points, on which the prediction is based, show little change in tumor volume. In these cases, it would be possible to add more data points to the model to improve the prediction; however, this comes at the cost of predicting further into RT delivery, which may be less clinically relevant. See Figure E12, which demonstrates the improved performance in this case when adding further data to the model. Further correlation with biological proliferation markers such as Ki-67 would be of interest; however, needle biopsy samples taken for histological diagnosis are not routinely tested for Ki-67, so this is not feasible outside of a prospective clinical trial.

Accurate prediction of tumor regression during RT has several potential benefits. We know that the RT plan created prior to treatment is rarely an exact representation of the dose delivered by the end of treatment caused by interfraction and intrafraction variation. Significant changes in anatomy can alter the planned dose distribution and may result in an increased dose to normal tissue or a decreased dose delivered to the target volume.[33,34] Predicting which tumors will significantly reduce in size could facilitate preplanned adaptive RT (ART) as opposed to reactive ART, which can cause delays in workflow both at the treatment unit and for the patient. A 2020 European survey highlighted that a change in ART technique, or increased use of ART in lung cancer RT, is a priority for many centers.[35] One study found that 90% of patients treated with ChemoRT displayed GTV

regression during their treatment, with a mean regression rate of 1.5% per fraction.[2] In that study, no trends were observed correlating tumor regression to age, histology, nodal status, initial tumor size, or chemotherapy.[2] This highlights the challenge in predicting those that may have clinically relevant regression, and yet the PSI model tested here, performed well across a heterogeneous population where these factors were unaccounted for.

An emerging clinical interest is the potential link between tumor volume regression and clinical outcomes such as LC or OS.[3-6] While this link has not been clearly established, a 2018 meta-analysis highlights the time of assessing GTV reduction because of one of several potential confounding factors.[4] More recently, a retrospective observational cohort study found that GTV volume following 40 to 50 Gy RT was a more robust predictor of OS than initial GTV,[6] a finding echoed in a second retrospective analysis.[36]

The potential benefit of the PSI model in this scenario is twofold. First, if regression is shown to be a reliable clinical predictor, advanced knowledge of the regression and the subsequent clinical response could facilitate a more informative conversation with the patient about the potential outcome and progress of their treatment. Second, because this work has shown that the model predicts regression accurately across a variety of prescriptions, future work can model an individual's response to an alternative prescription to test if the predicted response can be improved. Mid-treatment tumor volume reduction has been shown to predict clinical outcomes in NSCLC.[6] PSI-based dose adaptation has been considered in head and neck settings[9,12] and a PSI-optimized prescription could improve tumor regression and consequently increase loco-regional control in retrospective data.[25] Response-adapted RT is being investigated using other predictive approaches, including artificial intelligence,[25] but the appeal of the PSI model is its simplicity and utilization of data routinely collected in clinical practice.

While these findings are encouraging, further work is necessary to validate the model in this group. As stated, this is a TRIPOD Type 1a study describing model development,[26] because no external validation data set was available, and although the model has been tested in new data, this study established a new α_p parameter value. Additionally, the change in the model algorithm to PSO warranted a re-optimization of the model parameters rather than validation of the previous version of the model. Internal validation using leave-one-out cross-validation is an option when no independent data set is available; however, it is subject to limitations such as high variance in performance estimates or unstable predictions and may underestimate model performance.[37] External evaluation of these findings in a truly independent data set is essential to ensure meaningful validation.

The data included were gathered from 2 previous publications.[5,8] Delineation of tumor volumes in the lung is subject to uncertainties caused by the impact of respiratory motion and the scan acquisition parameters,[2,38] and no ground truth GTV delineation on CBCT or MVCT is available to carry out a statistical analysis of accuracy in these modalities. The GTV delineation was either manual by an expert observer using the planning CT as a guide[29] or deformation of the GTV onto the planning CT with manual adjustments by the expert observer.[5] For each group in this study, the imaging modality and delineation methodology were consistent for all measures, so any biases in GTV calculation would be systematic and the relative volumes in cc would remain acceptable to test these hypotheses. All data sets were anonymous tumor volume measurements, time points, and RT prescriptions only and therefore lacked associated clinical or demographic information. There may be confounding clinical factors or unknown biases when evaluating the relevance of tumor regression; in particular, the use, or lack thereof, of concurrent chemotherapy may influence the α parameter. In this analysis, we only sought to evaluate the performance of the PSI model and not evaluate the clinical importance of the predicted response. Because no clinical data were available regarding individual patients, our findings indicate that the PSI model works well in both the chemoRT and RT alone settings as we did not stratify patients based on this criterion.

Conclusions

The proliferation saturation model can predict, with a high degree of accuracy, NSCLC tumor volume regression in response to RT in 4 distinct dose fractionation schedules.

References

1. Knap MM, Hoffmann L, Nordsmark M, Vestergaard A. Daily cone-beam computed tomography used to determine tumour shrinkage and localisation in lung cancer patients. *Acta Oncol* 2010;49:1077-1084.
2. Michienzi A, Kron T, Callahan J, Plumridge N, Ball D, Everitt S. Cone-beam computed tomography for lung cancer - validation with CT and monitoring tumour response during chemo-radiation therapy. *J Med Imaging Radiat Oncol* 2017;61:263-270.
3. Brink C, Bernchou U, Bertelsen A, Hansen O, Schytte T, Bentzen SM. Locoregional control of non-small cell lung cancer in relation to automated early assessment of tumor regression on cone beam computed tomography. *Int J Radiat Oncol Biol Phys* 2014;89:916-923.
4. Käsmann L, Niyazi M, Blanck O, et al. Predictive and prognostic value of tumor volume and its changes during radical radiotherapy of stage III non-small cell lung cancer: a systematic review. *Strahlenther Onkol* 2018;194:79-90.
5. van Timmeren JE, van Elmpt W, de Ruysscher D, Reymen B, Hansen O, Brink C. Tumor regression during radiotherapy for non-small cell lung cancer patients using cone-beam computed tomography images. *Strahlenther Onkol* 2020;196:159-171.
6. Ostheimer C, Mäurer M, Ebert N, et al. Prognostic impact of gross tumor volume during radical radiochemotherapy of locally advanced non-small cell lung cancer-results from the NCT03055715 multicenter cohort study of the Young DEGRO Trial Group. *Strahlenther Onkol* 2021;197:385-395 Erratum in: *Strahlenther Onkol*. 2021;197:560-561.
7. Prokopiou S, Moros EG, Poleszczuk J, et al. A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation. *Radiat Oncol* 2015;10:159.

8. Sunassee ED, Tan D, Ji N, et al. Proliferation saturation index in an adaptive Bayesian approach to predict patient-specific radiotherapy responses. *Int J Radiat Biol* 2019;95:1421-1426.
9. Zahid MU, Mohamed ASR, Latifi K, et al. Proliferation saturation index to characterize response to RT and evaluate altered fractionation in head and neck cancer. *Appl Rad Oncol* 2021;10:32-39.
10. Zahid M, Glazar D, Brady R, et al. Proliferation saturation index to prospectively predict patient-specific responses to radiation in oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2019;105:E792.
11. Kutuva AR, Caudell JJ, Yamoah K, Enderling H, Zahid MU. Mathematical modeling of radiotherapy: impact of model selection on estimating minimum radiation dose for tumor control. *Front Oncol* 2023;13 1130966.
12. Zahid MU, Mohamed ASR, Caudell JJ, et al. Dynamics-adapted radiotherapy dose (DARD) for Head and neck cancer radiotherapy dose personalization. *J Pers Med* 2021;11:1124.
13. Pawlik TM, Keyomarsi K. Role of cell cycle in mediating sensitivity to radiotherapy. *Int J Radiat Oncol Biol Phys* 2004;59:928-942.
14. Enderling H, Anderson AR, Chaplain MA, Munro AJ, Vaidya JS. Mathematical modelling of radiotherapy strategies for early breast cancer. *J Theor Biol* 2006;241:158-171.
15. Gao X, McDonald JT, Hlatky L, Enderling H. Acute and fractionated irradiation differentially modulate glioma stem cell division kinetics. *Cancer Res* 2013;73:1481-1490.
16. Grimes DR. Estimation of the oxygen enhancement ratio for charged particle radiation. *Phys Med Biol* 2020;65:15NT01.
17. eong J, Oh JH, Sonke JJ, et al. Modeling the cellular response of lung cancer to radiation therapy for a broad range of fractionation schedules. *Clin Cancer Res* 2017;23:5469-5479.
18. Leder K, Pitter K, LaPlant Q, et al. Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell* 2014;156:603-616.
19. McClatchy DM, Willers H, Hata AN, et al. Modeling resistance and recurrence patterns of combined targeted-chemoradiotherapy predicts benefit of shorter induction period. *Cancer Res* 2020;80:5121-5133.
20. Scott JG, Fletcher AG, Anderson AR, Maini PK. Spatial metrics of tumour vascular organisation predict radiation efficacy in a computational model. *PLoS Comput Biol* 2016;12 e1004712.
21. Sung W, Hong TS, Poznansky MC, Paganetti H, Grassberger C. Mathematical modeling to simulate the effect of adding radiation therapy to immunotherapy and application to hepatocellular carcinoma. *Int J Radiat Oncol Biol Phys* 2022;112:1055-1062.
22. Wilson N, Drapaca CS, Enderling H, Caudell JJ, Wilkie KP. Modelling radiation cancer treatment with a death-rate term in ordinary and fractional differential equations. *Bull Math Biol* 2023;85:47.
23. Brady R, Enderling H. Mathematical models of cancer: when to predict novel therapies, and when not to. *Bull Math Biol* 2019;81:3722-3731.
24. Harshe I, Enderling H, Brady-Nicholls R. Predicting patient-specific tumor dynamics: how many measurements are necessary? *Cancers (Basel)* 2023;15:1368.
25. Caudell J, Echevarria M, Yang G, et al. OC-0106 Prospective trial of personalized fractionation in low-risk HPV positive oropharyngeal cancerroph. *Radiothe Oncol* 2023;182:S67-S68.
26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
27. Norton L, Simon R. Growth curve of an experimental solid tumor following radiotherapy. *J Natl Cancer Inst* 1977;58:1735-1741.
28. Tariq I, Chen T, Kirkby NF, Jena R. Modelling and Bayesian adaptive prediction of individual patients' tumour volume change during radiotherapy. *Phys Med Biol* 2016;61:2145-2161.
29. Woodford C, Yartsev S, Dar AR, Bauman G, Van Dyk J. Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images. *Int J Radiat Oncol Biol Phys* 2007;69:1316-1322.
30. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* 2004;33:261-304.
31. Park LJ, Park CH, Park C, Lee T. Application of genetic algorithms to parameter estimation of bioprocesses. *Med Biol Eng Comput* 1997;35:47-49.
32. Akaike H. Information Theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, eds. *Selected Papers of Hirotugu Akaike*. Springer; 1998:199-213.
33. Sonke JJ, Belderbos J. Adaptive radiotherapy for lung cancer. *Semin Radiat Oncol* 2010;20:94-106.
34. Hattu D, Mannens J, Öllers M, van Loon J, De Ruysscher D, van Elmpt W. A traffic light protocol workflow for image-guided adaptive radiotherapy in lung cancer patients. *Radiother Oncol* 2022;175:152-158.
35. Bertholet J, Anastasi G, Noble D, et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part II: Offline and online plan adaption for interfractional changes. *Radiother Oncol* 2020;153:88-96.
36. In Lee H, Kyung Choi E, Ssan Kim S, Seob Shin Y, Won Park J, Yeol Song S. Predictive value of primary tumor volume change during concurrent chemoradiotherapy in patients with unresectable stage III non-small cell lung cancer. *Radiother Oncol* 2024;198 110383.
37. Gronau QF, Wagenmakers EJ. Limitations of Bayesian leave-one-out cross-validation for model selection. *Comput Brain Behav* 2019;2:1-11.
38. Oechsner M, Chizzali B, Devecka M, Combs SE, Wilkens JJ, Duma MN. Registration uncertainties between 3D cone beam computed tomography and different reference CT datasets in lung stereotactic body radiation therapy. *Radiat Oncol* 2016;11:142.