Normal Tissue Complication Probability Modelling

Influence of Treatment Technique, Fractionation, and Dose Calculation Algorithm

Almut Troeller

Aus der
Klinik und Poliklinik für Strahlentherapie und Radioonkologie
der
Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. C. Belka

Normal Tissue Complication Probability Modelling: Influence of Treatment Technique, Fractionation and Dose Calculation Algorithm

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von

Almut Troeller

aus

Chemnitz

Jahr

2016

Gedruckt mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuerin: **Prof. Dr. rer. nat. Katia Parodi**
Zweitgutachter: **Prof. Dr. rer. nat. Olaf Dietrich**
Dekan: **Prof. Dr. med. dent. Reinhard Hickel**
Tag der mündlichen Prüfung: **12.07.2017**

Contents

Summary

Accurately estimating a patient's risk of experiencing side effects from radiation therapy is crucial for safe and effective treatment. Normal tissue complication probability (NTCP) models are tools that predict a patient's probability of experiencing side effects, depending on the dose received by organs at risk. The following thesis investigates the influence of treatment technique, fractionation and dose calculation algorithms on NTCP models, and shows that common NTCP models cannot be applied to different treatment techniques, fractionation schemes and dose calculation algorithms because they yield different model parameters depending on these factors.

In order to investigate how the treatment technique influences NTCP modelling results, NTCP model parameters for chronic gastrointestinal toxicities following prostate cancer treatment for patients treated with three dimensional conformal radiation therapy (3D-CRT) and intensity modulated radiation therapy (IMRT) are derived and compared. The 3D-CRT and IMRT patients are matched using propensity score matched pair analysis to reduce potential bias from confounding variables. It is shown that the models derived from the 3D-CRT and the IMRT populations were not cross-applicable. These results are unexpected, since NTCP model parameters for a clinical endpoint should ideally be independent of the treatment technique. Investigations into the potential sources of discrepancies between optimal model parameters for the two groups ruled out differences in follow-up time, imaging modalities used, known patient characteristics, and errors of planned versus delivered dose. Other potential sources of the discrepancies are discussed and may be related to the toxicity scoring, confounding patient characteristics not accounted for by matched pair analysis, and insufficiency of the NTCP models themselves.

Subsequently, the NTCP of developing radiation pneumonitis in lung cancer patients treated with hypofractionated radiation therapy (HFRT) is modelled. In order to account for different fractionation schemes, the linear quadratic (LQ) model and the low dose hyperradiosensitivity model (LDHRS) are used to adjust the physical dose accordingly. NTCP is modelled and compared using both biological models. The NTCP model parameters for HFRT using the LQ model differ from parameters that were previously derived for standard fractionated radiation therapy (SFRT) in the literature. Thus, it is demonstrated that NTCP models are not cross applicable for different fractionation schemes, even when accounting for differences in fractionation using the LQ model. Furthermore, novel NTCP parameters using the LDHRS model are provided and the feasibility of using the LDHRS model for modelling NTCP of ra-

diation pneumonitis for lung HFRT is demonstrated. It is discussed how the LDHRS model could potentially be used to find a common NTCP model for radiation pneumonitis in HFRT and SFRT lung cancer patients, although this will have to be verified in future studies.

Moreover, since it was observed in previous publications that HFRT in lung leads to increased incidences of toxicities related to the chest wall, NTCP model parameters for chest wall pain, rib fractures, and myositis are determined and discussed in this work.

Lastly, it is demonstrated that the choice of dose calculation algorithm significantly impacts dose volume parameters for regions of interest. Pencil beam (PB) and collapsed cone (CC) based dose calculation algorithms are compared for lung stereotactic body radiation therapy. It is shown that using CC to recalculate dose in lung plans created with PB leads to drastic differences in the planned dose in regions of interest. Since accurate dose distributions are vital in NTCP modelling, these discrepancies could lead to differences in NTCP model parameters.

Zusammenfassung

Die akkurate Einschätzung von Nebenwirkungswahrscheinlichkeiten in der Strahlentherapie ist essentiell um eine sichere und effektive Behandlung sicherzustellen. Normal tissue complication probability (NTCP) Modelle sind eine Methode für die Vorhersage der Wahrscheinlichkeit, dass ein Patient Nebenwirkungen erfährt in Abhängigkeit von der Dosis in Risikoorganen. Die vorliegende Arbeit befasst sich mit dem Einfluss der Behandlungstechnik, der Fraktionierung und des Dosisberechnungsalgorithmus auf NTCP-Modelle. Es wird gezeigt, dass verbreitete NTCP-Modelle nicht auf andere Behandlungstechniken, Fraktionierungsschemata und Pläne, die mit anderen Dosisberechnungsalgorithmen berechnet wurden, angewendet werden können, da die Modellparameter von diesen Faktoren abhängig sind.

Um zu untersuchen, inwiefern die Behandlungstechnik Ergebnisse der NTCP Modellierung beeinflusst, werden NTCP Modellparameter für chronische gastrointestinale Toxizitäten, welche nach der Behandlung von Prostatakarzinomen auftreten können, für dreidimensionale konformale Strahlentherapie (3D-CRT) und intensitätsmodulierte Strahlentherapie (IMRT) bestimmt und verglichen. Die 3D-CRT und IMRT Patiententendatensätze werden einander mittels eines Matchingalgorithmuses statistisch zugeordnet, um den Einfluss von potenziellen Störfaktoren wie beispielsweise unterschiedliche Populationscharakteristiken zu reduzieren. Es wird gezeigt, dass Modellparameter, die mittels des 3D-CRT Datensatzes bestimmt wurden, nicht auf den IMRT Datensatz angewendet werden können und umgekehrt. Diese Ergebnisse sind unerwartet, da die NTCP-Modellparameter für einen spezifischen klinischen Endpunkt idealerweise nicht von der Behandlungstechnik abhängig sein sollte. Durch weitere Nachforschung konnten unterschiedliche Nachsorgezeitspannen, Bildgebungsmethoden, bekannte Populationcharakteristiken und Diskrepanzen zwischen geplanter und abgestrahlter Dosis als potenzielle Ursachen für die unterschiedlichen Modellparameter ausgeschloßen werden. Weitere mögliche Ursachen für die Diskrepanzen zwischen den beiden Behandlungstechniken, einschließlich der Subjektivität der Bewertung von Toxizitäten, Patientencharakteristiken, die beim Matching nicht berücksichtigt werden konnten, und Modellbeschränkungen der NTCP Modelle, werden diskutiert.

Nachfolgend wird NTCP für Pneumonitis in Lungenkrebspatienten, welche mit hypofraktionierter Strahlentherapie (HFRT) behandelt wurden, modelliert. Dabei werden zwei biologische Modelle, das linearquadratische Modell und das low dose hyperradiosensitivity model (LDHRS), benutzt um die biologischen Auswirkungen verschiedener Fraktionierungsschemata zu berücksichtigen und die physikalische Dosis

entsprechend anzupassen. NTCP wird für beide Modelle modelliert und verglichen. Die Ergebnisse zeigen, dass NTCP-Modellparameter für HFRT, die mittels des LQ-Modelles bestimmt wurden, nicht mit den Werten, die in der Literatur für standardfraktionierte Strahlentherapie (SFRT) angegeben sind, übereinstimmen. Damit wird auch gezeigt, dass NTCP-Modelle nicht ohne Weiteres auf andere Fraktionierungsschemata angewendet werden können, wenn das LQ-Modell benutzt wird. Des Weiteren werden neue NTCP-Modellparameter für Fraktionskorrektur mittels des LDHRS-Modells bestimmt und die grundsätzliche Machbarkeit der NTCP-Modellierung mit dem LDHRS-Modell für hypofraktionierte Bestrahlung der Lungen wird demonstriert. Anschließend wird diskutiert, dass das LDHRS-Modell potenziell verwendet werden kann um ein gemeinsames Pneumonitis NTCP-Modell für HFRT uns SFRT in der Lunge zu finden. Dies muss jedoch in zukünftigen Studien verifiziert werden.

Des Weiteren wurde in früheren Studien festgestellt, dass HFRT in der Lunge zu einem Anstieg der Inzidenz von Toxitäten der Thoraxwand führt. Daher werden in der vorliegenden Arbeit NTCP-Modellparameter für Schmerzen in der Thoraxwand, Rippenfrakturen und Myositis bestimmt und diskutiert.

Zuletzt wird gezeigt, dass die Wahl des Dosisberechnungsalgorithmuses Dosisvolumenparameter für Tumore und Risikoorgane stark beeinflussen kann. Dosisberechnungen für Lungenstereotaxien basierend auf Pencil Beam (PB) und Collapsed Cone Algorithmen werden verglichen. Es wird gezeigt, dass die Nachrechnung von Plänen mit CC, die ursprünglich mit PB berechnet wurden, zu drastischen Unterschieden in der Plandosis führen können. Da die Qualität der NTCP-Modellierung auf akkuraten Dosisberechnungen beruht, können diese Dosisunterschiede zu unterschiedlichen NTCP-Modellparametern führen.

Acknowledgements

First and foremost, I would like to thank my thesis advisor Prof. Katia Paradi for taking the time to advise me, and for the opportunity to pursue my research endeavours.

This thesis could not have been done without the help of Dr. Matthias Söhn, who spent countless hours advising me, teaching me, and proofreading my work. He has been a role model for me as a scientist who truly cares about his work and colleagues. I am incredibly grateful that I had the privilege to work with him. Last, but not least, Dr. Söhn introduced me to living at 752 Pierce Street, where I had the time of my life and met many wonderful people.

Furthermore, I would like to thank the chief of physics at Beaumont Health, Dr. Di Yan for hosting me as an exchange research scholar, providing guidance, advice, financial support, opportunities to present my work at conferences and in scientific journals, and for the many cups of tea and coffee at Starbucks.

Also I would like to acknowledge the chief of physics at LMU Munich Dr. Michael Reiner, and the department head of the department of radiation oncology Prof. Claus Belka for their support of my work, financial support, and for graciously allowing me to travel to the United States to complete the research within this thesis.

I would like to especially thank Steffi Kantz for being an invaluable friend, supporting me throughout my time as a PhD student, proof reading papers and this thesis, her moral support, and for hosting me countless times. Furthermore, I am grateful to all physicists, physicians, dosimetrists, therapists and all other coworkers at LMU Munich and Beaumont Health, who have contributed to my thesis in one way or another, especially David Gersten, Derek Schulze, Raminder Sandhu, David Solis, Dan Ionascu, Ovidiu Marina, Leonid Zamdborg, Monica Abbate, and Andrea Lakasz.

Most importantly, I am immensely grateful to my husband Sean for his love, support and belief in me. Also, I cannot thank my German, American and Canadian families enough for all they have done for me: Silvia, Holger, Ronald, Helmar, Ursula, Hermann, Karsten, Barbara, Sharon, Patrick, Monica, Jon, Shayla, and Cassie.

Lastly I would like to thank Sally Kaplan, Michelle Zhang, Danielle Lack, Joy Jiewei Zamdborg, Sven Prüfer, and Stefan Prüfer for their friendship and the good times.

Acronyms

2D two dimensional.

2D-RT two dimensional radiation therapy.

3D three dimensional.

3D-CRT three dimensional conformal radiation therapy.

4D four dimensional.

4D-CT four dimensional computed tomography.

AC alternating current.

ADT androgen deprivation therapy.

AIC Akaike information criterion.

AICc corrected Akaike information criterion.

ART adaptive radiation therapy.

BED biologically effective dose.

BEV beam's eye view.

BTE Boltzmann transport equation.

CBCT cone beam computed tomography.

CC collapsed cone.

CDLR cutoff dose logistic regression.

cDVH cumulative dose volume histogram.

CI confidence interval.

CT computed tomography.

CTCAE Common Terminology Criteria for Adverse Events.

CTV clinical target volume.

DC direct current.

dDVH differential dose volume histogram.

DM diabetes mellitus.

DNA deoxyribonucleic acid.

DRR digitally reconstructed radiograph.

DVH dose volume histogram.

EPID electronic portal imaging device.

EUD equivalent uniform dose.

FU follow-up.

gEUD generalized equivalent uniform dose.

GI gastro intestinal.

GTV gross tumor volume.

HFRT hypofractionated radiation therapy.

ICRU International Commission on Radiation Units and Measurement.

IGRT image guided radiation therapy.

IMRT intensity modulated radiation therapy.

ITV internal target volume.

KERMA kinetic energy released per unit mass.

LDHRS low dose hyper-radiosensitivity.

LINAC linear accelerator.

LKB Lyman Kutcher Burman.

LL log likelihood.

LQ linear quadratic.

MC Monte Carlo.

MCWD mean chest wall dose.

MLC multi leaf collimator.

MLD mean lung dose.

MLE maximum likelihood estimation.

MRI magnetic resonance imaging.

MU monitor unit.

NCCN National Comprehensive Cancer Network.

NCI-CTC National Cancer Institute Common Toxicity Criteria.

NTCP normal tissue complication probability.

NTD normalized total dose.

OAR organ at risk.

OR odds ratio.

OTP Oncentra MasterPlan®.

PB pencil beam.

PET positron emission tomography.

PI portal image.

PTV planning target volume.

QUANTEC Quantitative Analyses of Normal Tissue Effects in the Clinic.

ROI region of interest.

RP radiation pneumonitis.

RT radiation therapy.

RTOG Radiation Therapy Oncology Group.

SBRT stereotactic body radiation therapy.

SFRT standard fractionated radiation therapy.

SWOG Southwest Oncology Group.

TERMA total energy released per unit mass.

TPS treatment planning system.

US ultra sound.

VMAT volumetric modulated arc therapy.

List of Figures

List of Tables

Chapter 1

Introduction

The successful treatment of cancer is one of the modern world's major medical challenges. According to the 2014 World Cancer Report by the World Health Organization, there were an estimated 14 million new cases of cancer worldwide in 2012, with lung, liver, stomach, colorectal and breast cancers as the leading cause of cancer deaths (Stewart *et al.*, 2014). Radiation therapy (RT) has been used to treat cancer since the early 20th century when the lethal effects of radiation on tissue were first discovered. Since then, radiation therapy has significantly improved and shown impressive outcomes, especially in combination with other treatment modalities such as surgery and chemotherapy. Today, local control and overall survival rate are significantly higher for many types of cancer than they were in the 1930s, when radiation therapy was limited to curing a small range of cancers such as skin or uterus cancer (Thariat *et al.*, 2012). In contrast, today a wider variety of cancers can be treated with RT such as breast cancer, lung cancer, prostate cancer, lymphomas, brain, head and neck and many other types. The overall survival rate 5 years after diagnosis for some cancers commonly treated with RT, for example prostate cancer, can be as high as 98.9% (Bethesda, 2014). While the main goal of RT is to irradiate the tumor and gain tumor control, this is achieved at the expense of normal tissue damage. Irradiation of normal tissues can lead to severe side effects that can impact the patient's quality of life or even have lethal consequences. Thus, it is imperative to spare normal tissues during treatment. Since uncertainties such as finite dose gradients, anatomical changes, movement, positioning errors, and microscopic disease prevent sufficient irradiation of tumor cells without irradiating normal tissues, it is crucial to RT that complication probabilities of organs at risk (OARs) are accurately assessed and guidelines for which doses OARs can tolerate are established. Until recent years, physicians heavily depended on published guidelines by Emami *et al.* (Emami *et al.*, 1991) in order to estimate normal tissue complication probability (NTCP) for OARs. However, Emami's guidelines are based on older treatment techniques using either homogeneous whole organ irradiation or homogeneous irradiation of partial organs. Furthermore, these values are based on physicians' clinical experience, rather than models and predictions derived from carefully collected and analyzed scientific data. Recent treatment plans are more complex and involve a high degree of heterogeneity. Thus more elaborate and objective mea-

sures for predicting complications are needed. NTCP modelling derives a model from treatment data and outcome of a large population that can then be applied to a patient to be treated. An NTCP model assigns a complication probability for a certain endpoint to an inhomogeneous dose distribution of an OAR. One of the first NTCP models was proposed by Lyman (Lyman, 1985) and extended by Kutcher and Burman (Kutcher and Burman, 1989) in the 1980s and 1990s. At that time, the models were limited due to the limited availability of data used, and the lack of computational power. Since then various NTCP models have been established and are successfully being used in the clinic (Marks *et al.*, 2010). However, the field of NTCP modelling also faces several challenges. Among these challenges are issues of model extrapolation to populations with different population characteristics, fractionation schemes, treatment techniques etc.

The aim of this thesis is to investigate some of the challenges and uncertainties that NTCP modelling faces, namely to investigate how treatment modalities, fractionation and the choice of dose calculation algorithms influence dose to regions of interest (ROIs) and NTCP models:

1. **Applicability of NTCP models to other treatment techniques**: Radiation therapy is a rapidly evolving field in which new treatment techniques are constantly incorporated into the clinical routine. In recent years intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) have been implemented in many clinical facilities. Since successful NTCP modelling requires accumulating data from large patient cohorts, many NTCP model parameters used in the clinic are derived from older treatment techniques, such as three dimensional conformal radiation therapy (3D-CRT) for which large patient populations are available for modelling. However, using NTCP parameters derived from a population treated with one treatment technique to patients treated with a different technique can be problematic. For example, the patient populations may differ in patient characteristics that influence the incidence of side effects, such as smoking or patient age. Differences in model parameters could also arise for instance because of advanced imaging technology being developed contemporaneously with newer treatment techniques. In chapter 5 of this thesis, the effect of different treatment modalities on the outcomes of NTCP modelling is investigated. Specifically, NTCP is modelled for two matched groups of patients treated with 3D-CRT and IMRT, respectively. The model parameters are compared and it is determined whether one set of model parameters is able to predict NTCP for the other and vice versa. Furthermore, the influence of confounding patient characteristics, actual delivered dose, and toxicity scoring bias on NTCP modelling is explored and discussed.

2. **Implications of fractionation schemes on NTCP models:** Fractionation of the dose delivered during radiation therapy was introduced in order to spare OARs. It is based on research indicating that with the ability to divide in an uncontrolled manner, tumor cells have lost some of their repair capabilities, while normal tissues retain the ability to repair. Delivering the full treatment dose in fractions

allows healthy tissues to recover while still destroying tumor cells. However, more recently, larger fractions are being delivered without markedly increasing normal tissue damage. Chapter 6 explores the effect of fractionation schemes on NTCP modelling of radiation pneumonitis. It investigates different biological models accounting for fractionation effects and their potential to be applied to alternative fractionation schemes. Furthermore, different fractionation schemes may lead to the incidence of different side effects which require novel NTCP models parameters. In section 6.3, models for side effects emerging from the chest wall are presented.

3. **Importance of dose calculation algorithms on organs at risk and tumor dose:** The basis of all NTCP models is the assumption that a patient's probability of experiencing side effects is a function of the dose delivered to tissue. Thus, in order to develop accurate NTCP models, it is of importance that the dose distribution used for modelling is correct. In practice, the dose delivered to the patient cannot be measured, but must be calculated. However, the dose calculation itself yields errors. In the past, most commercial treatment planning systems (TPSs) used pencil beam (PB) algorithms to approximately calculate dose. Since then more accurate methods of calculating dose, such as collapsed cone (CC) based algorithms and highly accurate Monte Carlo (MC) based methods have been implemented into clinically used TPSs. However, many dose constraints and dose response models for ROIs are derived from treatment plans calculated using PB algorithms. Thus, they do not necessarily reflect the best constraints and model parameters for plans calculated with CC or MC. It is therefore important to determine how the choice of the dose calculation algorithm affects the planned dose. In chapter 7 the effect of using PB and CC dose calculation algorithms on both normal tissues and tumor tissues is discussed.

Chapter 2 contains basic background information on the major physical principles used in RT including a brief introduction of medical linear accelerators, the interaction of X-rays with matter, and dose. In chapter 3, the treatment planning process including dose calculation algorithms is described and chapter 4 introduces methods used in NTCP modelling. The influences of the treatment technique, fractionation scheme, and the choice of dose calculation algorithms on NTCP models are investigated in chapter 5, chapter 6, and chapter 7, respectively. Finally, the conclusions of this thesis are discussed in chapter 8.

Part I

Background

Chapter 2

Physics of Radiation Therapy

In radiotherapy, malignant tumors are treated with ionizing radiation in order to destroy or inactivate tumor cells. This is achieved by the ionization of tumor deoxyribonucleic acid (DNA) by either photons, or charged particles such as electrons, protons or heavier ions. Ionization of the DNA strands, in turn, may lead to double strand breaks in the cell DNA, which are difficult to repair and can lead to cell death. While photons act mostly indirectly via ionization of water in the patient's body forming free radicals, charged particles may damage DNA directly.

This chapter focuses on the utilization of high energy photon beams that are produced by medical linear accelerators (LINACs) and then interact with tissue, thereby damaging or destroying cells. While other photon based techniques in radiation therapy (such as gamma rays from radioactive isotopes like ^{60}Co), and also charged particle radiation therapy are all important methods of treatment, all patients presented in this thesis were treated using photon beams from medical LINACs. Therefore, the thesis focuses on the production and interaction of photon beams from medical LINACs.

A picture of the front of a medical LINAC is shown in Figure 2.1. The photon beam produced by the LINAC travels from the LINAC head through air and eventually strikes the patient. In the patient the photons interact with the tissue matter and the beam is attenuated via scattering and absorption thereby producing high energy electrons. Thus, the photons transfer their energy to charged particles (electrons) which in turn transfer their kinetic energy to the tissue via excitation and ionization. As mentioned above, this energy can be sufficient to break or inactivate cells and is the basic means of destroying tumor cells in radiotherapy. The patient is positioned on a couch which can usually perform translational movements in three dimensions as well as rotate around at least one axis. In combination with the rotation of the gantry (from which the beam emerges), this enables the photon beam to irradiate the patient from virtually all angles.

In the following sections, X-ray production using a LINAC is described (2.1), followed by a brief description of the resulting photon beam (2.2). Interactions of X-ray photons with matter are then discussed (2.3), as well as ionization, excitation, energy deposition and absorbed dose (2.4). More detailed information can be found in the standard literature e.g. (Metcalfe *et al.*, 1997; Khan, 2010; Krieger, 2012).

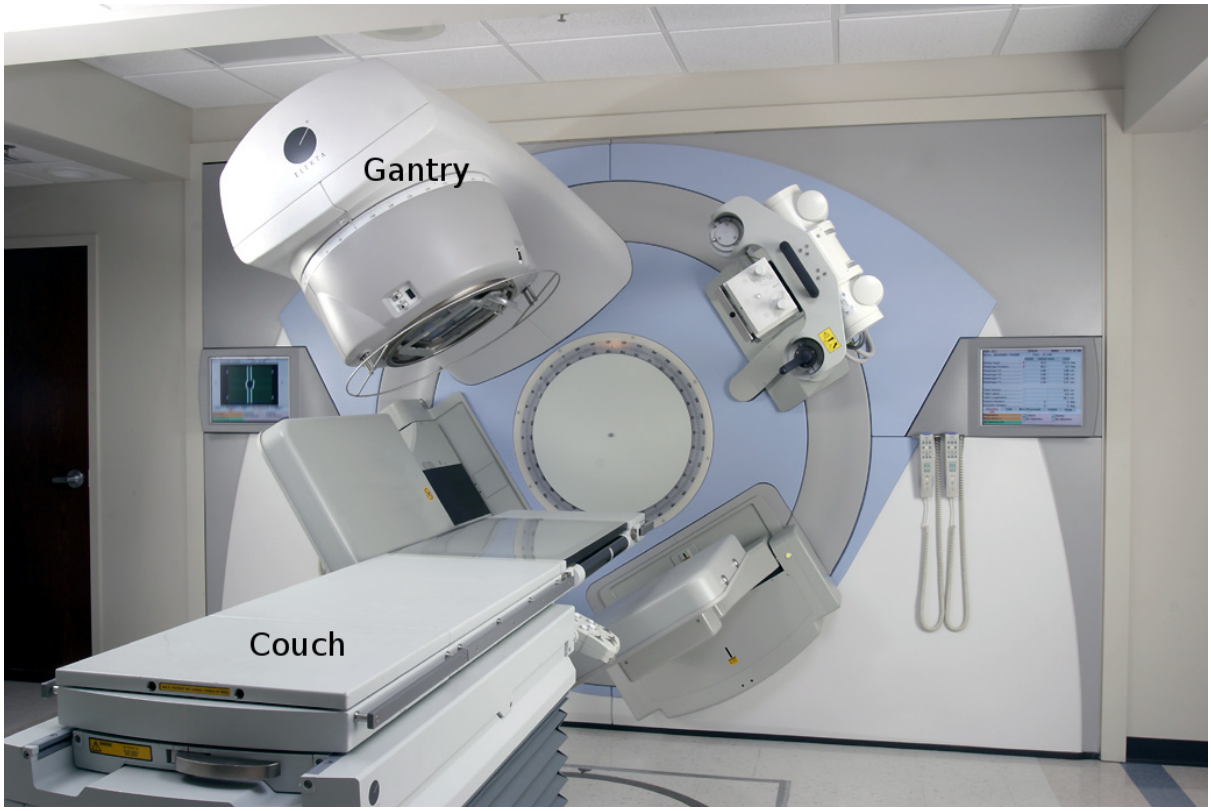


Figure 2.1: The front of a medical linear accelerator (LINAC) including the couch on which the patient is positioned. Image courtesy of Elekta (www.elekta.com, Stockholm, Sweden).

2.1 Photon Production with Medical Linear Accelerators

A medical LINAC produces X-ray photons by accelerating electrons which then strike a target. The electrons are accelerated in a high frequency electromagnetic field. A typical medical LINAC uses a magnetron or klystron to generate microwaves in a waveguide accelerator. An electron gun then inserts low energy electrons into the waveguide which are accelerated to very high velocities near the speed of light. When the high energy electrons leave the waveguide, they are usually bent into the final direction of the beam by a bending magnet and eventually hit a small target thereby producing bremsstrahlung X-rays. The X-rays then travel through the treatment head and eventually strike the patient. In the following the major components of LINACs are briefly discussed. A schematic diagram of a LINAC is shown in Figure 2.2a.

2.1.1 Electron Source

The **electron gun** is the LINAC's source of low speed electrons. It consist of a tungsten cathode and an anode in a vacuum tube. The cathode is heated to high temperatures ($T > 1000°$C) in order to produce free electrons via thermionic emission. These free electrons are then accelerated by the electric field between the cathode and the anode. Typical electron guns operate at 30-100 kV. At the anode, the electrons are inserted into

the acceleration waveguide of the LINAC.

2.1.2 Microwave Production

The microwaves in linear accelerators are produced by either a klystron or a magnetron. **Klystrons** do not generate microwaves but rather amplify an existing microwave. They consist of an electron gun that generates an electron beam which then passes through a buncher and a catcher cavity that both resonate at the frequency of the low power microwave to be amplified. The low power microwave is fed into the buncher cavity, where it generates an oscillating electric field in the path of the electron beam. Electrons that pass through the cavity become velocity modulated. That is, electrons that pass the cavity while the electric field points in the same direction as the electron flow will be decelerated, whereas electrons passing through the cavity when the electric field points in the opposite direction will be accelerated. The electrons then enter a drifting space in between the buncher cavity and the catcher cavity, where they form bunches due to the acceleration and deceleration they experienced in the buncher cavity. The distance between the cavities is designed such that the electrons pass through the second cavity when they are maximally bunched. At the catcher cavity, they induce a retarding electric field that decelerates the electron bunches, thereby producing high power microwave oscillations in the cavity. These microwaves are of the same frequency as the initial low power microwaves. In contrast to klystrons, **magnetrons** produce microwaves rather than amplifying existing microwaves. A magnetron consist of a cylindrical copper anode with several holes (cavities) and a heated cathode at the center of the cylinder. The cathode emits electrons via thermionic emission and a direct current (DC) electric field between the cathode and the anode combined with an external static magnetic field perpendicular to the magnetron cross section causes them to travel towards the anode in a circular path. The magnetron is tuned such that the electrons do not initially reach the anode but pass it closely in a circular motion and then move back towards the cathode. The holes in the anode are connected to the evacuated space in between the anode and the cathode via small slits that act as capacitors, while the cavities themselves form one loop induction coils. Thus the sections of the anode can act as parallel resonant circuits which resonate at microwave frequencies. When the spiralling electrons pass the cavity slits in the anode, they induce an alternating current (AC) electric field across the anode slits by transferring some of the energy of the DC field. This AC field in turn velocity modulates the electrons and bunches them based on the same principles as the klystron leading to a stronger AC field which generates microwaves in the anode cavities. In medical LINACs, magnetrons create pulsed microwaves of about 3000 MHz.

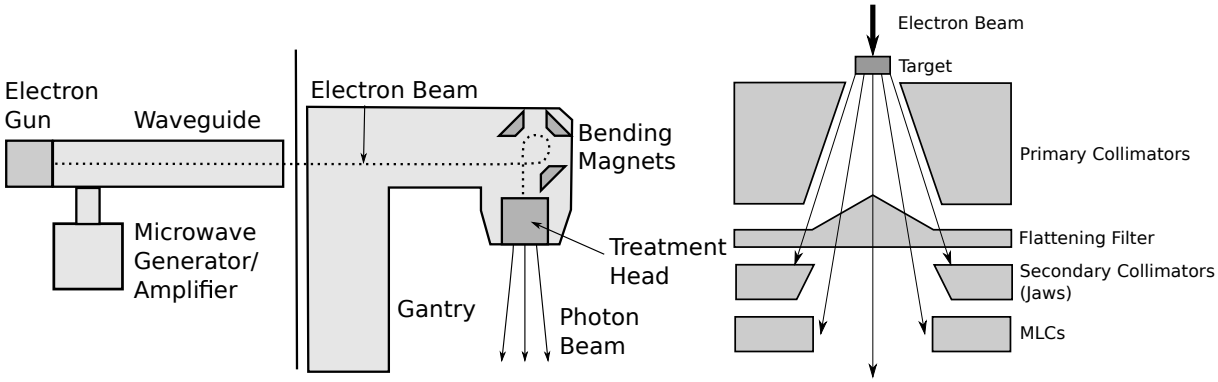
2.1.3 Acceleration and Bending Magnet

The microwaves produced by the klystron or magnetron are eventually fed into the accelerating waveguide of the medical LINAC where they are used to accelerate the

electrons inserted by the electron guns. The waveguide consist of several copper cavities in which the electrons are bunched and reach near speed of light velocity. Some current LINACs have travelling waveguides, where the microwaves enter on one side and leave on the other. A more efficient approach are standing waveguides where the microwaves are reflected at the end and create a standing wave. More details on the principles of waveguides in LINACs can be found for example in (Khan, 2010; Krieger, 2012). After the electron beam leaves the waveguide it needs to be bent towards the patient (see Figure 2.2a). This is accomplished by using a bending magnet. However, electrons of different energies follow different paths through the field of a bending magnet and thus the electron beam will diverge after deflection by a single $90°$ bending magnet. Therefore, either three $90°$ bending magnets bending the beam by $270°$, or two $44°$ magnets in combination with one $112°$ magnet that bend the beam along a snake like curve are typically used in LINACs to allow refocusing the electron beam.

2.1.4 Treatment Head

The treatment head consists of several components that shape and monitor the beam, such as the X-ray target, a flattening filter, ion chambers and collimators (Figure 2.2b). The **target** is made of tungsten and emits bremsstrahlung and characteristic X-rays when the high energy electrons collide with it. Bremsstrahlung is generated when the incoming electrons are deflected by fields of the target nuclei and experience deceleration. The consequent loss of kinetic energy is compensated by the production of photons. The angular distribution of photons depends on the energy of the electrons striking the target. At higher photon energies of at least a few hundred kV, the majority of photons are emitted in the forward direction while at lower energies (e.g. 100 kV) photons are emitted more laterally. Bremsstrahlung has a continuous spectrum with energies up to the initial kinetic electron energy since the electrons may lose all or part of their energy in the process. Just below the target is a fixed **primary collimator** that prevents leakage followed by a **flattening filter**. The flattening filter is used to obtain a uniform beam profile. However, the filter, usually composed of steel, copper or tungsten, leads to significant attenuation. Since in recent radiotherapy the beam fluence is often modulated, flattening filters may be omitted in favor of a higher dose rate. The flattening filter also leads to changes in the (angular) energy spectrum of the beam, because the mass attenuation coefficient depends on the energy of the photons. Lower energy photons are attenuated more easily than higher energy photons, which leads to an increase in the average energy of the photon beam. This effect is called 'beam hardening'. Beam hardening also occurs when the photon beam interacts with the patient's tissues. Below the flattening filter are **ion chambers** that monitor dose rate, beam symmetry and beam homogeneity. Below are four or two movable lead or tungsten jaws that can create a rectangular opening. In addition to the jaws, modern LINACs have **multi leaf collimators (MLCs)** in order to create irregular field shapes. The treatment had can be rotated along the central beam axis. Finally, the treatment head is surrounded by thick shielding that prevents X-rays from leaking.



(a) LINAC body.

(b) Treatment head of a medical LINAC.

Figure 2.2: Simplified schematic view of a medical linear accelerator (LINAC).

2.2 Photon Beam

The photon beam that leaves the LINAC is commonly described in terms of fluence and energy fluence. Fluence is defined as the number of photons N passing through a cross sectional unit area A

$$\Phi = \frac{\mathrm{d}N}{\mathrm{d}A}. \tag{2.1}$$

Since the X-ray beam is polyenergetic, the total fluence is an integral

$$\Phi = \int_0^{E_{\max}} \frac{\partial \Phi(E)}{\partial E} \mathrm{d}E \tag{2.2}$$

Energy fluence is the energy passing through $\mathrm{d}A$

$$\Psi = \frac{\mathrm{d}E}{\mathrm{d}A}. \tag{2.3}$$

Once the photon beam strikes a medium it will be attenuated. The interaction cross section σ describes the probability that a photon striking within a unit cross sectional area will interact; and the attenuation coefficient μ describes the attenuation of the photon beam. Since μ is dependent on the density of the material, the mass attenuation coefficient $\frac{\mu}{\rho}$ is often used, which depends on the atomic properties of the medium and the photon energy, and is independent of density.

2.3 Interaction of X-Ray Photons with Matter

There are three major types of interaction of X-rays with matter that attenuate the photon beam generated by the LINAC. The photons can be absorbed via (1) the photoelectric effect, and (2) pair production, or they can be scattered via (3) Compton scattering. Coherent scattering may also occur, but is important only at low energies and high Z materials and is not usually relevant for radiotherapy. Each interaction attenuates the photon beam and is associated with an interaction cross section and an attenuation coefficient.

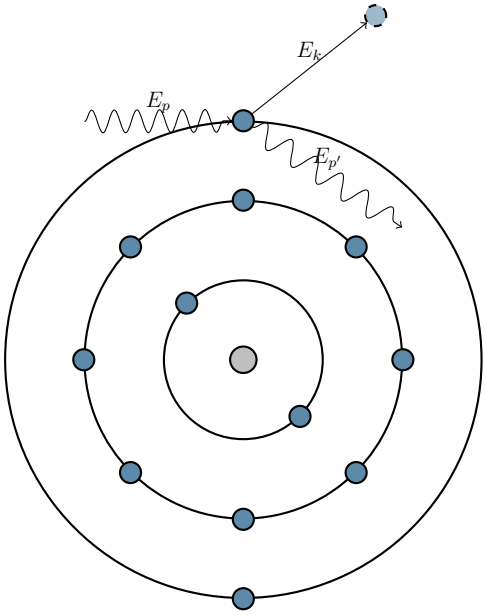


Figure 2.3: Schematic of Compton scattering.

2.3.1 Compton Scattering

Compton scattering occurs when the incoming photon interacts with an outer shell electron and transfers part of its energy to the electron (Figure 2.3). In order for Compton scattering to happen, the energy of the incident photon must be much larger than the binding energy of the electron, such that the electron can be assumed to be a free electron in first approximation. After the interaction with the photon, the electron acquires kinetic energy E_{kin}. The photon's energy after the collision can be calculated as

$$E_{\text{p}'} = E_{\text{p}} - E_{\text{kin}} - E_{\text{b}}, \tag{2.4}$$

where E_{p} is the photon's original energy, $E_{\text{p}'}$ is the photon's energy after the collision, and E_{b} is the electron's binding energy. As mentioned above, E_{b} is small compared to the other energies and can usually be neglected. The mass attenuation coefficient associated with Compton scattering depends on the photon energy and can be approximated as

$$\left(\frac{\mu}{\rho}\right)_{\text{compton}} \propto \frac{Z}{A} E_{\text{p}}^{-n}, \tag{2.5}$$

for photon energies between 0.2 MeV and 10 MeV, where Z is the atomic number, A is the atomic mass number and n varies between 0.5 and 1 (Krieger, 2012; Metcalfe *et al.*, 1997).

2.3.2 Coherent Scattering

Coherent scattering occurs when a photon is merely deflected in the vicinity of an atomic electron that can be considered bound. This form of scattering can be observed

when the electron's energy of ionization E_{ion} is much larger than the energy of the photon. The incident photon causes the electron to oscillate and the oscillating electron emits a photon at the same wavelength as the original photon, but it may be deflected by a small angle. As mentioned above, coherent scattering is negligible at energies above 100 keV and thus not of particular relevance to radiation therapy (Metcalfe *et al.*, 1997).

2.3.3 Photoelectric Effect

A photoelectric interaction occurs if a photon strikes a bound atomic electron (i.e. $E_{\text{ion}} \gg E_{\text{p}}$) and transfers all of its energy to the electron. If the electron is an inner shell electron, it will leave the shell and an outer shell electron may take its place releasing a new photon with a characteristic energy equal to the energy difference of the shells. The ejected electron obtains the kinetic energy E_{kin} which is equal to the original photon energy minus the electron's binding energy

$$E_{\text{kin}} = E_{\text{p}} - E_{\text{b}} \tag{2.6}$$

For energies around and below 0.1 MeV, the mass attenuation coefficient can be calculated as (Khan, 2010)

$$\left(\frac{\mu}{\rho}\right)_{\text{photo}} \propto Z^3 E_{\text{p}}^{-3}, \tag{2.7}$$

for high atomic numbers and

$$\left(\frac{\mu}{\rho}\right)_{\text{photo}} \propto Z^{3.5} E_{\text{p}}^{-3}, \tag{2.8}$$

for low Z materials (Khan, 2010; Krieger, 2012).

2.3.4 Pair Production

If a photon interacts with the electromagnetic field of a nucleus it can be absorbed and produce an electron positron pair. In order for pair production to occur, the energy of the photon E_{p} must be larger than twice the rest mass energy of electrons and positrons, i.e. $2m_0c^2 = 1.022$ MeV. After the interaction, the electron and positron share the energy difference between the photon energy E_{p} and the $2m_0c^2$ necessary for their generation, though not necessarily equally. On average, each of their kinetic energy is equal to half the difference between the photon energy E_{p} and twice the electron rest mass energy

$$\bar{E}_{\text{kin,e}^{\pm}} = \frac{E_{\text{p}} - 2m_o c^2}{2}, \tag{2.9}$$

where $\bar{E}_{\text{kin,e}^{\pm}}$ denotes the average acquired kinetic energy of the electron or positron. The positron will interact with an electron in the body and produce two 0.511 MeV

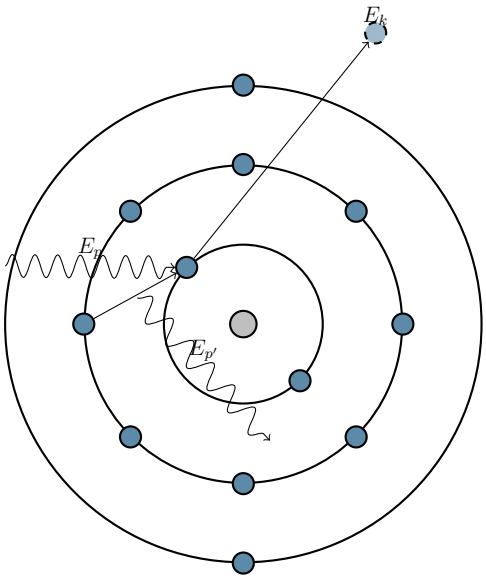


Figure 2.4: Schematic of photoelectric absorption.

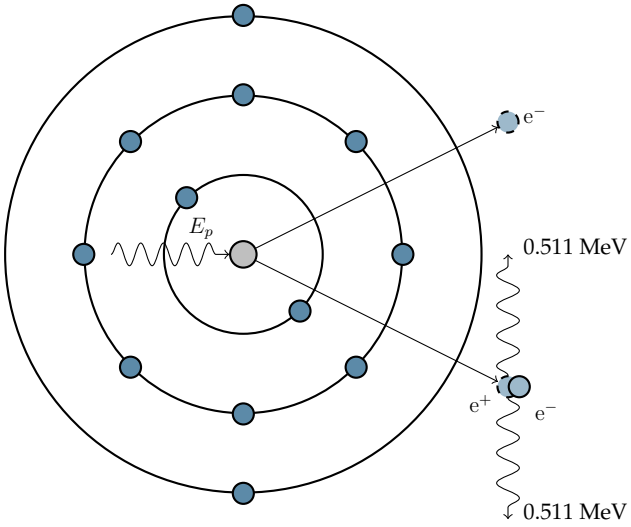


Figure 2.5: Schematic of pair production.

photons via electron-positron annihilation. The mass attenuation coefficient can be calculated as

$$\left(\frac{\mu}{\rho}\right)_{\text{pair}} \propto \frac{Z^2}{A} \log E_{\text{p}}, \tag{2.10}$$

until it saturates at high photon energies E_{p}.

2.4 Energy Deposition and Absorbed Dose

As described above the photons interact with tissue via Compton scattering, photoelectric interactions and pair production processes and transfer all or part of their energy to electrons that gain kinetic energy. The kinetic energy that is transferred by the

photons to charged particles is termed kinetic energy released per unit mass (KERMA)

$$K = \frac{\mathrm{d}E_{\mathrm{tr}}}{\mathrm{d}m} = \Psi \left(\frac{\mu_{\mathrm{tr}}}{\rho} \right), \tag{2.11}$$

where $\mathrm{d}E_{\mathrm{tr}}$ is the average energy transferred to charged particles as kinetic energy per interaction, and μ_{tr} is the energy transfer coefficient, which is defined as the fraction of photon energy transferred to kinetic energy of charged particles per unit thickness (Khan, 2010). The electrons then further transfer their kinetic energy via ionization and excitation. However, a small portion of the electrons' kinetic energy may radiate away as bremsstrahlung and does not contribute to local energy deposition. Thus the KERMA comprises two terms - collision KERMA (K_{col}) and radiative KERMA (K_{rad}).

$$K_{\mathrm{col}} = \Psi \frac{\mu_{\mathrm{ab}}}{\rho} = \frac{\mathrm{d}E_{\mathrm{tr}}}{\mathrm{d}m}(1-g), \tag{2.12}$$

$$K_{\mathrm{rad}} = \frac{\mathrm{d}E_{\mathrm{tr}}}{\mathrm{d}m} g, \tag{2.13}$$

where g is the fraction of energy that is radiated away as bremsstrahlung, and μ_{ab} is the absorption coefficient, which is defined as

$$\mu_{\mathrm{ab}} = \mu_{\mathrm{tr}}(1-g). \tag{2.14}$$

Absorbed dose is defined as the (mean) energy deposited per unit mass

$$D = \frac{\mathrm{d}\bar{E}}{\mathrm{d}m}, \tag{2.15}$$

and is responsible for damaging tumor cells and other tissues. Although the collision KERMA is close to absorbed dose for lower photon energy and both share the same unit, they differ for higher photon energies. This is because higher energy electrons may travel large distances before they finally deposit their kinetic energies in the medium. As most electrons are scattered in the direction of the beam, dose is displaced downstream of KERMA. Figure 2.6 shows a schematic of absorbed dose and KERMA as functions of depth and demonstrates this effect. Figure 2.7 shows a measured percentage depth dose curve in water for 6 MV. When secondary electron equilibrium is present, that is the number and energy distribution of electrons entering and exiting the volume is equal, the collision KERMA and absorbed dose can be considered to coincide.

Furthermore, contamination electrons can contribute to dose in the tissue. Contamination electrons are generated via scattering interactions of the X-ray photon beam with the components of the treatment head of the LINAC and can contribute to dose on the patient surface.

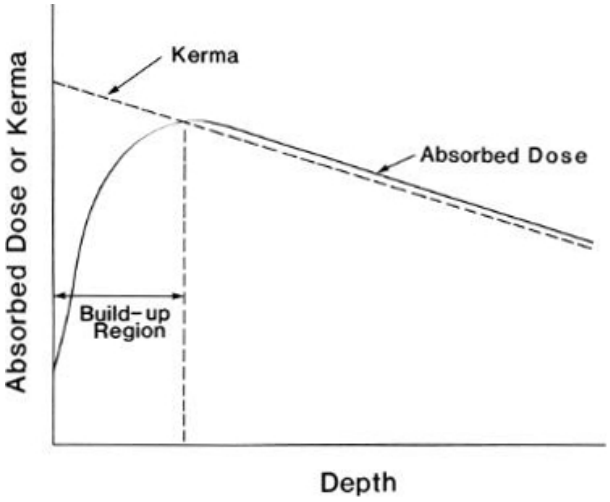


Figure 2.6: Absorbed dose and kerma. Figure by Khan *et al.* (Khan, 2010).

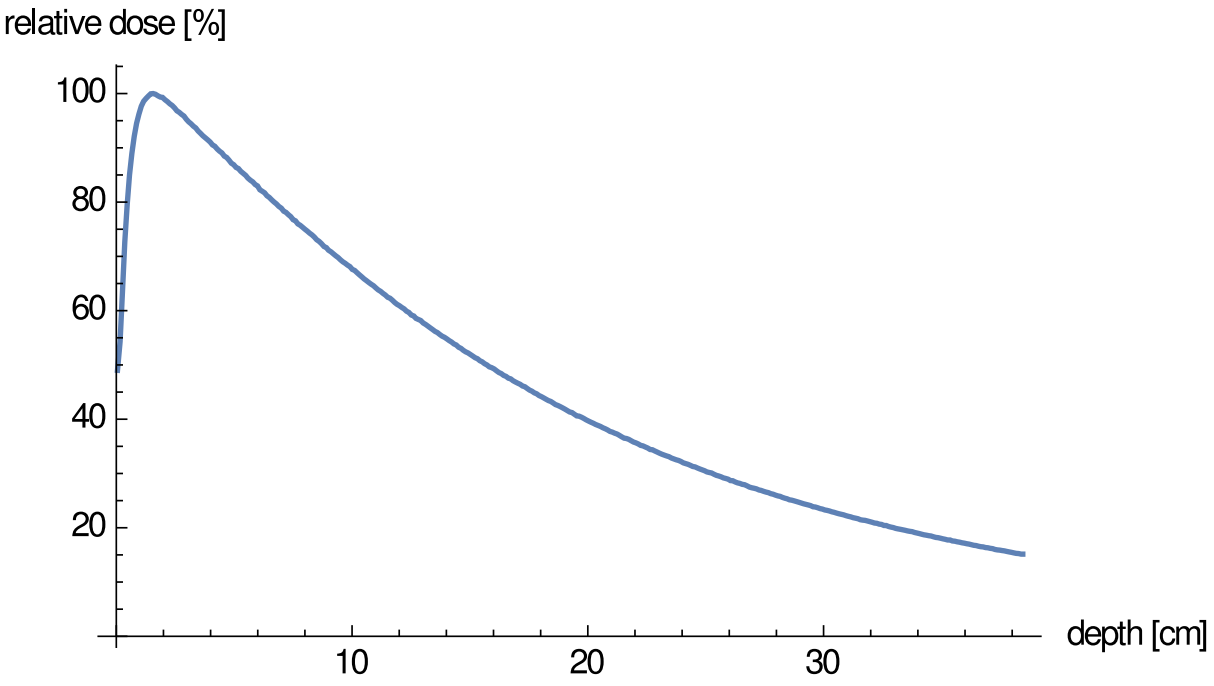


Figure 2.7: Percentage depth dose curve in water for 6 MV photons measured along the central axis of a 10 cm $\times$ 10 cm field.

Chapter 3

Treatment Planning

The treatment planning process begins with the so-called simulation, during which an initial computed tomography (CT) scan referred to as the planning CT is acquired, and skin marks or tattoos are placed in order to reproduce the position of the patient during the course of treatment. The planning CT is then used to delineate the region of interest (ROI) including the tumor and organs at risk (OARs). While the CT is usually the primary imaging modality for tumor delineation, other imaging modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) may be used to aid in the process. Margins are often added to the tumor contour to account for various types of uncertainties. A physician specifies the treatment dose and constraints to OAR, and medical physicists or dosimetrists design a treatment plan including the beam energy, beam angles, field shapes and weighting factors. The planning CT is also used to calculate the resulting dose distribution. Finally, the treatment plan is evaluated and approved by a medical physicist and a physician, often dosimetrically verified in a phantom, and delivered to the patient. The full treatment dose is usually not given in a single course, but instead delivered in several small doses called fractions (see subsection 3.4.2 for rationale). Thus, a typical treatment lasts up to several weeks with daily or nearly daily treatment fractions.

3.1 Target Definitions

After a planning CT scan and potentially additional images such as an MRI and a PET scan are acquired, the treatment target ROIs gross tumor volume (GTV), clinical target volume (CTV) and planning target volume (PTV) are defined and contoured on the planning CT scan. The concepts of these targets are defined by the International Commission on Radiation Units and Measurement (ICRU) (International Commission on Radiation Units and Measurements, 1999). The GTV contains the visible, macroscopic tumor area to be treated based on the planning CT and information obtained from other imaging modalities. However, microscopic disease may extend beyond the visible tumor mass in the planning scan. Since it is imperative for curative radiotherapy to kill all tumor cells, a margin is introduced, which creates a larger target volume that accounts for microscopic disease, referred to as the CTV. The PTV accounts for

geometrical uncertainties and is created by adding an additional margin to the CTV. The uncertainties can be systematic or random, and they can occur within one fraction (intra fractional) or between fractions (inter fractional). How far the CTV to PTV margins extend, depends on the treatment technique and clinical protocols used. Figure 3.1 demonstrates the GTV, CTV and PTV concepts. For tumors in areas where the position of the GTV and CTV is likely to change during a treatment fraction, an internal target volume (ITV) enclosing several CTVs may be utilized (Berthelsen, 2007). For example, ITVs are frequently employed for lung stereotactic body radiation therapy (SBRT), where several GTV contours for different breathing phases are delineated.

OARs are also contoured on the planning CT in order to be able to evaluate dose to normal tissues later in the treatment planning process. Since the PTV extends into normal tissues, a trade-off must be found between sufficient dose coverage of the PTV and limiting the dose to OAR. Thus, what constitutes optimal margins is subject to continuing research and the magnitude of the margins varies between institutions, treatment modality, anatomical location, and treatment protocols.

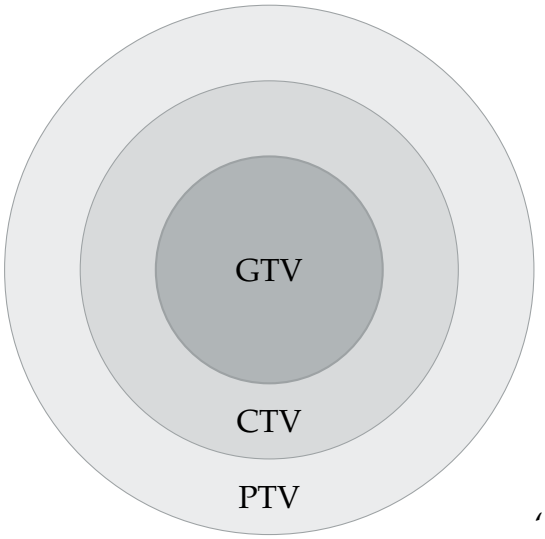


Figure 3.1: Concept of gross tumor volume (GTV), clinical target volume (CTV) and planning target volume (PTV).

3.2 Treatment Techniques

3.2.1 Early Beginnings and Two Dimensional Radiation Therapy

The first cancer patients were treated with radiation therapy in 1896 (e.g. (Despeignes, 1896)), very shortly after Röntgen had discovered X-rays in 1895. Until the 1930s, radiation therapy was delivered using Coolidge tubes or radium tubes. These devices were limited to 50-200 kV and only superficial tumors could be treated, as delivering dose to tumors seated more deeply would have exceeded the tolerance doses for the skin and subcutaneous tissues (see chapter 2, Figure 2.7). In the late 1940s, cobalt

60 units, which generated 1.17 MeV and 1.33 MeV photon beams and made it possible to spare cutaneous tissues, were first used. Starting in the 1950s, the invention of medical linear accelerators (LINACs), which could generate energies of 6-20 MV, paved the way for treatment of deeply seated tumors with doses of up to 60-70 Gy (Thariat *et al.*, 2012). At that time, the only images available for treatment planning were two dimensional (2D) bony radiographic images, which is why radiation therapy at that time is referred to as two dimensional radiation therapy (2D-RT). The fields used were rectangular in shape which led to exposure of large parts of OAR to radiation and thereby limiting the maximum dose to the tumor compared to contemporary treatment techniques. Custom made blocks that tailored the field shapes to the tumor were also available, but were impractical as they were difficult to produce and handle. Since the 1970s, radiation therapy has undergone significant improvements, such as the development of three dimensional conformal radiation therapy (3D-CRT), intensity modulated radiation therapy (IMRT) and image guided radiation therapy (IGRT), which will be discussed in the following.

3.2.2 Three Dimensional Conformal Radiation Therapy

The invention of CT in 1971 (Hounsfield, 1980; Beckmann, 2006) paved the way for a major improvement in radiation therapy. Beginning in the 1980s, the radiographic images used in 2D-RT were gradually replaced by CT images which allowed treatment planning in three dimensions and thus much more accurate delivery of dose. Radiation therapy was also further improved by the invention of multi leaf collimators (MLCs) (see section 2.1) in the 1990s, which made it possible to utilize fields that are much more tailored to the tumor anatomy and the surrounding tissue on a patient individual basis. Thus, 3D-CRT made further tumor dose escalation possible. While computer based treatment planning systems (TPSs) were available before 3D-CRT, the use of CT scans made the use of computers imperative. The availability of three dimensional (3D) information also necessitated the use of dose volume histograms (DVHs) which are now one of the main measures for evaluating treatment plans (see section 3.4.1 for more information on DVHs). In 3D-CRT, planning is performed forward, which means that a plan is designed by manually specifying the beam directions, beam weights and field shapes. The resulting dose is then calculated and evaluated and the fields are adjusted if necessary.

3.2.3 Intensity Modulated Radiation Therapy

Conventional 3D-CRT uses relatively uniform beam fluences throughout one field. IMRT is a newer technique that has been clinically used since the late 1990s and early 2000s and employs non-uniform fluence to deliver radiation. This is achieved by varying the leaf positions of the MLCs during the beam delivery, thereby creating many segments of different shapes for each beam. IMRT can be delivered as 'step-and-shoot' IMRT, where the beam is turned off when switching from one segment to the next, or dynamically as 'sliding window' IMRT where the collimator leaves move continu-

ously while the beam is on. More recently, techniques, such as volumetric modulated arc therapy (VMAT), that vary the leaf positions while the beam direction and dose rate are changing, have been introduced as well (Khan, 2010; Teoh *et al.*, 2011). IMRT allows for sparing OARs by better conformation of dose around the tumor and by delivering concave isodoses. A dose escalation study for prostate cancer by Zelefsky *et al.* shows that patients treated with IMRT had significantly less gastro intestinal (GI) toxicities than patients treated with 3D-CRT, despite the fact that IMRT patients were treated to higher doses (Zelefsky *et al.*, 2008). In contrast to 3D-CRT, the fluence and beam segments of IMRT plans are computer optimized by specifying treatment plan objectives and constraints, rather than specifying the field shapes manually. Objectives are the functions to be optimized. For example one objective of an IMRT plan may be to maximize the minimum dose in the tumor to ensure all tumor cells will be destroyed. Another example may be to minimize the mean dose to an OAR. Constraints, on the other hand, are conditions that must be satisfied by an optimal treatment plan. For instance, constraining the maximal acceptable mean dose to an OAR could avoid creating harmful treatment plans. The use of constraints in IMRT planning varies among institutions. While some institutions employ both objectives and constraints, others rely only on the use of objectives to optimize treatment plans.

3.2.4 Image Guided Radiation Therapy

IGRT refers to the use of imaging before or during treatment (Jaffray, 2012; Verellen *et al.*, 2008) and therefore is not a treatment method in itself, but is combined with other treatment modalities such as 3D-CRT and IMRT. The main goal of IGRT is to ensure the accurate delivery of dose to the tumor while also attempting to spare normal tissues from unnecessary irradiation (although IGRT requires additional imaging that may include ionizing radiation). During the course of treatment, imaging is used to reduce setup errors. IGRT images are most commonly acquired using electronic portal imaging devices (EPIDs), cone beam computed tomographies (CBCTs) and ultra sound (US), and are compared to the planning CT in order to locate the tumor and positioning the patient more accurately. This allows for the reduction of CTV to PTV margins and thus sparing critical structures (Guckenberger, 2011).

IGRT can be used for both online and offline correction strategies. For offline correction, data is gathered during several initial treatment sessions and used to determine the best position adjustments during the remaining course of treatment. Thus, offline corrections reduce systematic errors but cannot account for random errors. In online IGRT, the patient and tumor position is determined immediately before treatment, or even monitored during one treatment fraction, and adjustments to the position or beam are made immediately if necessary. Online strategies can reduce random errors, however their implementation into clinical use is difficult and requires more resources and effort (e.g. both hardware and software are needed to register the images). Furthermore, the additional daily imaging using CBCT raises concerns about damaging effects from imaging dose.

Historic Image Guidance, Portal Images, and Electronic Portal Imaging Devices

In the early beginnings of radiation therapy, image guidance was not generally available, although sporadic ideas may have existed (Verellen *et al.*, 2008). The low energies used in radiation therapy at that time did not allow enough radiation to pass through the patient to obtain images of sufficient quality. Accurate positioning was mainly ensured by immobilization of the patient and skin marks. After medical LINACs generating MV photon beams were invented, portal images (PIs) became an option for image guidance. They were obtained by placing radiographic films beneath the patient and recording 2D projections from the LINAC beam. Based on bony anatomy, this projection could then be compared to simulation X-ray images, or digitally reconstructed radiograph (DRR), which are 2D projections reconstructed from the planning CT scan. However, since the Compton effect is the dominant interaction of photons in the MV range with tissues, noise from scattering limited the quality of the images. PIs were still of comparatively poor quality, showing low soft tissue contrast. Furthermore, the films had to be developed and could not be used immediately. Since the 1980s, EPIDs, have gradually replaced PIs. EPIDs used more sensitive digital detectors rather than film and acquired digital images which made immediate setup corrections, automatic image analyses, and faster data sharing possible. EPIDs yield slight improvements over PI in terms of contrast and overall image quality, but are still subject to the limitations of using the MV beam for imaging and are restricted to two dimensions. Despite those shortcomings, they are valuable imaging tools and continue to be used in clinical practice today.

Cone Beam Computed Tomography

With the advent of CBCTs in 2000, it has also become possible to reconstruct 3D images and use them for image guidance. CBCTs consist of an X-ray source and a 2D digital detector. In contrast to EPIDs, a source mounted to the gantry at a $90°$ angle from the beam serves as an X-ray source in most CBCT devices, rather than the LINAC beam itself. The detector is placed $180°$ from the source. The X-ray source and detector can be rotated in order to obtain several 2D projections that are used to reconstruct a 3D image. The majority of CBCT devices use a kV source, although MV CBCTs that use the MV treatment source are also available (Srinivasan *et al.*, 2014). CBCTs were first used in dentistry and were later introduced and optimized for radiation therapy by Jaffray (Jaffray and Siewerdsen, 2000). CBCTs have several advantages: The images have a high spatial resolution and due to the use of kV radiation rather than MV, the images have less scattering artefacts. Since the images are 3D, they can be compared to the planning CT directly, without generating DRRs. This makes it possible to correct the patient's position with 6 degrees of freedom (3 translational, 3 rotational) if the couch is capable of performing these movements (Dhabaan *et al.*, 2012).

Other imaging modalities, such as US, CTs on rails, MRIs, stereoscopic X-rays, and fluoroscopy, are also available or under investigation for IGRT, but are less frequently used clinically.

3.2.5 Adaptive Radiation Therapy

The information obtained via IGRT cannot only be used to shift the patient position, but also in order to make changes to the treatment plan during the course of treatment, a concept called adaptive radiation therapy (ART) (Yan *et al.*, 1997). Adaptations to the treatment plan may be necessary for a number of reasons including anatomical changes (tumor shrinkage and deformation, weight loss) or functional changes (e.g. altered radiosensitivity because of hypoxia). ART can be done offline, online or potentially even real-time. During offline ART, images obtained during the first few treatment fractions are used to estimate tumor changes (movement and anatomical deformation) in order to create a new treatment plan which is adapted to altered contours and margins. This new plan is then used for the remaining treatment fractions (Yan *et al.*, 2000). The treatment plan may be adapted once, multiple times or not at all, depending on the individual patient and the available clinical resources. In online ART, the treatment plan is adapted based on information acquired immediately before a treatment fraction is being delivered. It is conceivable that treatment plans could even be adapted continuously during one treatment fraction (real-time ART). The implementation of ART is resource and labor intensive (images have to be registered to the planning CT, adapted plans have to be designed and evaluated, etc.). Yet, ART has evolved rapidly in the past 15 years and is increasingly utilized in the clinic (Li, 2011).

3.3 Dose Calculation

The dose distribution delivered to the patient cannot be measured directly, but must be calculated prior to treatment. There are several different methods of calculating this dose. Kernel based convolution/superposition methods rely on models, whereas Monte Carlo (MC) simulations and Boltzmann transport equation (BTE) solvers calculate dose using few approximations. There are three major categories of dose calculation algorithms used by contemporary treatment planning systems: (1) pencil beam (PB) algorithms, (2) collapsed cone (CC) algorithms, and (3) MC simulations (Khan, 2010; Ahnesjö and Aspradakis, 1999). BTE solvers are less commonly used clinically, but have recently been implemented into commercial TPS (Eclipse, Acuros XB, Varian Medical Systems, Palo Alto, CA, USA). In the following, these algorithms will be described briefly. However, it is to be noted that many variations and implementations of the algorithms exist and there may be considerable variation between implementations.

3.3.1 Kernel Based Superposition/Convolution Methods

All superposition/convolution algorithms use a similar approach for calculating dose. The first step is to generate an accurate beam model, which requires modelling the beam properties and energy spectrum based on machine characteristics. The beam

model can then be used to calculate the primary energy fluence $\Psi(\boldsymbol{r})$ before any tissue-beam interactions have occurred. The primary energy fluence and the mass attenuation coefficient $(\frac{\mu}{\rho})$ are used to determine total energy released per unit mass (TERMA) using equation

$$T(\boldsymbol{r}) = \Psi(\boldsymbol{r}) \left(\frac{\mu}{\rho}\right). \tag{3.1}$$

As described in chapter 2, the dose at any given position $\boldsymbol{r}$ consists of the primary dose and scattered dose from secondary particles such as Compton scattered electrons and photons from points $\boldsymbol{r'}$ and differs from collision kinetic energy released per unit mass (KERMA) and TERMA. Thus, in order to calculate dose, the TERMA is convolved with a dose spread function, also referred to as a point kernel

$$D(\boldsymbol{r}) = \int T(\boldsymbol{r'})\mathbf{A}(\boldsymbol{r} - \boldsymbol{r'})\mathrm{d}^3\boldsymbol{r'}. \tag{3.2}$$

The dose spread function $\mathbf{A}$ gives the dose distribution that results from a given primary interaction point $\boldsymbol{r'}$. It is typically pre-calculated using MC simulations (Mackie *et al.*, 1988). The convolution integral gives the dose at point r resulting from all primary interaction points $\boldsymbol{r'}$ that contribute to the dose in $\boldsymbol{r}$. The dose kernel $\mathbf{A}$ is generally calculated for water and needs to be corrected for electron density heterogeneities in the patient. Thus the kernels are scaled to the radiologic path length $\rho_{\boldsymbol{r}-\boldsymbol{r'}}(\boldsymbol{r} - \boldsymbol{r'})$ from the primary interaction site $\boldsymbol{r'}$ to $\boldsymbol{r}$. The radiologic path length is defined as the distance corrected for electron density relative to water (Khan, 2010). Scaling results in the following equation for the dose:

$$D(\boldsymbol{r}) = \int T(\rho_{\boldsymbol{r}}\boldsymbol{r'})A(\rho_{\boldsymbol{r}-\boldsymbol{r'}}(\boldsymbol{r} - \boldsymbol{r'}))\mathrm{d}^3\boldsymbol{r'}, \tag{3.3}$$

where $\rho_{\boldsymbol{r}}\boldsymbol{r'}$ is the radiologic path length from the source to $\boldsymbol{r'}$ (Khan, 2010). Full superposition/convolution calculations can be computationally expensive, because for a complete calculation, the dose contributions from all voxels to all other voxels would have to be considered. Childress *et al.* estimate 100^7 calculations are needed for a volume of 30x30x30 cm^3 and 3 mm resolution (Childress *et al.*, 2012). Thus various approximations, such as pencil beam algorithms and collapsed cone algorithms, have been proposed and implemented in order to reduce calculation times.

Pencil Beam

PB algorithms (Mohan *et al.*, 1986) assume that the beam consists of many infinitesimally small pencil shaped beams along the primary beam direction. Dose is calculated as a superposition of the corresponding dose deposition kernels (referred to as pencil beam kernels) and TERMA. The pencil beam kernels can be calculated using MC simulations for different energies, or they can be determined by integrating pre-calculated point kernels along the pencil beams. An example of a pencil beam kernel is shown in Figure (3.2). PB calculations are generally fast, especially in homogeneous media.

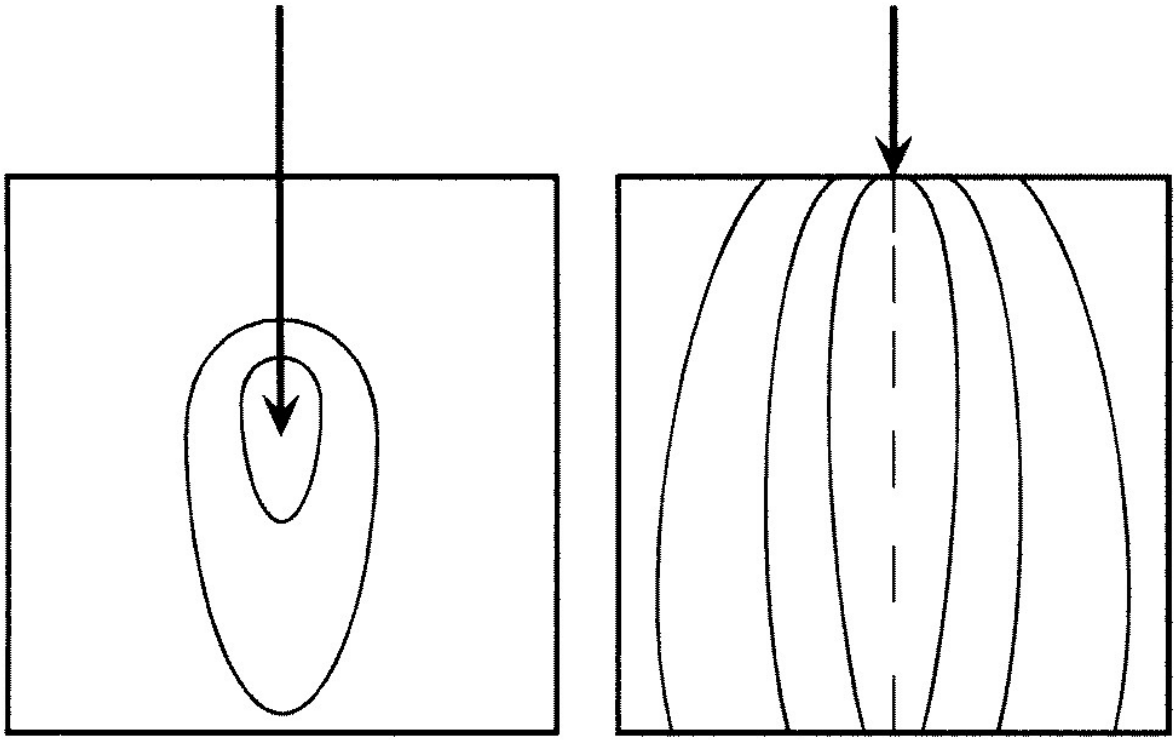


Figure 3.2: Dose deposition point kernel (left), pencil beam kernel (right). Figure by Ahnesjo and Aspradakis (Ahnesjö and Aspradakis, 1999).

Tissue inhomogeneities are considered by scaling the PB kernels based on equivalent path length, which is calculated from the total electron density between the point of incidence of the pencil beam and the voxel in which dose is deposited along the direction of the beam. The PB kernels are thus shortened for high densities and elongated for low electron densities. However, these corrections are only applied in the direction of the beam, whereas laterally all matter is assumed to be water. Thus, the PB lacks accuracy in inhomogeneous tissues, despite the density corrections, and secondary interactions other than in the beam direction are completely omitted.

Collapsed Cone

CC algorithms rely on the angular discretization of kernels into cones (Ahnesjö, 1989; Ahnesjö and Aspradakis, 1999; Childress *et al.*, 2012). They calculate the dose resulting from a primary interaction site $\boldsymbol{r}'$ in voxels within cones that have an apex at the primary interaction site voxel $\boldsymbol{r}'$. The entire dose in the cone is deposited only in the voxels along the central axis of the cone. Since the other voxels within the cone receive dose from neighboring cones, and each voxel may be included in several cones, this does not lead to large deviations. The accuracy of CC algorithms and also the calculation time depend on the number of cones used. Tissue inhomogeneities are accounted for by rectilinear rescaling of the dose deposition kernels.

In contrast to the PB algorithms, CC algorithms model lateral scattering and backscatter to some extent and yield improved dose calculations, especially in inhomogeneous

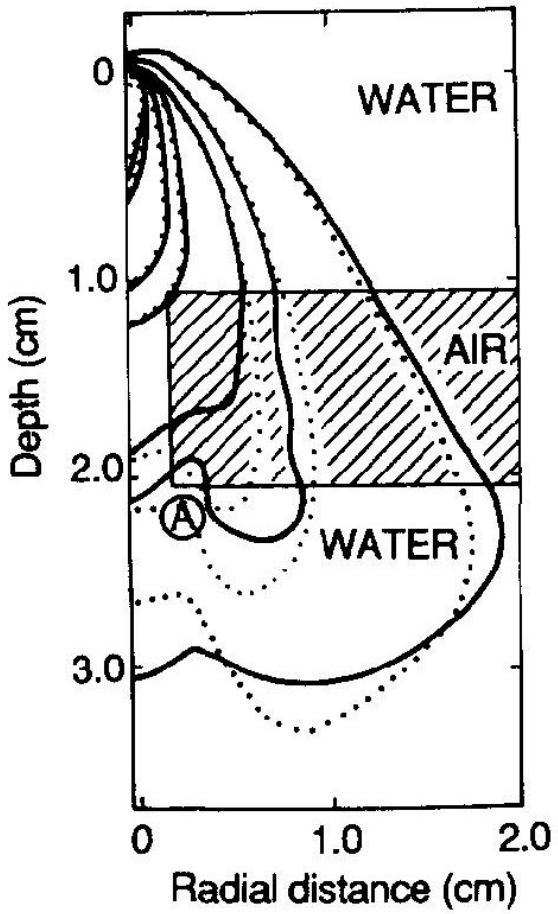


Figure 3.3: Dose deposition kernel with rectilinear rescaling (solid) and calculated with Monte Carlo (dotted). Figure by Woo and Cunningham (Woo *et al.*, 1990).

media. However, lateral scatter is still simplified, as CC algorithms neglect multiple scattering processes. Figure 3.3 shows a rectilinearly rescaled dose deposition kernel, compared to MC calculations.

3.3.2 Direct Methods

Monte Carlo Simulations

A MC simulation (Andreo, 1991; Ahnesjö and Aspradakis, 1999) is a stochastic method that is used for many parts of the dose calculation process such as for calculating the kernels used in the superposition/convolution methods. However, MC simulations can also be used to calculate dose directly without the use of kernels. MC algorithms simulate the trajectories and interactions of photons and contamination electrons from the LINAC head. Which interaction occurs for a given particle and where the interaction occurs is determined based on random numbers that are compared to interaction probabilities. The algorithms follow a particle and all secondary particles created by that particle until they are either absorbed, leave the area of interest, or cannot interact any more. This process is repeated for a large number of particles in order to obtain

statistically representative results. Typically, between 10^7 and 10^8 photons need to be simulated for clinical purposes.

MC simulations can be categorized into full MC simulations and fast MC simulations. Full MC simulations include simulating interactions in the LINAC head. They are computationally expensive, but if the parameters are well defined, full MC dose calculation algorithms are extremely accurate. Fast MC simulations use a pre-calculated virtual source model instead of simulating processes in the source and treatment head for each simulation. Thus, fast MC calculations are much faster and can be used for routine clinical purposes. MC simulation have been shown to agree well with measurements and are considered the gold standard of dose calculation algorithms.

Boltzmann Transport Equation

In contrast to the stochastic MC based dose calculation algorithms, BTE solvers are deterministic approaches to dose calculation. The BTE describes macroscopic properties of the interaction of photons and electrons with matter. Dose can be calculated by solving coupled systems of linear BTEs numerically. The accuracy of BTE algorithm depends on the variable discretizations. If the spatial and energy discretization is infinitesimally small, the solution of the BTE system converges to same solution as a MC simulation with infinitely many particles. For a review of BTE based dose calculation see (Kan *et al.*, 2013).

3.4 Treatment Plan Evaluation

3.4.1 Three Dimensional Dose Reduction

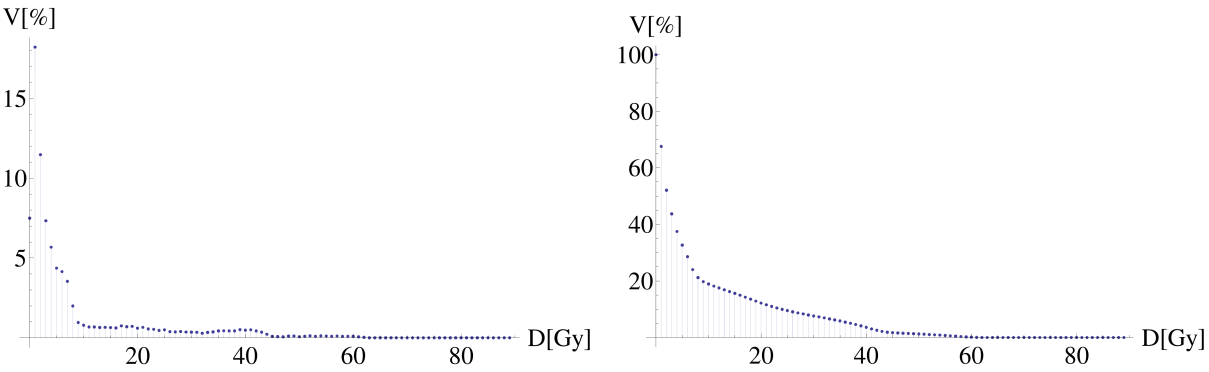
Depending on the resolution a typical CT scan may contain approximately 2.5×10^5 voxels per slice. In clinical practice it is not feasible to explicitly consider the dose in each of these voxels for assessment of the quality of a treatment plan. Thus, the 3D dose distribution within ROIs is often reduced to DVHs and the plan quality is evaluated based on simplified measures derived from those histograms (see Chapter 4). Differential dose volume histograms (dDVHs) are calculated by binning voxels with similar dose and recording their combined volume or frequency. Since there is considerable variation in organ sizes and other structures among patients, the DVH is often displayed in terms of relative volume rather than absolute volume:

$$\mathrm{dDVH}(D_i) = v_i, \tag{3.4}$$

where v_i is the relative volume of all voxels within dose bin D_i. In the clinic, however, knowing what volume percentage of a structure receives a certain dose or higher is often more relevant than knowing the dDVH. Thus, the DVH is usually displayed as a cumulative dose volume histogram (cDVH):

$$\mathrm{cDVH}(D_i) = 1 - \sum_{i|D>D_i} v_i. \tag{3.5}$$

Figure 3.4 shows a sample dDVH and its corresponding cDVH for the lungs.



(a) Differential DVH **(b)** Cumulative DVH

Figure 3.4: Differential and cumulative dose volume histograms of the lung.

3.4.2 Biological Effect of Physical Dose and Fractionation

Radiosensitivity, Cell Kill and Surviving Fraction

As described in the beginning of chapter 2, cells are destroyed by ionizing radiation damaging the cells' deoxyribonucleic acid (DNA). The exact relationship between the magnitude of dose received by tissues and the fraction of cells that survive continues to be under investigation. Several models describing the relationship have been proposed over time but the linear quadratic (LQ) model (Fowler, 1989; Hall and Giaccia, 2006) is the most established model and is currently being used clinically. Based on the LQ model, DNA can be damaged by a single interaction or two double strand breaks in the same DNA strand. While a single double strand break can usually be repaired by a cell, two double strand breaks can cause the cell to die. Lethal damage caused by single interactions, such as mutations of vital genes, are assumed to be linearly related to the dose D via αD, whereas lethal damage caused by two double strand breaks show a quadratic relationship with dose βD^2. The coefficients α and β refer to as the linear and the quadratic coefficient, respectively. The lethal events are assumed to be Poisson distributed with $\alpha D + \beta D^2$ which yields a surviving fraction SF

$$\mathrm{SF}(D) = \mathrm{e}^{-\alpha D - \beta D^2}. \tag{3.6}$$

The coefficients α and β are constants for a single cell type that need to be determined by experiment, but they vary in between different cell types. Already in the early beginnings of radiation therapy it was discovered that tissues differ in how severely they are affected by the same dose of ionizing radiation (Bergonié and Tribondeau, 1906), a characteristic referred to as radiosensitivity. The radiosensitivity of a cell depends on how rapidly that cell type divides, what stage of the cell cycle the cell is currently in, and many other factors. It is not a static property, but can be altered by external factors. For example, Read showed that oxygen is an important radio-sensitizer (Read, 1952). Thus, the radiosensitivity of cells and tissues can change during the course of treatment, for instance if a tumor becomes hypoxic.

Fractionation

In 1927, the effects of fractionation on tumor versus normal tissue were first discovered (Regaud and Ferroux, 1927). Nowadays, radiation is usually delivered in several fractions, rather than a single dose, as this has proven beneficial for sparing normal tissues and inflicting increased damage on tumor cells (Withers, 1975). A sufficient amount of time in between fractions allows normal tissue cells to repair in between treatments while tumor cells exhibit only little ability to efficiently repair. Another benefit of fractionation is the reassortment of tumor cells to the stage of the cell cycle in which they are more radiosensitive (Hall and Giaccia, 2006). Furthermore, fractionated radiotherapy allows hypoxic areas at the center of the tumor to become reoxygenated and thus increases cell kill in the tumor in the subsequent fraction. However, prolonged time between treatments can also counteract the latter effects, because it allows the tumor to regrow. Therefore, effective radiotherapy needs to optimize the time between adjacent fractions, depending on the tumor and surrounding tissue type. Standard fractionated radiation therapy (SFRT) is given in doses of 1.8 Gy to 2 Gy. The surviving fraction for fractionated treatment, according to the LQ model can be calculated as

$$\begin{aligned} \mathrm{SF}(d,n) &= \mathrm{SF}_1(d) \times \mathrm{SF}_2(d) \times \ldots \mathrm{SF}_n(d) \\ &= \mathrm{e}^{n(-\alpha d - \beta d^2)} \\ &\neq \mathrm{SF}(D), \end{aligned}$$

where d denotes the dose per fraction, n the number of fractions, and D the total dose. Since cell survival changes with fractionation, the physical dose calculated by the treatment planning system needs to be corrected for fractionation effects in many cases. A quantity called biologically effective dose (BED) can be used to compare different fractionation schemes:

$$\mathrm{BED}_{\mathrm{LQ}}\left(d, n, \frac{\alpha}{\beta}\right) = n\left(d + \frac{d^2}{\frac{\alpha}{\beta}}\right), \tag{3.7}$$

where d is the dose per fraction, n is the number of fractions delivered. Two fractionation schemes produce identical effects, if they produce the same BED. For clinical purposes the BED is sometimes normalized to 2 Gy per fraction, which is referred to as normalized total dose (NTD). The NTD is defined as

$$\mathrm{NTD}_{2,\mathrm{LQ}} = nd\left(\frac{1 + \frac{d}{\frac{\alpha}{\beta}}}{1 + \frac{2}{\frac{\alpha}{\beta}}}\right). \tag{3.8}$$

The ratio $\frac{\alpha}{\beta}$ marks the dose at which the linear component of cell kill equals the quadratic component (Hall and Giaccia, 2006) and, as mentioned above, is tissue specific. Large $\frac{\alpha}{\beta}$ (> 10 Gy) usually indicate tumor tissue or early effects in normal tissues, while small $\frac{\alpha}{\beta}$ indicate late complications of normal tissue. There are exceptions however, such as some prostate and breast tumors which have been shown to have $\frac{\alpha}{\beta}$ ratios as low as 1.5 Gy and 4 Gy, respectively (Brenner *et al.*, 2002; Zaorsky *et al.*, 2013; Whelan *et al.*,

2008). Figure 3.5 demonstrates the effect of fractionation on tumors in contrast to normal tissues based on the LQ model. While the surviving fraction in normal tissues is much higher if the total dose is given in fractions, the effect is much less pronounced in the tumor.

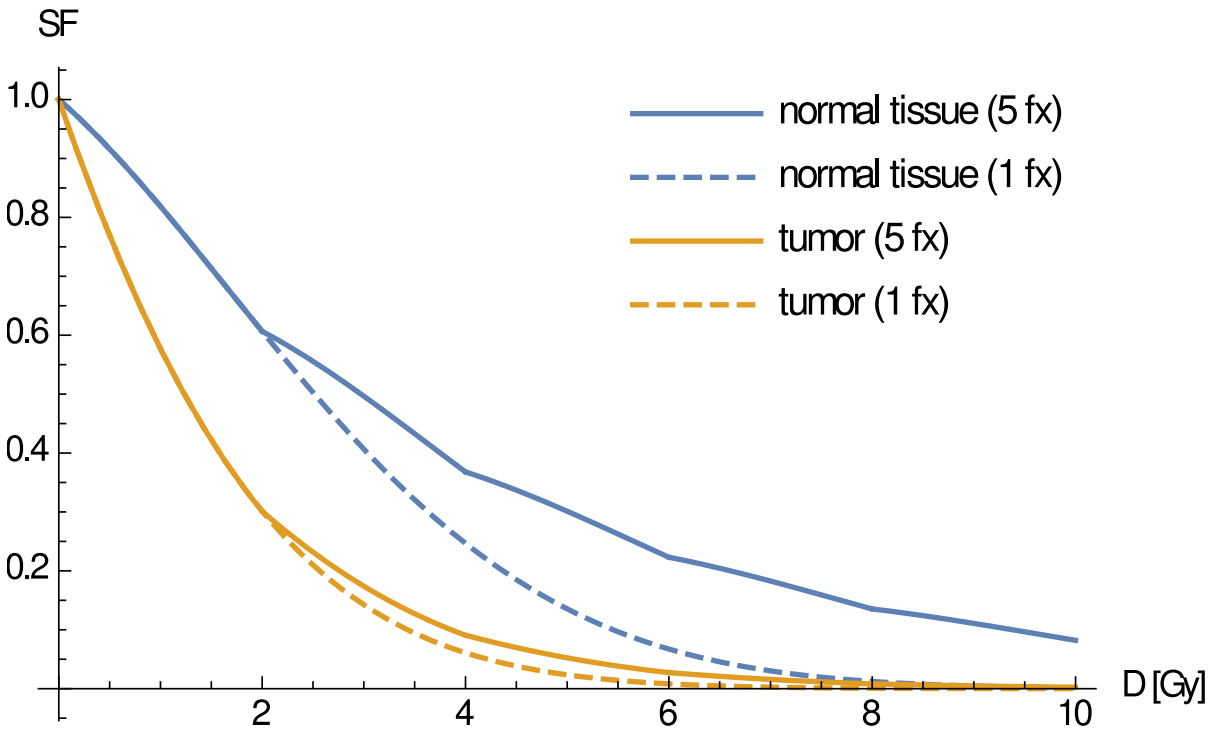


Figure 3.5: Surviving fraction SF after total dose D, given in one fraction of 10 Gy (dashed line) and 5 fractions of 2 Gy (solid line).

Hypofractionation

As mentioned above, SFRT is very effective for tumors with high $\frac{\alpha}{\beta}$ ratios. However, it is less effective for those tumors with unusually low $\frac{\alpha}{\beta}$ ratios (Brenner *et al.*, 2002; Whelan *et al.*, 2008). Thus, hypofractionated radiation therapy (HFRT), that is treatment in fractions larger than 2 Gy per fraction, has been proposed for breast and prostate tumors, and was studied in clinical trials with successful outcomes for breast cancer (Ray *et al.*, 2015; Zaorsky *et al.*, 2013). HFRT can reduce the number of fractions to less than ten fractions, and in some cases a single fraction may be used. Consequently, it reduces the overall treatment time, which leads to lower costs of treatment and is more convenient for the patient.

Due to continuing improvements in normal tissue sparing, HFRT is now also used to treat tumors with typical $\frac{\alpha}{\beta}$ ratios (≥ 10). This type of treatment was made possible by reducing treatment volumes (due to smaller margins), making it possible to irradiate less normal tissues. A common form of HFRT is stereotactic body radiation therapy (SBRT), which is a type of radiation treatment during which large doses per fraction are delivered to an extracranial treatment volume with high precision. The dose is commonly delivered in 1-5 fractions between 10 Gy and 20 Gy per fraction.

SBRT requires a high level of accuracy throughout the treatment planning process including stereotactic target localization using IGRT, beams from multiple angles, immobilization of the patient, and accurate dose calculation algorithms. SBRT was first implemented in order to improve local control of small inoperable lung tumors and has become standard of care for this type of tumor. SBRT is also used in other tumor sites, including lung, liver, pelvis and head and neck (Rubio *et al.*, 2013).

Chapter 4

Predicting Normal Tissue Complication Probability

As described in Chapter 1, adequately predicting a patient's probability of developing side effects from radiation is vital for improving radiation therapy. One method of estimating the risks associated with a treatment plan is normal tissue complication probability (NTCP) modelling. An NTCP model assigns a probability of a certain complication to a summary measure derived from a patient's dose distribution. For example, the probability of a patient experiencing radiation pneumonitis may be modelled as a function of the mean dose delivered to the lung. It is to be noted that NTCP models are specific to an endpoint rather than organ specific and several distinct side effects with distinct dose responses can occur from irradiation of the same tissue. In order to successfully use NTCP models in the clinic, the model parameters have to be determined by fitting the model to pre-existing data on dose distributions and the toxicities that resulted from them. NTCP modelling is performed in the following basic steps. First, the toxicity data and dose distribution, typically in form of dose volume histograms (DVHs), are collected for a large patient population. Using large patient populations is necessary because treatment plans are designed to keep the incidence of toxicities low. Thus, several hundreds of treatment plans that have been collected over years may be needed in order to provide a sufficient number of toxicities for NTCP modelling. After the data is collected, maximum likelihood estimation (MLE) is used to fit the data to an NTCP model and determine the optimal model parameters for the population. Once the parameters are determined, the fit of the model to the data should be verified. Ideally, the model should also be cross verified with another set of patients, but in practice the availability of data is often limited. Collecting a sufficient amount of patient data is both time consuming and costly. Older treatment plans may have to be recovered from archives and recalculated with newer treatment planning systems. Extra time has to be expended on properly evaluating and documenting toxicities, and toxicity scores of earlier patients may have to be re-scored if the scoring criteria have changed over time. Once the validity of the model is established, a patient's DVH can be plugged into the model to determine the patient's individual NTCP.

4.1 Toxicity Scoring

Toxicity scoring constitutes the basis of NTCP modelling. Toxicities, also called adverse events, are typically scored by physicians at several follow-up visits during and after treatment. However, some protocols rely on patient reported symptoms or a combination of patient and clinician reported symptoms (Gilbert *et al.*, 2015; Olsson *et al.*, 2015; Christodoulou *et al.*, 2014). The most common methods of grading toxicities is on a scale that reaches from absence of symptoms to lethal symptoms. Throughout this thesis, the Common Terminology Criteria for Adverse Events (CTCAE) scoring system was used (National Cancer Institute, 2006). Table 4.1 shows the general definition of the CTCAE grades. Grade 0 generally marks the absence of symptoms, while grade 1 is assigned to mild symptoms that do not need intervention. Toxicities of grade 2 and higher typically have a considerable impact on a patient's quality of life. However, not all grades are available for all adverse events. For example, some types of side effect cannot lead to death and therefore, grade 5 is not applicable to these toxicities. For the purpose of NTCP modelling, the toxicities need to be converted to binary values (0 or 1), and thus be pooled into two dichotomous groups (see 4.5). In many studies, grade 0 and grade 1 toxicities are pooled in one group (value 0), whereas toxicities of grade 2 and higher are pooled in the other group (value 1). The rational for this way of pooling is twofold: For one, grade 1 toxicities are mild and usually not clinically relevant. Furthermore, the distinction between grade 1 and grade 0 is often not reliable. Grade 1 toxicities may be asymptomatic except for microscopic evidence of tissue damage, and thus cannot be easily verified in each patient without expensive testing that is not clinically indicated. More details on toxicity scoring, especially bias arising from toxicity scoring, can be found in Section 5.2.1.

Grade	Description
1	Mild adverse event
2	Moderate adverse event
3	Severe adverse event
4	Life-threatening or disabling adverse event
5	Death related to adverse event

Table 4.1: Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 (National Cancer Institute, 2006).

4.2 Dose Volume Histogram Reduction

Although the DVH (see Chapter 3) constitutes a large reduction of the vast information contained in the three dimensional dose distribution, it still contains too much information for clinical use. When comparing two given DVHs it may not be clear which one is preferable over the other in terms of the probability of side effects. In fact, which one is preferable will depend on the specific endpoint associated with the

organ under consideration. Several methods have been proposed for a further reduction of the DVH to a single metric that can rank DVHs according to their probability of causing toxicities. Those metrics can be divided into two broad categories - (1) single point DVH reduction methods and (2) equivalent uniform dose (EUD) measures.

4.2.1 Single Point Dose Volume Histogram Reduction

Single point DVH reduction methods are summary measures that are points of the DVH curve. Prominently used examples are cutoff dose volumes such as V_{D_c}, the volume that receives a dose equal or higher than the cutoff dose D_c (Söhn, Yan, Liang, Meldolesi, Vargas and Alber, 2007):

$$V_{D_c} = V_{\text{cDVH}}(D_c) = \sum_{i|D_i \geq D_c} v_i, \tag{4.1}$$

where D_i is a dosebin of the DVH, v_i is the relative volume that received dose D_i, and $V_{\text{cDVH}}(D_c)$ is the volume of the cumulative DVH at D_c (see Equation 3.5). Other examples are the maximum dose (D_{max}) or minimum dose (D_{min}) received by a structure. These metrics are beneficial in clinical practice as they allow an easy evaluation of DVHs and side effect probabilities without further calculations. However, they may omit too much relevant information contained in the DVH in some cases.

4.2.2 Equivalent Uniform Dose Metrics

The second category of DVH reduction metrics is based on the concept of EUD and comprises predictors which take the entire dose distribution or DVH into account. The EUD is defined as the homogeneous dose irradiated to the entire organ, that would have the same effect on the organ, in terms of a certain endpoint, as a given heterogeneous dose distribution. If the local effect density of a given dose D is equal to $f(D)$, the total effect F of a heterogeneous dose distribution $\{D_k, v_k\}_{k=1,\dots,K}$ in a structure with K voxels

$$F(\{D_k, v_k\}_{k=1,\dots,K}) = \sum_k v_k f(D_k), \text{with} \sum_k v_k = 1, \tag{4.2}$$

where D_k is the local dose in voxel k and v_k is the relative volume of voxel k in the dose distribution. This effect is required to be equivalent to the effect of a uniform dose $D = \text{EUD}$

$$F(\{D_k, v_k\}_{k=1,\dots,K}) = f(\text{EUD}), \tag{4.3}$$

and the EUD is therefore

$$\text{EUD} = f^{-1}\left(F(\{D_k, v_k\}_{k=1,\dots,K})\right) = f^{-1}\left(\sum_k v_k f(D_k)\right). \tag{4.4}$$

The EUD can also be defined in terms of a DVH with dosebins D_i and relative volumes v_i:

$$\mathrm{EUD} = f^{-1}\left(F(\mathrm{DVH})\right) = f^{-1}\left(\sum_i v_i f(D_i)\right). \tag{4.5}$$

The true effect (density) is usually not known and could involve complex biological processes. However, several concepts have been proposed to model EUD. The most commonly used model for the effect density f for organs at risk (Niemierko, 1999) is the power law as used for the generalized equivalent uniform dose (gEUD) $f_{\mathrm{gEUD}}(D) = D^a$. The corresponding EUD of a DVH is then simply the p-norm with $p = a$.

$$\mathrm{gEUD}(\mathrm{DVH}) = \left(\sum_i v_i D_i^a\right)^{\frac{1}{a}}. \tag{4.6}$$

The gEUD is sometimes referred to as the effective dose (D_{eff}). Other models for the local effect of dose and EUD have been proposed as well, including mechanistic models, e.g. (Alber and Belka, 2006).

4.3 Volume Effect and Tissue Architecture

Although the gEUD concept is mainly algebraic and not derived from a biological background, its parameter a is sometimes interpreted as an indicator of the functional nature of the tissue. This parameter a is referred to as the volume-effect parameter, because it determines the importance of dose and volume on the effect on the tissue. Essentially, it determines whether a low dose to a large volume (a small) is more damaging than a high dose to a small volume (a large) or vice versa.

The gEUD converges to the maximum dose $D_{\max}$ for large a, and is equal to the mean dose for $a = 1$

$$\lim_{a\to\infty} \mathrm{gEUD}\left(\mathrm{DVH}, a\right) = D_{\max}(\mathrm{DVH}),$$
$$\lim_{a\to 1} \mathrm{gEUD}\left(\mathrm{DVH}, a\right) = D_{\mathrm{mean}}(\mathrm{DVH}).$$

A graph demonstrating DVHs representing gEUD for a lung DVH for several values of a is shown in Figure 4.1. In this context reference is often made to parallel or serial tissue architecture of organs. Organs that do not experience toxicity until a certain volume threshold of tissue damage is reached (such as liver and lung) are referred to as parallel organs. Organs that can tolerate small doses to large volumes, but cause toxicity as soon as a small volume receives a large dose are termed serial organs (e.g. spinal cord, rectum). However, dividing entire organs into serial and parallel without considering the specific endpoint can be misleading. For example, bleeding is generally a serial complication, while radiation pneumonitis is considered a parallel complication. Both parallel and serial complications may occur in the same organ. Although

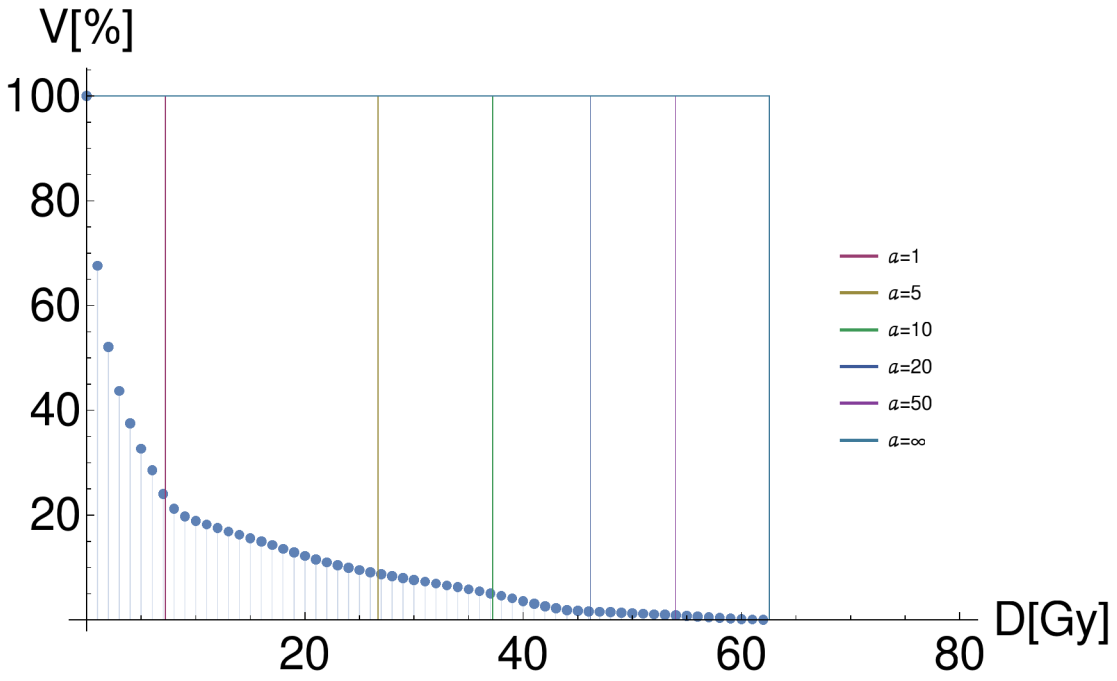


Figure 4.1: Generalized equivalent uniform dose (gEUD) for several values of the volume effect parameter a

characterization of tissues as serial or parallel can be useful, one must not forget that in reality the biological processes that determine tissue organization can be immensely complex.

4.4 Normal Tissue Complication Probability Models

When creating a treatment plan for a patient, one is interested in determining the probability NTCP_i that a patient with a DVH_i will experience a complication, or toxicity y_i. This probability is generally not known, and as pointed out above the biological processes behind it can be very complex. However, NTCP models can be used in order to provide an approximation for that probability. Numerous NTCP models have been proposed, ranging from simple models for homogeneous partial irradiation (Lyman, 1985) to complex ones incorporating biological mechanisms (Alber and Belka, 2006; Burman *et al.*, 1991; Lyman, 1985). Marks *et al.* (Marks *et al.*, 2010) established guidelines for heterogeneous three dimensional conformal radiation therapy (3D-CRT) within the framework of the Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) reports and presented the most commonly used NTCP models.

4.4.1 Cutoff Dose Logistic Regression Model

The cutoff dose logistic regression (CDLR) model is a logistic regression of the volume V_{D_c} receiving a cutoff dose D_c or higher (Equation 4.1)

$$\mathrm{NTCP}(D_c) = \frac{1}{1 + \exp\left(-\beta_0 - \beta_1 V_{D_c}\right)}. \tag{4.7}$$

The volume causing 50 % complication is the ratio of the logistic regression coefficients, $V_{D_c,50\%} = -\frac{\beta_0}{\beta_1}$. The logistic regression model can also be used with other metrics such as $\mathrm{D}_{\mathrm{mean}}$ and EUD, for example.

4.4.2 Lyman Equivalent Uniform Dose Model

The Lyman EUD model was introduced by Lyman, and extended by Kutcher and Burman (Lyman, 1985; Kutcher and Burman, 1989). It is also called the Lyman Kutcher Burman (LKB) model. The model assumes that NTCP is a sigmoidal function of the EUD:

$$\mathrm{NTCP}(\mathrm{EUD}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(\frac{-x^2}{2}\right) \mathrm{d}x, \text{ with} \tag{4.8}$$

$$t = \frac{\mathrm{EUD} - \mathrm{EUD}_{50}}{m\mathrm{EUD}_{50}}. \tag{4.9}$$

EUD_{50} denotes the EUD causing a 50% chance of experiencing a complication and m is a parameter that is inversely related to the slope of the NTCP curve at EUD_{50}. This model is sometimes written more conveniently as an error function

$$\mathrm{NTCP}(\mathrm{EUD}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\mathrm{EUD} - \mathrm{EUD}_{50}}{\sqrt{2}m\mathrm{EUD}_{50}}\right)\right). \tag{4.10}$$

4.5 Model Fitting

4.5.1 Maximum Likelihood Estimation

In order to determine the parameters of the NTCP models in Equation 4.7 or 4.9, the model must be optimized for a sample dataset. This is done using MLE or sometimes using other methods such as least square estimation. However, MLE should be the preferred method, because it is more generally applicable (Burnham and Anderson, 2004) and does not rely on binning of the binary toxicity data (Cox, 1989). In MLE the likelihood $\mathcal{L}$ of the model is maximized for a measurement of the set of toxicities $\{y_i\}_{i=1,\ldots,n}$, for n patients, where y_i is the toxicity caused by DVH_i. The y_i are dichotomous, that is they can have two values, either $y_i = 0$ or $y_i = 1$. If the NTCP model parameters denoted by $\boldsymbol{\theta}$ were known, the probability p of measuring a set of toxicities $\{y_i\}$ caused by $\{\mathrm{DVH}_i\}$ for n patients ($i = 1, \ldots, n$) could be calculated as

$$p(y_1, \ldots, y_n, \mathrm{DVH}_1, \ldots, \mathrm{DVH}_n \,|\boldsymbol{\theta}) = \prod_{i=1}^{n} \left(\mathrm{NTCP}_i(\mathrm{DVH}_i \,|\boldsymbol{\theta})\right)^{y_i} \left(1 - \mathrm{NTCP}_i(\mathrm{DVH}_i \,|\boldsymbol{\theta})\right)^{1-y_i},$$

(4.11)

since the values y_i are independent. That is, the probability is a function of the data given the parameters $\boldsymbol{\theta}$. However, the model parameters $\boldsymbol{\theta}$ are not known and need to be estimated by fitting known data to the NTCP model. Thus, rather than a function of the data given the parameters $\boldsymbol{\theta}$, a function of the parameters $\boldsymbol{\theta}$ given the data $y_i(\mathrm{DVH}_i)$ needs to be optimized. This function is called the likelihood $\mathcal{L}$:

$$\mathcal{L}(\boldsymbol{\theta}|\{y_i, \mathrm{DVH}_i\}) = \prod_{i=1}^{n} \mathcal{L}(\boldsymbol{\theta}|y_i, \mathrm{DVH}_i) \tag{4.12}$$

$$= \prod_{i=1}^{n} (\mathrm{NTCP}_i(\boldsymbol{\theta}|\,\mathrm{DVH}_i))^{y_i} (1 - \mathrm{NTCP}_i(\boldsymbol{\theta}|\,\mathrm{DVH}_i))^{1-y_i}. \tag{4.13}$$

To determine the optimal model parameters $\boldsymbol{\theta}_{\mathrm{opt}}$, the likelihood function is maximized with respect to $\boldsymbol{\theta}$ giving the MLE estimator:

$$\boldsymbol{\theta}_{\mathrm{opt}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}). \tag{4.14}$$

Since the logarithm preserves the position of the maximum, $\ln \mathcal{L}(\boldsymbol{\theta})$ is maximized rather than $\mathcal{L}(\boldsymbol{\theta})$ to simplify the calculation to a sum rather than a product:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \prod_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \mathcal{L}_i(\boldsymbol{\theta}). \tag{4.15}$$

The expression $\ln \mathcal{L}(\boldsymbol{\theta})$ is referred to as the log likelihood (LL).

4.5.2 Implementation

While the MLE estimator can be calculated analytically for simple problems in a closed expression, this is not possible for complex problems like NTCP modelling. All MLE estimators in this thesis were therefore calculated numerically using Mathematica (Wolfram Research Inc., Version 9.0-10.3, Champaign, IL) and software developed by the author based on C++.

4.6 Comparing Models

The likelihood value or log likelihood value of a model has no absolute value and is not comparable between datasets. However, within a given dataset, the value of the likelihood function at the optimal parameters can be used to compare different models if one model is a sub-model of the other (that is, one or more parameters are fixed). The likelihood ratio test (Pawitan, 2001) can be used in that case to determine if there is a significant improvement of the more advanced model (the full model) over the simpler version of the model (the reduced model). The likelihood ratio is computed as

$$\mathrm{LR} = 2\left(\ln \mathcal{L}_{\mathrm{full}} - \ln \mathcal{L}_{\mathrm{reduced}}\right), \tag{4.16}$$

where $\mathcal{L}_{\text{full}}$ is the maximum likelihood of the full model, whereas $\mathcal{L}_{\text{reduced}}$ is the maximum likelihood of the reduced model, in which at least one parameter is fixed. The likelihood ratio LR is then compared to the $100(1-\alpha)$th percentile of a chi-square distribution with df degrees of freedom $\chi^2_{1-\alpha}(df)$, where df is the difference in model parameters between the full model and the reduced model. If the value of LR is larger than $\chi^2_{1-\alpha}(df)$, there is significant improvement of the full model over the reduced model at a significance level of $100(1-\alpha)$ %.

Adding parameters to the model will also increase its likelihood of over-fitting the data and fitting noise. Akaike *et al.* published a criterion known as Akaike information criterion (AIC) to correct the log likelihood values for the number of parameters of the model (Akaike, 1974; Burnham, 2004). It is calculated as

$$\text{AIC}(\text{LL}, k) = 2k - 2\,\text{LL}, \tag{4.17}$$

where LL is the log likelihood $\ln \mathcal{L}(\theta_{\text{opt}})$ of the model and k is the number of model parameters. A lower AIC value indicates the superior model. If the sample size n is low, the AIC can be further corrected, using the corrected Akaike information criterion (AICc). The AICc is dependent on the model, but can be estimated by

$$\text{AICc}(\text{LL}, k, n) = \text{AIC} + \frac{2k(k+1)}{n-k-1}. \tag{4.18}$$

Just like the value of LL, neither the AIC, nor the AICc can determine the absolute quality of the NTCP model.

4.7 Goodness of Fit

Goodness of fit of a particular model can be determined using a Hosmer-Lemeshow goodness of fit test (Collett, 2003; Hosmer and Lemesbow, 1980). A Hosmer-Lemeshow statistics χ^2_{HL} is calculated from the data grouped into g groups:

$$\chi^2_{\text{HL}} = \sum_{i=1}^{g} \frac{(o_i - m_i\pi_i)^2}{m_i\pi_i(1-\pi_i)}, \tag{4.19}$$

where m_i is the total number of observations, o_i the observed number of successes, and π_i the average expected success probability in the i-th group. The goodness of fit p-value is obtained by comparing χ^2_{HL} to a χ^2 distribution with $g-2$ degrees of freedom. Obtaining a low p-value $(p < 0.05)$ signifies that the null-hypothesis, that the data stems from the same distribution as the theoretical model, can be rejected. The model is considered to yield an acceptable fit to the data if $p > 0.05$ (95% confidence) (Collett, 2003).

4.8 Standard Errors and Confidence Intervals

4.8.1 Variance Covariance Matrix

The standard errors and confidence interval (CI) of the maximum likelihood estimate $\boldsymbol{\theta}_{\text{opt}}$ can be calculated via the Hessian matrix $\mathbf{H}(\boldsymbol{\theta})$ of the log likelihood function (Pawitan, 2001)

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 \ln \mathcal{L}(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}, \tag{4.20}$$

assuming that the maximum likelihood estimator is normally distributed. From the Hessian matrix, the variance-covariance matrix $\mathbf{var}(\boldsymbol{\theta})$ can be calculated as the inverse of the negative of the Hessian matrix

$$\mathbf{var}(\boldsymbol{\theta}) = (-\mathbf{H}(\boldsymbol{\theta}))^{-1}. \tag{4.21}$$

Standard errors of the parameters $\boldsymbol{\theta}_{\text{opt}}$ are the square roots of diagonal elements of the variance covariance matrix evaluated at $\boldsymbol{\theta}_{\text{opt}}$. That is, the standard error estimate $\hat{\text{se}}$ of a parameter $\theta_{\text{opt},i}$ can be determined as

$$\hat{\text{se}}(\theta_{\text{opt},i}) = \sqrt{\text{var}_{i,i}(\boldsymbol{\theta}_{\text{opt}})} = \sqrt{\text{var}_{i,i}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\text{opt}}}} \tag{4.22}$$

and 95% CI for each parameter can be calculated as

$$\theta_{\text{opt},i} \pm 1.96\hat{\text{se}}(\theta_{\text{opt},i}). \tag{4.23}$$

4.8.2 Profile Likelihood Method

Another way to estimate CIs for $\boldsymbol{\theta}_{\text{opt}}$, without assuming normality of the maximum likelihood estimator, is using the profile likelihood method (Pawitan, 2001; Böhning *et al.*, 2008). If the parameter of interest is denoted by θ_i and the remaining parameters are $\boldsymbol{\theta}_j$, then the profile likelihood function is defined as

$$\mathcal{L}_{\text{prof}}(\theta_i) = \max_{\boldsymbol{\theta_j}} \mathcal{L}(\theta_i, \boldsymbol{\theta_j}). \tag{4.24}$$

To calculate $100(1-\alpha)\%$ CI, one has to find the interval of θ_i for which

$$\ln \mathcal{L}_{\text{prof}}(\theta_i) > \ln\left(\mathcal{L}(\boldsymbol{\theta}_{\text{opt}})\right) - \frac{\chi^2_{1-\alpha}(1)}{2}, \tag{4.25}$$

where $\chi^2_{1-\alpha}(1)$ is the $100(1-\alpha)$th percentile of a chi-square distribution with 1 degree of freedom.

4.8.3 Bootstrapping

Bootstrapping is another method for determining better standard errors of the maximum likelihood estimator, especially when the sample size is small (Efron and Tibshirani, 1998). Bootstrapping estimates the standard error by repeated analysis of resamples of the original dataset. Let $\{\text{DVH}_i, y_i\}, i = 1, \ldots, n$ be the set of patient DVHs

with the respective scored toxicities, and $\boldsymbol{\theta}_{\text{opt}}$ the ML estimates for the parameters $\boldsymbol{\theta}$. In order to obtain the standard error of $\boldsymbol{\theta}_{\text{opt}}$, a new set of n random samples $\{\text{DVH}_j, y_j\}$, $j = 1, \ldots, n$ is drawn from $\{\text{DVH}_i, y_i\}$ with replacement, which means the new DVHs are always drawn from the entire original set of DVHs, even if a particular DVH has already been drawn before. Since the new set of DVHs was drawn with replacement, some of the DVHs from the original set may occur twice in the new set while others may not be part of the new set at all. The MLE analysis is performed on that new set $\{\text{DVH}_j, y_j\}$ in exactly the same way as on the original dataset $\{\text{DVH}_i, y_i\}, i = 1, \ldots, n$, and a new estimate for $\boldsymbol{\theta}$ is recorded. This type of resampling is repeated K times, generating K sets of DVHs and their respective toxicities with n patients in each set. For each of these K sets a new estimate $\boldsymbol{\theta}_{\text{opt},k}$, $k = 1, \ldots, K$ is calculated and stored. The estimate of the standard error $\hat{\text{se}}(\boldsymbol{\theta}_{\text{opt}})$ of $\boldsymbol{\theta}_{\text{opt}}$ and CI of $\boldsymbol{\theta}_{\text{opt}}$ is then calculated as the standard deviation σ of the K values $\boldsymbol{\theta}_{\text{opt},k}$:

$$\hat{\text{se}}(\boldsymbol{\theta}_{\text{opt}}) = \sigma\left(\{\boldsymbol{\theta}_{\text{opt},k}\}_{k=1,\ldots,K}\right). \tag{4.26}$$

Alternatively, CI can be estimated by calculating the 5th and 95th percentile of $\{\boldsymbol{\theta}_{\text{opt},k}\}_{k=1,\ldots,K}$ (Burnham and Anderson, 2004).

Part II

Studies

Chapter 5

Influence of Treatment Technique

Publication: Comparison and Limitations of DVH based NTCP Models Derived from 3D-CRT and IMRT Data for Prediction of Gastrointestinal Toxicities in Prostate Cancer Patients Using Matched Pair Analysis [1]

5.1 Introduction

In this chapter, the applicability of normal tissue complication probability (NTCP) models derived from a patient population treated with a specific treatment technique to a second population treated with a different technique was investigated. Specifically, it was determined whether models derived from a population treated with three dimensional conformal radiation therapy (3D-CRT) can be applied to intensity modulated radiation therapy (IMRT) patients and vice versa. It has been shown that current NTCP models can work well if applied to the population and treatment-technique they were derived from (Michalski *et al.*, 2010). It is unclear, however, whether they are reliable predictors of toxicity when applied to a different treatment-modality (Michalski *et al.*, 2010), especially since the observed incidence of toxicities can differ vastly between treatment-techniques. The issue is of major relevance, as most NTCP model parameters are derived from 3D-CRT data (Michalski *et al.*, 2010), but are frequently used to estimate complications for intensity-modulated radiotherapy (Luxton *et al.*, 2004). While a single patient's NTCP will depend on the dose distribution received, ideally the NTCP model parameters for a specific organ/tissue should not depend on the actual dose-distribution delivered to that organ, regardless of treatment-technique. However, commonly used NTCP models potentially lack the complexity to resemble the underlying biological processes, or the model parameter estimates might be bi-

[1] Reprinted (adapted) with permission from *"Comparison and Limitations of DVH-Based NTCP Models Derived From 3D-CRT and IMRT Data for Prediction of Gastrointestinal Toxicities in Prostate Cancer Patients by Using Propensity Score Matched Pair Analysis"* by Troeller, A., Yan, D., Marina, O., Schulze, D., Alber, M., Parodi, K., Belka, C., Söhn, M.; Int. J. Radiat. Oncol. 91, 435–443. doi:10.1016/j.ijrobp.2014.09.046

ased by a cohort with little variability in dose levels (Hardcastle *et al.*, 2010). Furthermore, commonly used NTCP models are based on dose volume histogram (DVH)s which leave out spatial information. Bias can also arise from toxicity scoring, as neither the patient nor the physician performing the follow-up (FU) can be blinded to the technique. Also the evolution of improved image guidance contemporaneously with IMRT may play a role. The topic is further complicated by differences in patient characteristics inherent to the populations. In the following, a selection of possible factors influencing toxicity models are generally discussed, followed by an overview of matched pair analysis for comparing treatment groups. Then, a study investigating the topic is presented. In this study, NTCP model parameters for gastro intestinal (GI) toxicities were derived for a large, matched population of prostate cancer patients treated at a single institution with 3D-CRT or IMRT, in order to investigate discrepancies in toxicities between treatment-techniques. Best-fit parameters were calculated for all patients (3D-CRT+IMRT) as well as the 3D-CRT and IMRT sub-groups separately, and it was determined how well a model fits the data it was derived from and how well it fit the data it was not derived from. The impact of differences between the 3D-CRT and IMRT patient characteristics, image-guidance, bias of toxicities and other possible causes of differences in NTCP modelling results are discussed.

5.2 Factors Influencing Toxicity Modelling

5.2.1 Toxicity Scoring Bias

As pointed out in part I of this thesis, proper scoring of toxicities is crucial for NTCP modelling. However, although scoring systems such as the Common Terminology Criteria for Adverse Events (CTCAE) attempt to objectively evaluate toxicities, the scores are recorded by physicians based on their interpretation of patient-reported signs and symptoms, and thus may contain a certain bias. How side-effects are perceived may be influenced by the patient's and physician's knowledge of the treatment-technique. Such bias, unfortunately, cannot easily be excluded.

Furthermore, the follow-up time can play an important role in NTCP modelling. For example, if the follow-up time is not sufficiently long, patients who might have later experienced side effects of higher grades may have accidentally been categorized as grade 0. This can potentially change the NTCP model parameters and needs to be considered when comparing groups.

5.2.2 Errors of Planned Versus Delivered Dose

The dose calculated for a treatment plan differs from the actual dose delivered to the patient due to uncertainties such as motion and positioning. The magnitude of this error of the planned dose used to model NTCP may differ for treatment techniques. For example, both treatment techniques differ in dose gradients which may influence the impact of motion errors and positioning errors on the truly delivered dose in different ways. Furthermore, because of different technological advances such as image guided

radiation therapy (IGRT) contemporaneously with newer treatment techniques, the magnitude of the difference between planned and delivered dose may vary between treatment techniques. A more detailed discussion of these effects especially in reference to the study presented below, can be found in Section 5.6.

5.2.3 Confounding Patient Characteristics

When NTCP modelling is compared or applied to different patient populations, such as patient populations receiving different kinds of treatment, there may be confounding factors that need to be considered. For example, a history of smoking can influence how lung tissue reacts to radiation (Alsadius *et al.*, 2011; Solanki and Liauw, 2013). Consequently, if two populations differ significantly in how many of the patients smoke, NTCP models derived from such populations may differ as well. Many other endpoint dependent confounding factors have been identified for various tissues. For the GI toxicities investigated in this study, advanced age, diabetes and abdominal surgery prior to radiation therapy, hemorrhoids, androgen deprivation therapy (ADT), larger rectal volume, anticoagulants, and possibly smoking have been shown to influence the incidence of toxicities (Budäus *et al.*, 2012; Valdagni *et al.*, 2012).

5.3 Matched Pair Analysis

Ideally, two groups to be compared should be randomized to avoid bias from confounding variables. Unfortunately, this is not always possible for ethical or practical reasons. It would not be ethical, for example, to randomly expose patients to a certain type of treatment, when it is well known that another treatment is superior in treating their specific condition. In order to study and compare models for different groups, one therefore has to adjust for selection bias. This can be done using matched pair analysis to match pairs of patients from both populations with similar patient characteristics (Guo and Fraser, 2015; Rosenbaum and Rubin, 1983). If there are only few confounding variables that are categorical, it is possible to match patients exactly. That is, for each patient from one group, a patient with the exact same values of confounding variable is selected from the other group. However, in practice confounding variables are often continuous, or there is a multitude of confounding variables, and exact matching is not possible. In these cases, propensity score based matched pair analysis can be used.

Propensity score matching is performed in three steps:

1. **Propensity Score Calculation:** A measure for similarity in confounding variables, the propensity score, is calculated for each patient in the two groups g_1 and g_2.

2. **Matching:** Patients from group g_2 are matched to group g_1 based on the propensity score.

3. **Post Matching Analysis:** The datasets are analyzed regarding remaining differences in confounding variables and the match may be repeated with different parameters if significant differences remain between the matched groups.

5.3.1 Propensity Score

The propensity score P is defined as the conditional probability of a patient i being assigned to group 1 (i.e. $G_i = 1$, for example the group of patients with GI toxicities of grade ≥ 2) given the value of the confounding variables $\boldsymbol{x_i}$ (Rosenbaum and Rubin, 1983):

$$P(\boldsymbol{x_i}) = p(G_i = 1 | \boldsymbol{X_i} = \boldsymbol{x_i}). \tag{5.1}$$

The propensity score for each patient can be estimated by fitting the toxicity data of the entire patient population to a logistic regression model of the confounding variables $\boldsymbol{x_i}$:

$$P(\boldsymbol{x_i}) = \frac{1}{1 + \exp(-\boldsymbol{x_i}\boldsymbol{\beta_i})}, \tag{5.2}$$

where $\boldsymbol{\beta_i}$ are the regression coefficients. The $\boldsymbol{\beta_i}$ are determined by fitting the binary data (i.e. $G_i = 1$ if the patient was assigned to group 1 with toxicity grade ≥ 2, and $G_i = 0$ if the patient was assigned to group 2, with toxicity grade < 2) to the propensity score model using maximum likelihood estimation (MLE) (see Section 4.5). Once the logistic model parameters are determined, the propensity score for each patient can be calculated in both groups and be used as a similarity measure. Proper selection of which covariates should be included is important but non-trivial. Guo and Fraser note that although several methods have been proposed, no standard method of selecting covariates currently exists (Guo and Fraser, 2015).

5.3.2 Matching

After the propensity score has been calculated for each patient in groups g_1 (e.g. 3D-CRT) and g_2 (e.g. IMRT), pairs of patients from group g_1 and group g_2 that had similar probabilities of being in group G_1 or G_2 based on the confounding variables are matched based on their propensity score. Several matching methods/algorithms have been proposed, such as greedy matching, optimal matching, and fine matching which are discussed below. The matching process involves a trade-off between a minimal difference between the populations on the one hand and loss of patients on the other hand. The dataset may be reduced due to several factors: Some patients may not have a decent match in the other group, and similarly some control patients may never be used as a match.

Greedy Matching

The defining characteristic of greedy matching is that once a match has been selected for a specific patient, he or she will not be rematched. The matching patient will be

removed from the pool of possible matches for the remaining patients to be matched. Thus, greedy matching does not always select the best possible match as a match that was removed from the pool g_2 may be a better fit for another patient from g_1. Furthermore, greedy matching needs a relatively large overlap of the two groups. Advantages of greedy matching is that it fullfills the assumptions of multivariate analysis techniques (Guo and Fraser, 2015) and the ability to perform incomplete matches in which not every patient from group g_1 receives a match. There are several submethods of greedy matching, such as nearest neighbor matching, caliper matching or a combination of the two.

Optimal Matching

In contrast to greedy matching, optimal matching is a non-linear process which differs from greedy matching in that matches are constantly re-evaluated in favor of a possible closer match. A distance measure based on propensity scores is minimized over all possible pairwise matches. Optimal matching allows for variable matching in which each patient from the treatment group may be matched to a variable number of patients from the control group. Unlike greedy matching, the optimal match needs to be complete (Guo and Fraser, 2015).

Fine Balance Matching

Fine balance matching balances a variable rather than finding an optimal match for each patient. That is in fine balance matching, pairs of patients may be matched even if their propensity score is not similar as long as the matching variables are balanced overall (Rosenbaum, 2010).

5.3.3 Post Matching Analysis

The major component of post matching analysis is comparing the groups again and evaluate whether significant differences between the groups remain or have resolved. In the former case, matching could be repeated with more tightly restricted parameters at the expense of a smaller final dataset. If the matched sample shows no significant difference in the confounding variables the dataset can be used for modelling. However, in cases where the match is performed for further processing and analysis of the groups, a variety of post matching analysis methods exists and can be used.

5.4 Modelling Normal Tissue Complication Probability for Three Dimensional Conformal Radiation Therapy and Intensity Modulated Radiation Therapy Data

5.4.1 Patient Data

For this study, 1,115 patients treated for prostate cancer (1999-2010) were selected from a prospective prostate cancer database. Patients with a FU of at least 2 years were investigated for the occurrence of GI side-effects after radiotherapy. Of these patients, 457 were treated with conventional 3D-CRT, while 658 received IMRT. The endpoints considered in this study were chronic GI toxicities grade ≥ 2 as defined by the CTCAE v. 3.0. These include rectal bleeding, rectal pain, diarrhea, proctitis, fistulae and anal fissures.

5.4.2 Treatment Planning with Three Dimensional Conformal Radiation Therapy and Intensity Modulated Radiation Therapy

All patients were treated under an adaptive radiation therapy (ART) protocol (Yan *et al.*, 2000). The treatment plan is divided into an initial and an adapted treatment-plan. On the planning computed tomography (CT) acquired for each patient the initial planning target volume (PTV) was constructed by adding a 1 cm margin to the clinical target volume (CTV). The CTV consisted of either prostate or prostate and seminal-vesicles. A four-field-box was created and treated for the first five fractions for all patients. During the first week of treatment, daily portal images (PIs) and four CT scans were acquired. From these images the specific set-up error and mobility of each patient was estimated in order to construct a non-uniform, so called confidence-limited PTV with adapted margins. For the 3D-CRT patients an adapted-plan consisting of a new four-field-box was constructed, whereas for the IMRT patients an adapted IMRT plan was created, typically incorporating 5-7 beams. The adapted plans were used for treatment of all remaining fractions. Portal imaging was repeated once a week thereafter, to adjust patient setup. Starting in 2007, all PIs were replaced by cone beam computed tomography (CBCT), which affected 29 % of the IMRT patients. PI were matched to the digitally reconstructed radiograph (DRR) of the original planning CT via bony match, and and CBCTs were matched to the planning CT using bony anatomy. Prescribed doses ranged from 70.2 Gy to 82.8 Gy with a median of 75.6 Gy, delivered in 1.8 Gy fractions. The organs at risk (OARs) were outlined on the initial CT scan only. The rectum was contoured from the lower of the anal verge or ischial tuberosities to the higher of the sacroiliac joints or recto-sigmoid junction (Vargas *et al.*, 2005). A rectal wall contour was generated, by constructing a 3-4 mm ring inside the outer rectum contour. The bladder was contoured as a solid organ. The major dose-volume constraints were $V_{75.6\text{Gy}} \leq 30\%$ and $V_{82\text{Gy}} \leq 5\%$ for rectal wall, $D_{\max} \leq 85\text{Gy}$ and $V_{75.6\text{Gy}} \leq 50\%$ for bladder.

Composite dose-distributions, of the combined initial plan and the adapted plan,

and composite DVHs were calculated on the planning CT using Pinnacle's heterogeneity corrected adaptive-convolve algorithm (Philips Pinnacle 3, Versions 6.2-9.0, Philips Healthcare, Andover, MA). The same algorithm was used for all patients; the plans were re-calculated and exported retrospectively using Pinnacle scripts. Contours were created in Pinnacle. All CT scans were taken with 3 mm slice-thickness. DVHs were exported for rectal wall in dose-bins of 0.1 Gy.

5.4.3 Follow-Up

The FU protocol recommended examinations every three months for the first two years after treatment, every six months until five years after treatment and yearly thereafter. The FU data was acquired prospectively by the treating physicians and documented using CTCAE v.2.0 and CTCAE v.3.0. It was reviewed and amended to the CTCAE v. 3.0 by a single physician in 2011 - 2012 to ensure consistency of the toxicity scoring.

5.4.4 Application of Matched Pair Analysis

In order to account for most of the major possible differences in patient characteristics between the 3D-CRT and IMRT groups, a matched-pair analysis was performed on the datasets. Patients were investigated regarding several suspected confounding variables: age at diagnosis, diabetes mellitus (DM) and pelvic surgery prior to radiation therapy (RT), use of androgen deprivation therapy (ADT) prior to or during RT, rectal wall volume, PTV volume and National Comprehensive Cancer Network (NCCN) risk-groups. Data on smoking, hemorrhoids, abdominal surgery and hypertension was not available for most patients and was not included. The 3D-CRT and IMRT groups were matched based on the characteristics with low p-values ($p < 0.1$). Those characteristics included age, rectal wall volume, DM prior to treatment, and risk group. Greedy matching using a hybrid of exact matching (for the risk group) and propensity score matching (all other patient characteristics) was used to match the groups (see 5.3). Matching was implemented in R 2.15.1 (R-project.org, Vienna, Austria) using the "Matching" package (Sekhon, 2008). The patient characteristics before and after matching are presented in Table 5.1. After matching, the groups comprised 275 and 550 patients with 79 (28.7 %) and 43 (7.8 %) toxicities, respectively, and did not show significant difference in rectal wall volume, risk-group, age and diabetes (Table 5.1). ADT showed a significant difference after matching that was of small magnitude (26 %/19 %). ADT has been associated primarily with abdominal pain and diarrhea (Kumar *et al.*, 2005). The former was not part of this study, the latter occurred in less than 2% and 1% of all 3D-CRT and IMRT patients, respectively, and thus the matching was not repeated in favor of a larger number of patients.

Characteristic	3D-CRT	IMRT	p-value	3D-CRT matched	IMRT matched	p-value matched
Number of patients	457	658		275	550	
Time of treatment	1999-2006	2001-2010		1999-2006	2001-2010	
Median RxD, range	75.6 Gy 70.2-79.2 Gy	75.6 Gy 70.2-82.8 Gy	< 0.01	75.6 Gy 70.2-79.2 Gy	75.6 Gy 70.2-82.8 Gy	<0.001
Median FU, range	8.9 yrs 2.0-12.5 yrs	4.6 yrs 2.0-9.0 yrs	<0.001	8.8 yrs 2.0-12.5 yrs	4.7 yrs 2.0-9.0 yrs	< 0.001
GI-toxicity, grade ≥ 2	127 (27.8 %)	48 (7.3 %)	<0.001	79 (28.7 %)	43 (7.8 %)	<0.001
Age at Dx, range	72 yrs 51-87 yrs	70 yrs 45-86 yrs	< 0.001	71 yrs 51-87 yrs	71 yrs 45-86 yrs	0.46
Diabetes	67 (14.7%)	109/571 (19.1%)	0.06	41/275 (15%)	77/477 (16 %)	0.73
Past pelvic surgery	161/455 (35.4%)	134/352 (38.1 %)	0.48	89/274 (32%)	120/305 (39%)	0.1
Pre/During Tx ADT	106 (23.2 %)	131 (19.9 %)	0.21	72 (26%)	102 (19%)	0.02
Risk-Group†:			<0.001			1
low	230 (50.3%)	228 (34.7%)		105	210	
intermediate	168 (36.8%)	313 (47.6%)		129	258	
high	59 (12.9%)	117 (17.8%)		41	82	
Rectal wall volume	31.3 cm^3 $\pm$ 12.1 cm^3	24.7 cm^3 $\pm$ 7.5 cm^3	<0.001	25.7 cm^3 $\pm$ 7.6 cm^3	25.5 cm^3 $\pm$ 6.9	0.7
PTV volume	140.2 cm^3 $\pm$ 52.7 cm^3	140.3 cm^3 $\pm$ 50.4 cm^3	0.9	136.0 cm^3 $\pm$ 52.4 cm^3	142.0 cm^3 $\pm$ 51.3 cm^3	0.12

† Risk-groups are defined as low if PSA < 10, Gleason score ≤ 6 and T-stage $\leq$ T2a, intermediate if PSA < 20, Gleason ≤ 7 and T-stage $\leq$ T2b and high otherwise.

Table 5.1: Patient characteristics and follow-up (FU) time for three dimensional conformal radiation therapy (3D-CRT) and intensity modulated radiation therapy (IMRT), and p-values for difference between the two groups before and after matching.

5.4.5 Normal Tissue Complication Probability Models and Statistics

For this study, the Lyman equivalent uniform dose (EUD) model (Equation 4.9) and the cutoff dose logistic regression (CDLR) model (Equation 4.7) described in chapter 4 were used to model NTCP for matched 3D-CRT and IMRT patients and the combined group. The endpoint of the study were GI toxicities grade ≥ 2. The models were based on DVHs of the rectal wall contour. Additionally, NTCP for GI toxicities was modelled using full rectum DVHs. Furthermore, optimal NTCP model parameters were calculated for a subgroup of all IMRT patients that were treated before CBCTs were used (pre-CBCT).

The model parameters and standard-errors were determined by means of MLE

Lyman EUD model	a	EUD_{50}	m	LL
3D-CRT	4.76 (0.50, 15.64)	69.0 Gy (60.1Gy, 77.2Gy)	0.182(0.076, 0.639)	-157.4
IMRT	12.27 (1.50, ∞)	87.9 Gy (77.3Gy, 160Gy)	0.159 (0.071, 0.542)	-148.9
3D-CRT+IMRT	11.50 (4.01, 21.6)	75.5 Gy (71.7Gy, 79.1Gy)	0.089 (0.081, 0.186)	-324.6
QUANTEC	11.11 (7.14, 25.00)	76.9 Gy (73.7Gy, 80.1Gy)	0.13 (0.10, 0.17)	-
CDLR model	D_c	β_0	β_1	LL
3D-CRT	53.9 Gy (44.1Gy, 73.8Gy)	-2.54 (-3.49, -1.64)	0.037 (0.018, 0.057)	-157.6
IMRT	77.2 Gy (44.3Gy, 78.6Gy)	-2.89 (-3.38, -2.44)	0.049 (0.012, 0.085)	-147.9
3D-CRT+IMRT	38.8 Gy (24.9Gy, 46.1Gy)	-4.19 (-4.85, -3.62)	0.039 (0.030, 0.047)	-325.6

Table 5.2: Optimal normal tissue complication probability (NTCP) model parameters for the matched groups and their 95 % confidence intervals for the Lyman-EUD-model and the cutoff dose logistic regression (CDLR) model of three dimensional conformal radiation therapy (3D-CRT), intensity modulated radiation therpay (IMRT), and 3D-CRT+IMRT data for chronic gastrointestinal (GI) toxicities grade ≤ 2. Optimal parameters determined by QUANTEC (Michalski *et al.*, 2010) are shown for comparison.

(Collett, 2003; Cox, 1989; Jackson *et al.*, 1995), using rectal wall DVHs and their corresponding dichotomous toxicity data. The confidence intervals (CIs) for the model parameters were calculated using profile likelihoods. Goodness of fit of a particular model was determined using a Hosmer-Lemeshow goodness of fit test (Collett, 2003; Hosmer and Lemesbow, 1980). MLE and goodness of fit were performed using Mathematica (Wolfram Research Inc., Version 9.0, Champaign, IL) and an in-house developed software (C++).

5.5 Modelling Results

The mean DVHs, a typical dose distribution, and rectal wall DVHs of all patients are shown in Figure 5.1. The 3D-CRT and IMRT [2] DVHs contained significantly different volumes in each dose-bin, except for 5-6 Gy, and 83 Gy (Mann-Whitney, 1 Gy bins).

5.5.1 Normal Tissue Complication Probability Model Parameters

Rectal Wall

The model parameters for the Lyman EUD and the CDLR model for GI toxicities using the rectal wall are summarized in Table 5.2. According to the goodness of fit criteria, all models were satisfactory fits to the data from which they were derived ($p > 0.05$). When comparing models to the data they were not derived from, they yielded unsatisfactory fits ($p < 0.001$). NTCP curves derived from the 3D-CRT+IMRT dataset or the conformal data alone overestimated the NTCP for IMRT patients (Figure 5.2).

[2] All further reference to 3D-CRT/IMRT will allude to the matched groups.

In the same manner, curves derived from IMRT data or the combined dataset underestimated NTCP for 3D-CRT patients. Additionally, it was evaluated how well the data of each group fit the Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) recommended Lyman EUD NTCP model parameters (Michalski *et al.*, 2010). The QUANTEC model fit none of the datasets well. The p-values for 6 bins containing approximately the same total number of patients are shown in Table 5.3, but other groupings (3, 10, 20) showed similar statistical significance. For the CDLR model, the cutoff dose D_c yielding the best log likelihood is shown in Table 5.2. Significant log likelihood values could be obtained for a range of cutoff doses around those optimal values (Figure 5.3). The exact position of the optimal D_c varied within that range. The average difference in volume over all dose bins between the 3D-CRT and the IMRT DVHs was 5 % (0-17 %). The difference in mean EUD calculated for all 3D-CRT and IMRT DVHs for any a-value in the range of 6 - 20 was < 2 Gy.

Full Rectum

The optimal NTCP model parameters for NTCP modelling using the full rectum contours are shown in Table 5.4. The goodness of fit p-value shows that the model fits the data well (Table 5.5).

Pre-CBCT IMRT Patients

Additionally, the results for NTCP modelling using only the rectal wall DVHs of pre-CBCT IMRT patients are shown in Table 5.6 (optimal model parameters) and Table 5.5 (goodness of fit values). The model fits the data well, but does not fit the 3D-CRT data.

5.6 Discussion

NTCP was modelled with respect to GI toxicities for patients treated with 3D-CRT or IMRT and the estimated model parameters were compared. Matched-pair analysis was performed on the data in order to select sub-groups with comparable patient characteristics. Overall the difference between planned IMRT and 3D-CRT DVHs was relatively small. However, both before and after matching, the raw toxicities showed sizeable and significant ($p < 0.001$) differences between the 3D-CRT (27.8 % / 28.7 %) and the IMRT group (7.3 % / 7.8 %), which cannot be explained by those differences in dose alone. This creates the paradox situation that IMRT, conceived to reduce toxicities by sparing normal tissues from high doses, is successful in the former, but not by achieving the latter. Consequently, the purely dose-based NTCP modelling parameters derived from the 3D-CRT data do not describe NTCP for IMRT patients and other reasons for the discrepancies in toxicity need to be considered. In the following, other possible sources of bias between the groups, as causes of the nearly 4-fold decrease in IMRT toxicities, are discussed.

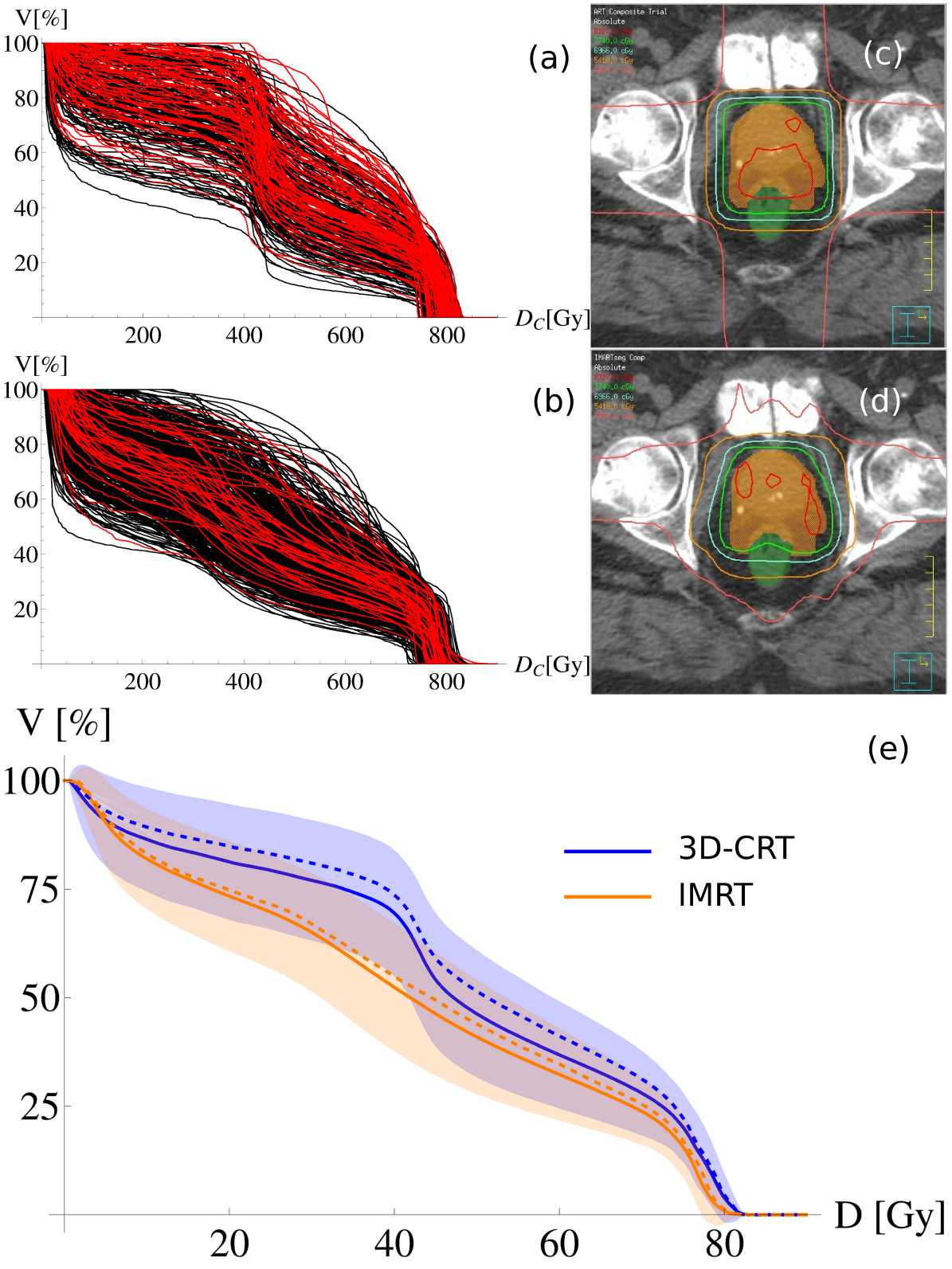


Figure 5.1: Left: (a) All three dimensional conformal (3D-CRT) and (b) intensity modulated radiation therapy (IMRT) dose volume histograms (DVHs), with patients experiencing gastrointestinal toxicities grade ≥ 2 marked red. Right: Typical 3D-CRT (c) and IMRT (d) dose-distribution. The red, green, blue, orange, light-red lines are the 81.3 Gy, 77.4 Gy,63.7 Gy, 51.2 Gy, 38.7 Gy iso-doses, respectively. Bottom: Mean DVH and standard deviation of all patients (solid/filled space), and patients with toxicities (dashed).

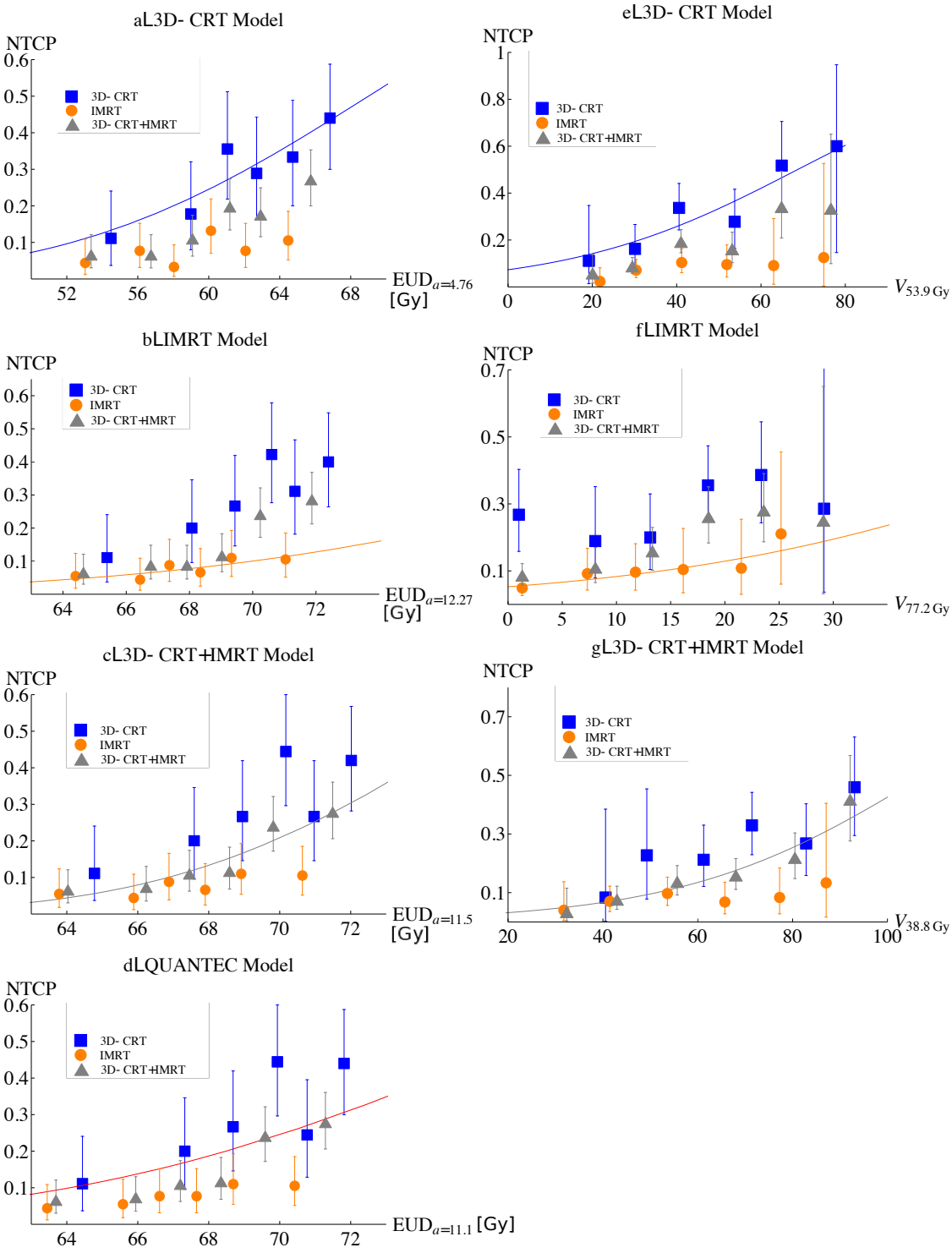


Figure 5.2: Normal tissue complication probability (NTCP) curves and data for the Lyman EUD model (a–d) and the cutoff dose logistic regression (CDLR) model (e–h). The blue, orange, and grey lines are the NTCP curves derived from 3D-CRT, IMRT, and 3D-CRT+IMRT, respectively. The squares, circles, and triangles are the binned 3D-CRT, IMRT, and 3D-CRT+IMRT data, respectively.

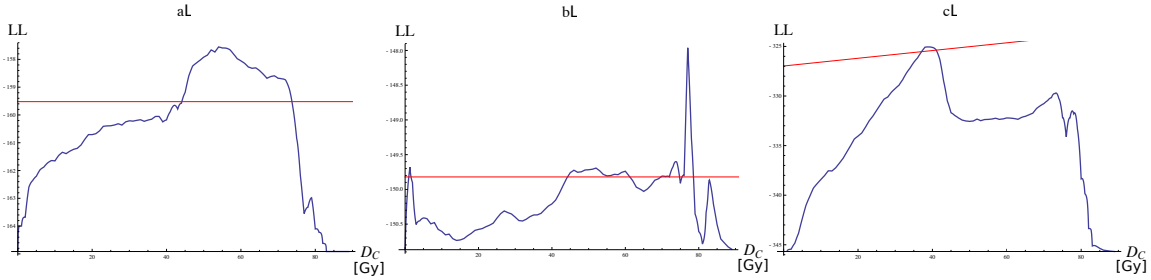


Figure 5.3: Optimal log-likelihood (LL) of the cutoff dose logistic regression (CDLR) model in dependence on the cutoff dose for (a) three dimensional conformal radiation therapy (3D-CRT) data, (b) intensity modulated radiation therapy (IMRT) data, and (c) 3D-CRT+IMRT data (steps of 1 Gy). The red line indicates the 95 % significance level.

3D-CRT Lyman-EUD Model	**pHL†**	**CRT cutoff-dose model**	**pHL†**
3D-CRT	0.75	3D-CRT	0.53
IMRT	<0.001	IMRT	< 0.001
3D-CRT+IMRT	< 0.001	3D-CRT+IMRT	< 0.001
IMRT Lyman-EUD model		**IMRT cutoff-dose model**	
3D-CRT	< 0.001	3D-CRT	< 0.001
IMRT	0.72	IMRT	0.1
3D-CRT+IMRT	< 0.001	3D-CRT+IMRT	< 0.001
3D-CRT+IMRT Lyman EUD model		**3D-CRT+IMRT cutoff dose model**	
3D-CRT	< 0.001	3D-CRT	< 0.001
IMRT	< 0.001	IMRT	<0.001
3D-CRT+IMRT	0.52	3D-CRT+IMRT	0.81
QUANTEC model			
3D-CRT	<0.01		
IMRT	<0.01		
3D-CRT+IMRT	<0.01		

†The model is considered to yield an acceptable fit to the data, if the goodness of fit value $p > 0.05$, (Collett, 2003).

Table 5.3: Hosmer-Lemeshow goodness of fit p-values for matched data.

Technique	Number of patients	GI toxicities grade ≥ 2	a	$\mathbf{EUD_{50}}$ (Gy)	m
3D-CRT	146	39 [26.7 %]	4.43 (5.37)	68.9 (6.6)	0.175 (0.135)
IMRT	546	43 [7.8 %]	15.33 (25.3)	93.3 (17.9)	0.186 (0.161)

Table 5.4: Optimal normal tissue complication probability (NTCP) parameters for the Lyman EUD model using full rectum dose volume histograms (DVHs). Standard errors for the parameters are shown in parenthesis

3D-CRT data (rectum)	p_{HL}	IMRT data (rectum)	p_{HL}	IMRT data (rectal wall, pre-CBCT)	p_{HL}
3D-CRT model (rectum)	0.11	IMRT model (rectum)	0.97	IMRT model (rectal wall, pre-CBCT)	0.39
3D-CRT model (rectal wall)	0.11	IMRT model (rectal wall)	0.80	IMRT model (rectal wall)	0.38
IMRT model (rectum)	< 0.001	3D-CRT model (rectum)	< 0.001	3D-CRT model (rectal wall)	< 0.001

Table 5.5: Goodness of fit values for the full rectum Lyman EUD model for three dimensional conformal radiation therapy (3D-CRT) and intensity modulated radiation therapy (IMRT) data, and the pre-CBCT IMRT data.

Technique	Number of patients	GI toxicities grade ≥ 2	a	$\mathbf{EUD_{50}}$ (Gy)	m
IMRT pre-CBCT	380	35 [9.2 %]	15.1 (15.9)	83.3 (7.3)	0.126 (0.07)

Table 5.6: Optimal Lyman EUD normal tissue complication probability (NTCP) model parameters for intensity modulated radiation therapy (IMRT) patients treated before cone beam computed tomography scans (CBCTs) were incorporated into the clinical routine (pre-CBCT). Standard errors for the parameters are shown in parenthesis.

5.6.1 Patient Characteristics

Many established patient characteristics related to GI toxicities (Budäus *et al.*, 2012) were accounted for using matched-pair analysis, or showed no significant difference between groups prior to matching. However, it is not feasible to account for all possible sources of bias. Partially this is because not all characteristics are available for every patient. Furthermore accounting for a larger number of characteristics would reduce the number of patients/toxicities available for modelling to an insufficient amount. Nevertheless, it is possible to roughly estimate the influence of various characteristics on toxicities. Among the characteristics that have been found to correlate with GI toxicities in multivariate studies with dose-volume parameters are pre-RT abdominal surgery (odds ratio (OR) = 4.7), hemorrhoids (OR = 2.6), ADT (OR = 0.6) and antihypertensive medication (OR = 0.38) (Kumar *et al.*, 2005; Fellin *et al.*, 2013). Assuming a hypothetical 5 % probability that a patient without abdominal surgery experiences a toxicity (less than the 7 % incidence in IMRT patients), it can be approximated that for the observed toxicity rates, 19 % of the IMRT patients in this study would have had abdominal surgery prior to RT, versus 160 % of the 3D-CRT patients. Although this is a hypothetical thought experiment and provides only a rough estimate, the unrealistic outcome suggests that even the most significant single patient characteristic of abdominal surgery is unlikely to cause the high discrepancy of toxicities between the two groups. Even a combination of these patient characteristics could not lead to the observed difference, as it is unlikely that they all differed to such a marked extent.

5.6.2 Image Guidance

Inaccurate estimation of delivered dose may affect modelling 3D-CRT and IMRT data. The treatment modalities have different dose-distributions, which could result in a different error of planned-versus-delivered dose in the two techniques (Schulze *et al.*, 2009), hindering the development of a common NTCP model. IMRT has gradually replaced 3D-CRT for prostate cancer since 2006 at Beaumont Health. During that time, not only have the treatment-modalities changed, but so has the use, extent and modality of image guidance. For example, patient positioning and mobility control have improved by replacing PIs with CBCTs. It is conceivable that the decreased proportion of IMRT patients experiencing toxicity is due to more accurate delivery of dose. However, both PI as well as CBCTs were registered to the planning CT using bony anatomy only, not soft tissue. Guckenberger *et. al* (Guckenberger *et al.*, 2006) have demonstrated little difference between PI and CBCT in that case. Thus the improvement due to positioning is likely only slight. Furthermore, with the advent of CBCT it became possible to compare bladder and rectum volumes to the planning CT before each fraction and adjust dietary instructions or administer fluid intake, if necessary. However, there was no official clinical regulation protocol and adjustments were performed only in extreme cases ($<$ 10 % of cases according to the Beaumont Health imaging specialist). In addition, 71 % of IMRT patients in this study were treated before CBCT was introduced clinically, which also excludes the adjustment of ART margins as a major

contribution.

Image guidance was further excluded as a source of bias, by the brief analysis that was performed using the pre-CBCT IMRT patients only. The optimal NTCP parameters (Table 5.6) did not change enough to satisfactorily explain discrepancies between 3D-CRT and IMRT model parameters with changes in image guidance. Goodness of fit p-values (Table 5.5) show that the NTCP model derived from pre-CBCT IMRT data is not applicable to 3D-CRT data.

5.6.3 Dose Volume Histograms

Thor *et al.* investigated the influence of daily organ motion on DVHs and subsequently on rectal toxicity (Thor *et al.*, 2013). They estimate the worst-case increase in the volume receiving a given dose between the planned and delivered rectal DVH for a given patient at approximately 25 %. If all 3D-CRT patients had such a worst-case increase in volume due to organ motion, whereas none of the IMRT patients did, this would result in a right-ward shift of the 3D-CRT NTCP curve shown in Figure 5.1(e) of 25 %. However, the required shift by goodness of fit is 35 %. It is unlikely that such a drastic systematic error would have been present in one set, but not the other set of patients.

5.6.4 Toxicity Scoring

As mentioned earlier in this chapter (see 5.2.1), it is difficult to exclude toxicity scoring as a source of bias. Double blind studies in which neither the patient nor the physician performing the FU know the treatment technique are not only impractical, but also ethically questionable. The patients in this study were generally seen by different resident physicians along with the treating physician at each follow-up visit, with an expected 4 follow-up visits during the first 2 years after RT.

5.6.5 Follow-Up Time

The difference in follow-up time of the groups is also notable. However, Kaplan-Meier curves for rectal toxicities in prostate patients show that approximately 75-80 % of complications typically occur within the first 2 years after the end of treatment (Mohammed *et al.*, 2012). Therefore, with a minimum follow-up of 2 years and a median of 4.7 years the FU time of the IMRT group, FU time is not expected to be a major source of bias between groups.

5.6.6 Limitations of Normal Tissue Complication Probability Models Used

One factor may be that the DVH leaves out spatial information contained in the three dimensional (3D) dose distribution that has been suggested to be relevant by some authors (Acosta *et al.*, 2013; Munbodh *et al.*, 2008; Peeters, Lebesque, Heemsbergen, van Putten, Slot, Dielwart and Koper, 2006) and has been shown to correlate with late

complications (Buettner *et al.*, 2009). Therefore, the two DVH based models used in this study may not be capable of correctly predicting side-effects for both 3D-CRT and IMRT with similar parameters. Furthermore, the NTCP models used in this study, especially the CDLR model, may show correlations between specific dose parameters and toxicities, where there is no actual causal connection (Söhn, Alber and Yan, 2007). The results of the cutoff dose analysis are influenced by the typical dose distributions in the patient groups and vary immensely between populations and the combination of the two populations (Figure 5.3). This suggests that the model in general may not be adequate in describing true causalities between toxicities and dose. Finally, as patients are treated within narrow dose-ranges and constraints, the data only allows modelling of a small section of the entire NTCP curve, inducing increased uncertainty in model parameters. The few toxicity events further impede NTCP modelling with respect to uncertainties of parameter estimates, especially in the IMRT population (Table 5.2).

5.6.7 Comparison with Other Studies

Lyman EUD Model

The NTCP parameter values derived for a, EUD_{50} and m from the 3D-CRT dataset for the Lyman EUD model using rectal wall were within the range of parameters found in the literature for rectal toxicities (Marks *et al.*, 2010; Söhn, Yan, Liang, Meldolesi, Vargas and Alber, 2007; Rancati *et al.*, 2004; Gulliford *et al.*, 2012; Peeters, Hoogeman, Heemsbergen, Hart, Koper and Lebesque, 2006; Fiorino *et al.*, 2009; Tucker *et al.*, 2004).

The QUANTEC study (Michalski *et al.*, 2010) determined optimal NTCP parameters for GI toxicities via a meta-analysis of several publications. The optimal parameters determined by QUANTEC did not fit the 3D-CRT data in this study, however, the QUANTEC parameters were within the CIs determined in this study.

Since the QUANTEC report and other literature include studies based on full rectum DVHs, NTCP modelling and analysis in this study was repeated using the entire rectum for all matched patients who had the full rectum contour available. The optimal NTCP models (Table 5.4) in this case differ little from those obtained with rectal wall. Thus discrepancies between the QUANTEC parameters and the parameters in this study are not due to using rectal wall rather than the full rectum. Furthermore, the models derived for full rectum contours were not cross-applicable between treatment techniques, which also excludes using rectal wall contours as a source of discrepancies between 3D-CRT and IMRT (Table 5.5).

Some publications for 3D-CRT modelled NTCP for rectal bleeding only, rather than all GI toxicities. For comparison, a smaller study for the Lyman EUD model, considering only rectal bleeding grade ≥ 2, was conducted previously for a sub-population of this study and resulted in similar conclusions regarding the discrepancy found between 3D-CRT and IMRT groups (Troeller *et al.*, 2012). Söhn *et al.* previously modelled a sub-group of the 3D-CRT population considered in this study with respect to rectal bleeding (Söhn, Yan, Liang, Meldolesi, Vargas and Alber, 2007) and obtained a = 11.9

$\pm$ 3.8, m = 0.108 $\pm$ 0.027, and EUD_{50} = 78.4 $\pm$ 2.1. The difference in parameters compared to this study may be due to the smaller patient population with considerably shorter follow-up of 1 year, and the more specific endpoint.

This study provides model parameters for IMRT patients for the Lyman EUD model. To the knowledge of the author, no other publications have modelled NTCP for IMRT patients at this point, likely due to the decreased incidence of toxicities in IMRT patients and the difficulty obtaining IMRT data. In general, optimal NTCP parameters vary significantly between studies and populations, and the parameters determined in this study are not meant to present a recommendation for treatment. On the contrary, they show that modelling results depend highly on the exact group of patients and treatment-modalities used.

CDLR Model

Results for the CDLR model confirm previous publications stating that volumes receiving doses of 40 Gy or higher are associated with GI toxicities (Söhn, Yan, Liang, Meldolesi, Vargas and Alber, 2007; Peeters, Hoogeman, Heemsbergen, Hart, Koper and Lebesque, 2006; Fiorino *et al.*, 2009; Tucker *et al.*, 2004; Jackson *et al.*, 2001). This study also provides model parameters for IMRT patients for the CDLR model. However, the best-fit NTCP curves are relatively flat and thus do not carry great predictive value. For example, 12 % of IMRT patients did not receive doses as high as the optimal cutoff dose of 77.2 Gy.

5.7 Conclusion

In this large single-institution cohort study, discrepancies between the incidence in GI toxicities in 3D-CRT versus IMRT patients were investigated, with special emphasis on NTCP models and their cross-applicability for predicting GI toxicity. The 3D-CRT derived optimal NTCP parameters fit poorly when applied to IMRT data, and vice versa. The reasons for the large discrepancies between toxicity incidences remain unclear. Matching the populations with respect to the most important clinical parameters did not resolve the differences that were found. Although, image guidance can further affect the extent of GI toxicity, these are unlikely to explain the difference between the two patient groups. Biases of toxicity scoring cannot be evaluated retrospectively. Limitations in the class of models explored in this study may result in the observed lack of fit, but would have to be further investigated and it is possible that other types of models could resolve the issue. However, NTCP models derived from 3D-CRT data are used at present in clinical practice. Given the limitations of this exploratory study, such models would likely not fit IMRT data well. Therefore, such 3D-CRT derived models should only be applied to IMRT patients with awareness of the issues arising from extrapolating models to other populations. Fortunately, the 3D-CRT based models overestimate toxicities for IMRT patients and do not lead to additional toxicities. They may however prevent dose escalation to increase tumor control, as they predict more toxicities than may actually be experienced by the patients.

Chapter 6

Influence of Fractionation

6.1 Introduction

As discussed in Chapter 3, Section 3.4.2, radiation therapy is usually delivered in several fractions rather than one single dose in order to allow normal tissues to recover between treatment fractions. Radiobiologically, fractionation is based on the linear quadratic (LQ) model of cell kill, and the difference in $\frac{\alpha}{\beta}$ ratios for tumors ($\geq$ 10 Gy) and normal tissue ($\simeq$ 3 Gy) that reflect the tumors' decreased ability to repair compared to healthy tissues. Standard fractionated radiation therapy (SFRT) is given in fractions of 1.8 Gy to 2 Gy. Various studies have investigated normal tissue complication probability (NTCP) models for SFRT for a multitude of structures and endpoints. Hypofractionated radiation therapy (HFRT), that is radiation therapy with more than 2 Gy per fraction, has been introduced into clinical practice more recently. Consequently, less studies on NTCP models for HFRT exist. This chapter investigates the impact of hypofractionation on NTCP models.

Since dose fractionation impacts the biological effect of physical dose on tissues, altered fractionation schemes affect NTCP models. Firstly, the fractionation scheme used can change a tissue's dose response to the physical dose delivered and consequently the incidence and severity of toxicities. Although fractionation can be accounted for by applying the LQ model, this model cannot be extrapolated to very low or high doses per fraction (see Section 6.2.2). Thus alternative models accounting for fractionation may have to be used. Also, HFRT treatments have different dose distributions. NTCP models derived from SFRT data cannot necessarily be extrapolated to HFRT and thus new NTCP models for HFRT need to be investigated. Secondly, depending on the fractionation scheme used, the type of toxicities observed may change. In the following, both of these effects will be investigated using the data of patients treated for lung cancer with stereotactic body radiation therapy (SBRT), which is a common form of HFRT (see 3.4.2).

In section 6.2, NTCP is modelled for pneumonitis after lung SBRT. Two different models accounting for fractionation effects were used: The classic LQ model and an alternative model, the low dose hyper-radiosensitivity (LDHRS) model. Model parameters are compared to SFRT and biological rationale and implications of the LDHRS

model are discussed.

While radiation pneumonitis is a major side effect during lung SFRT, chest wall and extrapulmonary soft tissue complications can emerge during HFRT. In section 6.3, NTCP models are created for toxicities related to the chest wall, including chest wall pain, myositis, and rib fractures.

6.2 Normal Tissue Complication Probability Modelling of Radiation Pneumonitis for Hypofractionated Radiation Therapy

Publication: Normal Tissue Complication Probability Modeling of radiation pneumonitis following hypo-fractionated radiotherapy [1]

6.2.1 Background

Radiation pneumonitis (RP) is a severe complication of radiotherapy of the lungs that can significantly impact a patient's quality of life or even have fatal consequences (Billiet *et al.*, 2014). Consequently, prediction and NTCP modelling of RP is important to ensure optimal treatment of lung cancer and other sites involving radiation to the lungs. Several studies have modelled NTCP of RP in SFRT of lung (Seppenwoolde *et al.*, 2003; Sonke *et al.*, 2011; Tucker *et al.*, 2008). In recent years, SBRT has been increasingly used for the treatment of early stage lung tumors, because of the increased biological effectiveness of large doses per fraction in tumor control (Beli *et al.*, 2010).

However, there is a concern that lung SBRT will lead to increased toxicity in normal tissue, although improved targeting may reduce RP (Ricardi *et al.*, 2015). A meta-analysis by Brock *et al.* showed symptomatic RP grade $\geq$2 ranging from 0 % - 29 % (6.5 % on average) in hypo-fractionated patients (Brock *et al.*, 2008). Although Brock *et al.* conclude that toxicities of SFRT and SBRT are roughly consistent, a just comparison would require that biologically effective dose (BED) to the lungs is taken into consideration. Since BED in the lungs can be drastically different in SBRT patients, it is also not certain that NTCP models derived from SFRT data can be extrapolated to SBRT. Studies that model NTCP for SBRT of lung are scarcer than those for SFRT, presumably due to SBRT being introduced into common clinical practice more recently. Borst *et al.* and Sonke *et al.* have modelled NTCP for hypofractionated data (Sonke *et al.*, 2011; Borst *et al.*, 2009). Both studies have modelled hypofractionated data based on the mean lung dose and have corrected dose for fractionation effects using the LQ model.

[1] Submitted to the International Journal of Radiation Oncology, Biology, Physics

6.2.2 Induced Repair and the Low Dose Hyperradiosensitivity Model

The LQ model described in section 3.4.2 is currently the standard model used to correct dose for fractionation effects. However, it has been shown that the LQ model is not accurate at very high and very low doses per fraction (Andisheh *et al.*, 2013; Hamilton *et al.*, 1996). The LDHRS model has been proposed as an alternative to the LQ model by Singh *et al.* to account for the effects of low doses per fraction (Singh *et al.*, 1994). The LDHRS model generally follows the assumptions of the LQ model, but additionally assumes that cell repair is not properly triggered at low doses per fraction ($\lessapprox 0.5$ Gy) which leads to increased damage at these doses. BED in the LDHRS model is calculated as

$$\mathrm{BED}_{\mathrm{LDHRS}} = nd\left(1 + \left(\frac{\alpha_s}{\alpha} - 1\right)\exp\left(-\frac{d}{d_c}\right) + \frac{d}{\frac{\alpha}{\beta}}\right), \tag{6.1}$$

where α, β are the coefficients of the linear and quadratic components of cell kill as defined in the LQ model (see equation 3.7), and n and d denote the number of fractions and dose per fraction, respectively. The parameter α_s determines the sensitivity to radiation in the low-dose per fraction regime, and d_c is the threshold dose for induced repair. Note that the LDHRS model converges to the LQ model for $\alpha_s \to \alpha$. Figure 6.1 shows the surviving fraction SF for the LDHRS model and the LQ model. As shown, the surviving fraction SF predicted by the LDHRS model is lower than predicted by the LQ model for low doses per fraction.

Similarly to the LQ model, normalized total dose (NTD) can also be defined for the LDHRS model

$$\mathrm{NTD}_{2,\mathrm{LDHRS}} = nd\left(\frac{1 + \left(\frac{\alpha_s}{\alpha} - 1\right)\exp(-\frac{d}{d_c}) + \frac{d}{\frac{\alpha}{\beta}}}{1 + \left(\frac{\alpha_s}{\alpha} - 1\right)\exp(-\frac{2}{d_c}) + \frac{2}{\frac{\alpha}{\beta}}}\right). \tag{6.2}$$

Hypersensitivity and induced repair has been shown to exist in vitro, mainly in tumor cells (Joiner *et al.*, 1996), but there is also evidence that the effect exists in human lung cell lines (Singh *et al.*, 1994). Hypersensitivity may exist in vivo, as was demonstrated for mouse skin (Joiner *et al.*, 1986).

6.2.3 Methods and Materials

This study compares Lyman and logistic regression NTCP models for predicting RP in hypo-fractionated lung data, including models based on mean lung dose (MLD), equivalent uniform dose (EUD) and cutoff dose volumes (V_{D_c}). Both the LQ model and the LDHRS model are used to correct for fractionation effects.

Patient Data

Lung dose volume histograms (DVHs) and radiation pneumonitis toxicity scores of 377 patients from 5 institutions participating in the Elekta lung research consortium treated between 2005 and 2011 were investigated in this study. All patients were

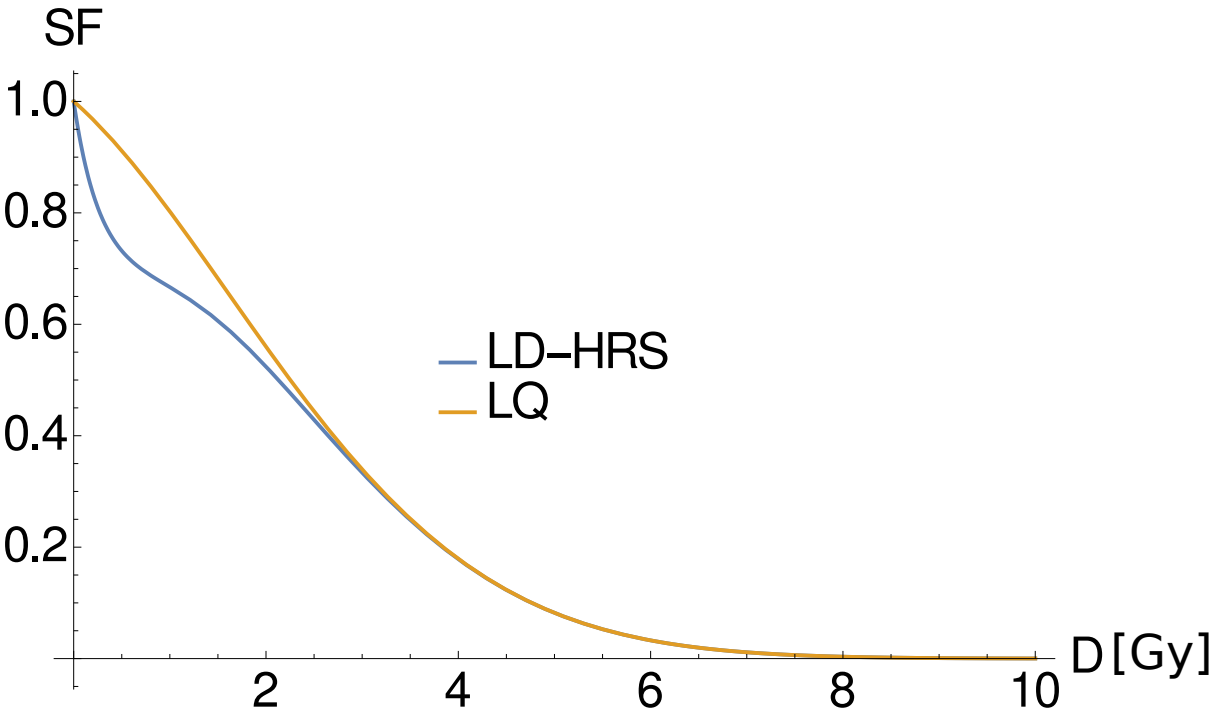


Figure 6.1: Surviving fraction SF after dose D, as predicted by the linear quadratic (LQ) model (orange) and the low dose hyperradiosensitivity (LDHRS) model (blue). The parameters used were α = 0.15 Gy^{-1}, β = 0.07 Gy^{-2}, α_s = 1.19 Gy^{-1}, and d_c = 0.58 Gy (Singh *et al.*, 1994).

treated with SBRT treatment schemes in 3-10 fractions with between 5 to 18 Gy per fraction up to a maximum total physical dose of 64 Gy (Table 6.2). The median prescription dose was 54 Gy and the median number of fractions was 3. Either three dimensional conformal radiation therapy (3D-CRT) or intensity modulated radiation therapy (IMRT) plans were used for treatment with 5-11 beams. Patients who had multiple treatments were excluded. For all patients, the gross tumor volume (GTV) was contoured based on planning computed tomography (CT) scans which were either average CTs, mid ventilation CTs or four dimensional computed tomography (4D-CT) scans. Internal target volumes (ITVs) were defined either by averaging the GTV contours from all 10 phases of a 4D-CT, combining maximum inspiration and expiration GTV contours, or as an average GTV on midventilation CTs. Three institutions did not extend the ITV to a clinical target volume (CTV), while in the remaining 2 institutions ITV to CTV margins of 4-7 mm were used. Planning target volume (PTV) margins were isotropic 5 mm in all institutions. All dose distributions were calculated with Pinnacle's collapsed cone algorithm (Philips Pinnacle^3, Philips Healthcare, Andover, MA). The lung DVHs were determined based on the contour of both lungs excluding the GTV. Radiation pneumonitis was scored based on the Common Terminology Criteria for Adverse Events (CTCAE) version 3.0. The endpoint of this study was radiation pneumonitis grade 2 and higher. The follow-up (FU) schedule recommended FU visits every 3 to 6 months during the first two years after treatment. Patients with a follow-up of less than 6 months were not included in this study.

Prescribed dose [Gy]	Number of fractions	Dose per fraction [Gy]	Number of patients
54	3	18	176
48	4	12	119
60	5	12	34
37.5	3	12.5	15
60	8	7.5	8
60	3	20	7
50	10	5	5
48	8	6	4
40	5	8	2
45	3	15	2
60	10	6	1
36	2	18	1
55	5	11	1
64	8	8	1
50	5	10	1

Table 6.1: Prescribed physical dose and fractionation.

Dose Volume Histograms

To account for the different fractionation schemes, all DVHs were converted to normalized total dose (NTD_2) for each dose bin using the LQ model (equation 3.8) and the LDHRS model (equation 6.2). The parameter d_c was fixed to 0.58 Gy for the LDHRS model. The parameters α and β were fixed to 0.15 Gy^{-1} and 0.07 Gy^{-2}, respectively for both the LQ and the LDHRS model, following parameters fitted to the LDHRS model by Singh *et al.* (Singh *et al.*, 1994).

Normal Tissue Complication Probability Modelling

NTCP was modeled using the Lyman model (Equation 4.9) and the cutoff dose logistic regression (CDLR) model (Equation 4.7). The Lyman model described in Chapter 4 was used with different summary measures X:

$$t = \frac{X - X_{50}}{mX_{50}},$$

where X was either $\mathrm{EUD}_{\mathrm{NTD,LQ}}$, $\mathrm{EUD}_{\mathrm{NTD,LDHRS}}$, $\mathrm{MLD}_{\mathrm{NTD,LQ}}$ or $\mathrm{MLD}_{\mathrm{NTD,LDHRS}}$. X_{50} denotes the respective X causing a 50 % chance of experiencing a complication. EUD was calculated using the generalized equivalent uniform dose (gEUD) using equation 4.6. The CDLR model was used with cutoff doses D_c converted to NTD (Söhn, Yan, Liang, Meldolesi, Vargas and Alber, 2007). Overall five different combinations of the two NTCP models with either the LQ model or the LDHRS model were investigated:

1. the Lyman model with the MLD calculated based on the LQ model (Lyman MLD LQ model),

2. the Lyman model with gEUD calculated based on the LQ model (Lyman EUD LQ model),

3. the Lyman model with the MLD based on the LDHRS model (Lyman MLD LDHRS model),

4. the CDLR model with V_{D_c} based on the LQ model (CDLR LQ model),

5. the CDLR model with V_{D_c} based on the LDHRS model (CDLR LDHRS model).

Optimal NTCP parameter estimates and confidence intervals (CIs) of the NTCP models for radiation pneumonitis were determined using maximum likelihood estimation (MLE) (section 4.5.1) and the profile likelihood method (section 4.8.2). Goodness of fit of data to the model was evaluated using the Hosmer-Lemeshow goodness of fit test (section 4.7). The NTCP models were compared with each other using the corrected Akaike information criterion (AICc) described by equation 4.18. The three nested Lyman models were further compared using the likelihood ratio test (section 4.6) to test if additional model parameters yield significant improvement. In the same manner, the two CDLR models were compared. All calculations were performed in Mathematica (Mathematica v. 9.0, Wolfram Research Inc., Champaign, IL).

6.2.4 Results

Patient Characteristics

28 (7.4%) of the patients investigated experienced pneumonitis of grade ≥ 2. Out of these 28 cases, 2 patients had pneumonitis grade 5, no patient had pneumonitis grade 4, 4 patients had pneumonitis grade 3, and 22 patients had pneumonitis grade 2. The median follow-up was 2.07 years. These data and other patient characteristics are summarized in Table 6.2.

Normal Tissue Complication Probability Models

The optimal NTCP model parameters are shown in Table 6.3. The Hosmer-Lemeshow goodness of fit values indicate that all five models yielded acceptable fits to the data (p_{HL} >0.05, 6 bins). The CDLR LDHRS model showed the best performance in terms of the log-likelihood compared to the Lyman EUD LQ, the Lyman MLD LQ model, Lyman MLD LDHRS model and the CDLR LQ model. When comparing the models with adjustments for the number of parameters used, the Lyman MLD LQ model shows the best AICc. A likelihood ratio test did not show a significant difference between the Lyman EUD LQ, the Lyman MLD LDHRS and the Lyman MLD LQ models. Furthermore, there was no significant difference between the nested CDLR models according to the likelihood ratio test. The optimal NTCP curves for all three Lyman models and binned data are shown in Figure 6.2. NTCP curves for the CDLR models are shown in Figure 6.3.

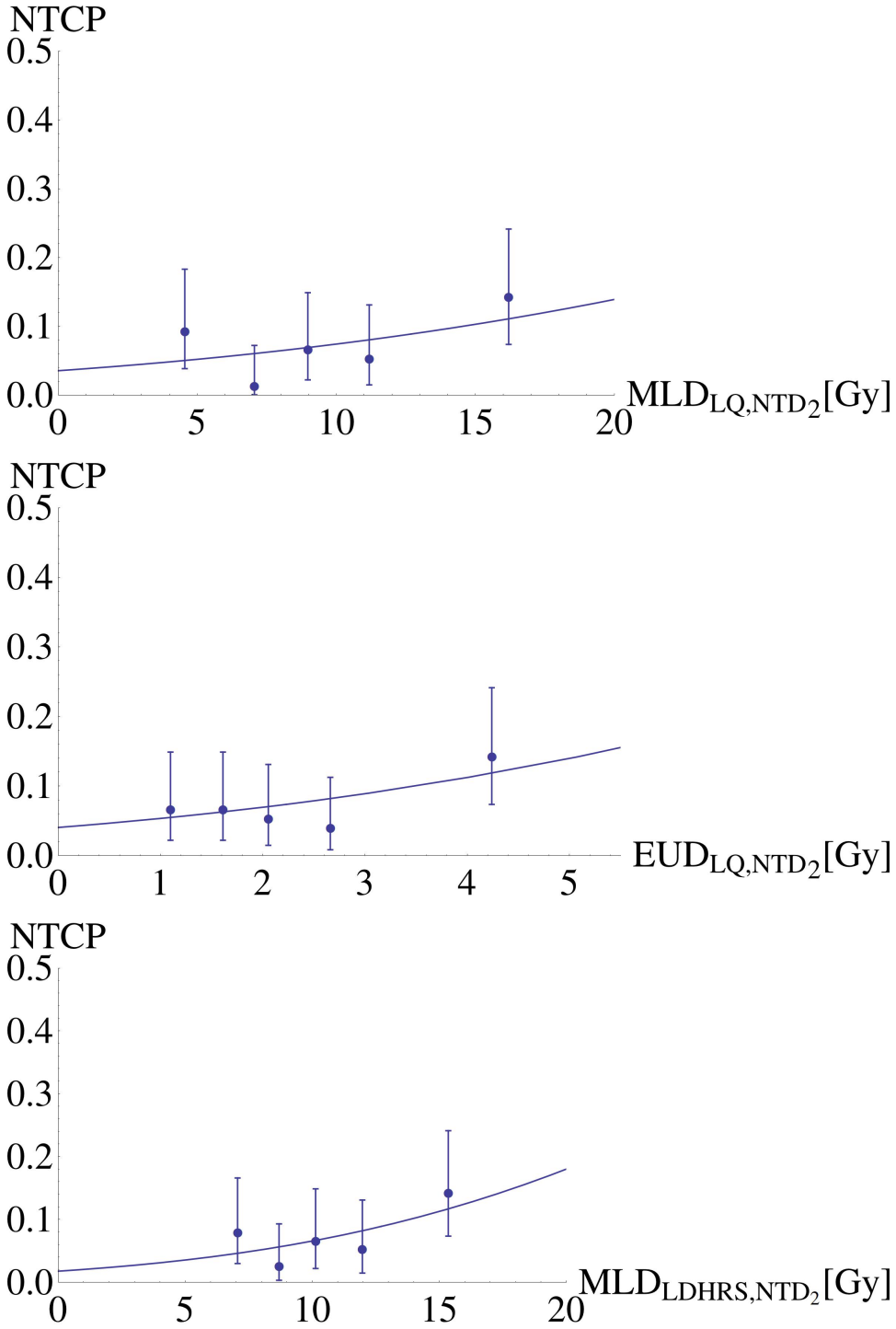


Figure 6.2: Optimal normal tissue complication probability (NTCP) curve and binned MLD/EUD data for the Lyman MLD LQ model (left), the Lyman EUD LQ model and the Lyman MLD LDHRS model. The error bars indicate the 95 % binomial confidence intervals from binning the data.

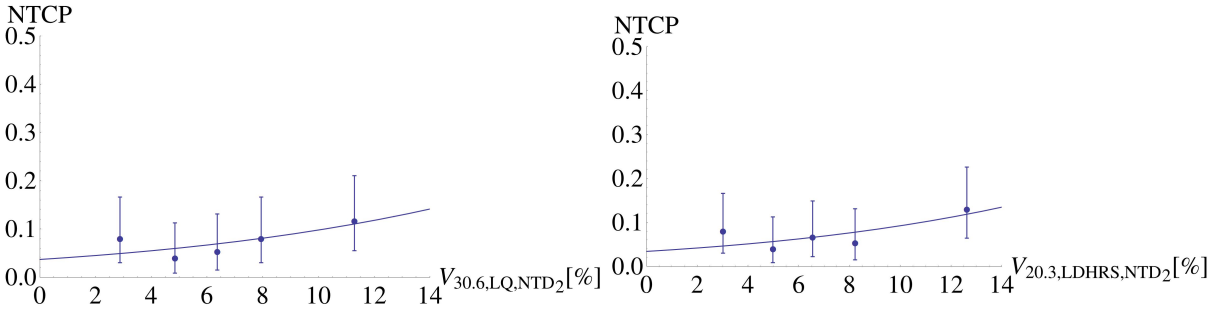


Figure 6.3: Optimal normal tissue complication probability (NTCP) curve and binned V_{D_c} data for the CDLR LQ model and the CDLR LDHRS model. The error bars indicate the 95 % binomial confidence intervals from binning the data.

6.2.5 Discussion

In this study NTCP was modelled for radiation pneumonitis after hypo-fractionated radiation therapy (RT) of lung cancer using various NTCP models. Two different cell survival models, (1) the LQ model and (2) the LDHRS model, were used to account for dose fractionation effects. The Lyman MLD LQ model showed the minimal AICc, due to a large likelihood and small number of parameters used in the model. However, the difference in AICc between the three Lyman models is small, which means that none of the models is preferable over the others in terms of AICc. The likelihood ratio test suggests that the additional parameters of the Lyman EUD LQ and the Lyman MLD LDHRS model do not lead to significantly improved model performance compared to the Lyman MLD LQ model. Similarly, the improvements in terms of log likelihood of the CDLR LDHRS model compared to the CDLR LQ model are not significant. However, although the LDHRS model does not yield significant improvement over the LQ model in terms of NTCP model fit, the use of the LDHRS model may enable the

Characteristic	
Age (median, range)	75.2 yrs (42yrs – 93.5 yrs)
Gender	199 female
	178 male
FU time (median, range)	25 months (6.2 months – 84.5 months)
Pneumonitis grade $\geq$2:	
0 or 1	349 (92.6 %)
2	22 (5.8 %)
3	4 (1.1 %)
4	0 (0.0 %)
5	2 (0.5 %)
Maximum tumor diameter (mean, SD)	2.59 cm $\pm$ 1.15 cm
Biopsy proven	253 (67.1 %)
Rx dose (median, range)	54 Gy (36 Gy – 64 Gy)

Table 6.2: Patient characteristics.

description of SBRT and SFRT with a single set of parameters, which will be discussed below.

Implications of using the LDHRS model compared to the LQ model

The LDHRS model may yield an alternative biological explanation for the low values of the volume-effect parameter, a, obtained for the Lyman EUD LQ model. The value for the volume effect parameter a used for EUD calculation determined in this study for SBRT was 0.3, whereas most studies investigating pneumonitis for SFRT report values close to or slightly larger than 1 (Seppenwoolde *et al.*, 2003; Martel *et al.*, 1994). It should be noted, however, that the upper confidence limit for a (0.0, 1.8) in this study includes those values. For low values of a (0<a<1) the EUD becomes smaller, or more influenced by low doses. This makes little sense in terms of the parallel versus serial tissue architecture interpretation of a that is cited in the literature (Li, Alber, Deasy, Jackson, Jee, Marks, Martel, Mayo, Moiseenko, Nahum and others, 2012). The LDHRS theory, however, may yield a better interpretation, since it assumes that low doses per fraction have a larger influence on cell survival, due to insufficient triggering of repair processes.

Figure 6.4 (a) and (b) show the theoretical ratio of the surviving fraction (SF) for different iso-doses in lung with various fractionation schemes to a standard fractionation scheme of 30 fractions and 2 Gy per fraction for the LQ and the LDHRS models, respectively. The LDHRS model predicts a surviving fraction for HFRT (i.e. $<$ 30 fractions) that is higher compared to the standard fractionation (30 fractions). At fractions used in the majority of SBRT cases in this study (3-5 fractions), the effect can be seen for total doses less than 20 - 25 Gy. The LQ model, on the other hand, always predicts

Model	Parameter estimates and CI	LL	AIC_c	p_{HL}
Lyman MLD LQ	MLD_{50,NTD_2} = 50.1 (28.8, ∞); m = 0.55 (0.44, 0.69)	-98.3	200.6	0.7
Lyman EUD LQ	a = 0.3 (0.0, 1.8); EUD_{50,NTD_2} = 13.2 (7.2, ∞); m = 0.57 (0.45, 0.71)	-97.9	201.8	0.4
Lyman MLD LDHRS	MLD_{50,NTD_2} = 35.4 (23.2,∞); m = 0.47 (0.35, 0.9); α_s = 4.9 (0.1, ∞)	-97.8	201.7	0.3
CDLR LQ	D_{c,NTD_2} = 30.6 Gy (0, ∞); β_0 = -3.27 (-3.93, -2.64); β_1 = 0.10 (0.01, 0.59)	-98.2	202.5	0.5
CDLR LDHRS	D_{c,NTD_2} = 20.3 Gy (10.4 Gy, 64.3 Gy); β_0 = -3.35 (-4.26, -2.51); β_1 = 0.11 (0.01, 0.38); α_s = 4.9 (0.1, ∞)	-97.2	202.4	0.4

Table 6.3: Optimal normal tissue complication probability (NTCP) model parameters, log-likelihood values, AICc values and Hosmer-Lemeshow goodness of fit p-values.

lower survival for the same dose delivered in smaller number of fractions compared to the standard fractionation. Considering the heterogeneous dose distribution (0 Gy - 60 Gy) in lung during typical SBRT, a large volume of normal lung tissue is irradiated with doses $\leq$ 25 Gy. Therefore, lung NTCP modelling using the LQ model would overestimate the BED and NTD corresponding to physical doses $\leq$ 25 Gy for SBRT, which would lead to an increase of the $\mathrm{MLD}_{50,\mathrm{NTD}_2}$, EUD_{50,NTD_2} and NTD corrected cut off doses obtained via NTCP modelling. At the same time, the LQ model may potentially underestimate the NTD for SFRT, as in SFRT more volume is irradiated with low doses per fraction. Thus, $\mathrm{MLD}_{50,\mathrm{NTD}_2}$, EUD_{50,NTD_2} and NTD corrected cut off doses could be closer than predicted by the LQ model. This could be an indication that both SBRT and SFRT can be described by the same lung NTCP model when LDHRS is used to calculate BED or NTD for lung tissue.

Comparison with Other Studies on Lung Stereotactic Body Radiation Therapy

Comparison of NTCP model results to other studies is generally difficult, because the populations may differ in a variety of different characteristics, and many studies use different values for the $\frac{\alpha}{\beta}$ ratio. Borst *et al.* modelled NTCP for 128 hypo-fractionated patients treated with 3D-CRT using the Lyman MLD model with NTD calculated with the LQ model and an $\frac{\alpha}{\beta}$ ratio of 3 Gy (Borst *et al.*, 2009). They obtained an $\mathrm{MLD}_{50,\mathrm{NTD}_2}$ of 19.6 Gy (95 %CI: 16 - 30 Gy) and m=0.43 (0.33 - 0.59) which is much lower than the 50.1 Gy (95 % CI: 28.8-∞ Gy) determined in this study. However, the study by Borst *et al.* uses the National Cancer Institute Common Toxicity Criteria (NCI-CTC) and Southwest Oncology Group (SWOG) criteria for scoring pneumonitis, rather than the CTCAE. Differences in scoring criteria can lead to differences in NTCP parameter estimates as was demonstrated by Tucker *et al.* (Tucker *et al.*, 2010). Furthermore, Borst *et al.* remark that their sample size is relatively small. They also determined a best fit cutoff dose and found a D_{c,NTD_2} of 50 Gy to yield the best fit, which is higher than the optimal cutoff dose of 30.6 Gy determined in this study based on the LQ model. This discrepancy however, can be attributed to the different $\frac{\alpha}{\beta}$ ratio of roughly 2.1 Gy used in this study, which was the $\frac{\alpha}{\beta}$ determined by Singh *et al.* (Singh *et al.*, 1994). When using an $\frac{\alpha}{\beta}$ ratio of 3 Gy an optimal cutoff dose of 57 Gy was obtained. Sonke *et al.* also modelled NTCP using the LQ model for SBRT patients who partially overlap with the group of patients considered in this study (Sonke *et al.*, 2011). They calculated $\mathrm{MLD}_{50,\mathrm{NTD}_2}$ = 45 Gy (95 % CI, 27-85 Gy), and m = 0.52 (95 % CI, 0.4-0.69), which is similar to the values determined in this study and also much larger than the results obtained by Borst *et al.* The two studies also come to different conclusions regarding whether hypo-fractionated and standard fractionated patients can be described with the same model. While Borst *et al.* do not find a significant difference between models derived from standard fractionated data, Sonke *et al.* concluded that there was a significant difference between SBRT and SFRT. However, neither of the two studies used matched samples of patients treated with SBRT and SFRT and therefore it is not clear whether the data is comparable. Differences in follow up or other covariates could possibly explain the difference in results. The small differences between NTCP pa-

rameters determined in the study performed by Sonke *et al.* and this study are due to different $\frac{\alpha}{\beta}$ ratios. When NTCP for the patients in this study was modelled with an $\frac{\alpha}{\beta}$ ratio of 3 Gy, which has been used by Sonke *et al.*, similar parameters were obtained ($\mathrm{MLD}_{50,\mathrm{NTD}_2}$ = 44.1 Gy; m = 0.55; LL = -98.3).

Comparison to Standard Fractionated Radiation Therapy

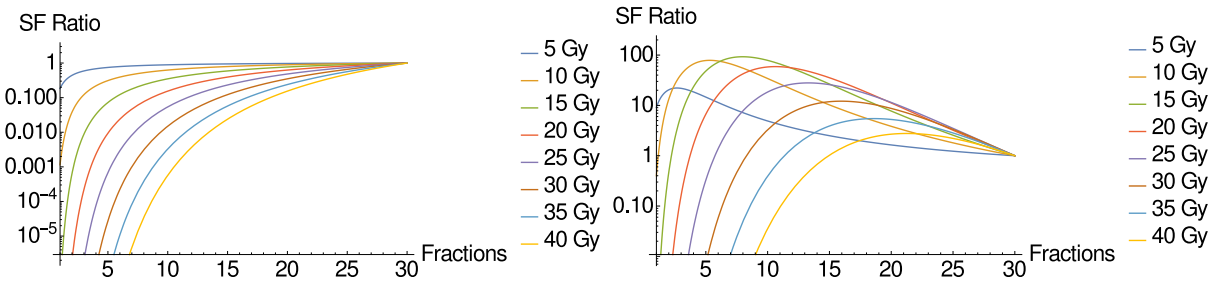
Several studies have shown that radiation pneumonitis in SFRT is best associated with $V_{20\mathrm{Gy}}$ and MLD (Kong *et al.*, 2006; Graham *et al.*, 1999; Claude *et al.*, 2004).

The optimal cutoff dose obtained for the LQ model is higher than the frequently cited $V_{20\mathrm{Gy}}$ for SFRT, but is within the range of other published data for lung SFRT (Yorke *et al.*, 2005) and $V_{20\mathrm{Gy}}$ is within the confidence interval determined in this study. When the LDHRS model was used, $V_{20.6\mathrm{Gy}}$ was found to be the optimal cutoff dose in this study.

Sonke *et al.* have found a $\mathrm{MLD}_{50,\mathrm{NTD}_2}$ of 28.5 Gy for a standard fractionated dataset using the LQ model (Sonke *et al.*, 2011). This is significantly lower than the $\mathrm{MLD}_{50,\mathrm{NTD}_2}$ determined for hypo fractionated patients in their and this study, when the LQ model was used for the NTD calculation. The $\mathrm{MLD}_{50,\mathrm{NTD}_2}$ determined for the LDHRS model (35.4 Gy), however, was much closer to typical MLD_{50} values determined for SFRT (Seppenwoolde *et al.*, 2003; Kwa *et al.*, 1998). This is an indication that the LDHRS model could resolve the issue of different MLD_{50} obtained for SFRT and SBRT and the LDHRS model may be capable of describing both treatment techniques with the same NTCP model. Since the LDHRS model would also affect the DVHs of standard fractionated patients, this is only an indication. In order to estimate how the LDHRS model may influence SFRT DVHs, the mean $\mathrm{MLD}_{\mathrm{NTD}_2}$ for all patients in this cohort were calculated with the LQ model and the LDHRS model, assuming that they were given in 30 fractions, using the optimal α_s = 4.6 Gy^{-1} determined in this study. The results were $\mathrm{MLD}_{\mathrm{NTD}_2,\mathrm{LQ},30}$ = 3.3 Gy $\pm$ 1.3 Gy, $\mathrm{MLD}_{\mathrm{NTD}_2,\mathrm{LDHRS},30}$ = 22.5 Gy $\pm$ 6.3 Gy. Thus, the MLD_{50} for standard fractionated patients is likely going to increase when modelled with the LDHRS model. However, for a general comparison of SFRT data and SBRT data, a matched pair analysis should be performed to ensure that the populations are comparable in terms of other patient characteristics such as treatment technique, smoking, pre-existing fibrosis, age, gender, race.

6.2.6 Conclusions

Both the LQ model and the LDHRS model yield acceptable NTCP models for radiation pneumonitis when fit to DVHs and toxicity data of hypo-fractionated lung RT. However, the LQ model uses less parameters and thus may be statistically preferable. The LDHRS model may offer a biological explanation for a possible increased influence of low doses per fraction on pneumonitis. Unfortunately, due to the large confidence intervals obtained in this study, the effect could not be verified. Furthermore, the LDHRS model may be able to describe SBRT and SFRT with the same model. In the future, further investigations on the validity of the LDHRS model are necessary and



(a) LQ model **(b)** LDHRS model

Figure 6.4: Ratio of surviving fractions of a total physical dose of X Gy, to X Gy given in 30 fractions as a function of the number of fractions, for X= 10 Gy, 15 Gy, 20 Gy, 25 Gy, 30 Gy, 35 Gy and 40 Gy for (a) the linear quadratic (LQ) model and (b) the low dose hyperradiosensitivity (LDHRS) model.

studies comparing hypo-fractionated NTCP to standard fractionated RT for matched datasets should be undertaken.

6.3 Normal Tissue Complication Probability Models for Chest Wall Related Toxicities

6.3.1 Background

The use of SBRT for the treatment of non-operable lung tumors has markedly improved local control in this patient population, and is now considered as a potential treatment option even for operable lung tumors. However, investigations are still underway to establish whether SBRT is sufficiently safe for normal tissues. Several institutions report an increased incidence of toxicities related to the chest wall, especially chest wall pain and rib fractures (Mutter *et al.*, 2012; Voroney *et al.*, 2009; Bongers *et al.*, 2011). Thus, various studies have sought to find dose volume predictors for chest wall toxicities after lung SBRT (Murray *et al.*, 2016; Thibault *et al.*, 2016; Mutter *et al.*, 2012; Voroney *et al.*, 2009; Bongers *et al.*, 2011; Andolino *et al.*, 2011). However, no standard dose volume constraints for treatment planning were established so far and it is not clear if more effort should be undertaken to spare the chest wall (Thibault *et al.*, 2016). The majority of aforementioned studies focus on cutoff dose volumes constraints, and NTCP models of chest wall toxicities are scarce. Modelling NTCP for chest wall toxicities can be challenging, as the chest wall and ribs are not typically contoured for lung SBRT, and thus have to be contoured retrospectively. Furthermore, chest wall toxicities are not always documented properly. The aim of this study was to model NTCP for chest wall toxicities in lung SBRT patients using four different NTCP models.

6.3.2 Methods and Materials

Patient Data

In this study, DVHs of 70 patients treated for early stage non small cell lung cancer or limited pulmonary metastases were evaluated to create NTCP models for toxicities related to irradiation of the chest wall. All patients were treated with SBRT between 2006 and 2012. 25 patients were prescribed 60 Gy to the tumor in 5 fractions of 12 Gy and the remaining 45 patients were prescribed 48 Gy given in 4 fractions. The endpoint of this study were chest wall toxicities grade $\geq$ 2, which included chest wall pain, myositis (muscle inflammation) and rib fractures. All toxicities were scored using the CTCAE version 3.0. Patients with FU times of less than 6 months were excluded from the study.

The GTV was contoured on breathing phase 0 of a 4D-CT of the thorax with 10 phases and then propagated to the other 9 breathing phases. The 10 contours were merged to form an ITV and then a 4 mm margin was added to create the CTV contour. The PTV was contoured by adding a 5 mm margin to the CTV. Three chest wall contours (CW1cm, CW2cm, CW3cm) were created in Pinnacle (Philips Pinnacle^3, Philips Healthcare, Andover, MA) using a semi automated in-house Pinnacle script. CW1cm, CW2cm, and CW3cm were defined as 1 cm, 2 cm and 3 cm rings extending radially outward from the lung and chest wall interface (Figure 6.5). The dose matrix and

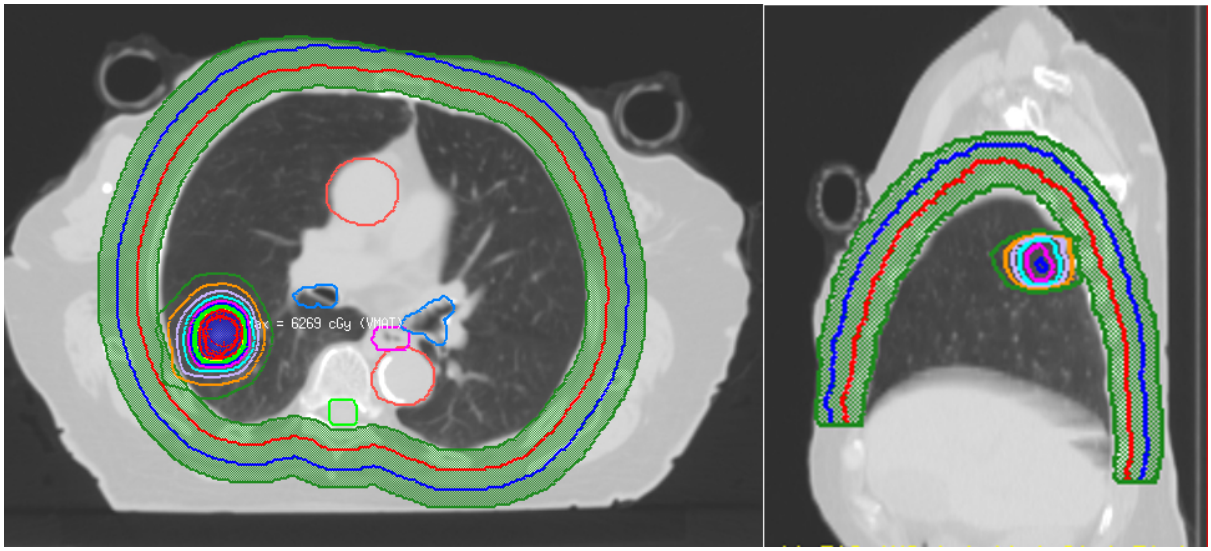


Figure 6.5: Chest wall contours CW1cm (red), CW2cm (blue) and CW3cm (green) extending radially outward from the chest wall lung interface (green).

DVHs were calculated using Pinnacle's collapsed cone (CC) algorithm. Mathematica (Wolfram Research, Inc., Mathematica, Version 9.0.1, Champaign, IL (2013)) was then used to convert DVHs to BED using the LQ model (Equation 3.7) with an $\frac{\alpha}{\beta}$ ratio of 3 Gy. The CW1cm contour was then used for NTCP modelling as dose volume parameters in this contour were shown to be most strongly correlated with chest wall toxicities (Mangona *et al.*, 2013).

Normal Tissue Complication Probability Models

The Lyman model (Equation 4.9) and the CDLR model (Equation 4.7) were used to model NTCP. The summary measures used were mean chest wall dose (MCWD) and EUD for the Lyman model (Lyman MCWD model, Lyman EUD model), and V_{D_c} for the CDLR model (CDLR model). All of these summary measures were calculated from DVHs that were first converted to BED. Additionally, for the CDLR model, V_{D_c} calculated using physical dose was used (CDLR PHYS). The optimal model parameters were determined using MLE as described in Section 4.5.1. CIs were calculated using the likelihood profile method (Section 4.8.2). Goodness of fit was evaluated using the Hosmer Lemeshow goodness of fit test (Equation 4.19).

6.3.3 Results

Patient Characteristics

Of the 70 patients used in this study, 25 (35.7 %) experienced chest wall toxicities of grade ≥ 2. The median FU was 2.3 years with a range of 0.5 years to 6.5 years. The median age was 77 years with ages ranging from 57 years to 93 years. 41 patients were female and 29 patients were male.

Model	Parameter estimates and CI	LL	AIC_c	p_{HL}
Lyman MCWD	$MCWD_{50,BED}$ = 23.8 (16.8 Gy, 59.0 Gy); m = 0.93 (0.58, 2.32)	-42.5	89.2	0.7
Lyman EUD	a = 2.15 (0.40, ∞); $EUD_{50,BED}$ = 72.2 Gy (7.7 Gy, ∞); m = 0.77 (0.41, 2.04)	-42.2	90.7	0.6
CDLR	$D_{c,BED}$ = 212.9 Gy (22.2 Gy, 301.1 Gy); β_0 = -1.24 (-2.96, -0.55); β_1 = 0.34 (0.06, 0.60)	-41.9	90.3	0.3
CDLR PHYS	D_c = 39.9 Gy (8.4 Gy, 57.8 Gy); β_0 = -1.18 (-3.20, -0.47); β_1 = 0.19 (0.02,0.40)	-43.0	92.4	0.3

Table 6.4: Optimal normal tissue complication probability (NTCP) model parameters, log-likelihood values, AICc values and Hosmer-Lemeshow goodness of fit p-values for chest wall toxicities (chest wall pain, rib fractures, myositis) grade ≥ 2.

Normal Tissue Complication Probability Models

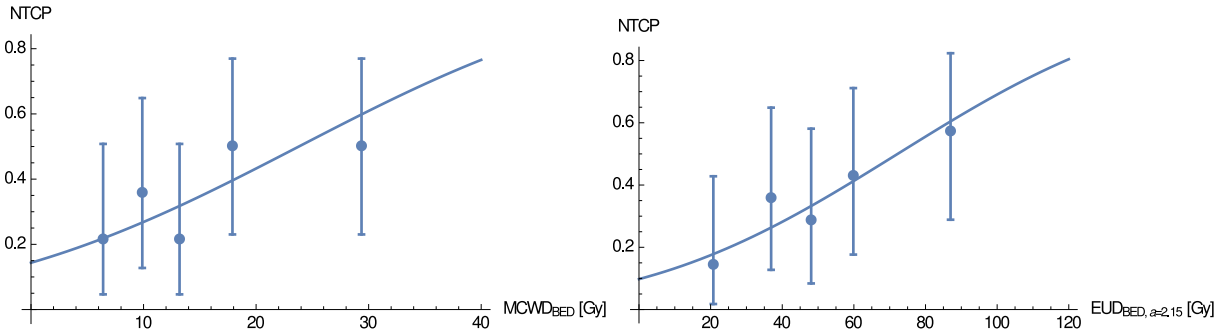
The optimal NTCP model parameters, CIs, log likelihood, AICc and Hosmer-Lemeshow goodness of fit p-values are shown in Table 6.4. The optimal NTCP curves together with the binned toxicity and DVH data are shown in Figure 6.6a, Figure 6.6b, Figure 6.7a, and Figure 6.8a for the Lyman MCWD, the Lyman EUD, the CDLR and the CDLR PHYS model, respectively. According to the Hosmer Lemeshow goodness of fit test, all models fit the data well ($p_{HL} > 0.05$). The Lyman MCWD model had the lowest AICc, however the AICc values of all models were similar. Table 6.5 shows the physical dose values corresponding to the BED values given in 4 and 5 fractions.

Model	BED	5 fractions	4 fractions
CDLR	212.9 Gy	49.5 Gy	44.9 Gy
Lyman MCWD	23.8 Gy	12.8 Gy	11.9 Gy
Lyman EUD	72.2 Gy	26.3 Gy	24.0 Gy

Table 6.5: Conversion of biologically effective dose (BED) with $\frac{\alpha}{\beta} = 3$ Gy to physical dose given in 5 and 4 fractions.

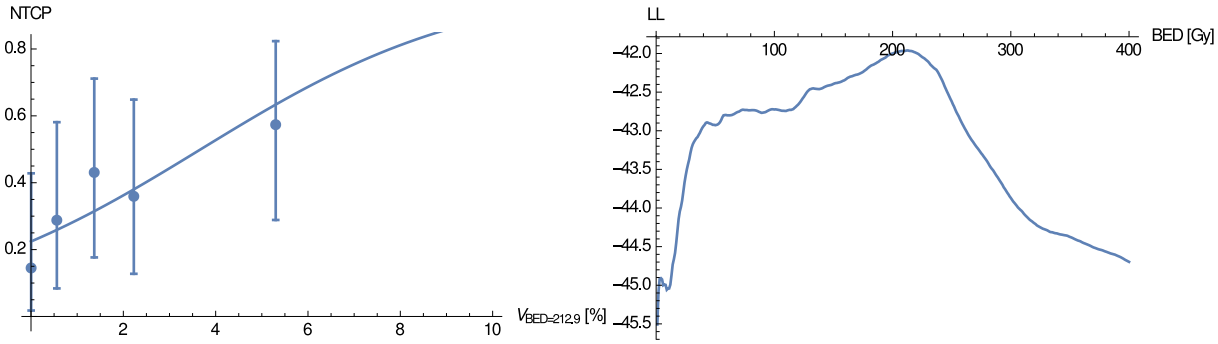
6.3.4 Discussion

In this study, NTCP models for chest wall toxicity grade ≥ 2 were created using CW1cm DVHs. All derived models fit the data well. The CDLR model using BED DVHs obtained the highest likelihood, but when corrected for the number of parameters using AICc, the Lyman MCWD model showed the best performance. The CDLR



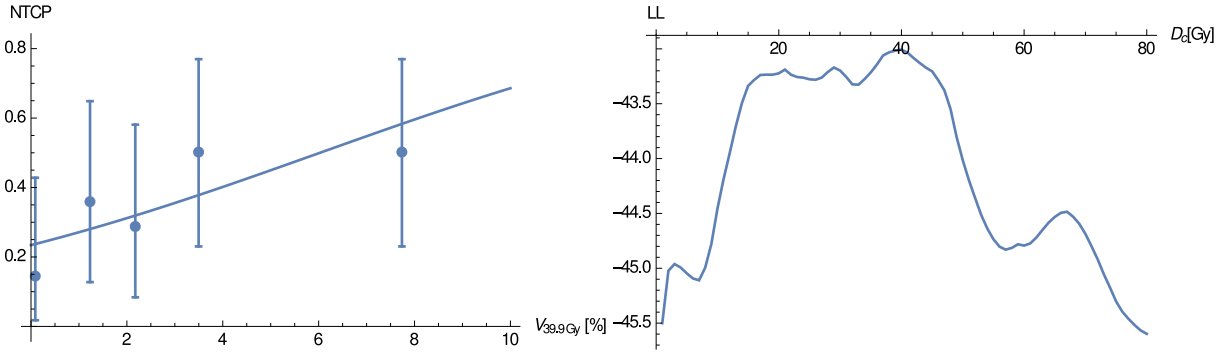
(a) Lyman MCWD model **(b)** Lyman EUD model

Figure 6.6: Optimal normal tissue complication probability (NTCP) curve and binned MLD and EUD data for the Lyman model. The error bars indicate the 95 % binomial confidence intervals from binning the data.



(a) CDLR model **(b)** Log likelihood LL versus D_c

Figure 6.7: Optimal normal tissue complication probability (NTCP) curve and binned V_{D_c} data for the CDLR model using BED. The error bars indicate the 95 % binomial confidence intervals from binning the data.



(a) CDLR PHYS model **(b)** Log likelihood LL versus D_c

Figure 6.8: Optimal normal tissue complication probability (NTCP) curve and binned V_{D_c} data for the CDLR model using physical dose (CDLR PHYS). The error bars indicate the 95 % binomial confidence intervals from binning the data.

model using physical dose (CDLR PHYS) performed worst both in terms of log likelihood and AICc. The AICc values for all models, except the CDLR PHYS model were similar.

For the Lyman EUD model, an optimal value for the volume effect parameter a of 2.15 was obtained, which suggests that chest wall toxicities are potentially subject to a volume effect. Since the CI for the volume effect parameter a is large (0.4, ∞), the value will have to be confirmed in future studies with more patients. This study is a preliminary study and is limited by the small number of patients used for the study, leading to large CIs. However, the incidence of chest wall toxicities was relatively high and thus allowed for NTCP modelling.

Several studies have investigated dose volume predictors for chest wall toxicities. Mutter *et al.* investigated predictors for chest wall pain in SBRT patients receiving 40-60 Gy in 3-5 fractions (Mutter *et al.*, 2012). They report an estimated 2 year actuarial incidence of chest wall toxicities of 39% which is similar to the incidence of 35.6 % in this study. Mutter *et al.* found the volume receiving a physical dose of 30 Gy or higher to be the best predictor for chest wall, while this study found a physical cutoff dose of 39 Gy to yield the largest log likelihood. However, $V_{30\mathrm{Gy}}$ is within the CI obtained in this study and the log likelihood profile of the cutoff dose D_c shows a plateau between 20 Gy and 40 Gy, indicating that values in that whole range may all yield acceptable NTCP models (Figure 6.8b). Thus, the results obtained in this study seem roughly consistent with Mutter *et al.*

Mutter *et al.* converted DVHs to BED using $\frac{\alpha}{\beta}$ ratios ranging from 2 - 10 Gy, and conclude that there is no advantage of biologically corrected dose over physical dose. This study, in contrast, showed that the CDLR BED model had superior AICc over the CDLR phys model with physical dose. However, there were several differences between the two studies that could lead to discrepancies. Mutter *et al.* used the NCI-CTC scale, and extended the chest wall contour by 2 cm and 3 cm, while this study used CTCAE for toxicity grading and focused on CW1cm. Currently, there is no widely accepted clinical standard on how to contour the chest wall and ribs and future studies will have to determine appropriate contouring.

Thibault *et al.* investigated patient characteristics associated with chest wall toxicities. They found female gender, osteoporosis, a tumor position adjacent to the chest wall (within 5 mm), and several DVH parameters ($\mathrm{D}_{\mathrm{max}}$, $D_{1\mathrm{cm}^3}$, $D_{2\mathrm{cm}^3}$, $D_{5\mathrm{cm}^3}$, V_{30Gy}, V_{40Gy}, V_{50Gy}) to be predictive for rib fractures in a univariate analysis (Thibault *et al.*, 2016). After multivariate analysis, only tumor position adjacent to the chest remained and they concluded that no DVH parameter was able to predict rib fractures independently. Their results indicate that patient characteristics may need to be taken into account when comparing SBRT populations and when modelling larger cohorts in the future.

Andolino *et al.* determined predictors for rib fractures and chest wall pain using the CTCAE for SBRT patients treated with SBRT for lung and liver cancer (Andolino *et al.*, 2011). They used the median effective concentration dose–response model, which correlates a cutoff dose volume with the risk of toxicities occurring. They determined that the volume of chest wall receiving 25 Gy - 60 Gy physical dose is associated with

a significantly higher incidence of chest wall toxicities, which is consistent with the results obtained in this study. However, in contrast to this study, they used CW3cm and scored all chest wall toxicities including grade 1.

Murray *et al.* report evidence that small hot spots are associated with chest wall pain grade ≥ 2. They looked for predictors for CTCAE grade ≥ 2 chest wall pain in SBRT patients treated with 55 Gy in 5 fractions and found that tumor size and D_{max} received by 1 cm^3 of the chest wall and tumor size were predictive of chest wall pain.

6.3.5 Conclusion

In this study, NTCP was modelled for chest wall toxicities grade ≥ 2 after lung SBRT. The Lyman model based on the mean dose to the chest wall performed the best in terms of AICc, but all BED based models fit the data well and had similar AICc values, except for the model based on physical dose. Correcting for fractionation by using BED DVHs rather than physical DVHs to model NTCP is advantageous. However, future studies need to determine the most appropriate values for the $\frac{\alpha}{\beta}$ ratio for chest wall toxicities. Furthermore, future prospective investigations with larger patient populations should be undertaken to confirm the results of this study and also to establish a consensus on how to contour the chest wall. Since NTCP models predicting chest wall toxicities are scarce, this preliminary study could give valuable information on the dose response of chest wall toxicities and potentially aid in finding acceptable clinical dose constraints for the chest wall during lung SBRT.

Chapter 7

Influence of Dose Calculation Algorithms

Publication: Stereotactic radiotherapy of intrapulmonary lesions: comparison of different dose calculation algorithms [1]

7.1 Background

The goal of normal tissue complication probability (NTCP) modelling is to create a relationship that describes the probability of a patient experiencing side effects as a response to the radiation dose received by a structure. Thus, in order to obtain an accurate model, it is of importance that the dose distributions used for modelling reflect the actual dose delivered to the patient. Improvements in radiation therapy (RT), such as image guided radiation therapy (IGRT) and adaptive radiation therapy (ART), have reduced setup errors and errors resulting from anatomical changes. Thus these and other advances in radiation therapy have reduced the difference between planned dose and delivered dose. Another factor in correctly assessing the dose delivered to the patient is the choice of the dose calculation algorithm used to calculate the planned dose. As described in Chapter 3, there are three main types of algorithms used in contemporary RT: pencil beam (PB) based algorithms, collapsed cone (CC) based algorithms, and Monte Carlo (MC) simulations. These algorithms differ in how accurately they calculate dose. The PB based algorithms used to serve as the standard algorithms in treatment planning due to their calculation speed. However, it is well known that the PB algorithms lack accurate modelling of lateral scattering and backscatter, because they neglect tissue density inhomogeneities other than in the direction of the beam and assume water density instead (see Section 3.3). In contrast, CC and MC include more accurate inhomogeneity corrections/modelling (Ahnesjö and Aspradakis,

[1] Reprinted (adapted) with permission from *"Stereotactic radiotherapy of intrapulmonary lesions: comparison of different dose calculation algorithms for Oncentra MasterPlan®"* by Troeller, A., Garny, S., Pachmann, S., Kantz, S., Gerum, S., Manapov, F., Ganswindt, U., Belka, C., Söhn, M., 2015. Radiat. Oncol. 10. doi:10.1186/s13014-015-0354-3

1999). The use of more accurate dose calculation algorithms, such as CC and MC, has increased during the past few years due to increased speed of the algorithms and availability of advanced computational technology. A popular choice for dose calculation are now CC based algorithms that *approximately* model lateral scattering and backscatter via density dependent anisotropic rescaling of pre-calculated point kernels for effects from secondary particle transport. MC simulations, which *explicitly* model particle interactions and transport in the patient, are generally considered the gold standard of dose calculation.

Differences in dose calculation accuracy in the patient between the different algorithms can be expected to be most predominant in anatomic regions with large local tissue density inhomogeneities and density discontinuities such as for small lung lesions surrounded by low density lung tissue. When stereotactic body radiation therapy (SBRT) of intrapulmonary lesions was first introduced into routine clinical treatment, most institutions calculated dose distributions using dose calculation programs based on PB type algorithms. However, as indicated above, it is now widely and clinically accepted that dose calculation algorithms of PB type do generally not provide enough accuracy for dose calculation in inhomogeneous tissue (Krieger and Sauer, 2005; Vanderstraeten *et al.*, 2006). Recent recommendations for the implementation of SBRT of lung cancer usually include the use of CC or MC algorithms. However, historically prescription schemes were based on clinical experience with PB. Recent guidelines such as the International Commission on Radiation Units and Measurement (ICRU) recommend similarly strict requirements for treatment plans calculated with CC or MC as they do for PB (ICRU-62, Bethesda, Md, USA, 1999). Because of the insufficient modelling of lateral scattering and backscatter, the dose to the tumor is usually overestimated when using PB and the same plan may indicate lower dose when recalculated with CC (Aarup *et al.*, 2009; Dobler *et al.*, 2006). This will lead to treatment plans with changes in field size and monitor units (MUs) delivered, as well as changes in dose to organ at risk (OAR) when creating a treatment plan based on a CC dose calculation algorithm, without proper adjustment of the prescription.

In the following, differences between dose distributions and derived dose volume parameters calculated with PB and CC are investigated, which are the two dose calculation algorithms implemented in the clinical three dimensional conformal radiation therapy (3D-CRT) treatment planning system (TPS) used at the LMU Munich, Oncentra MasterPlan® (OTP). In order to quantify differences between dose distributions calculated with PB and CC in patients treated at LMU Munich, treatment plans of patients that received SBRT of intrapulmonary lesions were examined, since these patients usually have smaller tumors and may show large discrepancies between dose calculated with PB and CC.

7.2 Methods

The treatment plans of 17 patients treated for small intrapulmonary lesions between 2008 and 2010 were retrospectively evaluated (Table 7.1). For each patient, three com-

Patient	Maximum tumor diameter (cm)	Total volume GTV (ccm)	Tumor location	Central/peripheral
1	1.6	2.0	RLL	peripheral
2	1.6	3.5	RUL	central
3	4.1	41.2	RUL	peripheral
4	3.1	12.1	LLL	peripheral
5	2.8	6.4	R	central
6	1.5	1.4	RUL	peripheral
7	1.2	4.6	RLL	peripheral
8	4.5	3.8	R	central
9	1.2	1.2	R	central
10	2.6	15.2	RLL	peripheral
11	3.0	17.0	LUL	central
12	2.5	4.4	RUL	peripheral
13	2.5	27.0	RLL	peripheral
14	3.9	27.9	RLL	peripheral
15	4.6	61.6	RLL	central
16	1.8	4.7	RUL	peripheral
17	4.0	36.2	LUL	peripheral

Table 7.1: Tumor size, volume and location for all patients. For the purpose of this study tumors that were directly attached to the mediastinum or pleura were defined as peripheral and all others (fully surrounded by low density lung tissue) as central. RUL, RLL, LLL, LUL, and R refer to right upper lobe, right lower lobe, left lower lobe, left upper lobe and right, respectively.

puted tomography (CT) scans (3 mm slice distance) were acquired prior to treatment: a free-breathing scan, a maximum inhale and a maximum exhale scan. The scans were then imported to the OTP treatment planning system (Version 4.2, Nucletron). The gross tumor volume (GTV) was contoured on each of the CT scans and an internal target volume (ITV) was defined by forming the union of all GTVs. The planning target volume (PTV) was created from the ITV, by adding a 7 mm margin laterally and 9 mm in the cranial-caudal direction. For all patients, a cumulative physical dose of up to 60 Gy was prescribed to the PTV in single fractions of 7.5 Gy, such that the 80% isodose line completely covered the PTV. All dose calculations were performed on the free-breathing CT scan without density overrides. The 3D-CRT, un-modulated treatment plans were designed for a Siemens Oncor treatment unit with a 10 mm leaf width multi leaf collimator (MLC). Mixed photon energy was used (6 and 15 MV) incorporating 7 beams on average. The beam model was verified against experiment and accepted for clinical use. The dose distributions of the original treatment plans were calculated using OTP's enhanced PB algorithm. For the purpose of this study, all plans were re-calculated with the enhanced CC algorithm implemented in OTP ($CC_{\text{re-calc}}$), without altering the field shape, field size, beam setup, or monitor units. Furthermore two additional plans calculated with CC were created for each patient: A plan that resembles what would currently be considered acceptable for a patient in the LMU clinic (CC_{clin})

and a stricter second plan (CC_{cov}) that obtained coverage similar to the original PB plan (PB). The former constitutes a clinically accepted compromise that may formally violate the PTV coverage criteria mentioned above in low density parts of the PTV where full dose build-up is difficult to realize for physical reasons. The latter was created to demonstrate how dose to OARs would change in the hypothetical scenario of switching from a PB algorithm to a CC algorithm without adapting the criteria for an acceptable plan.

For the CC_{clin} and CC_{cov} plans MUs as well as field size and shape had to be adjusted accordingly. For the CC_{cov} the fields were opened in beam's eye view (BEV) until coverage was achieved, while for the CC_{clin} the distance between the jaws and the PTV-outline were not to exceed 1 cm in BEV. All four plans per patient were then compared to each other. Parameters used for comparison were volume coverage (V_{60Gy}), mean dose (D_{mean}), dose received by 99% and 1% of the volume (D_{99} and D_1), minimum dose (D_{min}), and maximum dose (D_{max}) for the GTV, ITV and PTV. Estimated coverage in dependence on GTV volume was also considered. Furthermore D_{mean}, V_{20Gy} and V_{10Gy} to the lungs, D_{mean} to the heart and D_{max} to the spinal cord and esophagus were evaluated. Both lungs were contoured as one organ including the PTV. The adjusted CC_{clin} and CC_{cov} treatment plans were further analyzed regarding alterations of MUs and field size compared to the original PB plan. The influence of tumor position and movement on coverage was investigated. For the purpose of this study tumors that were directly attached to the mediastinum or pleura were defined as peripheral and all others as central. Statistical significance of the differences between parameters calculated with PB and parameters calculated with $CC_{re\text{-}calc}$, CC_{clin} and CC_{cov} was determined using the Wilcoxon signed-rank test ($p < 0.05$ signifies significant difference). Correlation of the difference in target coverage between PB and CC (i.e. $V_{60Gy,PB} - V_{60Gy,CC}$) with the target region of interest (ROI) volumes was determined using Spearman's rank correlation coefficient ρ. The significance of the correlation was verified with Spearman's rank test, with the null hypothesis H_0 that ROI volume and difference in coverage are independent and alternative hypothesis H_a that they are dependent. The analysis of the parameters and statistical analyses was performed in Mathematica (Wolfram Research, Inc., Mathematica, Version 9.0.1, Champaign, IL, 2013).

7.3 Results and Discussion

7.3.1 Original Pencil Beam Plan versus the Original Plan Recalculated with Collapsed Cone ($\mathrm{CC_{re-calc}}$)

For the target structures GTV, ITV and PTV, all evaluated parameters ($\mathrm{V_{60Gy}}$, $\mathrm{D_{mean}}$, $\mathrm{D_{min}}$, $\mathrm{D_{max}}$, D_{99}, D_1) were significantly lower ($p < 0.001$) for the plans that were recalculated based on the CC algorithm ($\mathrm{CC_{re\text{-}calc}}$) as opposed to the original ones incorporating the PB algorithm (PB). All OAR parameters considered in this study ($\mathrm{p} < 0.01$), except $\mathrm{D_{max}}$ to the spinal cord, were statistically significantly lower when the PB plan was recalculated with CC. However the differences regarding heart and spinal cord were of relatively small magnitude and may not be clinically relevant. The results for all structures are shown in Table 7.2. The GTV, ITV and PTV coverage with 60 Gy and higher as well as D_{99} and $\mathrm{D_{max}}$ in dependence on the GTV, ITV and PTV volume for both PB and $\mathrm{CC_{re\text{-}calc}}$ treatment plans are shown in Figure 7.1 and Figure 7.2. For the PTV, a larger discrepancy between PB and CC coverage is significantly correlated with a smaller ROI volume ($\rho_{\mathrm{PTV}} = -0.77$, $p_{\mathrm{PTV}} < 0.001$), while for the GTV and ITV the correlation coefficient is less pronounced and not significant ($\rho_{\mathrm{ITV}} = -0.43$, $p_{\mathrm{ITV}} = 0.08$, $\rho_{\mathrm{GTV}} = -0.18$, $p_{\mathrm{GTV}} = 0.49$). Large difference of $\mathrm{D_{min}}$, D_{99}, $\mathrm{D_{max}}$ and D_1 between PB and CC is significantly correlated with ROI volume for all three target ROIs GTV, ITV and PTV. There was considerable variability of the dose volume parameters of GTV, ITV and PTV between patients, likely due to tumor volume and position. As expected the differences were more pronounced for tumors completely surrounded by lung tissue and not adherent to pleural or mediastinum tissues. For example, the difference in average GTV relative volume covered by 60 Gy or more was 3.5% for peripheral tumors compared to 6.1 % for centrally located tumors. Presumably, this is due to the fact that the density of the chest wall and mediastinal tissue is closer to the density of water, than the low density lung tissue. Therefore the PB algorithm, which assumes water density laterally, estimates dose more accurately in tumors partially attached to denser tissue than in those surrounded mainly by air. In one extreme case of a central tumor with very small volume, the GTV $\mathrm{V_{60Gy}}$ was reduced from 100% to 26%, $\mathrm{D_{mean}}$ from 74.2 Gy to 57.9 Gy and D_{99} from 71.6 Gy to 49.9 Gy. The PTV coverage was tremendously reduced from 96% to 3.6%. The dose volume histograms (DVHs) of this patient are shown in Figure 7.3. Dose distributions of this patient and another less extreme case with larger tumor volume close to the posterior chest wall are shown in Figure 7.4. No correlation was found between tumor movement (which was quantified by the ratio of ITV volume to GTV volume) and difference in GTV coverage (Figure 7.1 bottom, right, $\rho = 0.12$, $p = 0.63$). However, this result may be limited to this study and a more rigorous analysis is necessary to evaluate the impact of tumor motion on dose distribution in general, for example by performing a four dimensional (4D) dose calculation study similar to Guckenberger *et al.* (Guckenberger *et al.*, 2007).

ROI	Parameter	PB	$CC_{re-calc}$	CC_{cov}	CC_{clin}
GTV	V_{60Gy} (%)	99.8 ± 0.9	86.6 ± 26.5	99.6 ± 1.0	99.4 ± 1.5
	D_{mean} (Gy)	73.7 ± 1.6	67.0 ± 4.4	72.8 ± 1.5	72.5 ± 1.6
	D_{min} (Gy)	63.1 ± 11.0	51.4 ± 9.6	61.3 ± 8.5	59.5 ± 9.6
	D_{99} (Gy)	67.9 ± 7.1	56.8 ± 7.6	65.5 ± 5.4	64.2 ± 6.3
	D_{max} (Gy)	76.8 ± 1.4	71.6 ± 4.0	76.6 ± 1.3	76.6 ± 1.3
	D_1 (Gy)	76.4 ± 1.4	71.3 ± 3.9	76.3 ± 1.3	76.4 ± 1.3
ITV	V_{60Gy} (%)	99.6 ± 1.7	79.8 ± 30.4	98.4 ± 4.3	97.3 ± 5.7
	D_{mean} (Gy)	73.3 ± 1.7	64.8 ± 5.2	71.0 ± 2.1	70.6 ± 2.3
	D_{min} (Gy)	60.6 ± 12.4	47.7 ± 11.0	57.9 ± 11.7	56.3 ± 11.7
	D_{99} (Gy)	66.0 ± 10.5	53.2 ± 9.9	62.1 ± 8.6	60.7 ± 9.4
	D_{max} (Gy)	77.0 ± 1.5	71.6 ± 4.0	76.6 ± 1.3	76.6 ± 1.3
	D_1 (Gy)	76.7 ± 1.4	71.1 ± 4.0	76.2 ± 1.2	76.2 ± 1.2
PTV	V_{60Gy} (%)	95.1 ± 1.9	55.7 ± 27.0	90.9 ± 12.5	85.5 ± 15.3
	D_{mean} (Gy)	69.9 ± 1.3	60.0 ± 5.7	67.4 ± 2.6	66.6 ± 3.1
	D_{min} (Gy)	43.4 ± 4.9	35.8 ± 3.9	49.2 ± 4.4	46.4 ± 5.2
	D_{99} (Gy)	55.0 ± 1.9	44.3 ± 5.3	56.7 ± 3.7	54.2 ± 4.5
	D_{max} (Gy)	77.0 ± 1.5	71.6 ± 4.0	76.6 ± 1.3	76.6 ± 1.3
	D_1 (Gy)	76.5 ± 1.4	70.7 ± 4.1	75.8 ± 1.2	75.8 ± 1.2
Lungs	V_{20Gy}(%)	7.6 ± 4.3	6.8 ± 4.2	9.6 ± 5.0	8.9 ± 4.9
	V_{10Gy} (%)	14.3 ± 5.7	13.4 ± 6.3	17.4 ± 6.6	16.3 ± 6.5
	D_{mean} (Gy)	5.3 ± 2.1	5.0 ± 2.0	6.6 ± 2.3	6.1 ± 2.3
Heart	D_{mean} (Gy)	3.0 ± 3.5	2.9 ± 3.5	3.6 ± 3.5	3.4 ± 3.5
Spinal cord	D_{max} (Gy)	8.3 ± 8.8	8.1 ± 8.5	10.9 ± 9.6	10.3 ± 9.6
Esophagus	D_{max} (Gy)	17.6 ± 9.7	16.9 ± 8.9	19.6 ± 10.6	19.3 ± 10.3

Table 7.2: Mean dose and volume parameters and standard deviations for targets and organs at risk for pencil beam (PB) and collapsed cone (CC) calculations.

7.3.2 Pencil Beam Plan versus Clinical Collapsed Cone Plan (CC_{clin}) and Collapsed Cone Plan with Similar Coverage (CC_{cov})

For the GTV there was no statistically significant difference in D_{mean}, D_{min}, D_{max}, D_{99}, and D_1 when comparing PB to the plan with similar coverage (CC_{cov}). By design there was no difference in coverage (V_{60Gy}). For the PTV there was a significant difference in D_{mean}, D_{min}, D_{99} and D_1 but not in terms of coverage and D_{max}. In order to achieve similar coverage, MUs delivered and field size in X and Y direction had to be adjusted. Total MUs and average field size in X and Y direction were significantly larger in the CC_{clin} and CC_{cov} plans ($p < 0.001$, Table 7.3). Fields were more often adjusted

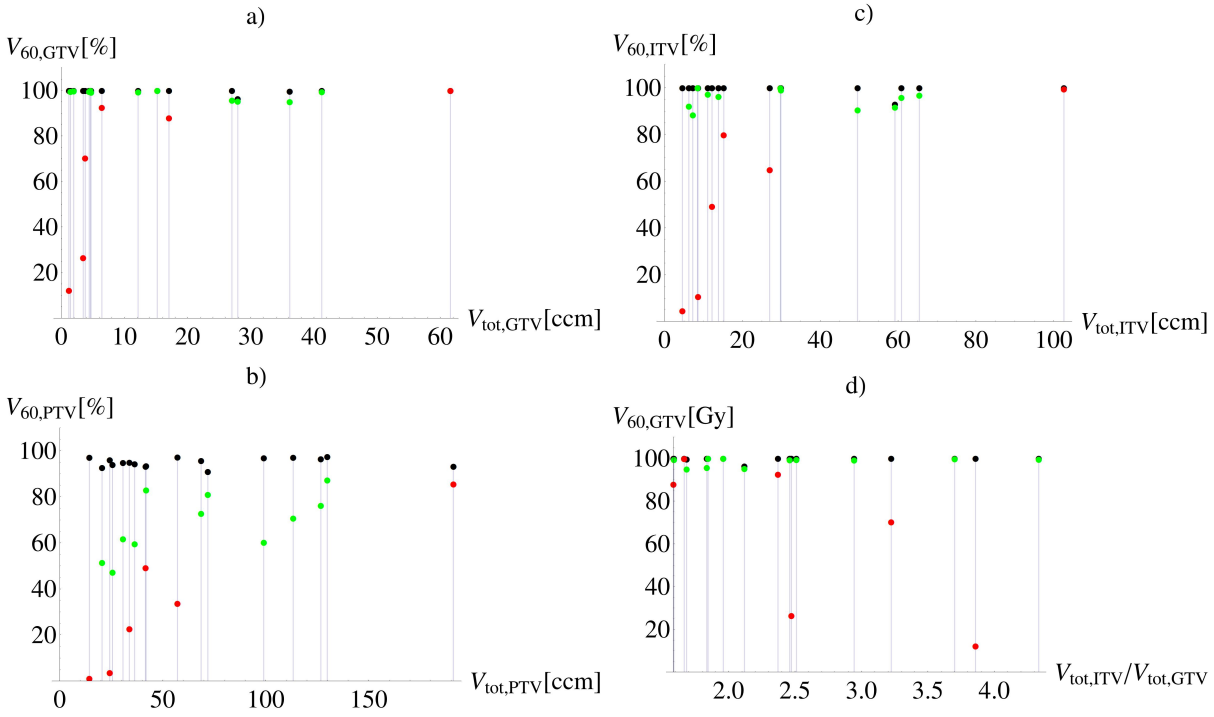


Figure 7.1: Target coverage versus target volume for a) GTV, b) ITV and c) PTV. Figure c) shows GTV coverage versus the ratio of ITV and GTV volume which is indicative of tumor movement. The black markers indicate the original values calculated with pencil beam (PB), while the red and green markers show the recalculated collapsed cone ($CC_{re\text{-}calc}$) patients with central tumors (red) and tumors attached to mediastinum or pleura (green).

craniocaudally than laterally. In the LMU current clinical practice the enhanced CC algorithm is used to calculate dose, and the field size is allowed to increase up to a distance of 10 mm between the jaws and the outer PTV margin in BEV (compared to 5 mm in former plans calculated with PB) to achieve more adequate calculated GTV coverage. The clinically acceptable plans CC_{clin} therefore had acceptable GTV coverage that was not different from coverage calculated with the original PB. However, CC_{clin} plans had decreased D_{min} and D_{99} for the GTV. The field size is usually not increased enough to also achieve the same planned PTV coverage as with PB, especially in areas mainly consisting of air. Thus, by construction, PTV values do not reach values of the original PB plan for CC_{clin} plans.

For OARs, all evaluated parameters were significantly larger in the CC_{cov} plan than

	PB	**CC_{cov}**	**CC_{clin}**
Average relative change of total MU (%)	-	5.0 ± 5.6	5.8 ± 5.9
Average field size X (cm)	4.6 ± 1.3	5.0 ± 1.3	4.9 ± 1.3
Average field size Y (cm)	6.5 ± 1.9	7.2 ± 1.9	6.9 ± 2.0

Table 7.3: Dose and volume parameters for targets and organs at risk for pencil beam and collapsed cone calculations.

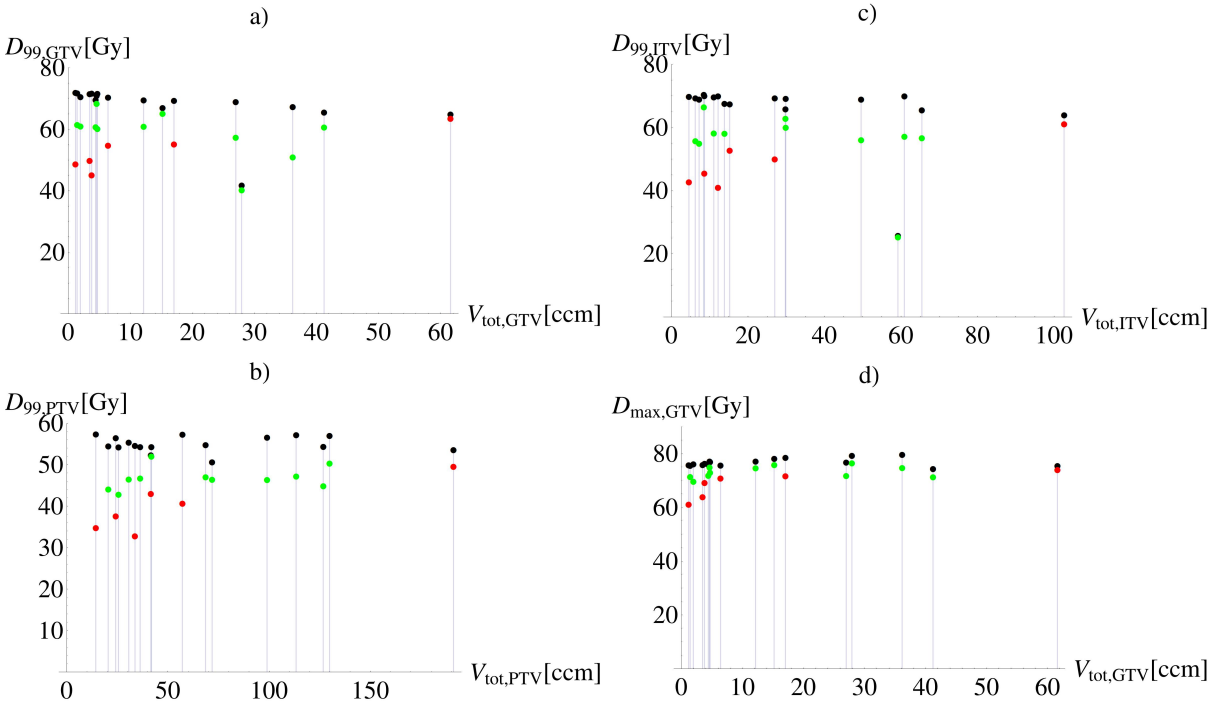


Figure 7.2: D_{99} versus the total absolute volume for a) GTV, b) ITV and c) PTV. Figure d) shows the maximal dose versus the total GTV volume. The black markers indicate the original values calculated with pencil beam (PB), while the red and green markers show the recalculated collapsed cone ($CC_{re\text{-}calc}$) patients with central tumors (red) and tumors attached to mediastinum or pleura (green).

PB and consequently in $CC_{re\text{-}calc}$. This is due to the increased monitor units and field sizes necessary to achieve similar calculated target coverage for the CC_{cov} plan compared to the PB plan. The results for the CC_{cov} and $CC_{re\text{-}calc}$ plans imply that when switching dose calculation algorithms from PB to CC without changing the criteria for an acceptable treatment plan, the OARs will be exposed to increased dose. This is because when using CC the same plan quality in terms of reported target coverage (especially in the PTV) can only be achieved by increasing MUs and field size, which leads to higher dose to OAR. Although the CC calculated dose is closer to the truly delivered dose and thus preferable over PB, many OAR constraints originate from experience with PB. While some of the available recommendations for lung SBRT may suggest adaptation of the dose prescription according to the type of dose-calculation algorithm used that is not nearly always the case for clinical recommendations (Hurkmans *et al.*, 2009).

The difference between dose estimates by using PB and CC in target structures is especially pronounced in the PTV, because the PTV contains tumor tissue as well as air in the planning CT and thus includes large variation of tissue density.

Our data is in the range of previously reported works by Haedinger *et al.* (Haedinger *et al.*, 2005), who found a decrease in D_{mean} of 11.2 % whereas 14.2 % was found in this study. However, a much more drastic change in average PTV coverage of 39.4 volume % was found compared to 7.1 volume % found by Haedinger *et al.*, who used the

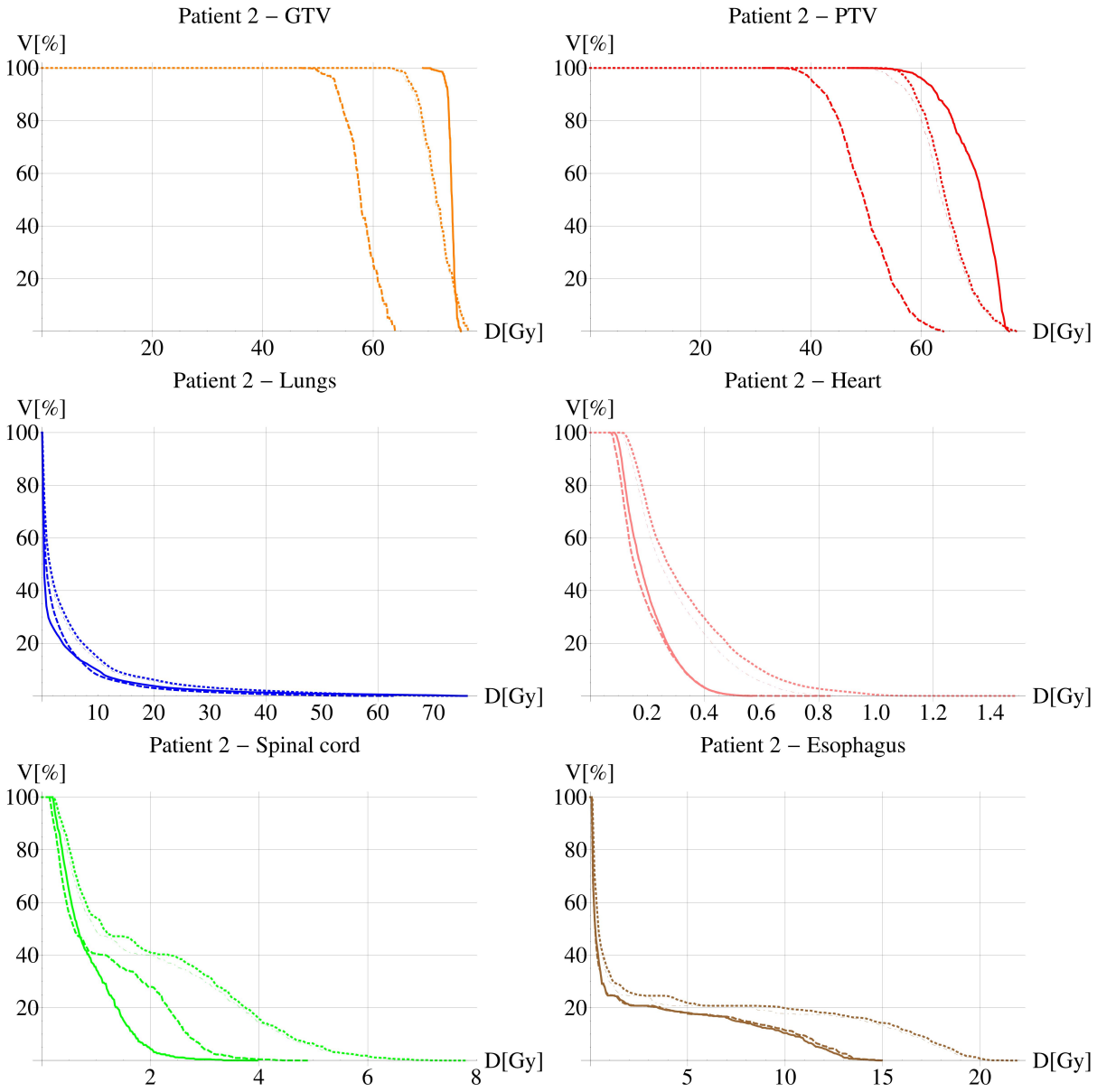


Figure 7.3: Dose volume histograms of GTV, PTV, lungs, heart, spinal cord and esophagus for a patient with extreme differences between PB and CC distributions. The solid, dashed, dotted and dashed-dotted lines represent the PB, $CC_{re\text{-}calc}$, CC_{cov} and CC_{clin} plans respectively.

Helax-TMS treatment planning system. It is therefore essential to keep stressing the importance of the choice of suitable dose calculation algorithms. Although there are a variety of studies that have investigated implications of using PB type versus CC type dose calculation algorithms, e.g. (Haedinger *et al.*, 2005; Traberg Hansen *et al.*, 2005), most publications on the matter consider the PTV and clinical target volume (CTV) mainly or exclusively. However, Guckenberger *et al.* (Guckenberger *et al.*, 2007) have shown that the 4D dose calculated over all breathing phases in the GTV is similar to the dose in the GTV in one single phase for 3D-CRT plans (end-exhalation, end-inhalation, or mid-ventilation phase). This means that the GTV may really be the rel-

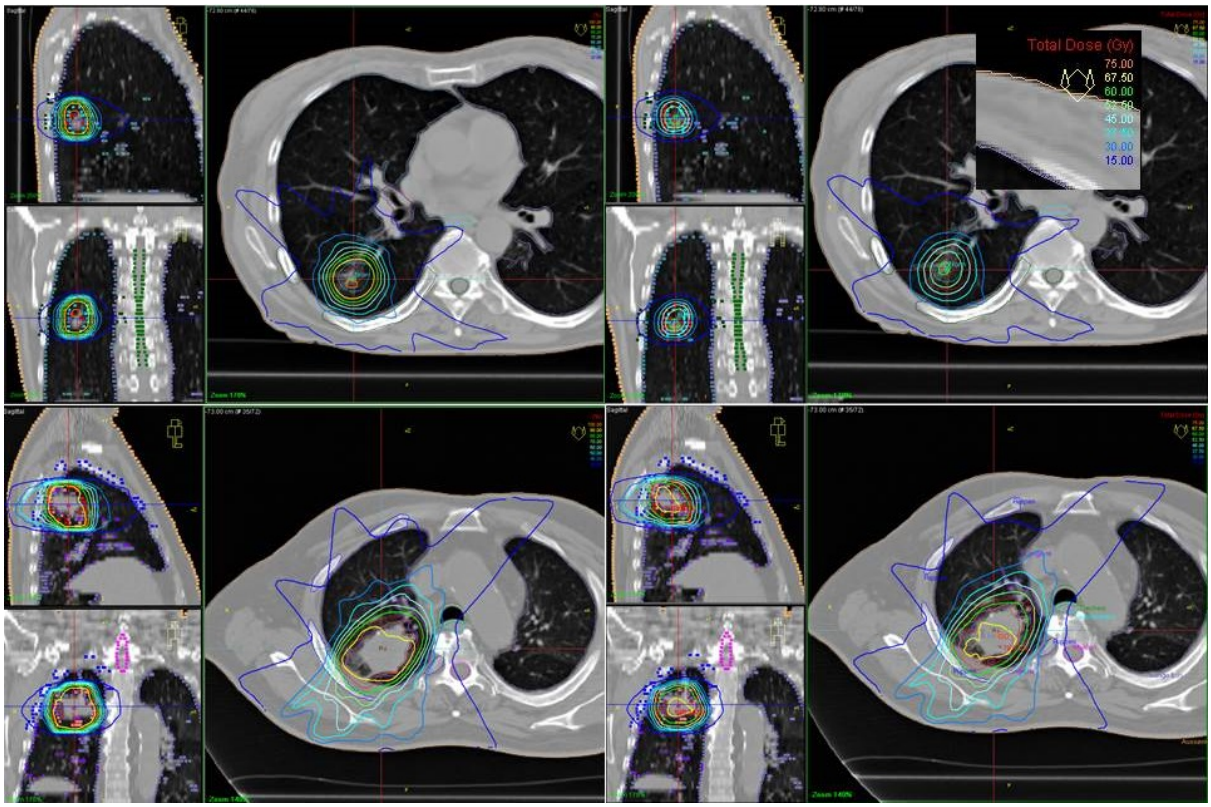


Figure 7.4: Sagittal, coronal and transversal CT view of two sample patients. The top slices show the most extreme case (Patient 2) due to the small tumor volume, whereas the bottom slices depict a less extreme case. Pencil beam (PB) is shown on the left and the plans recalculated with collapsed cone ($CC_{re\text{-}calc}$) on the right. The orange, yellow, bright green, light green, light blue, medium blue, turquoise and dark blue lines represent the 75 Gy, 67.5 Gy , 60 Gy, 52.5 Gy, 45 Gy, 37.5 Gy, 30 Gy and 15 Gy isodoses, respectively.

evant ROI for evaluating such differences resulting from dose calculation algorithms assuming that the fields are opened enough to allow coverage of the GTV in all phases. The present study therefore also includes data for the GTV. Aarup and Dobler (Aarup *et al.*, 2009; Dobler *et al.*, 2006) demonstrated discrepancies in mixed-density phantom studies. They showed that PB algorithms tend to overestimate the target dose, while CC and MC seemed to provide more reliable data compared to measurements. Latifi *et al.* demonstrated that there was a significantly higher rate of reoccurrence when SBRT plans were planned using a PB algorithm for dose calculation than if CC was used (Latifi *et al.*, 2014).

The performance of dose calculation algorithms in lung depends on the use of different patient models (Sikora *et al.*, 2009) in order to account for target motion in the presence of large density inhomogeneities (i.e. static CTs , average CTs, CTs with density overrides, maximum intensity projection) (Bradley *et al.*, 2006; Pan *et al.*, 2013). For the purpose of this study, dose calculation was performed only on a single static planning CT, which is the current clinical practice at LMU Munich for lung SBRT treatment planning using OTP. Furthermore, it should be noted that both the quality of the

beam model and the TPS-specific implementation of the head model can influence the performance of different dose calculation algorithms and impact the calculated dose distribution. The results obtained in this study may therefore be specifically useful for OTP users. For the purpose of this study PB and CC beam models were used that were carefully verified with respect to base data measurements prior to clinical use at LMU Munich.

The large discrepancies resulting from use of different dose calculation algorithms are of special importance when multi-institutional studies are performed. A comparative planning study for the JCOG 0403 protocol showed notable differences between the participating institutions for D_{max}, D_{min}, D_{95}, and the homogeneity index of the PTV, although target definitions and target dose constraints were the same. These inter-institutional deviations were mainly attributed to the different choice dose calculation algorithms used in the institutions (Matsuo *et al.*, 2007). Even if only CC algorithms are used, their quality may depend on the exact implementation of the algorithm in the treatment planning system, and be specific to the release version (Huang *et al.*, 2013; Vanderstraeten *et al.*, 2006). The use of a wide variety of dose prescription modes in stereotactic radiotherapy leads to additional in-transparency when comparing data (Onishi *et al.*, 2007).

More recently started studies require tissue density heterogeneity correction. For Radiation Therapy Oncology Group (RTOG) 0236, SBRT conformal treatment plans that met the dosimetric compliance criteria recommended for RTOG 0813 were generated using XiO/superposition, and then recalculated using MC. Tissue density heterogeneity correction was applied in the initial calculations. V_{20Gy} increased on average by 18% in the MC plans (Li, Galvin, Harrison, Timmerman, Yu and Xiao, 2012).

Although CC algorithms predict dose more accurately than PB algorithms, they can nevertheless deviate from measured dose and dose calculated with MC. Several studies investigated deviations of MC algorithms from measured dose or dose calculated with MC, e.g. (Krieger and Sauer, 2005; Kry *et al.*, 2013; Vanderstraeten *et al.*, 2006). Krieger *et al.* (Krieger and Sauer, 2005) and Kry *et al.* (Kry *et al.*, 2013) found good agreement between MC and measured dose. Thus, in order to fully evaluate the implications of using PB and CC algorithms in the clinic, a comparison with MC would be necessary.

7.4 Conclusion

The use of different calculation algorithms leads to significant changes not only in dose, but also in field size and MUs delivered, if similar target coverage criteria are applied for treatment planning. This has to be taken into account for treatment planning and comparison of data concerning side effects as well as local control in radiotherapy of intrapulmonary lesions. Using CC improves the accuracy of dose calculation in the tumor. However, attempting to cover the parts of the PTV and ITV that mainly consist of air, caused a higher strain on the lung and OARs in this study. In cases with borderline acceptable OAR exposition, OAR plans should be evaluated cautiously. If

data concerning safety margins, OAR limits, maximum or mean dose and PTV/ITV coverage of radiotherapy plans in stereotactic radiotherapy are compared, it is of interest to state which dose calculation algorithm was used. Future projects will include comparison of the resulting CC and PB dose distributions to the MC gold standard, as dose calculated with CC itself can deviate from measured values. Also evaluation of accumulated 4D dose may be necessary to estimate the actual increase of tumor control probability that can be achieved by trying to obtain PTV coverage with CC.

Chapter 8

Conclusions

Since the beginnings of radiation therapy, sparing healthy tissues from the harmful effects of the radiation used to destroy tumor cells has been a vital component of the treatment process. Therefore the need to accurately predict the probability of side effects already became apparent early in the evolution of radiotherapy.

Normal tissue complication probability (NTCP) models provide a probability of a patient experiencing side effects as a function of the dose received by the respective structure. Reliable NTCP models can be useful tools in clinical practice for evaluating treatment plans. Today, a multitude of studies have been undertaken to model NTCP for different treatment sites, structures and endpoints. Yet many challenges remain in the field of NTCP modelling.

This thesis aimed at exploring challenges and uncertainties in NTCP modelling, as well as evaluating how organs at risk dose estimates compare. The choice of treatment technique was shown to influence NTCP model parameters, and that models derived from patients treated with different treatment techniques have limited cross-applicability. The use of NTCP models for patients treated with hypofractionated radiation therapy (HFRT) was also investigated and discussed. It was determined that parameters for HFRT differ from values reported for standard fractionated radiation therapy (SFRT) in the literature, when the conventional linear quadratic (LQ) model was used to account for fractionation. Furthermore, an alternative model for correcting dose for fractionation effects was investigated, which has the potential to describe HFRT and SFRT with the same set of model parameters. Finally, the effects of different dose calculation algorithms on dose to regions of interest were explored, since NTCP models depend on the accuracy of the underlying dose distributions. It was shown that calculated dose can differ significantly between algorithms.

Cross-Applicability of Normal Tissue Complication Probability Models Between Treatment Techniques

The results of this work indicate that NTCP models cannot generally be extrapolated to treatment techniques they were not derived from. This was demonstrated by an NTCP

modelling study using patients treated for prostate cancer. Optimal NTCP model parameters for gastrointestinal (GI) toxicities derived from three dimensional conformal radiation therapy (3D-CRT) patient data and intensity modulated radiation therapy (IMRT) patient data were calculated and compared. It was shown that the models derived from the two populations were not cross-applicable. These results were unexpected, since NTCP model parameters for an endpoint should ideally not depend on the treatment technique.

Comparison of NTCP models derived from different populations is generally difficult because of inherent differences in confounding patient characteristics of the patient cohorts. This study is the first to systematically compare NTCP models for 3D-CRT and IMRT patients. A novel approach was followed with the use of propensity score matched pair analysis between the 3D-CRT and IMRT patients on multiple patient characteristics. This allowed for a more fair comparison of the two groups, and reduction of potential confounding bias.

Subsequent investigations into the source(s) of discrepancies between optimal model parameters for the two groups ruled out differences in follow-up time, imaging modalities used, known patient characteristics and errors of planned versus delivered dose. Other potential sources of the discrepancies could be related to the scoring of toxicity, confounding patient characteristics not accounted for by matched pair analysis, and insufficiency of the models themselves. It is possible, albeit unlikely, that the patient characteristics that were not recorded were responsible for the different model parameters observed for 3D-CRT and IMRT patients and observer bias in toxicity scoring could not be ruled out retrospectively. Finally, the models themselves may not be capable of describing the two populations with the same model parameters. The models used in this work are phenomenological models rather than mechanistic, and although they may work well when fitted to data from one cohort, are potentially not capable of describing other patient cohorts. Furthermore, they are based on dose volume histograms which omit spatial information contained in the 3D dose distribution that may be necessary for accurate NTCP modelling.

As part of this study, a set of optimal NTCP parameters for GI toxicities in IMRT patients was determined. Although NTCP for GI toxicities in 3D-CRT has been studied intensively, there are few studies on NTCP models for GI toxicities in IMRT patients. This is due to limited data available on patients treated with IMRT, and the low incidence of GI toxicities in the IMRT population. Yet despite the lack of validated NTCP models, IMRT is increasingly used in the clinic. Therefore, the IMRT parameters presented in this thesis provide a basis for future work, however, they do not constitute clinical recommendations without validation in other patient cohorts. Fortunately, the NTCP model parameters derived from the 3D-CRT data overestimated NTCP for IMRT patients, and thus do not pose a risk of increasing side effects when applied to IMRT patients.

The Effects of Hypofractionation on Normal Tissue Complication Probability Models

This work also applied NTCP modelling to radiation pneumonitis in HFRT lung cancer patients. The LQ model was used to correct for the fractionation scheme and the optimal NTCP model parameters were determined. However, the resulting HFRT parameters differ from the reported parameters for SFRT in the literature. The actual optimal model parameters for HFRT suggest that HFRT patients experience pneumonitis at higher doses than SFRT patients. This demonstrates that NTCP models are not cross applicable for different fractionation schemes, even if accounting for fractionation using the LQ model.

In addition to the LQ model, the low dose hyperradiosensitivity (LDHRS) model was used to account for hypofractionation. The LDHRS model offers a biological explanation for the observed decrease in incidence of toxicities seen in the HFRT patients. The LDHRS model can potentially be used to find a common NTCP model for radiation pneumonitis in HFRT and SFRT lung cancer patients that describes both populations with the same model parameters. However, further data are needed to confirm this in the future. This was the first study to use the LDHRS model for NTCP modelling and it was shown that it is feasible to use the model for modelling NTCP for radiation pneumonitis in HFRT of the lungs.

Furthermore, NTCP model parameters for chest wall toxicities including chest wall pain, rib fractures and myositis were determined and discussed in this thesis. These toxicities occur frequently in patients who underwent lung stereotactic body radiation therapy (SBRT), and there are no published studies which investigate this to the best of the author's knowledge. Since SBRT is becoming a standard treatment method for certain forms of lung cancer, it is important to establish a predictive model in order to minimize the morbidity of treatment.

The Impact of Dose Calculation Algorithms on Estimated Dose

In this thesis, the choice of dose calculation algorithm is shown to significantly impact dose volume parameters for both the target volumes and OARs. This is especially true for non-homogeneous tissues. Specifically, using Collapsed Cone (CC) to recalculate dose in lung plans created with pencil beam (PB) can lead to drastic differences in planned dose for regions of interest. When attempting to achieve the same coverage criteria with CC that were achieved with PB, field sizes and monitor units had to be significantly adjusted. These differences could undoubtedly lead to differences in parameters for tumor control probability (TCP) and NTCP models. Thus, the studies in this thesis demonstrate the importance of stating the dose calculation algorithm when presenting NTCP model parameters, especially in the lung. Furthermore, it indicates the importance of using dose calculation algorithms that model scattering more accu-

rately, such as CC or Monte Carlo for calculation of dose distributions used in NTCP modelling for lung radiation therapy.

Final Remarks

The findings in this thesis show that NTCP models are sensitive to treatment technique and fractionation. They also show that the choice of dose calculation algorithms can significantly alter dose to regions of interest and will thus have an impact on NTCP models derived from such dose distributions. Therefore, NTCP models should be carefully validated before extrapolating the model parameters to other populations that differ in treatment technique, fractionation scheme and the dose calculation algorithm used. This becomes especially relevant as the field of radiation therapy rapidly evolves, creating new treatment approaches and incorporating new technologies.

Despite these limitations, NTCP models are invaluable tools in the field of radiation oncology. They provide evidence based, data-driven, quantitative, and objective metrics that can guide clinicians in identifying the risk of side effects. This information, together with TCP, can provide the clinical team with a risk and benefit estimate of a given treatment plan. The NTCP models derived in this work for GI toxicities in IMRT patients, and pneumonitis and chest wall toxicities in lung SBRT patients, have helped create clinical guidelines for constraints to OARs in these types of treatment.

In the future, research pertaining to NTCP modelling may focus on studies aiming to further investigate sources of model uncertainties, and extend the models such that they can be adjusted to individual clinical scenarios, including the treatment technique, fractionation and other factors. When applied appropriately, NTCP models are undoubtedly a crucial tool in radiation therapy, with great potential to improve treatment outcomes, minimize side effects and improve the patient's quality of life.

List of Publications

Journal Publications

Troeller A, Yan D, Marina O, Schulze D, Alber M, Parodi K, Belka C, and Söhn M (2015). Comparison and Limitations of DVH-Based NTCP Models Derived From 3D-CRT and IMRT Data for Prediction of Gastrointestinal Toxicities in Prostate Cancer Patients by Using Propensity Score Matched Pair Analysis, *International Journal of Radiation Oncology, Biology, Physics* 91, 435–443. doi:10.1016/j.ijrobp.2014.09.04.

Troeller A, Garny S, Pachmann S, Kantz S, Gerum S, Manapov F, Ganswindt U, Belka C, Söhn M (2015). Stereotactic radiotherapy of intrapulmonary lesions: comparison of different dose calculation algorithms for Oncentra MasterPlan®. *Radiation Oncology* 10:51. doi:10.1186/s13014-015-0354-3.

Zehentmayr F, Söhn M, Exeli A, Wurstbauer K, **Tröller A**, Deutschmann H, Fastner G, Fussl C, Steininger P, Kranzinger M, Belka C, Studnicka M, Sedlmayer F (2015). Normal tissue complication models for clinically relevant acute esophagitis ($\geq$ grade 2) in patients treated with dose differentiated accelerated radiotherapy (DART-bid). *Radiation Oncology* 10(1):121. doi:10.1186/s13014-015-0429-1.

Kantz S, Söhn M, **Troeller A**, Reiner M, Weingand H, Alber M, Belka C, Ganswindt U (2015). Impact of MLC design and IMRT technique in meningioma and head-and-neck treatments. *Radiation Oncology* 10:184.

Abstracts

Troeller A, Söhn M, Grills I, Guckenberger M, Belderbos J, Sonke J, Hope A, Werner-Wasik M, Xiao Y, Yan D (2015) TH-AB-304-06: Investigation of Fractionation Issues in NTCP Modeling of Pneumonitis: An Analysis of Common NTCP Models for Hypo-Fractionated and Standard-Fractionated Data, *Medical Physics* 42(6):3701. doi:10.1118/1.4926121.

Troeller A, Söhn M, Grills I, Ionascu D, Guckenberger M, Klement R, Belderbos J, Sonke J, Hope A, Bissonnette J, Werner-Wasik M, Xiao Y, Yan D (2014). Are Lung

NTCP Models Different In Hypofractionated And Standard Fractionated Radiotherapy?, *International Journal of Radiation Oncology, Biology, Physics* 90(1):S41-S42. doi:10.1016/j.ijrobp.2014.05.162.

Walter F, Schoenecker S, Freislederer P, Nikolajek K, **Troeller A**, Söhn M, Belka C (2014). The First Clinical Observations for the Application of the Catalyst (TM)-System for the Judgement of the Patient's positioning during Pelvic Radiotherapy. *Strahlentherapie und Onkologie*.

Walter F, Schoenecker S, Freislederer P, Nikolajek K, **Troeller A**, Söhn M, Belka C (2014). Clinical Experience with the use of Catalyst (TM)-System to Assess the Overlay Accuracy in Patients with Breast Cancer. *Strahlentherapie und Onkologie*.

Lutz C, **Tröller A**, Møller D, Khalil A, Söhn M, Hoffmann L, Alber M (2014). Model-free investigation of the dose-volume-response of radiation pneumonitis by principal component analysis. *ESTRO 33, Vienna*.

Troeller A, Söhn M, Yan D (2013). Normal Tissue Complication Probability Modeling of Chronic Hematuria in Prostate Cancer Patients Treated with 3D Conformal Radiotherapy. *International Journal of Radiation Oncology, Biology, Physics* 87(2):S688. doi: http://dx.doi.org/10.1016/j.ijrobp.2013.06.1824.

Meyer K, Baschnagel A, **Troeller A**, Yan D, Chen P, Krauss D (2013). Dosimetric Correlation with Patterns of Local Failure in Locally Advanced Larynx and Hypopharynx Cancers, *International Journal of Radiation Oncology, Biology, Physics* 87(2):S452. doi: http://dx.doi.org/10.1016/j.ijrobp.2013.06.1192.

Troeller A, Söhn M, Yan D (2013). TU-G-108-09: Can NTCP Models Derived From 3D-CRT Data Be Used to Predict Gastrointestinal Toxicities for IMRT Prostate Cancer Patients?, *Medical Physics* 40, 455. doi: http://dx.doi.org/10.1118/1.4815458.

Troeller A, Söhn M, Yan D (2012). Comparison of Normal Tissue Complication Probability (NTCP) Model Outcomes for Conformal Radiotherapy (CRT) versus Intensity Modulated Radiotherapy (IMRT) of Prostate Cancer, *International Journal of Radiation Oncology, Biology, Physics* 84(3):S77. doi: http://dx.doi.org/10.1016/j.ijrobp.2012.07.305.

Troeller A, Söhn M, Yan D (2012). SU-E-T-248: An Extended Generalized Equivalent Uniform Dose Accounting for Dose-Range Dependency of Radiobiological Parameters. *Medical Physics* 39, 3760. doi: http://dx.doi.org/10.1118/1.4735314.

Bibliography

Aarup L R, Nahum A E, Zacharatou C, Juhler-Nøttrup T, Knöös T, Nyström H, Specht L, Wieslander E and Korreman S S (2009). The effect of different lung densities on the accuracy of various radiotherapy dose calculation methods: Implications for tumour coverage, *Radiotherapy and Oncology* **91**(3), 405–414.

Acosta O, Drean G, Ospina J D, Simon A, Haigron P, Lafond C and de Crevoisier R (2013). Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy, *Physics in Medicine and Biology* **58**(8), 2581–2595.

Ahnesjö A (1989). Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media, *Medical Physics* **16**(4), 577–592.

Ahnesjö A and Aspradakis M M (1999). Dose calculations for external photon beams in radiotherapy, *Physics in Medicine and Biology* **44**(11), R99–155.

Akaike H (1974). A new look at the statistical model identification, *Automatic Control, IEEE Transactions on* **19**(6), 716–723.

Alber M and Belka C (2006). A normal tissue dose response model of dynamic repair processes, *Physics in Medicine and Biology* **51**(1), 153–172.

Alsadius D, Hedelin M, Johansson K A, Pettersson N, Wilderäng U, Lundstedt D and Steineck G (2011). Tobacco smoking and long-lasting symptoms from the bowel and the anal-sphincter region after radiotherapy for prostate cancer, *Radiotherapy and Oncology* **101**(3), 495–501.

Andisheh B, Edgren M, Belkić D, Mavroidis P, Brahme A and Lind B K (2013). A comparative analysis of radiobiological models for cell surviving fractions at high doses, *Technology in Cancer Research & Treatment* **12**(2), 183–192.

Andolino D L, Forquer J A, Henderson M A, Barriger R B, Shapiro R H, Brabham J G, Johnstone P A, Cardenes H R and Fakiris A J (2011). Chest Wall Toxicity After Stereotactic Body Radiotherapy for Malignant Lesions of the Lung and Liver, *International Journal of Radiation Oncology, Biology, Physics* **80**(3), 692–697.

Andreo P (1991). Monte Carlo techniques in medical radiation physics, *Physics in Medicine and Biology* **36**(7), 861–920.

Beckmann E C (2006). CT scanning the early days, *The British Journal of Radiology* **79**(937), 5–8.

Beli I, Koukourakis G, Platoni K, Tolia M, Kelekis N, Kouvaris J, Syrigos C, Mystakidou K, Varveris C and Kouloulias V (2010). Hypofractionated radiotherapy in non small cell lung cancer: a review of the current literature, *Reviews on Recent Clinical Trials* **5**(2), 103–111.

Bergonié J and Tribondeau L (1906). De Quelques Résultats de la Radiotherapie et Essai de Fixation d'une Technique Rationnelle, *Comptes-Rendus des Séances de l'Académie des Sciences* **143**, 983–985.

Berthelsen A K (2007). What's new in target volume definition for radiologists in ICRU Report 71? How can the ICRU volume definitions be integrated in clinical practice?, *Cancer Imaging* **7**(1), 104–116.

Bethesda (2014). SEER Cancer Statistics Factsheets: Prostate Cancer., *National Cancer Institute* .

Billiet C, Peeters S and De Ruysscher D (2014). Focus on treatment complications and optimal management: radiation oncology, *Translational Lung Cancer Research* **3**(3), 187–191.

Bongers E M, Haasbeek C J, Lagerwaard F J, Slotman B J and Senan S (2011). Incidence and Risk Factors for Chest Wall Toxicity After Risk-Adapted Stereotactic Radiotherapy for Early-Stage Lung Cancer, *Journal of Thoracic Oncology* **6**(12), 2052–2057.

Borst G R, Ishikawa M, Nijkamp J, Hauptmann M, Shirato H, Onimaru R, van den Heuvel M M, Belderbos J, Lebesque J V and Sonke J J (2009). Radiation pneumonitis in patients treated for malignant pulmonary lesions with hypofractionated radiation therapy, *Radiotherapy and Oncology* **91**(3), 307–313.

Bradley J D, Nofal A N, El Naqa I M, Lu W, Liu J, Hubenschmidt J, Low D A, Drzymala R E and Khullar D (2006). Comparison of helical, maximum intensity projection (MIP), and averaged intensity (AI) 4D CT imaging for stereotactic body radiation therapy (SBRT) planning in lung cancer, *Radiotherapy and Oncology* **81**(3), 264–268.

Brenner D J, Martinez A A, Edmundson G K, Mitchell C, Thames H D and Armour E P (2002). Direct evidence that prostate tumors show high sensitivity to fractionation (low alpha/beta ratio), similar to late-responding normal tissue, *International Journal of Radiation Oncology, Biology, Physics* **52**(1), 6–13.

Brock J, Ashley S, Bedford J, Nioutsikou E, Partridge M and Brada M (2008). Review of Hypofractionated Small Volume Radiotherapy for Early-stage Non-small Cell Lung Cancer, *Clinical Oncology* **20**(9), 666–676.

Budäus L, Bolla M, Bossi A, Cozzarini C, Crook J, Widmark A and Wiegel T (2012). Functional outcomes and complications following radiation therapy for prostate cancer: a critical analysis of the literature, *European Urology* **61**(1), 112–127.

Buettner F, Gulliford S L, Webb S, Sydes M R, Dearnaley D P and Partridge M (2009). Assessing correlations between the spatial distribution of the dose to the rectal wall and late rectal toxicity after prostate radiotherapy: an analysis of data from the MRC RT01 trial (ISRCTN 47772397), *Physics in Medicine and Biology* **54**(21), 6535–6548.

Burman C, Kutcher G J, Emami B and Goitein M (1991). Fitting of normal tissue tolerance data to an analytic function, *International Journal of Radiation Oncology, Biology, Physics* **21**(1), 123–135.

Burnham K P (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection, *Sociological Methods & Research* **33**(2), 261–304.

Burnham K P and Anderson D R, eds (2004). *Model Selection and Multimodel Inference*, Springer New York, New York, NY.

Böhning D, Kuhnert R and Rattanasiri S (2008). *Meta-analysis of binary data using profile likelihood*, CRC Press, Boca Raton.

Childress N, Stephens E, Eklund D and Zhang M (2012). Mobius3D White Paper: Dose Calculation Algorithm, *Mobius Medical Systems, LP Rev. 0* .

Christodoulou M, McCloskey P, Stones N, Bayman N, Burt P, Chittalia A, Harris M, Lee L, Pemberton L, Sheikh H, Swindell R and Faivre-Finn C (2014). Investigation of a Patient Reported Outcome tool to assess radiotherapy-related toxicity prospectively in patients with lung cancer, *Radiotherapy and Oncology* **112**(2), 244–249.

Claude L, Pérol D, Ginestet C, Falchero L, Arpin D, Vincent M, Martel I, Hominal S, Cordier J F and Carrie C (2004). A prospective study on radiation pneumonitis following conformal radiation therapy in non-small-cell lung cancer: clinical and dosimetric factors analysis, *Radiotherapy and Oncology* **71**(2), 175–181.

Collett D (2003). *Modelling binary data*, Chapman & Hall/CRC, Boca Raton.

Cox D R (1989). *Analysis of binary data*, number 32 *in* 'Monographs on statistics and applied probability', 2nd ed edn, Chapman and Hall, London ; New York.

Despeignes V (1896). Observation concernant un cas de cancer de l'estomac traité par les rayons Röntgen., *Lyon Med. J.* **82**, 428–430.

Dhabaan A, Schreibmann E, Siddiqi A, Elder E, Fox T, Ogunleye T, Esiashvili N, Curran W, Crocker I and Shu H K (2012). Six degrees of freedom CBCT-based positioning for intracranial targets treated with frameless stereotactic radiosurgery, *Journal of Applied Clinical Medical Physics* **13**(6).

Dobler B, Walter C, Knopf A, Fabri D, Loeschel R, Polednik M, Schneider F, Wenz F and Lohr F (2006). Optimization of extracranial stereotactic radiation therapy of small lung lesions using accurate dose calculation algorithms, *Radiation Oncology (London, England)* **1**, 45.

Efron B and Tibshirani R J (1998). *An introduction to the bootstrap*, number 57 *in* 'Monographs on statistics and applied probability', nachdr. edn, Chapman & Hall, Boca Raton, Fla.

Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrider J E, Shank B, Solin L J and Wesson M (1991). Tolerance of normal tissue to therapeutic irradiation, *International Journal of Radiation Oncology, Biology, Physics* **21**(1), 109–122.

Fellin G, Rancati T, Fiorino C, Vavassori V, Antognoni P, Baccolini M, Bianchi C, Cagna E, Borca V C, Girelli G, Iacopino B, Maliverni G, Mauro F A, Menegotti L, Monti A F, Romani F, Stasi M and Valdagni R (2013). Long term rectal function after high-dose prostatecancer radiotherapy: Results from a prospective cohort study, *Radiotherapy and Oncology* .

Fiorino C, Valdagni R, Rancati T and Sanguineti G (2009). Dose-volume effects for normal tissues in external radiotherapy: pelvis, *Radiotherapy and Oncology* **93**(2), 153–167.

Fowler J F (1989). The linear-quadratic formula and progress in fractionated radiotherapy, *The British Journal of Radiology* **62**(740), 679–694.

Gilbert A, Ziegler L, Martland M, Davidson S, Efficace F, Sebag-Montefiore D and Velikova G (2015). Systematic Review of Radiation Therapy Toxicity Reporting in Randomized Controlled Trials of Rectal Cancer: A Comparison of Patient-Reported Outcomes and Clinician Toxicity Reporting, *International Journal of Radiation Oncology, Biology, Physics* **92**(3), 555–567.

Graham M V, Purdy J A, Emami B, Harms W, Bosch W, Lockett M A and Perez C A (1999). Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC), *International Journal of Radiation Oncology, Biology, Physics* **45**(2), 323–329.

Guckenberger M (2011). Image-guided Radiotherapy Based on Kilovoltage Cone-beam Computed Tomography — A Review of Technology and Clinical Outcome, *European Oncology & Haematology* **07**(02), 121.

Guckenberger M, Meyer J, Vordermark D, Baier K, Wilbert J and Flentje M (2006). Magnitude and clinical relevance of translational and rotational patient setup errors: a cone-beam CT study, *International Journal of Radiation Oncology, Biology, Physics* **65**(3), 934–942.

Guckenberger M, Wilbert J, Krieger T, Richter A, Baier K, Meyer J and Flentje M (2007). Four-dimensional treatment planning for stereotactic body radiotherapy, *International Journal of Radiation Oncology, Biology, Physics* **69**(1), 276–285.

Gulliford S L, Partridge M, Sydes M R, Webb S, Evans P M and Dearnaley D P (2012). Parameters for the Lyman Kutcher Burman (LKB) model of Normal Tissue Complication Probability (NTCP) for specific rectal complications observed in clinical practise, *Radiotherapy and Oncology* **102**(3), 347–351.

Guo S and Fraser M W (2015). *Propensity score analysis: statistical methods and applications*.

Haedinger U, Krieger T, Flentje M and Wulf J (2005). Influence of calculation model on dose distribution in stereotactic radiotherapy for pulmonary targets, *International Journal of Radiation Oncology, Biology, Physics* **61**(1), 239–249.

Hall E J and Giaccia A J (2006). *Radiobiology for the radiologist*, Lippincott Williams & Wilkins, Philadelphia.

Hamilton C S, Denham J W, O'Brien M, Ostwald P, Kron T, Wright S and Drr W (1996). Underprediction of human skin erythema at low doses per fraction by the linear quadratic model, *Radiotherapy and Oncology* **40**(1), 23–30.

Hardcastle N, Davies A, Foo K, Miller A and Metcalfe P E (2010). Rectal dose reduction with IMRT for prostate radiotherapy, *Journal of medical imaging and radiation oncology* **54**(3), 235–248.

Hosmer D W and Lemesbow S (1980). Goodness of fit tests for the multiple logistic regression model, *Communications in Statistics - Theory and Methods* **9**(10), 1043–1069.

Hounsfield G N (1980). Nobel Award address. Computed medical imaging, *Medical Physics* **7**(4), 283–290.

Huang J Y, Eklund D, Childress N L, Howell R M, Mirkovic D, Followill D S and Kry S F (2013). Investigation of various energy deposition kernel refinements for the convolution/superposition method, *Medical Physics* **40**(12), 121721.

Hurkmans C W, Cuijpers J P, Lagerwaard F J, Widder J, van der Heide U A, Schuring D and Senan S (2009). Recommendations for implementing stereotactic radiotherapy in peripheral stage IA non-small cell lung cancer: report from the Quality Assurance Working Party of the randomised phase III ROSEL study, *Radiation Oncology (London, England)* **4**, 1.

International Commission on Radiation Units and Measurements, ed. (1999). *Prescribing, recording, and reporting photon beam therapy*, number 62 *in* 'ICRU report', International Commission on Radiation Units and Measurements, Bethesda, Md.

Jackson A, Skwarchuk M W, Zelefsky M J, Cowen D M, Venkatraman E S, Levegrun S, Burman C M, Kutcher G J, Fuks Z, Liebel S A and Ling C C (2001). Late rectal bleeding after conformal radiotherapy of prostate cancer. II. Volume effects and dose-volume histograms, *International Journal of Radiation Oncology, Biology, Physics* **49**(3), 685–698.

Jackson A, Ten Haken R K, Robertson J M, Kessler M L, Kutcher G J and Lawrence T S (1995). Analysis of clinical complication data for radiation hepatitis using a parallel architecture model, *International Journal of Radiation Oncology, Biology, Physics* **31**(4), 883–891.

Jaffray D A (2012). Image-guided radiotherapy: from current concept to future perspectives, *Nature Review Clinical Oncology* **9**(12), 688–699.

Jaffray D A and Siewerdsen J H (2000). Cone-beam computed tomography with a flat-panel imager: initial performance characterization, *Medical Physics* **27**(6), 1311–1323.

Joiner M C, Denekamp J and Maughan R L (1986). The use of 'top-up' experiments to investigate the effect of very small doses per fraction in mouse skin, *International Journal of Radiation Biology and Related Studies in Physics, Chemistry, and Medicine* **49**(4), 565–580.

Joiner M C, Lambin P, Malaise E P, Robson T, Arrand J E, Skov K A and Marples B (1996). Hypersensitivity to very-low single radiation doses: its relationship to the adaptive response and induced radioresistance, *Mutation Research* **358**(2), 171–183.

Kan M W K, Yu P K N and Leung L H T (2013). A Review on the Use of Grid-Based Boltzmann Equation Solvers for Dose Calculation in External Photon Beam Treatment Planning, *BioMed Research International* **2013**, 1–10.

Khan F M (2010). *The physics of radiation therapy*, 4. ed edn, Wolters Kluwer [u.a.], Philadelphia.

Kong F M, Hayman J A, Griffith K A, Kalemkerian G P, Arenberg D, Lyons S, Turrisi A, Lichter A, Fraass B, Eisbruch A, Lawrence T S and Ten Haken R K (2006). Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): predictors for radiation pneumonitis and fibrosis, *International Journal of Radiation Oncology, Biology, Physics* **65**(4), 1075–1086.

Krieger H (2012). *Grundlagen der Strahlungsphysik und des Strahlenschutzes*, Studium, 4., überarb. und erw. aufl edn, Springer-Spektrum, Wiesbaden.

Krieger T and Sauer O A (2005). Monte Carlo- versus pencil-beam-/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom, *Physics in Medicine and Biology* **50**(5), 859–868.

Kry S F, Alvarez P, Molineu A, Amador C, Galvin J and Followill D S (2013). Algorithms Used in Heterogeneous Dose Calculations Show Systematic Differences as Measured With the Radiological Physics Center's Anthropomorphic Thorax Phantom Used for RTOG Credentialing, *International Journal of Radiation Oncology, Biology, Physics* **85**(1), e95–e100.

Kumar R J, Barqawi A and Crawford E D (2005). Adverse events associated with hormonal therapy for prostate cancer, *Reviews in urology* **7 Suppl 5**, S37–43.

Kutcher G J and Burman C (1989). Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method, *International Journal of Radiation Oncology, Biology, Physics* **16**(6), 1623–1630.

Kwa S L, Lebesque J V, Theuws J C, Marks L B, Munley M T, Bentel G, Oetzel D, Spahn U, Graham M V, Drzymala R E, Purdy J A, Lichter A S, Martel M K and Ten Haken R K (1998). Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients, *International Journal of Radiation Oncology, Biology, Physics* **42**(1), 1–9.

Latifi K, Oliver J, Baker R, Dilling T J, Stevens C W, Kim J, Yue B, DeMarco M, Zhang G G, Moros E G and Feygelman V (2014). Study of 201 Non-Small Cell Lung Cancer Patients Given Stereotactic Ablative Radiation Therapy Shows Local Control Dependence on Dose Calculation Algorithm, *International Journal of Radiation Oncology, Biology, Physics* **88**(5), 1108–1113.

Li J, Galvin J, Harrison A, Timmerman R, Yu Y and Xiao Y (2012). Dosimetric verification using monte carlo calculations for tissue heterogeneity-corrected conformal treatment plans following RTOG 0813 dosimetric criteria for lung cancer stereotactic body radiotherapy, *International Journal of Radiation Oncology, Biology, Physics* **84**(2), 508–513.

Li X A (2011). *Adaptive radiation therapy*, CRC Press, Boca Raton, Fla.

Li X A, Alber M, Deasy J O, Jackson A, Jee K W K, Marks L B, Martel M K, Mayo C, Moiseenko V, Nahum A E and others (2012). The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPMa), *Medical Physics* **39**(3), 1386–1409.

Luxton G, Hancock S L and Boyer A L (2004). Dosimetry and radiobiologic model comparison of IMRT and 3D conformal radiotherapy in treatment of carcinoma of the prostate, *International Journal of Radiation Oncology, Biology, Physics* **59**(1), 267–284.

Lyman J T (1985). Complication probability as assessed from dose-volume histograms, *Radiation Research. Supplement* **8**, S13–19.

Mackie T R, Bielajew A F, Rogers D W and Battista J J (1988). Generation of photon energy deposition kernels using the EGS Monte Carlo code, *Physics in Medicine and Biology* **33**(1), 1–20.

Mangona V, Grills I, Lindamood K, Schulze D, Aneese A, Ionascu D, Marina O, Robertson J and Yan D (2013). Predictors of Chest Wall and Extrapulmonary Soft Tissue Toxicity After Stereotactic Lung Radiation Therapy: A Dosimetric Analysis, *International Journal of Radiation Oncology, Biology, Physics* **87**(2), S43.

Marks L B, Yorke E D, Jackson A, Ten Haken R K, Constine L S, Eisbruch A, Bentzen S M, Nam J and Deasy J O (2010). Use of normal tissue complication probability models in the clinic, *International Journal of Radiation Oncology, Biology, Physics* **76**(3 Suppl), S10–19.

Martel M K, Ten Haken R K, Hazuka M B, Turrisi A T, Fraass B A and Lichter A S (1994). Dose-volume histogram and 3-D treatment planning evaluation of patients with pneumonitis, *International Journal of Radiation Oncology, Biology, Physics* **28**(3), 575–581.

Matsuo Y, Takayama K, Nagata Y, Kunieda E, Tateoka K, Ishizuka N, Mizowaki T, Norihisa Y, Sakamoto M, Narita Y, Ishikura S and Hiraoka M (2007). Interinstitutional variations in planning for stereotactic body radiation therapy for lung cancer, *International Journal of Radiation Oncology, Biology, Physics* **68**(2), 416–425.

Metcalfe P, Kron T and Hoban P (1997). *The physics of radiotherapy X-rays from linear accelerators*, Medical Physics Pub., Madison, Wis.

Michalski J M, Gay H, Jackson A, Tucker S L and Deasy J O (2010). Radiation Dose–Volume Effects in Radiation-Induced Rectal Injury, *International Journal of Radiation Oncology, Biology, Physics* **76**(3), S123–S129.

Mohammed N, Kestin L, Ghilezan M, Krauss D, Vicini F, Brabbins D, Gustafson G, Ye H and Martinez A (2012). Comparison of acute and late toxicities for three modern high-dose radiation treatment techniques for localized prostate cancer, *International Journal of Radiation Oncology, Biology, Physics* **82**(1), 204–212.

Mohan R, Chui C and Lidofsky L (1986). Differential pencil beam dose computation model for photons, *Medical Physics* **13**(1), 64–73.

Munbodh R, Jackson A, Bauer J, Schmidtlein C R and Zelefsky M J (2008). Dosimetric and anatomic indicators of late rectal toxicity after high-dose intensity modulated radiation therapy for prostate cancer, *Medical Physics* **35**(5), 2137–2150.

Murray L, Karakaya E, Hinsley S, Naisbitt M, Lilley J, Snee M, Clarke K, Musunuru H B, Ramasamy S, Turner R and Franks K (2016). Lung stereotactic ablative radiotherapy (SABR): dosimetric considerations for chest wall toxicity, *The British Journal of Radiology* **89**(1058), 20150628.

Mutter R W, Liu F, Abreu A, Yorke E, Jackson A and Rosenzweig K E (2012). Dose–Volume Parameters Predict for the Development of Chest Wall Pain After Stereotactic Body Radiation for Lung Cancer, *International Journal of Radiation Oncology, Biology, Physics* **82**(5), 1783–1790.

National Cancer Institute (2006). *Common Terminology Criteria for Adverse Events v3.0.*, NIH publication, Department of Health and Human Services, National Institutes of Health, National Cancer Institute.

Niemierko A (1999). A generalized concept of equivalent uniform dose, *Medical Physics* .

Olsson C E, Alsadius D, Pettersson N, Tucker S L, Wilderäng U, Johansson K A and Steineck G (2015). Patient-reported sexual toxicity after radiation therapy in long-term prostate cancer survivors, *British Journal of Cancer* **113**(5), 802–808.

Onishi H, Shirato H, Nagata Y, Hiraoka M, Fujino M, Gomi K, Niibe Y, Karasawa K, Hayakawa K, Takai Y and others (2007). Hypofractionated stereotactic radiotherapy (HypoFXSRT) for stage I non-small cell lung cancer: updated results of 257 patients in a Japanese multi-institutional study, *Journal of Thoracic Oncology* **2**(7), S94–S100.

Pan T, Riegel A C, Ahmad M U, Sun X, Chang J Y and Luo D (2013). New weighted maximum-intensity-projection images from cine CT for delineation of the lung tumor plus motion, *Medical Physics* **40**(6), 061901.

Pawitan Y (2001). *In All Likelihood Statistical Modelling and Inference Using Likelihood.*, OUP Oxford, Oxford.

Peeters S T H, Hoogeman M S, Heemsbergen W D, Hart A A M, Koper P C M and Lebesque J V (2006). Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling, *International Journal of Radiation Oncology, Biology, Physics* **66**(1), 11–19.

Peeters S T H, Lebesque J V, Heemsbergen W D, van Putten W L J, Slot A, Dielwart M F H and Koper P C M (2006). Localized volume effects for late rectal and anal toxicity after radiotherapy for prostate cancer, *International Journal of Radiation Oncology, Biology, Physics* **64**(4), 1151–1161.

Rancati T, Fiorino C, Gagliardi G, Cattaneo G M, Sanguineti G, Borca V C, Cozzarini C, Fellin G, Foppiano F, Girelli G, Menegotti L, Piazzolla A, Vavassori V and Valdagni R (2004). Fitting late rectal bleeding data using different NTCP models: results from an Italian multi-centric study (AIROPROS0101), *Radiotherapy and Oncology* **73**(1), 21–32.

Ray K, Sibson N and Kiltie A (2015). Treatment of Breast and Prostate Cancer by Hypofractionated Radiotherapy: Potential Risks and Benefits, *Clinical Oncology* **27**(7), 420–426.

Read J (1952). Mode of addition of x-ray doses given with different oxygen concentrations, *The British Journal of Radiology* **25**(294), 336–338.

Regaud C and Ferroux R (1927). Discordance entre les effects des rayons X sur les testicules et la peau, implications pour le fractionnement de la dose, *Compt. Rend. Soc. Biol.* **97**, 431–434.

Ricardi U, Badellino S and Filippi A R (2015). Stereotactic radiotherapy for early stage non-small cell lung cancer, *Radiation Oncology Journal* **33**(2), 57.

Rosenbaum P R (2010). *Design of observational studies*, Springer series in statistics, Springer, New York, NY.

Rosenbaum P R and Rubin D B (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika* **70**(1), 41–55.

Rubio C, Morera R, Hernando O, Leroy T and Lartigau S E (2013). Extracranial stereotactic body radiotherapy. Review of main SBRT features and indications in primary tumors, *Reports of Practical Oncology & Radiotherapy* **18**(6), 387–396.

Schulze D, Liang J, Yan D and Zhang T (2009). Comparison of various online IGRT strategies: The benefits of online treatment plan re-optimization, *Radiotherapy and Oncology* **90**(3), 367–376.

Sekhon J S (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for R, *Journal of Statistical Software, Forthcoming* .

Seppenwoolde Y, Lebesque J V, de Jaeger K, Belderbos J S, Boersma L J, Schilstra C, Henning G T, Hayman J A, Martel M K and Ten Haken R K (2003). Comparing different NTCP models that predict the incidence of radiation pneumonitis, *International Journal of Radiation Oncology, Biology, Physics* **55**(3), 724–735.

Sikora M, Muzik J, Söhn M, Weinmann M and Alber M (2009). Monte Carlo vs. Pencil Beam based optimization of stereotactic lung IMRT, *Radiation Oncology* **4**(1), 64.

Singh B, Arrand J E and Joiner M C (1994). Hypersensitive response of normal human lung epithelial cells at low radiation doses, *International Journal of Radiation Biology* **65**(4), 457–464.

Solanki A A and Liauw S L (2013). Tobacco use and external beam radiation therapy for prostate cancer: Influence on biochemical control and late toxicity, *Cancer* **119**(15), 2807–2814.

Sonke J, Scheenstra A, Grills I, Hope A, Guckenberger M, Werner-Wasik M, Belderbos J, Kestin L, Bissonnette J and Yan D (2011). NTCP Modeling for Radiation Pneumonitis after SBRT for Malignant Pulmonary Lesions: Results of a Multi-institutional Analysis, *International Journal of Radiation Oncology, Biology, Physics* **81**(2), S28–S29.

Srinivasan K, Mohammadi M and Shepherd J (2014). Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review, *Polish Journal of Radiology* **79**, 181–193.

Stewart B W, Wild C, International Agency for Research on Cancer and World Health Organization (2014). *World cancer report 2014*, International Agency for Research on Cancer ; Distributed by WHO Press, Lyon; Geneva.

Söhn M, Alber M and Yan D (2007). Principal component analysis-based pattern analysis of dose-volume histograms and influence on rectal toxicity, *International Journal of Radiation Oncology, Biology, Physics* **69**(1), 230–239.

Söhn M, Yan D, Liang J, Meldolesi E, Vargas C and Alber M (2007). Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models, *International Journal of Radiation Oncology, Biology, Physics* **67**(4), 1066–1073.

Teoh M, Clark C H, Wood K, Whitaker S and Nisbet A (2011). Volumetric modulated arc therapy: a review of current literature and clinical use in practice, *The British Journal of Radiology* **84**(1007), 967–996.

Thariat J, Hannoun-Levi J M, Sun Myint A, Vuong T and Gérard J P (2012). Past, present, and future of radiotherapy for the benefit of patients, *Nature Reviews Clinical Oncology* **10**(1), 52–60.

Thibault I, Chiang A, Erler D, Yeung L, Poon I, Kim A, Keller B, Lochray F, Jain S, Soliman H and Cheung P (2016). Predictors of Chest Wall Toxicity after Lung Stereotactic Ablative Radiotherapy, *Clinical Oncology* **28**(1), 28–35.

Thor M, Apte A, Deasy J O and Muren L P (2013). Statistical simulations to estimate motion-inclusive dose-volume histograms for prediction of rectal morbidity following radiotherapy, *Acta Oncologica (Stockholm, Sweden)* **52**(3), 666–675.

Traberg Hansen A, Petersen J B, Høyer M and Christensen J J (2005). Comparison of two dose calculation methods applied to extracranial stereotactic radiotherapy treatment planning, *Radiotherapy and Oncology* **77**(1), 96–98.

Troeller A, Soehn M and Yan D (2012). Comparison of Normal Tissue Complication Probability (NTCP) Model Outcomes for Conformal Radiation Therapy (CRT) Versus Intensity Modulated Radiation Therapy (IMRT) of Prostate Cancer, *International Journal of Radiation Oncology, Biology, Physics* **84**(3), S77.

Tucker S L, Cheung R, Dong L, Liu H H, Thames H D, Huang E H, Kuban D and Mohan R (2004). Dose-volume response analyses of late rectal bleeding after radiotherapy for prostate cancer, *International Journal of Radiation Oncology, Biology, Physics* **59**(2), 353–365.

Tucker S L, Jin H, Wei X, Wang S, Martel M K, Komaki R, Liu H H, Mohan R, Chen Y, Cox J D and Liao Z (2010). Impact of Toxicity Grade and Scoring System on the Relationship Between Mean Lung Dose and Risk of Radiation Pneumonitis in a Large Cohort of Patients With Non–Small Cell Lung Cancer, *International Journal of Radiation Oncology, Biology, Physics* **77**(3), 691–698.

Tucker S L, Liu H H, Liao Z, Wei X, Wang S, Jin H, Komaki R, Martel M K and Mohan R (2008). Analysis of Radiation Pneumonitis Risk Using a Generalized Lyman Model, *International Journal of Radiation Oncology, Biology, Physics* **72**(2), 568–574.

Valdagni R, Vavassori V, Rancati T, Fellin G, Baccolini M, Bianchi C, Cagna E, Gabriele P, Mauro F, Menegotti L, Monti A F, Stasi M and Fiorino C (2012). Increasing the

risk of late rectal bleeding after high-dose radiotherapy for prostate cancer: the case of previous abdominal surgery. Results from a prospective trial, *Radiotherapy and Oncology* **103**(2), 252–255.

Vanderstraeten B, Reynaert N, Paelinck L, Madani I, De Wagter C, De Gersem W, De Neve W and Thierens H (2006). Accuracy of patient dose calculation for lung IMRT: A comparison of Monte Carlo, convolution/superposition, and pencil beam computations, *Medical Physics* **33**(9), 3149–3158.

Vargas C, Yan D, Kestin L L, Krauss D, Lockman D M, Brabbins D S and Martinez A A (2005). Phase II dose escalation study of image-guided adaptive radiotherapy for prostate cancer: use of dose-volume constraints to achieve rectal isotoxicity, *International Journal of Radiation Oncology, Biology, Physics* **63**(1), 141–149.

Verellen D, Ridder M D and Storme G (2008). A (short) history of image-guided radiotherapy, *Radiotherapy and Oncology* **86**(1), 4–13.

Voroney J P J, Hope A, Dahele M R, Purdy T, Franks K N, Pearson S, Cho J B, Sun A, Payne D G, Bissonnette J P, Bezjak A and Brade A M (2009). Chest Wall Pain and Rib Fracture after Stereotactic Radiotherapy for Peripheral Non-small Cell Lung Cancer, *Journal of Thoracic Oncology* **4**(8), 1035–1037.

Whelan T J, Kim D H and Sussman J (2008). Clinical experience using hypofractionated radiation schedules in breast cancer, *Seminars in Radiation Oncology* **18**(4), 257–264.

Withers H R (1975). The Four R's of Radiotherapy, *in Advances in Radiation Biology*, Vol. 5, Elsevier, pp. 241–271.

Woo M K, Cunningham J R and Jezioranski J J (1990). Extending the concept of primary and scatter separation to the condition of electronic disequilibrium, *Medical Physics* **17**(4), 588–595.

Yan D, Lockman D, Brabbins D, Tyburski L and Martinez A (2000). An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer, *International Journal of Radiation Oncology, Biology, Physics* **48**(1), 289–302.

Yan D, Vicini F, Wong J and Martinez A (1997). Adaptive radiation therapy, *Physics in Medicine and Biology* **42**(1), 123–132.

Yorke E D, Jackson A, Rosenzweig K E, Braban L, Leibel S A and Ling C C (2005). Correlation of dosimetric factors and radiation pneumonitis for non–small-cell lung cancer patients in a recently completed dose escalation study, *International Journal of Radiation Oncology, Biology, Physics* **63**(3), 672–682.

Zaorsky N G, Ohri N, Showalter T N, Dicker A P and Den R B (2013). Systematic review of hypofractionated radiation therapy for prostate cancer, *Cancer Treatment Reviews* **39**(7), 728–736.

Zelefsky M J, Levin E J, Hunt M, Yamada Y, Shippy A M, Jackson A and Amols H I (2008). Incidence of late rectal and urinary toxicities after three-dimensional conformal radiotherapy and intensity-modulated radiotherapy for localized prostate cancer, *International Journal of Radiation Oncology, Biology, Physics* **70**(4), 1124–1129.

Declaration

Eidesstattliche Versicherung

Troeller, Almut

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

Normal Tissue Complication Probability Modelling
Influence of Treatment Technique, Fractionation, and Dose Calculation Algorithm

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Muenchen, 13.07.2017

Ort, Datum

Unterschrift Doktorandin/Doktorand