Ingrid Elisabeth Tveten

Dosimetric Predictors of Toxicity in Prostate Cancer Radiotherapy

From Per-Organ to Per-Voxel Analysis

NTNU
Norwegian University of
Science and Technology

Ingrid Elisabeth Tveten

Dosimetric Predictors of Toxicity in Prostate Cancer Radiotherapy

From Per-Organ to Per-Voxel Analysis

Master's thesis in Applied Physics and Mathematics
Supervisor: Signe Danielsen
Co-supervisor: Kathrine Røe Redalen and Kajsa Fridström
June 2021

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

Preface

This master's thesis is submitted for the degree of Applied Physics and Mathematics at the Norwegian University of Science and Technology (NTNU), Trondheim. The work was carried out during the spring of 2021.

The work is based on imaging data and dosimetric data from the RIC study, which aimed to investigate whether reduced safety margins in prostate cancer radiotherapy could lead to a reduction in patient-reported side effects.

I want to thank my team of encouraging, talented supervisors for guiding me through the long and very educational process of writing a master's thesis. A big thank you goes to my main supervisor, Signe Danielsen. I appreciate your sharing of knowledge and experience, and thank you for being available for questions and always enthusiastic about my work. Thank you to Ph.D. candidate and co-supervisor Kajsa Fridström for providing me with data. I appreciate our discussions on all that may have gone wrong and your insight from the clinical perspective. Finally, I want to thank co-supervisor Kathrine Røe Redalen. Thank you for providing me with the opportunity to attend a conference – even when a pandemic is raging! I appreciate your valuable feedback and the genuine interest you show for your students.

I couldn't have asked for more supportive friends. I want to thank my four house-mates for a wonderful year and for making our house a home – who knew spending time at home could be this much fun? Thank you to my classmates and "Cake Fridays" at biophysics for providing energy and motivation at the end of a long week. Finally, a thank you to my family for always being supportive of what I do, even though you are not physicists and probably don't understand why I am so interested. I appreciate your love and encouragement.

Ingrid Elisabeth Tveten
Trondheim, June 11, 2020

Abstract

Objective: As prostate cancer radiotherapy (RT) improves, the survival rate increases and more patients may have to live longer with complications that can potentially reduce quality of life (QoL). New and improved methods for modelling of normal tissue complications from RT are needed to predict outcomes and personalize treatment. The purpose of this project was to create a framework that can be used to analyze the relationship between RT dose distributions and patient-reported outcome measures (PROMs) on a per-voxel level.

Materials/methods: A programming framework was developed in Python. The program was used to evaluate dose distributions and associated PROMs up to 36 months after RT in the RIC study. The RIC study was a **R**andomized trial of daily cone-beam (study arm A) vs standard **I**mage-guided RT (study arm B) in prostate **C**ancer, that sought to investigate whether reduced safety margins in prostate cancer RT could lead to a reduction in complications. First, dose-volume histograms (DVHs) were compared between study arms, and between patients with and without complications. A framework for voxel-based analysis (VBA) was implemented and evaluated by comparison to a traditional normal tissue complication probability (NTCP) model, the Lyman-Kutcher-Burman (LKB) model. The LKB model for NTCP was implemented and its parameters were fitted to the study data. Finally, the patients receiving daily image verification were registered to a template patient and the resulting dose distributions were compared for patients with and without complications.

Results: Evaluation of DVHs showed that patients in arm A received higher doses to all organs at risk (OARs) than patients in arm B. Compared to patients without complications, patients who experienced rectal complications had larger relative volumes of the rectal wall irradiated with doses up to 30 Gy for all patients in the study, and up to 65 Gy for patients receiving daily verification. Fitting of the LKB parameters to the RIC study data yielded a tolerance dose for 50 % complication rate of 78.9 Gy, $n = 0.007$ and $m = 0.053$. The mean Dice coefficient between the registered rectums and the template was 0.39. For patients where the registration was reliable (Dice ≥ 0.5), a region in the inferior, anterior rectum received significantly higher dose for patients with complications than for patients without complications.

Discussion: Since reduced safety margins mean less dose is delivered outside the target, lower dose to OARs of patients having daily verification was expected. The low-dose dependence of rectal complications found when comparing DVHs is at odds with literature suggesting a high-dose dependence of rectal complications. On the contrary, the LKB parameters showed fair agreement with literature values. The literature is how-

ever mostly based on clinician-reported outcome. The implemented image registration algorithm gave highly variable results, and should be further improved by tuning the registration parameters. Nevertheless, per-voxel analysis of registered dose distributions revealed a subregion in the rectum possibly related to rectal complications up to 18 months after RT. However, results are inconclusive and investigations in larger datasets are needed.

The analysis should be extended to time points later than 36 months since late complications may not yet be evident at 36 months. With further improvements in the registration quality, the tools developed for voxel-based analyses have the potential to identify radiosensitive regions in tissue and predict treatment outcome. Such tools can be used for improved treatment planning for both prostate and other cancer types.

Sammendrag

Bakgrunn: Bedre strålebehandling mot prostatakreft gjør at overlevelsesraten øker. Dette fører til at flere pasienter lever lenger med bivirkninger som potensielt kan redusere livskvalitet (QoL). Nye og forbedrede metoder for modellering av komplikasjoner i normalvev som følge av strålebehandling behøves for å kunne forutsi behandlingsutfall og sørge for persontilpasset behandling. Hensikten med dette prosjektet var å utvikle et programmeringsrammeverk som kan benyttes til voxel-basert analyse (VBA) av sammenhengen mellom dosefordeling og pasientrapporterte utfallsmål (PROMs) i stråleterapi.

Materialer/metoder: Et programmeringsverktøy har blitt utviklet i Python. Programmet ble brukt til å evaluere dosefordelinger samt PROMs inntil 36 måneder etter stråleterapi fra RIC-studien (a **R**andomized trial of daily cone-beam vs standard **I**mage-guided radiotherapy in prostate **C**ancer). RIC-studien undersøkte hvorvidt reduserte marginer i stråleterapi av prostata (studiearm B) kunne føre til reduksjon av bivirkninger sammenliknet med standard marginer (studiearm A). Dose-volum-histogrammer (DVH-er) ble sammenliknet mellom studiearmene, og for pasienter med og uten bivirkninger. Et rammeverk for VBA ble implementert og testet ved sammenlikning med en tradisjonell modell for sannsynlighet for normalvevskomplikasjoner (NTCP). NTCP er et estimat på sannsynligheten for bivirkninger etter strålebehandling og kan potensielt brukes som klinisk beslutningsstøtte. Lyman-Kutcher-Burman(LKB)-modellen for NTCP ble implementert og parametrene ble estimert for data fra RIC-studien. Videre ble pasientene som fikk daglig bildeverifikasjon registrert til en anatomisk pasientmal, og de registrerte dosefordelingene ble sammenliknet for pasienter med og uten bivirkninger.

Resultater: DVH-sammenlikninger viste at pasienter i arm A fikk høyere dose til alle risikoorganer (OAR) enn pasienter i arm B. Pasienter med rektumbivirkninger hadde et større relativt bestrålt volum av rektumveggen enn pasienter uten bivirkninger. Sammenhengen ble funnet for doser opp til 30 Gy når alle pasientene ble vurdert samlet, og opp til 65 Gy for pasientene som fikk daglig bildeverifikasjon. Tilpasning av LKB-modellen gav TD50 = 78.9 Gy, $n = 0.007$ og $m = 0.053$. TD50 er estimert toleransedose som fører til komplikasjoner for 50 % av pasientene, den lave verdien for n indikerer en seriearkitektur for rektum, altså at organet er sensitivt for høye doser, mens m indikerer hvor bratt dose-respons-kurven er. Gjennomsnittlig Dice-koeffisient for registrerte rektum sammenliknet med den anatomiske malen var 0.39. For pasienter med en pålitelig registrering (Dice $\geq$ 0.5) ble det identifisert et område i fremste, nedre del av rektum som fikk betydelig høyere dose blant de pasientene som opplevde bivirkninger.

Diskusjon: Siden reduserte marginer fører til mindre bestråling utenfor målvolumet,

var lavere dose til OAR for pasienter med reduserte marginer som ventet. Sammenhengen mellom lave doser og rektumbivirkninger som ble funnet ved DVH-sammenlikning samsvarer i liten grad med litteratur som antyder sammenheng mellom høydosebestråling og bivirkninger. Derimot viser LKB-parametrene rimelig god overensstemmelse med litteraturverdier. Den implementerte bilderegistreringsalgoritmen produserte varierende resultater og bør forbedres ved å tilpasse registreringsparametrene. Likevel indikerte VBA at et område i nedre del av rektum kan være relatert til rektumbivirkninger inntil 18 måneder etter strålebehandling. Resultatene er usikre og analyse av større datasett behøves.

Videre undersøkelser bør inkludere evaluering av utfall senere enn 36 måneder etter behandling siden senskader ikke nødvendigvis har fremtrådt på dette tidspunktet. Med videre forbedring av registreringsalgoritmen kan det utviklede verktøyet for VBA ha potensiale til å identifisere strålesensitive områder i vev. Dette kan brukes til å forutsi behandlingsutfall og forbedre behandlingsplanlegging både for prostata og andre krefttyper.

Contents

Acronyms

ASGD Adaptive Stochastic Gradient Descent.

CBCT Cone-Beam Computed Tomography.

CCS Common Coordinate System.

CRT Conformal Radiotherapy.

CS Coordinate System.

CT Computed Tomography.

CTCAE Common Terminology Criteria for Adverse Effects.

CTV Clinical Target Volume.

DSB Double-Strand Breaks (DNA).

DSC Dice Similarity Coefficient.

DVH Dose-Volume Histogram.

EORTC European Organization for Research and Treatment of Cancer.

FOV Field of View.

FSU Functional Subunit.

gEUD generalized Equivalent Uniform Dose.

GI Gastrointestinal.

GU Genitourinary.

HD Hausdorff Distance.

IBDM Image-Based Data Mining.

IGRT Image Guided Radiation Therapy.

IMRT Intensity-Modulated Radiation Therapy.

LKB Lyman-Kutcher-Burman (model).

LQ Linear-Quadratic (Model).

MI Mutual Information.

ML Machine Learning.

MMI Mattes' Mutual Information.

NN Nearest-Neighbour.

NTCP Normal Tissue Complication Probability.

OAR Organ At Risk.

PROMs Patient-Reported Outcome Measures.

PTV Planning Target Volume.

QLQ-C30 EORTC Core Quality of Life Questionnaire.

QoL Quality of Life.

QUFW94 Questionnaire Umeå Fransson Widmark 1994.

RBS Rectal Bother Score.

RIC A **R**andomised, Two Centre Trial on Daily Cone-beam vs Standard Weekly Orthogonal **I**mage Guided Radiotherapy (IGRT) for Prostate **C**ancer.

ROI Region of Interest.

RT Radiotherapy.

RTOG Radiation Therapy Oncology Group.

SSB Single-Strand Breaks (DNA).

SSD Sum of Squared Differences.

TCP Tumour Control Probability.

TPS Treatment Planning System.

VBA Voxel-Bases Analysis.

VMAT Volumetric Modulated Arc Therapy.

1 Introduction

The prostate is the most common location for incidence of cancer in Norwegian men, with an average of over 5000 new cases annually in the period 2015–2019 [1]. Prostate cancer is normally treated with surgery, radiotherapy, hormone therapy, or a combination of these. The treatment is often successful. Early diagnosis and improved treatment has contributed to the 5-year survival rate increasing from less than 70 % to over 95 % in the last 25 years.

With increasing survival rates, prostate cancer patients may have to live longer with the potentially painful side effects of treatment. Examples of such side effects after radiotherapy are rectal bleeding, changes in bowel habits, and urinary or erectile dysfunction. Such effects may have an impact on the QoL of the patients, and effort to spare normal tissue is therefore made in the treatment planning process. Much data is available on the relationship between dose and adverse effects, but most studies use clinician-reported measures to assess the outcome. Inclusion of PROMs may help fully capture patient experiences when assessing normal tissue effects [2].

Modern treatment techniques such as intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) help spare normal tissue by conforming the radiation dose to the target and delivering lower doses to surrounding tissue. These techniques allow adaptation of the treatment to the anatomy of each patient, based on the location of the tumour and the OARs. In order to fully exploit these advantages, accurate data is needed on how much dose is acceptable and what regions to spare.

Whole-organ models considering only the dose-volume dependence of normal tissue have been extensively researched. These models often assume that organs are heterogeneous with regards to their sensitivity to radiation. However, it has been shown that dose to some regions of e.g. the rectum or bladder is more correlated with adverse outcome [3, 4]. Correspondence has also been found between dose to the rectal wall and PROMs [5]. Nevertheless, new methods need to be developed and tested in order to disentangle the complicated relationships between normal tissue effects and the amount and location of dose.

The overall aim of the project was to develop a programming framework in Python that can be used for analysis of radiotherapy dose plan data on a per-voxel level. Such tools can serve as decision support in the treatment planning process; They provide a basis for outcome predictions of the suggested treatment and ensure optimal dose distributions regarding both tumour control and risk of complications.

Data from a study on prostate cancer radiotherapy, the RIC study (A **R**andomised, Two Centre Trial on Daily Cone-beam vs Standard Weekly Orthogonal **I**mage Guided

Radiotherapy (IGRT) for Prostate Cancer), was used for the development and testing of the framework. The study sought to investigate whether reduced safety margins in radiotherapy could lead to a reduction in side effects. In the process of developing the framework, the dose-volume data from the study was evaluated and related to PROMs. The voxel-based analysis was compared to and traditional DVH analytics and NTCP models. Existing models for NTCP provide the basis for established tolerance levels, and are based on clinician-reported outcome measures. The use of PROMs in such models is however of increasing interest and was tested with data from the RIC study.

2 Theory

This chapter presents the necessary theory for the project. The chapter begins with a brief introduction to concepts and clinical goals in radiation therapy (with curative intent), before the reader is given some background on basic radiobiology, which is part of the rationale for these clinical goals. The theory in these sections is based on [6] and [7], unless otherwise noted. Then, normal tissue effects and ratings used in the follow-up of radiotherapy patients are discussed, with a focus on radiotherapy for prostate cancer. Lastly, the purpose and process for modelling normal tissue complication probability is presented. An introduction is given to commonly used and some more recent models, with emphasis on voxel-based analysis.

Sections 2.1 to 2.3 are loosely based on the author's specialization project from the fall of 2020.

2.1 Overview of Radiotherapy and Treatment Planning

2.1.1 Clinical Goals in External Beam Radiotherapy

In radiotherapy, the treating particles may be photons or ions such as electrons or protons. The focus of this thesis is on external beam RT with photon beams.

The photons deposit energy in the tissue that contributes to the destruction of DNA in the cells, which limits the cells' reproductive potential and leads to cell death. These mechanisms will be presented in Section 2.2. The amount of energy delivered is measured in units of Gray (Gy), where $1\,\text{Gy} = 1\,\text{J/kg}$.

In curative RT, the intention of the treatment is to irradiate the tumour or cancerous tissue with a high enough dose to kill all cancerous cells capable of regrowing the tumour, which can be quantified by the tumour control probability (TCP). Simultaneously, clinicians are trying to compromise the health of normal tissue as little as possible, since irradiation of normal tissue may also lead to cell death. Significant radiation dose to normal tissue may lead to complications, and the probability of side effects can be quantified by the NTCP.

Curative RT is a compromise between TCP and NTCP, as illustrated by the TCP and NTCP dose-response curves in Figure 2.1. To be able to achieve sufficient tumour control without compromising normal tissue, the two curves need to be separated so that tumour response is achieved at a lower dose than adverse outcomes. The separation of the curves is termed the therapeutic window, and describes the window of opportunity where the
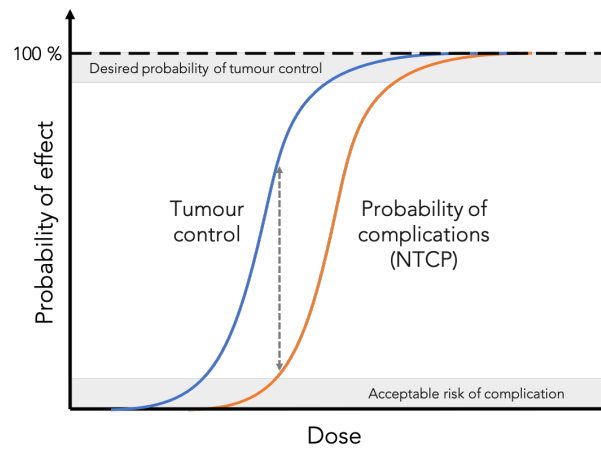
Figure 2.1: Dose-response curves for tumour (blue) and normal tissue (orange). The vertical dashed gray line indicates the therapeutic window, which describes the relative curative effect of a treatment to its probability of causing complications. The top and bottom gray boxes indicate the desired probability of tumour control (curative effect) and the acceptable complication probability, which may be 10-15 %.

response of the tumour and of normal tissue can be differentiated enough to achieve clinical goals.

To increase the therapeutic window, i.e. increase probability of cure relative to probability of complications, one may sensitize the tumour to radiation (make the tumour respond to lower doses of radiation), effectively shifting the tumour response curve to the left and increasing the therapeutic window. Some such mechanisms (fractionation and oxygenation) will be presented in Section 2.2.

2.1.2   The Treatment Planning Process

The treatment process in radiotherapy can be summarized in four steps: imaging of the patient, volume delineation, treatment planning and treatment delivery. The process is illustrated in Figure 2.2. The theory presented in this section is based in part on ICRU Report 83 [8] and Podgorsak [9].
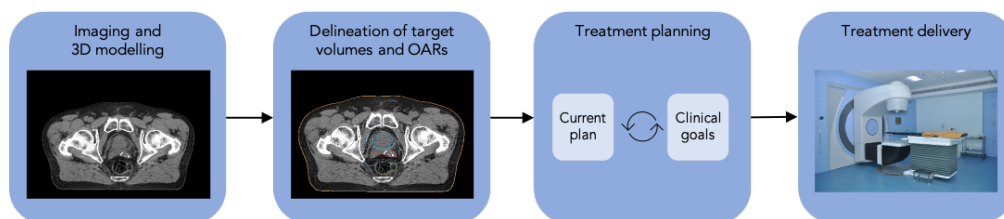
Figure 2.2: Schematic overview of the main steps in the treatment planning process; imaging to evaluate tumour volume and location, delineation of target volumes and OARs, the treatment planning process (iterative inverse planning), and treatment delivery.

Fixation and Imaging

The first step in the treatment process is fixation and imaging of the patient. In conventional 3D treatment planning, a computed tomography (CT) scan (the planning CT) is acquired of the patient while immobilized in the desired treatment position. Since de-
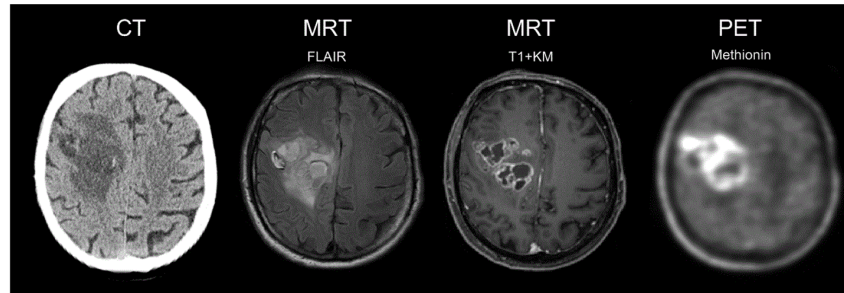
Figure 2.3: CT, MRI and PET imaging have different types of contrast and can be used to analyze structural and metabolic characteristics of tissue such as the brain, and to visualize tumours. Image from OncoRay (https://www.oncoray.de/research/image-guided-high-precision-radiotherapy).

lineation of volumes, dose calculation, and planning is performed based on this CT, the patient must be in the same position during all treatment sessions as in the planning CT. Implanted fiducial markers in the tumour and tattoos on the skin can help guide the alignment and ensure reproducibility.

MR and/or PET images of the patient are often also acquired since these can provide soft tissue contrast and metabolic information, as illustrated by Figure 2.3. The improved contrast and functional information may be helpful in volume delineation and treatment planning.

Delineation of Volumes

A clinician outlines target volumes and other regions of interest on the planning CT. An example of a slice of a CT scan with organ and target volume delineations is shown in Figure 2.4. The clinical target volume (CTV) consists of the demonstrable extent of the tumour (from images, pathological/histological samples, or clinical examination) with a margin of a few mm to cm. The margin is meant to account for microscopic disease around the visible tumour that needs to be eradicated to achieve the wanted curative effect.

Since the location of the tumour may change relative to the patient geometry during the treatment period (e.g. due to breathing or variation in degree of rectal/bladder filling) a setup margin is added to the CTV. This setup margin is meant to account for machine tolerances and the total geometrical variation in the setup. The final volume is called the planning target volume (PTV), and is often defined relative to the CTV: PTV = CTV + setup margin. The PTV is often defined in the coordinate system of the treatment machine, and is the basis for treatment planning. The PTV is often the volume for which dose is prescribed, as it is assumed that the PTV will cover the CTV even when variations in patient geometry and setup occur. The ICRU report [8] also defines an internal target
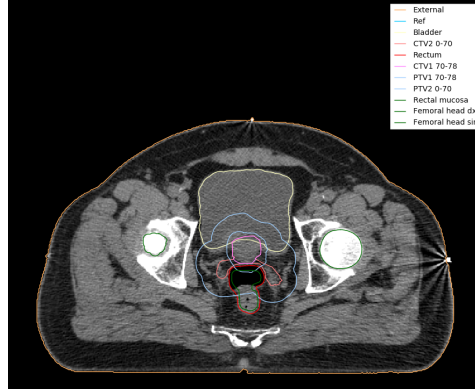
Figure 2.4: Example of delineation of target volumes and organs at risk in the pelvic region for a prostate cancer patient. The figure is generated from project data for illustrative purpose.

volume, which consists of the CTV with an internal margin to account for geometrical variations in the patient. However, the internal target volume is not relevant in this report since the it is contained in the PTV.

In addition, organs at risk (OAR) that overlap or are located near the PTV, or that may be at risk of being irradiated, are delineated. The planning process often involves minimizing the dose to these regions by specifying tolerance levels, and therefore the dose to these organs must be calculated. Other structures or regions of interest (ROI) to which the clinician wants to keep track of the delivered dose can also be delineated.

Treatment Planning

In modern radiotherapy the treatment plan is found by inverse planning. A clinician defines a set of clinical goals for each patient (the origin and definition of these goals will be presented in Section 2.1.4). These goals can be the prescription dose to the PTV or dose-volume constraints on the OARs.

An objective function that measures to which degree the clinical goals are satisfied by a treatment plan is defined based on the clinical goals. An example of such a function measuring whether homogeneous dose is delivered to the target volume is

$$F_T = \frac{1}{N_T} \sum_{i=1}^{N_T} (D_i - D_i^{(\mathrm{P})})^2,$$

where D_i is the delivered (planned) dose to the ith voxel, $D_i^{(\mathrm{P})}$ is the prescription dose to the ith voxel, and N_T is the number of target voxels. If all voxels receive the prescribed dose, the value of the objective function F_T is zero. Any voxels receiving higher or lower dose will contribute to the objective function [6].

Similarly, dose to the OARs should be minimized. An objective function of the form

$$F_{\mathrm{OAR},k} = \frac{1}{N_k} \sum_{i=1}^{N_k} D_i$$

can be used for OARs, since any dose to the organ contributes to $F_{\text{OAR},k}$. Here, k denotes each risk organ to be included in the objective function, and N_k is the number of voxels in the risk organ.

Other types of objectives may be defined and can generally can be of varying importance. Therefore each objective function F_j may be assigned a weight w_j. The total objective function is thus a weighted sum of the objectives,

$$F_{TOT} = \sum_j w_j F_j.$$

The objective function is minimized through an iterative process, either until the the optimal treatment plan for the defined goals is found, or until no further optimizations can be achieved. In the latter case the clinician may have to adjust the constraints if a satisfactory plan could not be found.

Treatment Delivery

When a satisfactory treatment plan has been found and approved, the patient can be treated. In external beam radiotherapy, the radiation is delivered by the use of a linear accelerator (linac) that can rotate and irradiate the patient from many angles. In tradi-
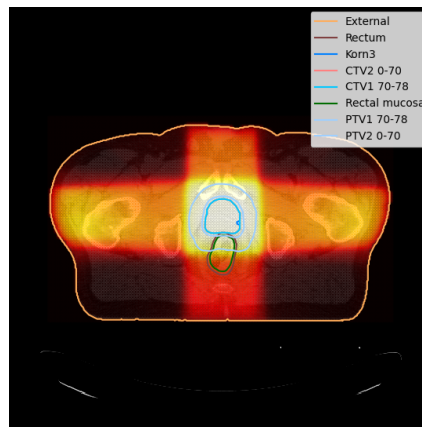
Figure 2.5: Illustration of the use of a four-field box technique, which is common in 3D-CRT.

tional 3D-conformal radiotherapy (3D-CRT), the dose is delivered from a few angles, as illustrated by the four-field box technique shown in Figure 2.5. The dose is often given in 20-35 fractions of 2 Gy (or near 2 Gy) over the span of 4-6 weeks.

As mentioned previously, the patient must be in the same position during all treatment sessions to ensure that the dose is delivered to the desired location. Further, the geometry of the patient and location of internal structures relative to the PTV (which is often defined in the machine coordinate system) can vary. Both these factors contribute to uncertainty in the delivery of dose. Image-guided radiotherapy (IGRT) attempts to minimize the variations in where the dose is delivered by imaging the tumour and comparing these images to those used in planning. The positioning of the patient may then be adjusted to ensure accuracy in the dose delivery. It is then generally assumed that the delivered dose equals the planned dose.

2.1.3 Dose-Volume Histograms

Dose-volume histograms offer a convenient way of summarizing the dose distribution to a three-dimensional volume. Cumulative DVHs are the most relevant in a treatment planning perspective and will be the focus of this section. As the name implies, DVHs
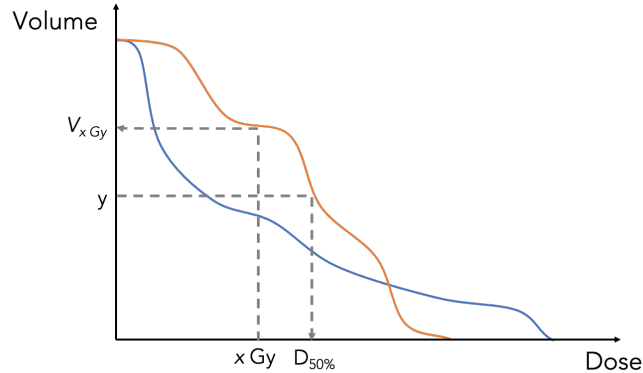
Figure 2.6: Example of a cumulative DVH, where each point shows the volume (y-axis) that was irradiated above certain dose (x-axis). Two different DVH curves are shown. By following the gray arrows one may read out dose or volume characteristics from the DVH.

provide information on the volume of a structure that has received a certain dose, as is shown by the two example DVH curves in Figure 2.6. Each dose-volume point (x, y) on the curve shows the volume y that received a dose above or equal to x, usually denoted $V_{x\text{Gy}} = y$. In a similar fashion, the value of $D_{y\%}$ indicates the minimum dose of the $y\%$ highest dose to structure.

DVHs may show either absolute or relative volume. However, for practical purposes the relative volume is often plotted since structures have widely different (absolute) volumes and this allows the volumes to be plotted on the same axis. The dose-volume characteristics are easy to calculate and provide a simple means of comparison between different treatment plans. The DVH characteristics are widely used clinically for this purpose.

2.1.4 Dose-Volume Constraints

As introduced in Section 2.1.1, a successful radiotherapy treatment plan is one that acquires the necessary level of tumour control, while giving as low a dose as possible to the OARs and normal tissue surrounding the tumour. In theory, one could achieve a very high degree of tumour control by continually increasing the dose. However, increasing dose to the tumour is inevitably linked to elevated doses to normal tissue, and consequentially to increased risk of tissue complications. In particular, dose to OARs should be limited as much as possible.

Emami et al. [10] published a paper where tolerance doses for several organs were presented. The tolerance doses TD5 and TD50 are the whole-organ doses supposedly leading to 5 % and 50 % complication rate within 5 years after treatment, and such

tolerance doses were presented for irradiation of 1/3, 2/3 and the whole volume of the organ.

However, the tolerance of most normal tissues does not depend on a single threshold. It also depends on the volume irradiated. In the QUANTEC issue of the *International Journal of Radiation Oncology*Biology*Physics (IJROBP)*, updated tolerance doses based on a large-scale literature review were presented for several organs (e.g. rectum [11], bladder [12], and penile bulb [13]). Rather than tolerance doses for just 1/3, 2/3 and the whole organ, the QUANTEC reviewers present the dose-volume constraints (given as DVH characteristics) that are statistically most linked to unwanted complications. The QUANTEC constraints are widely used as guidelines in clinical treatment.

2.2 Basic Radiobiology

2.2.1 Mechanisms for Cell Damage by Ionizing Radiation

Cell damage in radiotherapy occurs because the DNA of the cells is damaged by the radiation. The high-energy particles used in treatment (in this case photons will be considered) have the ability to ionize the DNA molecule, thereby inducing breaks in the DNA strands. If the DNA damage is irreparable, the cell cannot proliferate and will die. There are two main mechanisms by which ionizing radiation can damage DNA. These are illustrated in Figure 2.7. Direct damage occurs then the incoming photons directly ionize the DNA molecule. Indirect damage occurs when the photons instead ionize wa-
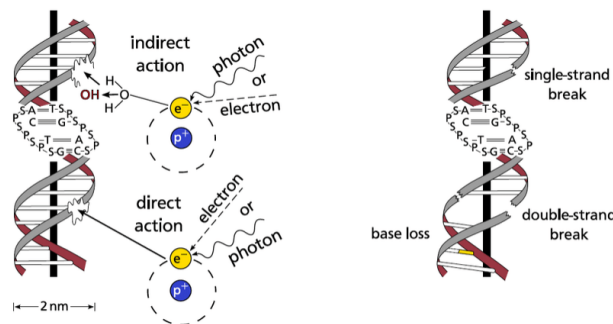
Figure 2.7: Schematic of direct and indirect DNA damage (left), and the types of radiation damage induced (right). Figure from [14].

ter (or other molecules), generating reactive oxygen species that induce the DNA strand breaks.

DNA damage can broadly be divided into single-strand and double-strand breaks. A photon dose of $1\,\mathrm{Gy}$ typically induces approximately 10^5 ionizations, around 1000 single-strand DNA breaks (SSB), and around 20-40 double-strand breaks (DSB) per cell. However, due to the large repair capacity of normal cells, only about 30 % of cells will be killed by this damage [15]. Single-strand breaks can often be repaired since the undamaged DNA strand can be used as a template from which the broken DNA strand may be copied. However, double strand breaks are often irreparable, leading to cell death. Cancerous cells do not have the same repair capacity as normal cells, even for single-strand

breaks. This effect may be exploited in radiotherapy by giving the treatment in smaller fractions, which gives the normal cells time to repair DNA. This will be described further in Section 2.2.3.

2.2.2   Dose-Response Curves and the Linear-Quadratic Model

As mentioned in the previous section, a photon dose of 1 Gy induces 1000 SSBs and 20-40 DSBs per cell, killing approximately 30 % of normal cells. Since the number of DNA breaks is related to the delivered dose, cell survival must also be dependent on the dose. By using cell assays to quantify the cell survival after irradiation with a certain dose,
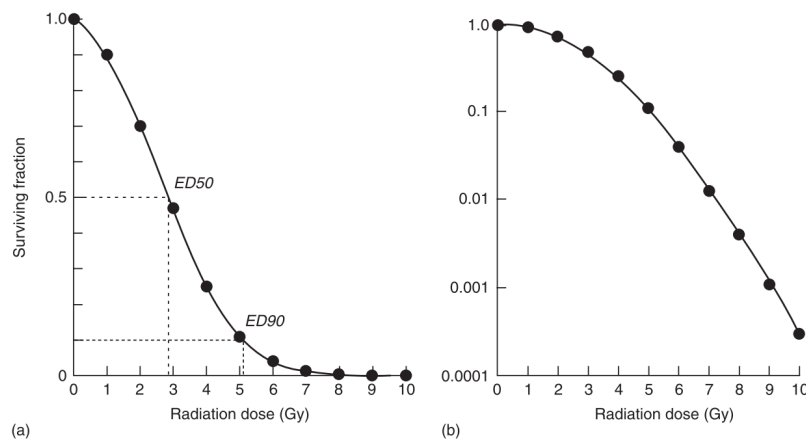
Figure 2.8: Typical cell survival curve for irradiated cells on a linear (left) and logarithmic (right) scale. Figure from [15].

it is possible to generate dose-response curves that show the surviving fraction of cells. Examples of cell survival curves plotted on a linear and logarithmic scale are shown in Figure 2.8. Since a very large degree of tumour cell kill is required for tumour control, curves are typically plotted on a logarithmic scale to better investigate the cell survival at low survival [15].

The cell survival curve is characterized by a "shoulder region" for low radiation doses and a near linear region for higher doses. The surviving fraction SF of normal tissue cells is accurately approximated by the linear-quadratic (LQ) model,

$$SF = e^{-\alpha D - \beta D^2}. \tag{2.1}$$

Here, D is the delivered dose, and α and β are parameters describing the radiosensitivity of the cells. The logarithmic equation takes the form

$$-\log(SF) = \alpha D + \beta D^2 = \alpha D \left(1 + \frac{D}{\alpha/\beta} \right). \tag{2.2}$$

We see that α relates linearly to the dose, whereas the curviness of the survival curve depends on the ratio α/β. There is a mechanistic relation to DNA damage in that αD can be said to relate to single-track ionization events (SSBs), whereas the βD^2 part of

the equation relates to double-track events (two SSBs within a certain distance on the DNA strand generate a DSB).

In normal tissue with high repair capacity, the majority of SSBs are repaired before another ionization can generate a DSB. The α for normal tissues therefore tends to be lower than for cancerous cells, resulting in a lower α/β ratio. The survival curve for normal cells will therefore be more curved, with a larger surviving fraction at low doses, and smaller surviving fraction at higher doses. This effect, as mentioned previously, is favourable in radiotherapy, as will be described next.

2.2.3   The Five Rs of Radiotherapy

The five Rs of radiotherapy (or radiobiology), namely repair, redistribution, repopulation, reoxygenation and radiosensitivity, are the rationale for fractionated RT [16, 17]. They describe important factors that influence how biological tissue responds to radiation.

Repair refers to the repair of sublethal DNA damage following irradiation. As mentioned, normal tissue tends to have higher surviving fractions than cancerous cells at low doses due to higher repair capacity. Therefore, with fractionation of dose, normal tissue is given time to repair before the next fraction, which leads to relative sparing of normal tissue.

Cells are more susceptible to radiation-induced damage during some parts of the cell cycle. These cells will be preferentially killed off, leaving the remaining cells somewhat synchronized in the cell cycle. Redistribution (or reassortment) explains how the remaining cells will have moved after some time into a part of the cell cycle in which they are more prone to be damaged by radiation. This sensitizes the remaining cell population to radiation and is another justification for fractionated RT.

Repopulation refers to cell proliferation. Repopulation of tumour cells happens more quickly than that of normal tissue cells, which limits the duration of the treatment to avoid uncontrollable tumour growth.

Reoxygenation is another reason behind fractionated RT. Due to generation of reactive species damaging DNA indirectly, oxygen is a powerful radiosensitizer. Oxygenated parts of the irradiated volume will have a greater cell kill than hypoxic regions. This is particularly relevant in tumours, which tend to have defective vasculature and necrosis at their centres where oxygen cannot reach. Fractionation allows the tumour to reoxygenate between successive treatments and increases the tumour cell kill.

The fifth R, introduced by Steel et al. [17], is radiosensitivity. It has been shown that tumour cells from different types of cancers, or within a tumour, inherently have different sensitivity to radiation. This results in varying steepness of the survival curve, an effect which is enhanced at low dose rates.

2.2.4   Tissue Morbidity and Organ Architecture

The complication probability in normal tissue depends on the delivered dose and the volume of organ or tissue that has been irradiated [15]. However, different organs and tissue types have varying responses to variations in dose and irradiated volume.

Avoidance of normal tissue complications depends on the survival of a sufficient number of cells in the tissue. In particular, tissue-specific cells that are responsible for carrying out function (e.g. neurons in the brain or cells involved in gas exchange in the lungs) must be present in order to maintain organ function. Withers et al. introduced the idea of a functional subunit (FSU), as "the largest unit of cells capable of being regenerated from a surviving clonogenic cell without loss of the specified function" [18]. The FSUs are independent, and the number of FSUs that are inactivated (sterilized) by radiation exposure depends on the dose and dose rate, their radiosensitivity and other parameters [15].

However, the clinical consequences depend not only on the number of FSUs that are inactivated, but on their organization in the tissue or organ. Serially organized organs, similar to electrical components in series, may lose function if only one or very few FSUs are sterilized. They have a binary response to radiation exposure and tend to show damage only if the dose exceeds a threshold.

On the contrary, parallel organs exhibit a strong volume dependence, meaning that the degree of damage tends to increase with the volume of organ irradiated. The organ function can be maintained in some organs even though a significant percentage of the FSUs are sterilized. In reality, no tissue has a purely serial or parallel architecture, and therefore most organs have a radiation response somewhere between the two extremes.

2.3   Normal Tissue Effects after Prostate Cancer RT

2.3.1   Acute and Late Toxicity

Normal tissue complications are often divided into two groups; acute or early radiation effects, and late effects. The occurrence of late or early effects depends on the tissue type and on the dose delivered to the tissue [7]. Early effects appear soon after radiation exposure (within a few weeks) [9]. The damage is often transient in nature, and the tissues are not necessarily permanently damaged. Late effects can appear months to years after the treatment (exposure), and are usually more chronic than early effects.

2.3.2   Typical Side Effects and Organs at Risk

Since RT of the prostate involves irradiation of the pelvic region, the main OARs are those located near the prostate or in the field of view (FOV) of the radiation beams; examples are the rectum and anal canal, urinary bladder and urethra, and the penile bulb. The main focus in this report will be on rectal side effects.

Radiotherapy for prostate cancer may cause gastrointestinal (GI) toxicity in patients, and is related to irradiation of the rectum and anal canal. Acute side effects may be increased stool frequency or change in stool texture [11], as well as rectal bleeding. Late side effects may include change in storage capacity resulting in frequent bowel movements, as well as stool urgency, leakage and pain/discomfort. Irradiation of the anal canal may also cause fecal incontinence.

Irradiation of the genitourinary (GU) organs may cause side effects such as incontinence or increased urinary frequency or urgency, among others. In addition, RT may

influence sexual health. Although erectile dysfunction has many causes, studies report that up to 50 % of males receiving radiotherapy develop radiation-induced erectile dysfunction [19]. Generally, patients may experience a change in bowel, urinary or sexual habits, and the effects can significantly impact QoL.

2.3.3 Normal Tissue Complication Categories and Rating

The theory in this section is based on [15] and the discussion in [20].

The reporting of side effects is important to be able to evaluate treatment schemes. The reporting should be standardized to provide a reliable basis for comparison of different treatments. Traditionally, side effects have been reported by clinicians on the basis of a clinical examination. For outcome reporting after prostate cancer RT, the Radiation Therapy Oncology Group (RTOG) provides objective criteria that are often used to report toxicity [11]. Other objective criteria are provided by the Common Terminology Criteria for Adverse Effects (CTCAE), developed by the National Cancer Institute in the US. These criteria are commonly scored from 0 to 5, where 0 means no effect, while outcome is rated to 5 if the side effect causes death.

Some studies assess and report side effects based on PROMs, which is defined as an outcome reported directly by patients themselves and not interpreted by an observer. These outcomes may include assessments of health status, QoL or experienced symptoms [21]. Approval of a treatment based solely on PROMs was first done for prostate cancer. The European Organization for Research and Treatment of Cancer (EORTC) developed a questionnaire where patients report on 30 questions on health-related QoL and general symptoms, the EORTC Core Quality of Life questionnaire (QLQ-C30) questionnaire [22]. For prostate cancer, the Quationnaire Umeå Fransson Widmark (QUFW94) [23] is sometimes used. The QUFW94 is commonly known as the Prostate Cancer Symptom Scale since it inquires about prostate-specific symptoms. The questionnaire can be used to report on four areas of potential side effects;(GI) toxicity, urinary problems, rectal function and problems, and sexual function.

2.4 NTCP Modelling

DVHs are normally used for evaluation of treatment plans, but comparison of dose plans and prediction of toxicity solely based on the dose-volume information found in DVHs is difficult. The calculation of an NTCP value given a treatment plan could help indicate the risk of complication and simplify the process of comparing treatment plans. This section will outline the general method for DVH-based estimation of normal tissue complication probabilities and present the commonly used Lyman-Kutcher-Burman model. An overview of modern NTCP models that go beyond the DVH-based approach is also given.

2.4.1 A Generalized NTCP Modelling Process

NTCP modelling describes the process of correlating the delivered treatment to the outcome (complications) that the patient experiences. Previously, efforts have been made

to develop reliable models capable of representing the three-dimensional dose distribution information in terms of one or a few comparable parameters. In NTCP models, the
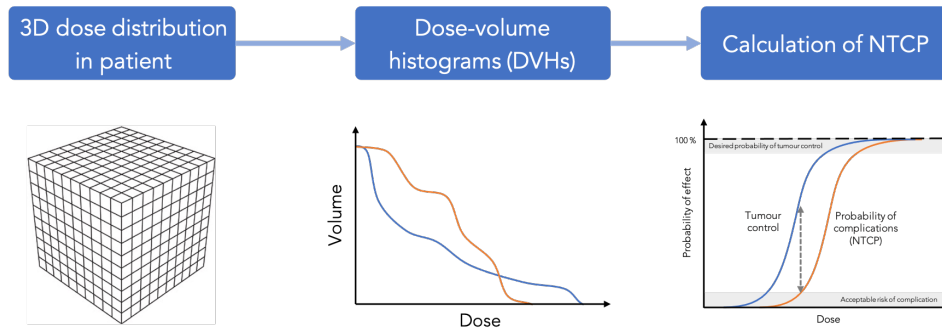
Figure 2.9: Schematic representation of the NTCP estimation process. The 3D dose distribution in a patient is summarized in DVHs for each structure of interest. An NTCP model correlates the DVHs to the outcome/complications in a group of patients and the output is a prediction of the NTCP.

three-dimensional dose distribution is often summarized in a DVH, as shown in Figure 2.9. DVHs were described in a previous section. When the dose distribution or the dose-volume relationship is known, the calculation of DVHs is straightforward. In the process of estimating the NTCP, these two-dimensional DVHs are summarized by a single measure, termed DVH reduction. DVH reduction can be done in a variety of ways, but several NTCP models use a form of volume-weighted average dose or effective irradiated volume to summarize inhomogeneous dose distributions.

Since the distribution of patient responses is assumed to be normally distributed within a (sufficiently large) population, an NTCP curve (the cumulative probability distribution) is generally believed to have a sigmoid shape. It is typically characterized by a tolerance dose TD50 yielding complication in 50 % of cases, and by the standard deviation of the distribution, which determines the steepness of the NTCP curve. One such NTCP curve is described by the Lyman-Kutcher-Burman model.

2.4.2   The Lyman-Kutcher-Burman Model

In 1985, John Lyman published "Complication Probability as Assessed from Dose-Volume Histograms" [24], where he presented a method for evaluating complication probability from DVHs. This probability could then easily be used in the optimization of treatment plans and to support clinical decisions.

Lyman argued that the normal tissue response should follow a bell-shaped curve, and that it depends both on dose and irradiated volume. The complication probability is the cumulative distribution function, given by the integral

$$\text{NTCP} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2} \, \mathrm{d}t. \tag{2.3}$$

The parameter t is given by

$$t = \frac{D - \text{TD50}\left(\frac{V}{V_{\text{ref}}}\right)}{m \cdot \text{TD50}\left(\frac{V}{V_{\text{ref}}}\right)}, \tag{2.4}$$

where D is the dose (the volume is assumed to be uniformly irradiated) and m is a parameter representing the steepness of the dose-response curve (which is determined by the standard deviation of the probability distribution). V/V_{ref} is the partial volume that is irradiated with the dose D. $\text{TD50}(V/V_{\text{ref}})$ is the tolerance dose that leads to complications in 50 % of the patient population, as described in Section 2.1.4.

In Lyman's model, the tolerance dose follows a power-law relationship given by

$$\text{TD}(V/V_{\text{ref}}) = \text{TD50}(1)\left(\frac{V}{V_{\text{ref}}}\right)^{-n}. \tag{2.5}$$

The parameter n relates the tolerance dose at partial irradiation, $\text{TD50}(V/V_{\text{ref}})$, to the tolerance dose at whole-organ irradiation, $\text{TD}(1)$.

Kutcher and Burman proposed a generalization of the model to inhomogeneous dose distributions by converting the DVH into an equivalent DVH [25]. This is done by calculating a generalized equivalent uniform dose (gEUD) to the entire organ volume, which is equivalent to the inhomogeneous dose distribution. TD50 is thus defined as the whole-organ dose that produces a 50 % complication rate. The gEUD can be found by summing over all volume subunits and doses to the subunits, weighted by n, as given by the formula

$$\text{gEUD} = \left(\sum_i v_i d_i^{1/n}\right)^n. \tag{2.6}$$

Here, $v_i = V_i/V_{\text{tot}}$ is the relative volume of subunit i, and d_i is the dose to the subunit. The parameter n (as above) is a weighting factor for the dose dependence, and ranges from $n \ll 1$ (the high-dose region is weighted more) to $n \sim 1$ (the low-dose region is the most statistically significant).

The gEUD is substituted for the dose D in (2.4), and the TD50 parameter in the equation is thus the whole-organ tolerance dose.

2.4.3 Modern NTCP Models

Modern delivery methods in radiotherapy, such as IMRT and VMAT, deliver highly conformal radiation doses to the target. Normal tissue is spared to a larger degree than earlier, and discerning the amount of dose to OARs that is likely to produce complications is challenging since there are subtle differences in the delivered dose.

Recently, focus has turned towards NTCP models that consider the spatial location of the dose, and that thus try to investigate local dose response with a finer resolution than that of whole-organ models. Such models have been demonstrated to be able to relate spatial location of the dose to e.g. GI and GU toxicity after prostate cancer RT [4, 5, 26, 27], and to radiation-induced acute dysphagia and trismus after RT in head and

neck cancers [28, 29]. Dose delivery outside the target has also been related to lower rate of failure of treatment in a Dutch prostate cancer trial [30, 31]. A brief overview of some recent methods to investigate spatial dose-response relationships is given here.

Since the introduction of the LKB model, there have been many advances in the field of NTCP modelling. For instance, dose-surface maps and dose-surface histograms have been used to evaluate toxicity of hollow organs such as the bladder or rectum [4, 32–34]. Dose-surface maps attempt to visualize not just relationships between dose and volume, but also the relation of toxicity to the spatial location of the dose. Different regions of an organ may have different dose sensitivity, an effect which is not considered when using DVHs for treatment evaluation. Buettner et al. [26] found that including morphological rather than just volumetric aspects of the dose distribution improved the prediction power of a model for radiation-induced rectal bleeding.

A further step in this direction is the development of voxel-based models, which correlate the outcome with the dose to each voxel. This approach will be further explained in the next section. Such models have been shown to be able to identify specific regions that correlate with failure of treatment [35] or incidence of rectal bleeding [36] after prostate cancer RT. Mylona et al. [4] investigated urinary dysfunction after prostate cancer RT and found that dose to certain regions of the bladder and urethra were correlated with urinary toxicity.

Image-based data mining (IBDM) is an umbrella term for techniques that aim to extract relationships and patterns from raw image data [37]. Several of the studies mentioned use techniques from IBDM to obtain predictions of spatial dose-effect relationships [4, 35]. Machine learning (ML) methods are also being used to predict RT-related toxicity [38]. Methods from ML can find patterns in large amounts of data. Several studies use ML for normal tissue complication prediction based on for instance genomic data [39] or other clinical variables [40]. Such methods are likely to become more common in the near future, as ML gains foothold in the medical fields.

2.5 Voxel-Based NTCP Modelling

Dosimetric VBA in radiotherapy consists of three main steps. The first step is to perform a spatial normalization of the patients to a common anatomical template, so-called image registration. Second, the dose distributions are propagated from the native space to the common template. Finally, the necessary statistical analysis on the normalized dose distributions are carried out [36, 41]. This section will introduce the background needed for understanding the methods used in the project.

The following notation is used in this section: Vector notation $\vec{x}$ for points $\vec{x} = (x_1, x_2, \ldots, x_n)$ in n dimensions. A transform is denoted by a bold symbol $\mathbf{T}$, while an image is denoted by a capital letter I.

2.5.1 Image Registration for Spatial Normalization

Image registration is the process of finding the transform between two images. The reference image is usually termed the *fixed* image, $I_F(\vec{x})$, whereas the image to be aligned is termed the *moving* image, $I_M(\vec{x})$. Here, $\vec{x}$ denotes the position of a point in the image.

Examples of the use of image registration in medical image analysis are intra-subject multi-modal registration, such as registering a PET or MR image to a CT image, or inter-subject registration to spatially align a population of anatomies and perform statistical analyses.

For inter-patient registration, the fixed image (or reference image) to which the other images are registered can be a phantom or a carefully selected reference patient. This patient geometry then represents the common coordinate system (CCS). The images are aligned in the CCS by finding a transform $\mathbf{T}(\vec{x})$, which, when applied to a point $\vec{p}$ in the moving image, will transform $\vec{p}$ into a point $\vec{q}$ in the fixed image space.

The parameters of the transform form a parameter space. To assess the correspondence between the fixed and the moving images, a metric is defined to measure the similarity of the fixed image $I_F(\vec{x})$ and the transform of the moving image $I_M(\mathbf{T}(\vec{x}))$. The similarity metric forms the basis for a cost function C. Minimizing C in the transform parameter space yields the transform that maximizes the similarity between the images. There are many algorithms developed to perform image registration. Some important image registration components will be described in the following section.

2.5.2 Registration Components

Transforms

The transform $\mathbf{T}(\vec{x})$ can be of varying complexity. The transform used for registration should be selected carefully to ensure that it is suitable for each registration problem [42]. Generally, the transforms can be divided in two groups; linear and elastic (non-rigid) transformations.

Linear transforms include translation, rotation, and scaling transforms, or affine transformations. An affine transformation is a linear transform composed of translation, rotation, scaling, stretching, and shearing [42].

Linear transformations by definition preserve linearity, i.e. lines in the original image are transformed to lines in the resulting image [43]. These transforms are therefore not necessarily able to warp anatomical images with local differences, such as those that may be found between subjects. However, if the object to be registered is mostly rigid, a linear transform may be sufficient and appropriate for registration. Linear transforms have only a few transform parameters, and are fairly quick to compute.

Elastic (non-rigid) transformations, on the other hand, can locally warp the moving image so that it aligns with the reference image. A good elastic registration may therefore be more suitable than a linear transformation for some inter-patient registration applications. However, use of nonlinear transformations is more computationally expensive and may in some cases lead to worse registration accuracy [43].

To lessen the requirement for computational resources, a simple transform with a smaller parameter space is often used to initialize more complex transforms with (very) large parameter spaces. Achieving a fair registration with the simple transforms means less time must be spent searching the large, complex parameter spaces. This can increase the speed of computation.

Image Pyramids and Multi-Resolution Strategy

An image pyramid is the successive smoothing and downsampling of an image, and is often used in image registration. Each downsampling lowers the resolution of the image, while the smoothing step ensures that important features are preserved. The lower resolution of the images in the pyramid reduces computational complexity without compromising precision [44].

Similarity Measures and Metrics

This subsection is based on Goshtasby [43] unless otherwise noted.

Like the transform to be used, a suitable similarity metric S must be chosen according to the registration problem. Generally, the similarity measure between images can be feature-based or intensity-/voxel-based.

Feature-based registration uses corresponding points or landmarks in the images to measure the degree of alignment [42]. These methods require a pre-processing step that may be manual (e.g. manual contouring, selection of corresponding anatomical locations or fiducial markers) or automatic (e.g. auto-segmentation). Any errors in the pre-processing step may propagate to the registration and influence the end result. If we know the point-to-point correspondence between each pair of points $\vec{p}$ in the moving image and $\vec{q}$ in the fixed image, we may use the sum of squared distances (SSD) similarity metric

$$S = -\sum_i \|\vec{p} - \mathbf{T}(\vec{q})\|^2$$

as basis for the cost function C. The negative sign is present since the best registration is the one that minimizes $\sum_i \|\vec{p} - \mathbf{T}(\vec{q})\|^2$, thereby maximizing S.

Intensity-based registration circumvents the issues relating to definition of corresponding points, as mentioned in relation to feature-based registration. It instead directly measures the amount of shared information in two images, and is suited for multi-modal registration (i.e. between images of different modalities).

The SSD may be used as a voxel-based similarity metric in a slightly modified form as

$$S_{SSD} = -\frac{1}{N} \sum_{\vec{x} \in \Omega_F} \|I_M(\mathbf{T}(\vec{x})) - I_F(\vec{x})\|^2, \tag{2.7}$$

where Ω_F is the set of points that the fixed image is defined on, and N is the number of voxels in the overlapping region. However, this is only appropriate when the image intensities are corresponding, such as with mono-modal registration and similar image settings.

For multi-modal registration, or in images where the intensities cannot necessarily be compared directly, the Mutual Information (*MI*) metric can be used. *MI* measures how much the information from one random variable tells about the information on another random variable [42]. Since information is related to entropy, we can define the *MI* in terms of the entropies $H(I_F)$ and $H(I_M)$ of the fixed and moving images, respectively:

$$H(I_F) = -\sum_{a \in I_F} p(a) \log[p(a)] \qquad H(I_M) = -\sum_{b \in I_M} p(b) \log[p(b)]. \tag{2.8}$$

Here, $p(a)$ is the probability that a voxel in I_F has intensity a, and $p(b)$ is the probability that a voxel in I_M has intensity b. The joint entropy of the images is

$$H(I_F, I_M) = -\sum_{a \in I_F} \sum_{b \in I_M} p(a, b) \log[p(a, b)], \tag{2.9}$$

where $p(a, b)$ is the probability that a voxel in the overlapping image region between I_F and I_M has intensity a and b, respectively. The mutual information is given in terms of entropies as

$$\mathcal{S}_{MI}(I_F, I_M) = H(I_F) + H(I_M) - H(I_F, I_M), \tag{2.10}$$

and should be maximal when the images are aligned. A special formulation of the *MI*, called Mattes' mutual information (MMI), is sometimes used in image registration applications [42, 45].

Optimization

Image registration is an optimization problem where the objective is to maximize the similarity (metric) between the images. In non-rigid registration schemes, a penalty term may be added to penalize large deformations.

Gradient descent optimizers estimate the transform parameters by iteratively searching the parameter space for the optimal solution. Since the similarity metric $\mathcal{S}$ is a function of the transform parameters, then (assuming $\mathcal{S}$ is differentiable) the optimal solution can be found by following the gradient until convergence. A special case is the adaptive stochastic gradient descent (ASGD) method, which is robust and is often used in image registration [46].

Resampling and Interpolation

This paragraph is based on Goshtasby [43] and the ITK Software Guide [47].

Generally, the transform **T** mapping points from one image to another will not map points to grid positions. Therefore, the transformed image needs to be resampled on the fixed image grid. This is done by interpolating the intensity value of the transformed points at the grid positions.

Several interpolation methods exist. Nearest-neighbour (NN) interpolation assigns a voxel in the fixed grid the intensity value of the nearest voxel in the mapped image. Linear interpolation (bilinear (2D), trilinear (3D)) finds the voxel intensity at the grid position by performing a linear interpolation between neighbouring voxels in each direction. B-spline interpolation uses B-spline approximations of order N to estimate the image intensity. B-spline interpolation of order 0 is nearly identical to NN interpolation.

2.5.3 Application to Dosimetric Analyses in RT

To be able to directly compare the dose distributions delivered to a population of patients, the anatomies need to be spatially aligned. It can be assumed that each anatomical (e.g. CT) image has a spatially corresponding dose distribution in the native image space. Once the anatomical images have been aligned, the resulting deformation

field can be used to warp the dose distribution that is paired with each moving image. Once the dose distribution is resampled in the CCS, a point-by-point comparison can be performed.

2.5.4   Quantitative Evaluation of Registration/Delineation Match

Evaluation of the image registration result is critical to be able to assess the uncertainties relating to the spatial normalization. If the alignment is poor, it will consequently influence the results of the subsequent analyses.

Two common metrics for evaluation of the overlap and distance between two structures or volumes are the Dice similarity coefficient (DSC) and Hausdorff distance (HD), respectively. The DSC and HD are presented here.
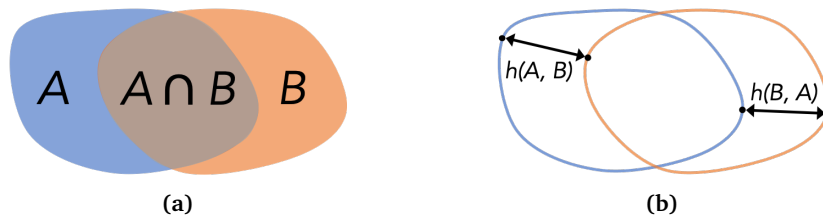
Dice Similarity Coefficient

To evaluate the match between two structures (point sets), metrics measuring the overlap between the structures are often used. A common such metric is the DSC. The DSC measures the overlap between two volumes, as illustrated in Figure 2.10(a). The DSC for two volumes A and B is defined as

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|}. \tag{2.11}$$

The value of the metric ranges from 0, corresponding to no overlap between structures, to 1, meaning the volumes overlap completely.

The DSC is dependent on the structure size. For a given mismatch (such as an offset), the DSC is smaller for small volumes, and can thus be sensitive when small volumes are compared.

Figure 2.10: Illustration of **(a)** the Dice similarity coefficient (DSC) and **(b)** Hausdorff distance (HD) similarity measures.

Hausdorff Distance

A common distance metric is the Hausdorff distance, which measures the maximum distance between two point sets A and B. The HD measures both the maximum distance from the set A to the nearest point in set B, and vice versa, as seen in Figure 2.10(b). The metric is defined as the largest of these two directed distances. Mathematically it is given by

$$\text{HD}(A, B) = \max\big(h(A, B), \ h(B, A)\big), \tag{2.12}$$

where

$$h(A,B) = \max_{\vec{a} \in A} \min_{\vec{b} \in B} \|\vec{a} - \vec{b}\| \tag{2.13}$$

is the directed Hausdorff distance between A and B. The $\| \cdot \|$ is the Euclidean distance between each pair of points.

The HD is sensitive to outliers, and generally increases with the size of the measured structures. A percentile of the Hausdorff distance, e.g. the 95% HD is sometimes used to limit the sensitivity to outliers, but has not been considered in this project.

3 Materials and Methods

3.1 The RIC Study

The data analyzed in this project has been obtained through the RIC study (A **R**andomised, Two Centre Trial on Daily Cone-beam vs Standard Weekly Orthogonal **I**mage Guided Radiotherapy (IGRT) for Prostate **C**ancer), which was carried out by the St. Olavs and Ålesund Hospitals. Inclusion criteria for the study included diagnosis of intermediate or high-risk prostate cancer. A total of 260 patients were included in the study, and the patients were treated with external-beam radiotherapy in the period between 2012 and 2015.

The study investigated side effects in patients treated with curative RT up to 78 Gy given in 39 fractions of 2 Gy. The primary aim of the study was to determine whether daily position verification with cone-beam CT (CBCT) and reduced PTV safety margins (arm B) may reduce rectal side effects compared to weekly CT position verification and standard safety margins (arm A). In addition to the primary outcome of rectal side effects, outcomes such as patient survival, freedom from biochemical recurrence, quality of life, and side effects related to urinary and sexual function were assessed.

In both study arms, the CTV1 70-78 was defined as the prostate with any suspected extra tumour growth into the seminal vesicles, as described by clinical findings. The CTV2 0-70 was defined as the prostate and 1 or 2 cm of the seminal vesicles (for medium- and high-risk patients, respectively). The PTV1 70-78 was defined as the CTV1 without margins. In arm A with standard PTV margins, the PTV2 0-70 was defined as the CTV2 with a 15 mm margin in all directions, except in the posterior direction where a 10 mm margin was applied. For patients in arm B, the PTV2 0-70 with reduced margins was defined as the CTV2 with a 7 mm margin in all directions.

The study was approved by the Regional Committee for Medical and Health Research Ethics (reference no. 2011/710).

3.1.1 Treatment and Treatment Planning

Patients were treated with a 15 MV photon beam. The PTV2 0-70 received doses up to 70 Gy, given in 35 fractions of 2 Gy daily, and a four-field box technique was used (two lateral, one anterior, and one posterior beam). Then, a boost dose consisting of four fractions of 2 Gy was given to the PTV1 70-78, consisting of the CTV1 with a 3 mm margin. For the boost dose a five-field technique was used (two lateral, one posterior, and two anterior-oblique beams). The treatment technique was the same in both treatment

arms, however the treatment margins for the PTV2 0-70 differed.

The dose requirement for the PTV was that the volume should receive between 95 and 107 % of the prescribed dose. For the rectum, the dose limit was defined as 60 Gy to half the rectal circumference. No dose limit was defined for other organs at risk.

3.1.2   Patient Follow-Up

The patients were followed every 6 months for the first two years, then annually until 5 years post-treatment. The last follow-up will be at 10 years post-treatment. Follow-up consists of a clinical exam, lab tests, and physician-evaluated side effects (CTCAE v4.0). Further, patients self-reported on quality of life and side effects from treatment by responding to two questionnaires. The questionnaires were distributed to patients at inclusion (month 0, at the start of RT), as well as at 2 months (end of RT), 5, 12, and 18 months after inclusion, annually until 5 years after inclusion, and at 10 years after inclusion.

Patient-Reported Outcome Measures

The assessment of side effects in the RIC study was done using PROMs from two questionnaires; the EORTC QLQ-C30 [22] and the QUFW94 [23], both presented in Section 2.3.3 and available in Appendix A. Questions are scored on a scale from 0 to 10, where 0 means "no problem"/"very good function", and 10 means "very many problems"/"very bad function". Only the QUFW94 questionnaire will be addressed in this project.

Rectal Bother Score as Primary Outcome

A combined measure of five items from the QUFW94 questionnaire was summarized as the rectal bother score (RBS) in Tøndel et al. [48]. The following items are included in the RBS (QUFW94 question numbers in parentheses): overall bother from bowel symptoms (47), stool frequency (48), stool leakage (50), planning of toilet visits (51), and overall bother from bowel symptoms (58). These questions are scored on a numerical scale from 0 to 10, except question 48, which is the number of toilet visits per day.

Tøndel et al. [48] compared RBS scores between the two treatment arms to evaluate acute rectal toxicity. They calculated the RBS score by averaging scores for each of the five items, summing these averages, and dividing by five. The group found no differences in RBS between the two study arms, and no significant differences in the other outcome measures.

3.2   Python Code

The code written as part of the project is available from a public repository on the author's GitHub: `https://github.com/ingridtv/NTCP-model`.

## 3.3 Implementation of Core Functionality in Python

This section will describe the workflow in the process from DICOM dose plans to dose data that may be used for analyses.

Dose plans were exported from the treatment planning system and stored on an external hard drive, with one folder for each patient. Dose plans for the 0-70 and 70-78 Gy treatments were exported separately and stored in separate folders. Dose files (DICOM type: RT Dose) were exported for each beam. The 0-70 and 70-78 Gy plans are referred to as *base* and *boost* dose groups in the program, and the same labelling will be used here.

The most important packages used to implement the core functionality of the program is `numpy`, used for array handling and operations, and `scipy`, used for statistical analysis and curve fitting.For DICOM handling and data extraction, `pydicom` and `dicompyler-core` were used. The `matplotlib` package was used for visualization and plotting.

3.3.1 Initialization of a `Patient` Object

The first step in the program is to initialize a `Patient` object. During initialization the `DICOMLoader` class reads DICOM files from the patient's folder, sorts the files by DICOM type (e.g. CT image, RT Dose, RT Struct), and stores the data as `pydicom Datasets`. This is done for each of the base and boost folders. Next, the total dose distribution is found by summation of doses from each of the beams. The 3D dose matrix for each of the dose `Datasets` is read into a `DoseMatrix` object. The `DoseMatrix` class (see Appendix **??**) has functionality for addition of dose distributions. It also stores information such as the location of voxels (in patient coordinates) and the dose units.
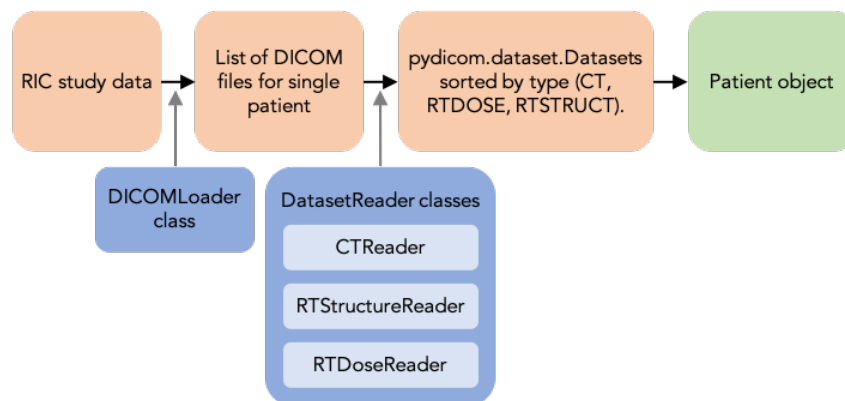
Figure 3.1: Schematic of the initialization of a `Patient` object (green) with data extracted from files.

The DICOM file of type RT Struct contains information on the delineated structures for the patient. Data on each of the delineated structures is read from the RT Struct file and stored as a `ROIStructure` object (see Appendix **??**) that contains several `ROISlices`, as well as the structure name and identifier (an integer number). Each `ROISlice` contains

contour data for the structure. In sum, the `ROIStructure` class defines the outline of a structure in the patient coordinate system.

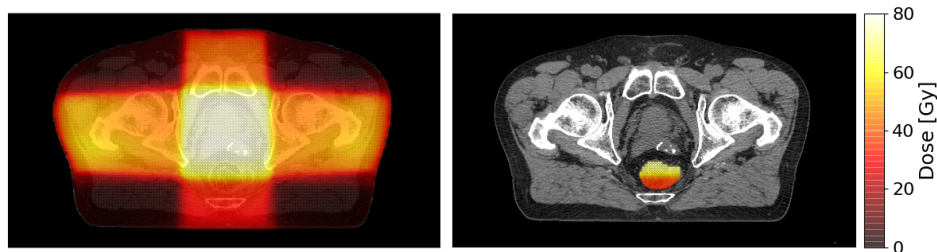### 3.3.2   `DoseMatrix` Masking

The `DoseMatrix` and `ROIStructure` classes can interact through *masking* of dose matrices. Dose matrices are masked to extract the dose distribution for a single organ or structure from the whole dose distirbution. Masking means that all dose points outside the outline
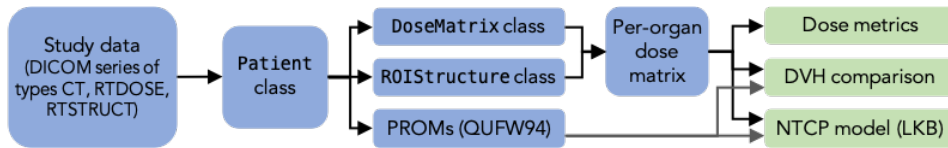
Figure 3.2: Schematic of the processing and use of study data for different analyses. From DICOM data series, the `Patient` class was used to extract CT images, dose distributions, and structures of interest. This data, as well as PROMs, were used in analysis of dose and DVH data and NTCP modelling with the LKB model.

defined by the `ROIStructure` are set as invalid, leaving only the points within the outline. This is done by creating a masking matrix of Booleans with the same dimensions as the dose matrix. An example of a masked dose distribution is shown in Figure 3.3.

Figure 3.3: An example of masking of a `DoseMatrix`. The unmasked `DoseMatrix` (left) is masked by the rectum `ROIStructure`, effectively removing dose points outside the contour.

By iterating through the contour points in the `ROIStructure`, mask matrix elements are set to `True` if they are within the contour, and `False` if not. The algorithm for generating a mask matrix from a DoseMatrix and a ROIStructure is given in Listing 3.1.

Code listing 3.1: Pseudo code for generating mask matrices

```
1    mask_matrix = array(shape = shape(DoseMatrix))
2
3    For each z-plane in the DoseMatrix
4        If the ROI is in the z-plane
5            contour_path = matplotlib.path.Path(contour) generates a
6              closed path of the points
7            contour_path.contains_points() with the (x, y) points in the
8              DoseMatrix gives an array A of Booleans with shape (Nx*Ny)
```

```
 9          Reshape A to match the shape of the z-plane
10          mask_matrix[z] = A
11      Else
12          mask_matrix[z] = array of 'False' (no points are within contour)
```

Finally, the `numpy.ma` module (`ma` for "masked array") is used to mask the dose array of the `DoseMatrix` with the mask matrix. The resulting `DoseMatrix` can then easily be used for per-organ analyses.

3.3.3   Rectal Wall Extraction

As the rectal wall was not available as a separate contour from the treatment planning system, an algorithm to extract the rectal wall dose matrix from the rectum and rectal mucosa dose matrices was devised. The rectal mucosa is the mucosal layer lining the inside of the rectal wall. In the rectal wall extraction, the idea is to consider the rectal wall to be the area inside of the rectum contour and outside the interior of the rectal mucosa.

(a)   (b)
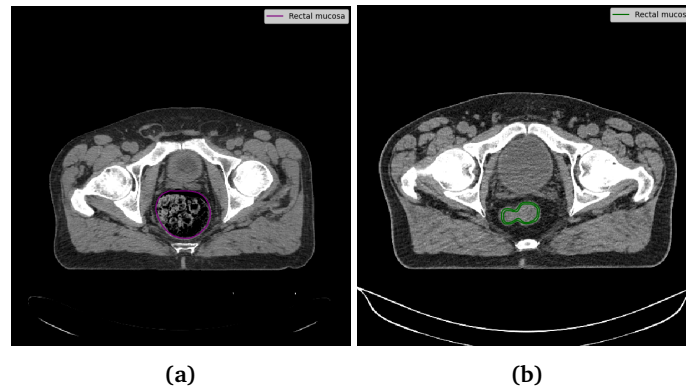
Figure 3.4: Sample of contours of the rectal mucosa for two patients. The rectal mucosa is contoured following one of two main approaches by different clinicians: Either the outer contour of the rectal mucosa is outlined (left), or the rectal mucosa is delineated as a closed contour (right). Patients were chosen to demonstrate the variation in contours.

The rectal mucosa is contoured in two different ways for the various patients. Some patients have the structure contoured along the exterior of the rectal mucosa as in Figure 3.4(a), while for other patients the contour encloses the circumference of the rectal mucosa, but not its contents (Figure 3.4(b)). Further, the contours may be convex or have concave regions.

To handle both these methods of delineation, an algorithm for extracting rectal wall mask and dose data was developed and is given in Listing 3.2 below.

Code listing 3.2: Pseudo code for rectal wall extraction from rectum and rectal mucosa

```
1   For structure in {rectum, rectal mucosa}
2
3       ROI = ROIStructure for 'structure'
4
5       For each ROISlice in ROI
6           Get the contour points in the slice
7           Use scipy.spatial.ConvexHull() to find the convex hull
```

```
 8              If contour adjustment is given
 9                  If structure = rectum
10                      Expand the convex hull contour by 'margin' (parameter)
11                  Else if structure = rectal mucosa
12                      Shrink the convex hull contour by 'margin'
13
14          Generate mask matrix from ROI and a DoseMatrix
15
16      Rectal wall matrix = logical XOR(rectum, rectal mucosa) of mask matrices
```

The result is one mask matrix (of Booleans) for each of the rectum and rectal mucosa
ROIs. Assuming that the points within each ROI is set to 'True' in the mask_matrix, the
rectal wall mask is found as the logical XOR[1] of the two ROI structure matrices.

The contour adjustment, if opted for, can be both is performed as described in Listing
3.3:

Code listing 3.3: Pseudo code for contour expansion or shrinking

```
 1      Find the displacement (dx, dy) of the centroid from the origin
 2      Get the vertices of the convex hull
 3
 4      For each point in the structure
 5          Translate the point by (-dx, -dy)
 6          Get polar coordinates (rho, theta) of the points
 7
 8          If mode = 'absolute'
 9              rho = rho + margin (margin can be positive or negative, for
10                expanding or shrinking the contour)
11          If mode = 'relative'
12              rho = rho * margin (again, margin can be positive or negative)
13
14          Get cartesian coordinates (x, y) of the points
15          Translate the point back by (dx, dy), corresponding to the
16            original position of the centroid
```

The rectal mucosa was shrunk according to the above algorithm, and the radius made
smaller by 1 mm following the reports on rectal wall and rectal mucosa thicknesses in
[49]. The rectum circumference was also expanded by 1 mm to ensure that all rectal
mucosa points were contained within the rectum contour.

3.3.4   Extraction of Dose-Voxel Data and DVHs

Dose-per-voxel data for a structure is extracted by taking a masked DoseMatrix and
using the numpy.ma.compressed() function to flatten the DoseMatrix.dose array, leav-
ing only the unmasked points. The resulting array contains one dose value for each
$2 \times 2 \times 3$ mm voxel in the structure. The dose-voxel data is written to a text file so it may
be used further (the processes of initializing patients and masking of dose matrices is
somewhat time consuming, taking an estimated 10-15 seconds on a laptop computer).

The masked and flattened dose-voxel data can be used directly in the dicompyler-
core.dvh.DVH class, which provides a simple interface for extraction of DVH statistics,
plotting of DVHs, and conversion between differential/cumulative and absolute/relative
volume DVHs. An example of a relative volume DVH for the rectum and PTV, as plotted
by the DVH class is shown in Figure 3.5.

---

[1]The logical XOR gives the set of points contained in one of the two ROI structures, but not in both of
them.
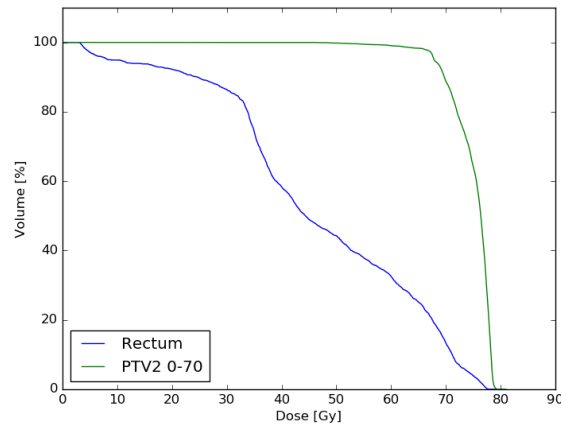
Figure 3.5: The `dicompylercore.dvh.DVH` class is used to extract DVH statistics from the dose-voxel data and plot DVHs. An example is shown here for the rectum and PTV2 0-70.

The DVHs calculated using the program were controlled by visual comparison to the DVHs generated in the treatment planning system (TPS) for a randomly selected subset of patients ($N = 17$) where each treatment arm was represented.

3.4 DVH Analysis

A two-sided Student's t-test was used to test for statistical differences between DVHs in two groups of patients. The null hypothesis is that the mean DVH values (volumes) for the ith dose bin for groups 1 and 2 ($\overline{V}_i^{(1)}$ and $\overline{V}_i^{(2)}$, respectively) do not differ for the given dose bin (H$_0$: $\overline{V}_i^{(1)} = \overline{V}_i^{(2)}$), while the alternative hypothesis is that their values differ (H$_1$: $\overline{V}_i^{(1)} \neq \overline{V}_i^{(2)}$).

DVH values were compared between the two groups in dose bins of 0.1 Gy. The significance level required for rejection of the null hypothesis was set to $p < 0.05$.

3.4.1 Comparison Between Treatment Arms

DVHs were divided in two groups based on the patients' assigned treatment arms. For all OARs and the target volume PTV1 70-78, the mean DVH in each group was calculated and each dose bin i was compared as described above. The PTV1 70-78 was chosen out of the four target volumes (CTVs and PTVs) since it has the strictest dose requirement.

3.4.2 Comparison Based on Outcome

Based on the PROMs at 36 months, DVHs were divided in two groups; those patients exceeding a set threshold (categorized as patients with complication), and those not exceeding the threshold (categorized as patients without complications). Dose bins were compared between the two groups using the method described above.

The tests were performed in three patients groups: All patients pooled, and arm A and B considered separately. The process was performed for the composite RBS and repeated for each of the single-item measures in the RBS, all considered at 36 months after RT. The threshold was set to RBS $= \{2.5, 3.5, 4.5\}$.

Set limitation that comparisons not considered if any of the groups consisted of fewer than 5 DVHs? This was the case for high threshold values, corresponding to severe complications.

3.5   LKB Model and Curve Fitting

The LKB model as described in Section 2.4.2 was implemented in the program. The calculation of gEUD as defined in Eq. (2.6) was used as the DVH reduction method.

The voxel-dose data was to be fitted to the RBS as defined in Section 3.1.2. The threshold for RBS was set to 2.5. Patients with RBS above the threshold had NTCP set to 1 (has complication), while patients exceeding the threshold had NTCP set to 0 (no complication). The threshold approach for curve fitting was chosen since it is commonly seen in literature when other complication ratings are used (see e.g. the QUANTEC review for rectal side effects [11]). The threshold of 2.5 was chosen since it yielded a similar rate of complications as other studies (see [11]).

Table 3.1: Input to the `curve_fit` function was initial guesses for the optimization of LKB parameters, as well as lower and upper bounds on the parameters in the model. The QUANTEC recommendations are chosen as initial guesses.

| Parameter	Initial guess	Lower bound	Upper bound
TD50 [Gy]	76.9	0.0	150.0
n [-]	0.09	0.007	1.0
m [-]	0.13	0.005	1.0

Optimization of the LKB model parameters TD50, *n* and *m* was performed using the `curve_fit` function from the `scipy.optimize` module. Initial guesses were set equal to the QUANTEC parameters. These initial guesses and bounds on the parameter values were as given in Table 3.1. The `curve_fit` function returns the estimated parameter values, as well as a covariance matrix. The diagonals in the matrix gives the variance of the parameters. The standard deviation of the parameters is found by taking the square root of the diagonal elements.

3.6   Voxel-Based Model

3.6.1   Number of patients

Since the VBA aims to explain spatial dose differences, it is important to be confident that the planned dose used for analysis is as equal to the delivered dose as possible. Therefore, only the patients in arm B ($N = 116$), where daily image verification was
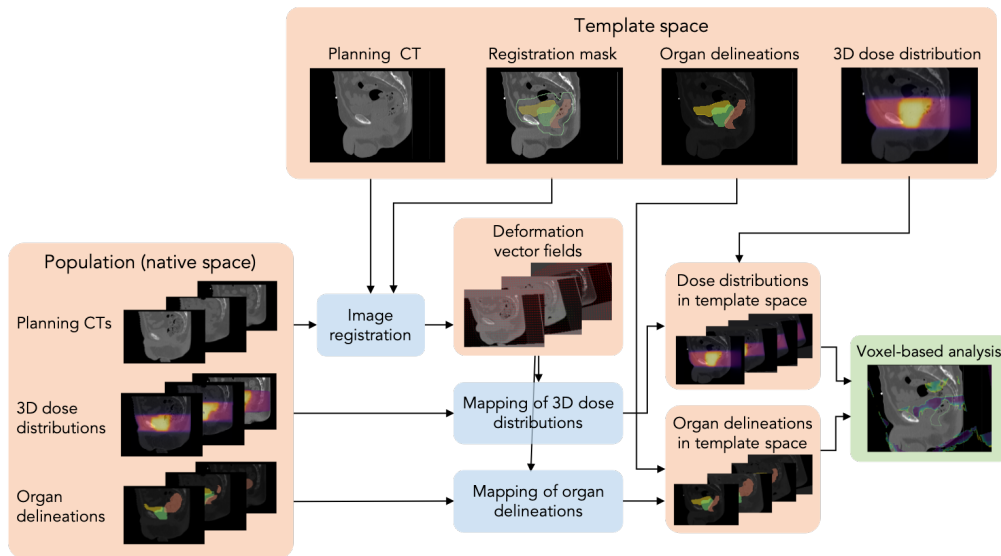
Figure 3.6: Flowchart showing the image registration and analysis process

performed, were deemed suitable for VBA. Of these patients, one patient lacked the bladder contour and was excluded from further analysis, leaving $N = 115$ patients.

To choose a suitable anatomical template, 15 patients were inspected visually to find an "average-looking anatomy" (this will be explained further in the section on template selection).

To determine the transform parameters to use for registration, a subset of 6 patients were registered to the anatomical template. Then, $N = 114$ patients were registered to the anatomical template using the transform of choice. Only patients with a DSC $\geq$ 0.5 between the registered rectum and rectum template were included in the statistical analysis.

3.6.2 Spatial Normalization

In this section, the term PTV will refer to the PTV2 0-70 (defined in 3.1) unless otherwise noted. Before any registration procedures, an automatic initial alignment between the images was made; Each 3D CT image was translated so that the centre of the image was located at the origin.

Structural Model of Patient

The structural model for the patients is determined by finding an "average" patient anatomy that can serve as the anatomic template to which the remaining patients will be registered. To determine a suitable anatomical template, a subset of $N = 15$ patients were visually inspected using 3D Slicer. For each patient, the following elements were noted:

- their apparent size (is the patient slim, average, or larger),
- rectal filling (degree of filling and whether air is present), and

- bladder filling (degree of filling).

Patients with a significant amount of air in the rectum were avoided. The most representative anatomy (average size and rectal/bladder filling) was chosen as the anatomical template.

Registration Software

SimpleElastix [50] was used for image registration. SimpleElastix is an open-source software extension of the SimpleITK [51] image processing software. The library also provides access to the image registration algorithms contained in Elastix [52].

Selecting a suitable mask to focus the registration

A registration mask defines the region of interest for the registration, and samples are drawn from the area defined by the registration mask.

A registration mask was constructed as the union of the rectum, PTV, and bladder contour masks, with a margin. The margin was added by using a `SimpleITK.BinaryDilate` filter with a dilation radius of r (corresponding to r multiplied by the pixel spacing in each direction).

    (a)   **(b)**
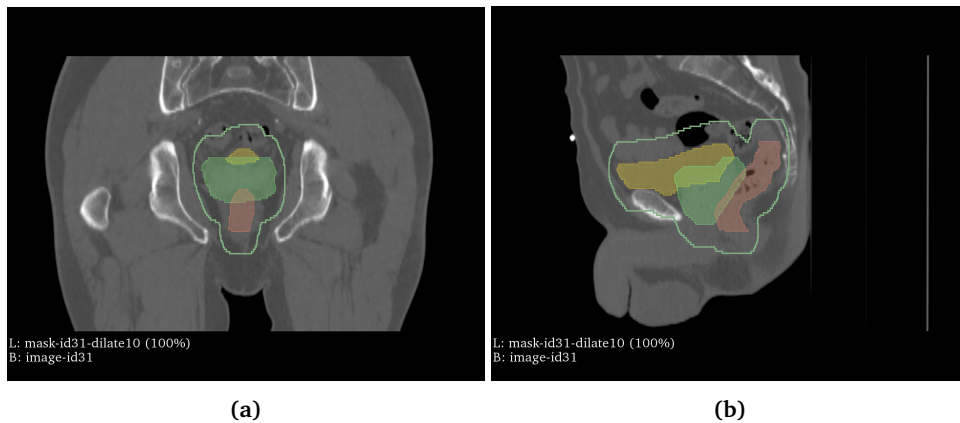
Figure 3.7: **(a)** Coronal and **(b)** sagittal views of the mask that was used to focus the registration (green line). The mask is overlaid a CT image of the anatomical template with the three contours: rectum (red), bladder (yellow) and PTV (green). The mask was the union of the bladder, rectum and PTV, with a dilation radius $r = 10$.

Four dilation radii were tested using the default registration scheme in SimpleElastix to register four patients to the template; $r = 0, 10, 20, 30$. Each registration was then evaluated by visual inspection in 3D Slicer.

- For $r = 0$, no bone structures were contained in the rectum mask.
- For $r = 10$, the mask included part of the pubic bone (the lowest, most anterior part of the pelvic bone).
- For $r = 20$, the mask extended to the top of the FOV of the CT image of the template patient. In the left-right direction, the hip joint was included in the mask.

- For $r = 30$, the mask covered nearly all of the patient.

A margin of $r = 10$ yielded the best registration result.

Selection of a Suitable Registration Scheme

The default registration scheme in SimpleElastix is a multi-resolution, multi-transform registration. Three transforms of increasing complexity - rigid, affine, and B-spline (deformable) transformations - are applied sequentially to align the moving image to the fixed image.

Four different registrations schemes were attempted for a subset of patients (the first 6 patients in arm B by patient ID). These are:

1. default registration scheme (rigid - affine - B-spline), see Appendix C.1
2. default registration scheme with increased maximum number of iterations (from 256 to 512), see Appendix C.2
3. affine transform, see Appendix C.3
4. affine + B-spline transform, see Appendix C.4

All the registration schemes use the similarity metric `AdvancedMattesMutualInformation` for the rigid and affine transforms. For the deformable registration in the last step of the default registration schemes, a metric `TransformBendingEnergyPenalty` is used additionally to penalize transforms with large deformations. The metric weights in this step are 1:1.

For each registration scheme, the overlap between the resulting rectum contour and the template rectum contour was assessed by visual inspection in 3D Slicer, as well as by using DSC. The mean DSC and HD for the rectum, bladder and PTV for the 4 methods is presented as a part of the results.

The affine registration scheme had the highest DSC for the rectum, and was therefore chosen for registration with the rest of the patients.

Inter-Patient Registration Pipeline

For each patient, the registration pipeline is as follows:

1. Read DICOM series for the patient as a 3D volume
2. Use a fixed image mask to focus the registration
3. Using an `ElastixImageFilter`, calculate the transform of the patient image into the anatomic template
4. Using a `TransformixImageFilter`, calculate the deformation field of the resulting transform
5. Transform the dose distribution correspondingly
6. Transform each contour map (rectum, PTV, and bladder) correspondingly to ensure overlap between the

   - transformed CT image and transformed contour,
   - transformed rectum, bladder, and PTV contours, and the template structures.

Before this step, the `FinalBSplineInterpolationOrder` of the `TransformixImageFilter` was set to 0 (corresponding to NN interpolation, as mentioned in 2.5.2) to ensure the binary nature of the contour maps are kept.

3.6.3   Evaluation of Image Registration

Quantitative Similarity Metrics

To assess the quality of the image registration, the contours for the rectum, bladder and PTV were propagated from the native CS to the CCS using the deformation field found as described in Section 3.6.2. The template contours Ω_T and the propagated contours $\Omega_i^{(\text{CCS})}$ for each patient i were compared using the DSC and the (100 %) HD.

DVH Comparison Before and After Registration

The DVH statistics $D98\%$, D_{mean} and $D2\%$ for each patient were compared before and after registration to ensure DVHs from transformed contour masks and dose distributions corresponded with the original DVHs.

DVH comparisons were visualized for four patients:

- DVH from transformed structure and dose distribution vs DVH in native coordinates
- DVH from transformed structure vs. DVH from template structure, both using the transformed dose distribution

Visualization of Rectal Alignment

For each of the registered rectums, the average of the binary mask (points within the rectum were set to 1 and points outside the mask were set to 0) was plotted. The average of the masks thus represents the probability that a the registered rectum is present in each voxel. The probability distribution was plotted together with the rectum template for all registered patients, and for those with DSC ≥ 0.45 and DSC ≥ 0.5.

3.6.4   Voxel-Based Statistical Analysis

Since all images are resampled to the CCS defined by the template CT, and thus have the same shape and location, it is easy to perform a voxel-by-voxel comparison. To illustrate how the process described in this section can be used to create a dosimetric voxel-based model, a comparison of dose between patients with and without complications was carried out.

Based on the registration, a cut-off of DSC $\geq$ 0.5 for the rectum was needed to be included in the statistical analysis. This yielded a subset of 32 patients. For these patients, the mean DSC and range was $0.56(0.50-0.73)$, while the mean HD and range was $22.9\,\text{mm}(13.6-38.9)$.

Patients were divided into two groups based on their experienced rectal toxicity rating (RBS) at 5, 12, 18, 24 and 36 months after RT. Of the 32 patients with the best Dice scores, 24 patients had answered all the five items in the RBS and was included

in the analysis. A voxel-wise t-test was performed between the dose distributions of the two groups, and the dose difference and p-values were plotted over the template CT image to be able to identify possible radiosensitive regions within the rectum.

4 Results

4.1 DVH Analysis

To analyze the dose to the rectal wall, a method for finding the rectal wall from the rectum and rectal mucosa contours was developed. An example of the results of this procedure is presented in Section 4.1.1.

Mean DVHs between two groups of patients were compared using a two-sided Student's t-test. First, DVHs were compared between treatment arms. The results are presented in Section 4.1.2. Second, rectal wall DVHs were compared for patients with and without complications. Results of these analyses are presented in Section 4.1.3 and 4.1.4.

4.1.1 Rectal Wall Extraction

The rectal wall was not delineated as a separate volume, so a method to extract the rectal wall dose matrix was developed to be able to analyze the rectal wall DVHs. An example of the resulting rectal wall masked dose is shown for one CT slice for a patient in Figure 4.1.

4.1.2 Comparison Between Treatment Arms

OARs

The mean DVHs of each of the treatment arms were compared for five OARs (rectum, rectal mucosa, rectal wall, bladder and penile bulb) using a two-sided t-test. Plots of the mean DVHs are shown in Figure 4.2. The shaded areas illustrate the standard deviation. The p-values from the t-test are plotted together with the mean DVHs, and the dashed line in each plot represents the significance level ($p < 0.05$).

On average, patients in arm A had a larger relative volume irradiated than patients in arm B. This was true for doses between $2\,\mathrm{Gy}$ and $75\,\mathrm{Gy}$ for all OARs ($p < 0.001$). No significant difference was found in the 0-2 Gy and 75-78 Gy region. The mean absolute volumes of the organs were found not to be significantly different between treatment arms.

Target Volumes

Delineated volumes for the four target volumes (CTV/PTV 0-70/70-78) were compared and the results are given in Table 4.1. Plots of the mean DVHs are presented in Figure
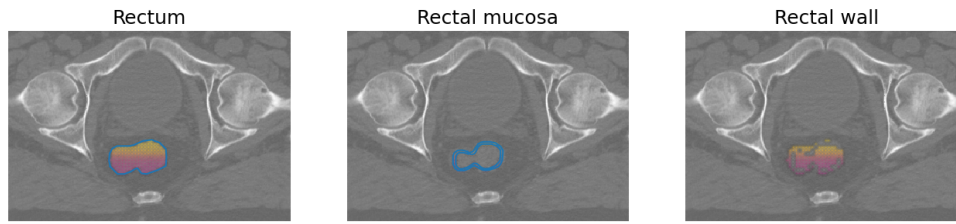
(a) Original delineations and rectal wall masking
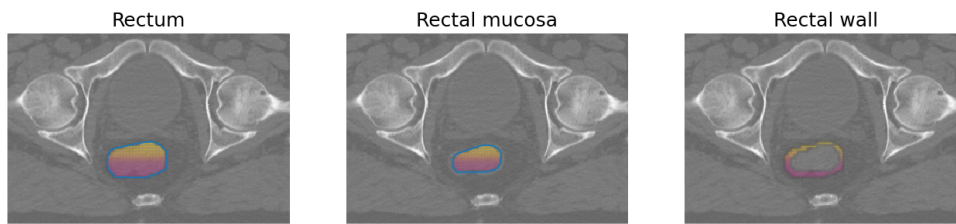
(b) After application of `ConvexHull` approach

Figure 4.1: Extraction of the rectal wall dose matrix using the convex hull approach (bottom) is compared to the original result (top). Using the convex hull approach, only the rectal wall is contained in the dose matrix, not the contents of the rectum.

(a) Rectum   **(b)** Rectal mucosa   **(c)** Rectal wall

(d) Bladder   **(e)** Penile bulb
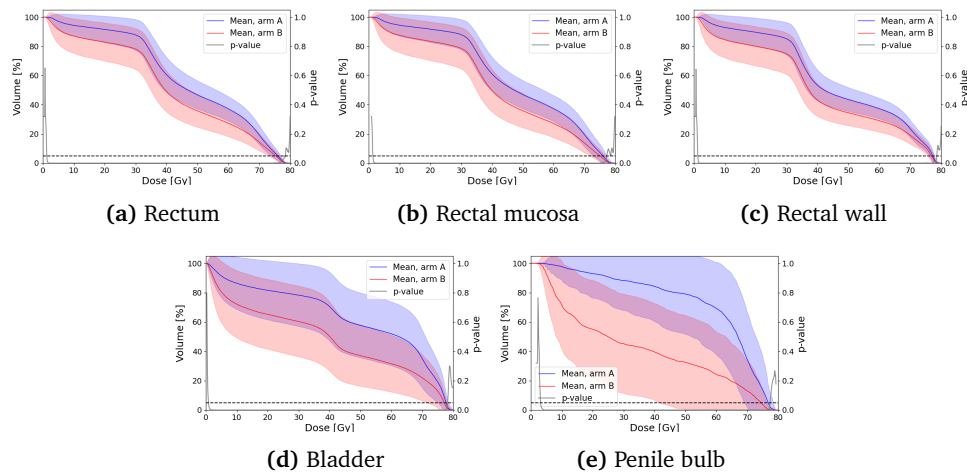
Figure 4.2: The mean DVHs and standard deviation (shaded areas) for each of the treatment arms are plotted against dose for five OARs. The p-values are plotted on the secondary (right) axis, and the dashed black line indicates a significance level of $p < 0.05$.

4.3. The shaded areas illustrate the standard deviation. The mean delineated volume of the PTV2 0-70 was significantly larger in arm A (with standard margins) than in arm B (with reduced PTV margins). The mean absolute volumes of the other target volumes were found not to be significantly different between treatment arms.

The DVHs for the PTV1 70-78 are shown in Figure 4.3, while figures for the CTV1

Table 4.1: Comparison of the mean volume V between treatment arms for each of the target volumes.

Organ	Arm A		Arm B		Volume	
	N	V [cm^3]	N	V [cm^3]	Difference	p
PTV2 0-70	119	280.0	116	132.2	9.2	< 0.001
PTV1 70-78	116	64.6	111	63.7	8.4	0.774
CTV2 0-70	116	49.2	111	49.6	0.2	0.898
CTV1 70-78	118	41.9	115	41.3	17.3	0.799

70-78, CTV2 0-70 and PTV1 0-70 are presented in Appendix B.1. On average, patients in arm A and B received very similar doses to the target volumes. The PTV2 70-78 had

**(a)** PTV1 70-78 (whole DVH)   **(b)** PTV1 70-78 (doses 70 to 78 Gy)
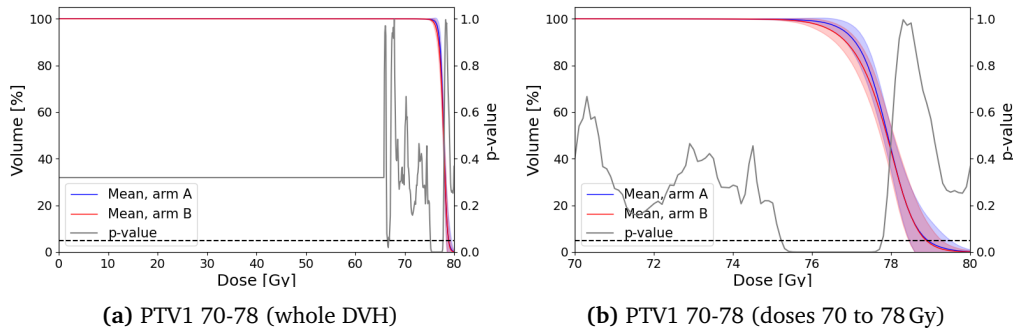
Figure 4.3: The mean DVHs and standard deviation (shaded areas) for each of the treatment arms are plotted against dose for (a) the whole PTV1 70-78 DVH, and (b) The same DVH with doses from 70 to 78 Gy. The p-values are plotted on the secondary axis, where the dashed black line indicates a significance level of $p < 0.05$.

a slightly larger V75Gy–V78Gy in arm A than in arm B, meaning that the PTV coverage is slightly less in arm B than in arm A. However, PTV coverage was sufficient in both treatment arms.

4.1.3 Rectal Wall DVH Comparison Based on RBS

For patients in both arms, in treatment arm A (weekly verification and standard margins), and in treatment arm B (daily CBCT verification and reduced margins), DVHs were compared between patients with and without adverse outcome. As described in Section 3.4.2, the outcome measured was RBS with varying thresholds. The results for a cut-off value of RBS ≥ 2.5 (all patient groups) are shown in Figure 4.4, and for RBS ≥ 4.5 in arm A is shown in Figure 4.5. Results for RBS ≥ 3.5 and 4.5 can be found in Appendix B.

The numbers of patients having RBS ≥ 2.5 were 27/170 (15.9 %), 16/82 (19.5 %), and 11/88 (12.8 %) in both arms, and arm A and B, respectively. In arm A, 7/82 patients (8.5 %) had RBS ≥ 4.5.
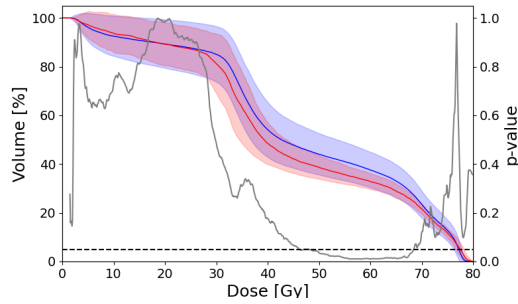
(a)

(b)

(c)
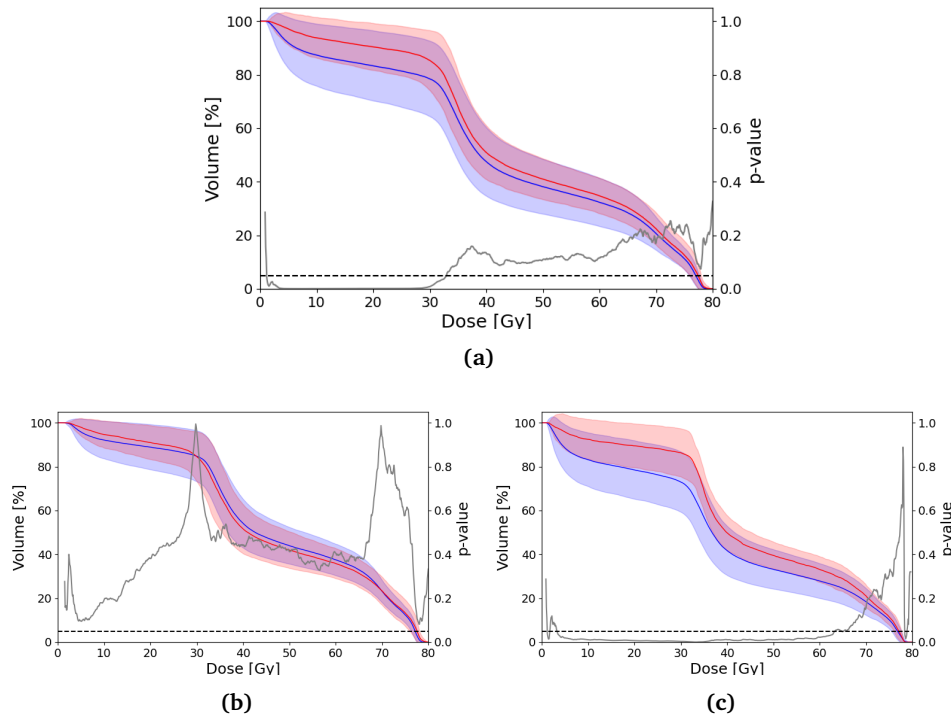
Figure 4.4: Mean DVHs and standard deviation (corresponding shaded regions) for the rectal wall are plotted for patients with (red) and without (blue) adverse outcome. The cut-off value was set to RBS ≥ 2.5. DVH comparison was performed for patients in **(a)** both arms pooled, **(b)** arm A and **(c)** arm B. The gray lines shows p-values from a two-sided t-test, and is plotted on the secondary (right) axis. The dashed black line indicates a significance level of $p < 0.05$.

As seen in Figure 4.5, for arm A and a cut-off of RBS ≥ 4.5, patients without complications had a significantly higher V50Gy–V70Gy than the patients with complications.

4.1.4   Rectal Wall DVH Comparison for RBS and Single PROMs

To investigate the influence of various single-item measures on rectal bother score, mean DVHs were compared between patients with (RBS ≥ 2.5) and without complications for the rectal wall. The results are presented in Figure 4.6.

Of the single-item PROMs, overall bother from symptoms (Figure 4.6(b)) is the only item not showing significant dose difference between patients with and without complications. For planning of toilet visits, stool frequency and stool leakage (Figure 4.6(d)-(f)), patients exceeding the threshold had significantly larger V2Gy–V70Gy. V5Gy–V35Gy were significantly larger for patients rating that symptoms limit their daily activity.

Figure 4.5: Mean DVHs and standard deviation for the rectal wall for patients in arm A with (red) and without (blue) adverse outcome. The cut-off value was RBS ≥ 4.5. The gray line shows *p*-values from a two-sided *t*-test (plotted on the secondary axis), and the dashed black line indicates a significance level of $p < 0.05$.
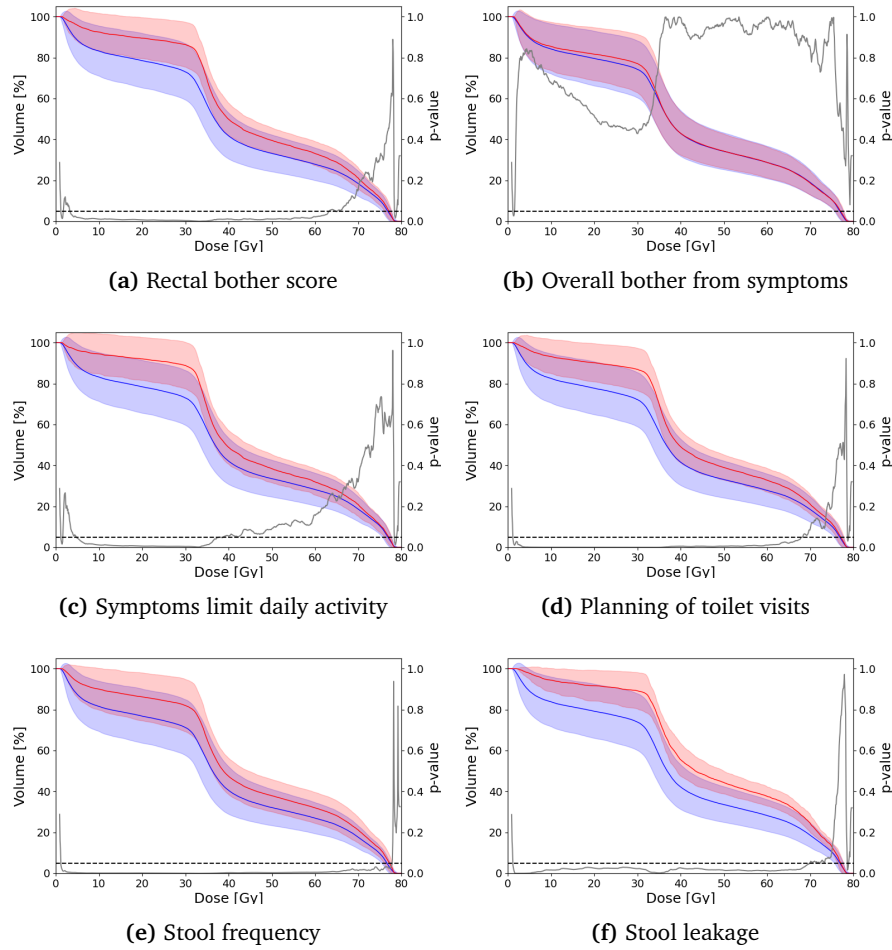
4.2 LKB Model Parameters

A total of 174 treatment plans had dose data and PROMs for the rectum available. The outcome measure used in the LKB model was rectal bother score 36 months after the end of RT, calculated from patient-reported outcome (the QUFW94 questionnaire) as described in Section 3.1.2. For curve fitting, a threshold for complication of RBS ≥ 2.5 was set. Of the 174 patients, 28 patients (16.1 %) exceeded the threshold and were categorized as having rectal complications.

The three parameters TD50, *n* and *m* were fitted to the data using the function `scipy.optimize.curve_fit`. The parameter *n* maps the distribution of dose-per-voxel data for a patient to a gEUD value, and NTCP values for each patient were placed along the x-axis according to the gEUD value from the estimated *n*. The optimal values for the parameters, together with the calculated or estimated uncertainties, are given in Table 4.2. The fitted curve, together with the patient outcomes (0 or 1) and the curve given

Table 4.2: Optimized parameters for the LKB model for the rectum. The outcome measured was rectal bother score with a threshold for complication of RBS ≥ 2.5. 28/174 patients (16.1 %) exceeded the threshold. The optimal value ± estimated standard deviation is given. Recommended values from the QUANTEC review [11] are provided for reference.

Parameter	curve_fit	QUANTEC
TD50 [Gy]	78.9 ± 3.9	76.9
n [−]	0.007 ± 0.014	0.09
m [−]	0.053 ± 0.043	0.013

by QUANTEC parameters is plotted in Figure 4.7.

(a) Rectal bother score

(b) Overall bother from symptoms

(c) Symptoms limit daily activity

(d) Planning of toilet visits

(e) Stool frequency

(f) Stool leakage

Figure 4.6: Mean DVHs and standard deviation (corresponding shaded regions) for the rectal wall are compared for patients with (red) and without (blue) adverse outcome. Patients in study arm B were included in the analysis, and the cutoff value for the outcome was set to RBS ≥ 2.5. The p-values from a two-sided t-test are plotted on the secondary (right) axis, and the dashed black line indicates a significance level of $p < 0.05$.
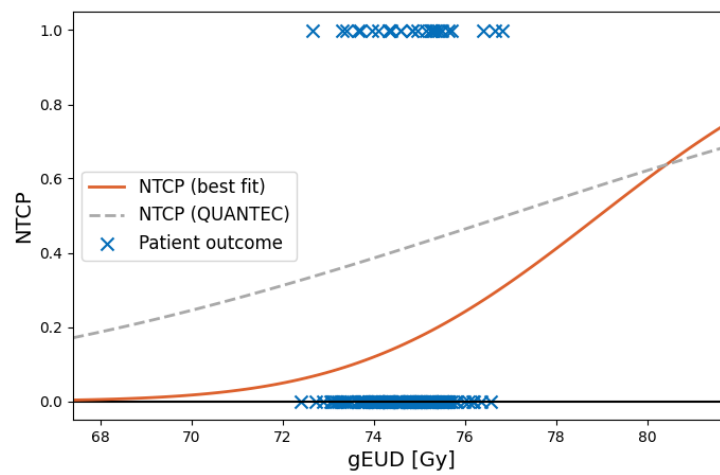
Figure 4.7: The fitted LKB model curve (red) is plotted together with the QUANTEC curve (dashed, gray) and the patient outcomes (blue crosses). Patients outcomes are plotted with the gEUD found with the fitted parameters on the horizontal axis. Patients with RBS less than the threshold are given a value of 0 (no complication), while patients exceeding the threshold had a score of 1 (complication).

4.3 Evaluation of Image Registration

4.3.1 Selection of transform

The DSC and HD found for the four initially tested transforms (for six patients) are presented in Table 4.3.

Table 4.3: Mean DSC and HD for the bladder, rectum, and PTV2 0-70, are given for the four transforms initially tested.

Registration scheme		Affine	Affine + B-Spline	Default	Default, 512 iterations
DSC [−]	Rectum	0.40	0.40	0.25	0.20
	Bladder	0.68	0.62	0.61	0.58
	PTV2 0-70	0.70	0.61	0.67	0.65
	Total	0.60	0.54	0.51	0.48
HD [mm]	Rectum	31.1	39.5	35.0	36.0
	Bladder	22.7	33.1	24.3	26.4
	PTV2 0-70	16.6	27.8	21.2	23.2
	Total	23.5	33.5	26.8	29.0

Based on these results, the affine transform parameter scheme was chosen for registration for the rest of the patients.

4.3.2 Similarity Metrics

Each of the rectum, bladder and PTV2 0-70 were compared to the corresponding structures of the template. The distribution of DSC and HD for each structure is visualized in
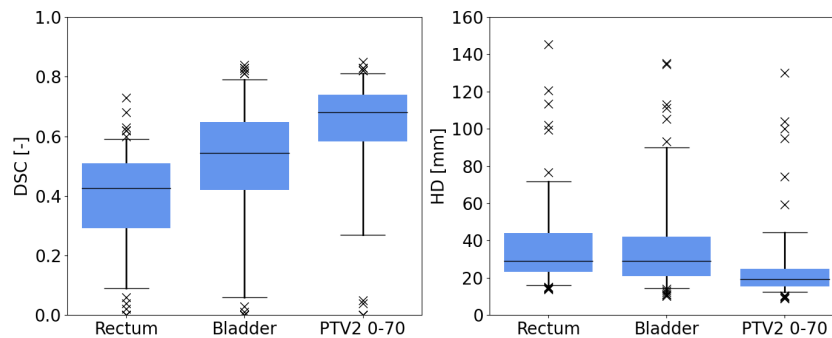
Figure 4.8: The boxplots show the DSC (left) and HD (right) of the anatomical template structures and the structures that have been registered to the CCS. Boxes represent the interquartile range, lines mark the median. Whiskers represent the 5th and 95th percentile. Outliers are indicated by dots.

the boxplots in Figure 4.8, and a summary of the metrics for each structure is presented in Table 4.4.

Table 4.4: Statistics for DSC and HD for the rectum, bladder and PTV 0-70, after structures have been registered to the CCS where the corresponding anatomical template is found.

	Rectum		Bladder		PTV 0-70	
	DSC	HD	DSC	HD	DSC	HD
Minimum	0.00	13.6	0.00	10.1	0.00	8.9
Median	0.42	29.0	0.55	29.0	0.68	19.4
Mean	0.39	37.4	0.52	36.9	0.63	23.8
Maximum	0.73	145.4	0.84	135.2	0.85	130.2
STD	0.16	22.8	0.20	24.7	0.18	18.5

4.3.3 Visual Comparison of Rectal Alignment

To visualize the rectum alignment after registration, the mean of the binary rectum masks were plotted. The mean rectums are shown in Figure 4.9. Note that for the group with rectum DSC $\geq$ 0.45, two patients were excluded after visual inspection.

Since the binary rectum maps contain ones (1) where the rectum is present, and zeros (0) where not present, the voxel-wise mean of the binary rectum maps yields the

(a) All patients ($N = 116$)   **(b)** DSC $\geq$ 0.45 ($N = 49$)   **(c)** Rectum DSC $\geq$ 0.5 ($N = 33$)
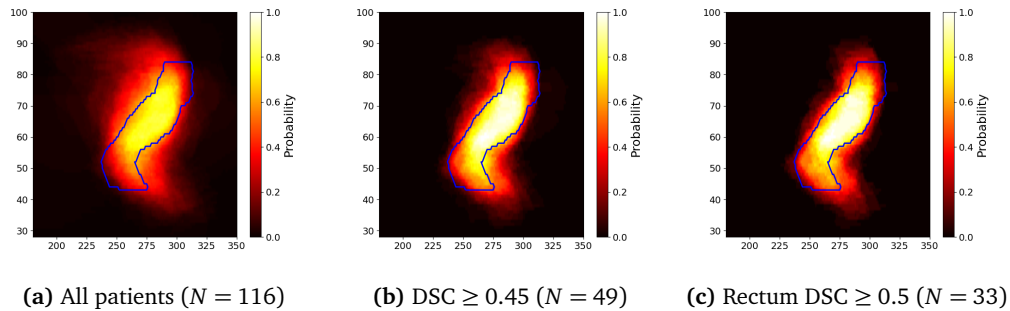
Figure 4.9: Visual representation of the rectum location after registration, where the value at each location equals the fraction of patients that have the rectum present in that voxel. The anterior direction is to the left in the images. The rectum template is shown in blue. For the group with rectum DSC $\geq$ 0.45, two patients were excluded after visual inspection.

fraction of rectums that are present in each voxel. This can be interpreted as a probability of finding the rectum in that voxel after registration.

4.3.4   DVH Comparison Before/After Transformation

DVH Statistics Before/After Transformation

For each patient, the DVH statistics D98% (near-minimum), D_{mean}, and D2% (near-maximum) were calculated before and after registration (i.e. in the native and common coordinate systems, respectively). The difference between the values before and after registration were calculated, and the minimum, median, mean and maximum difference is presented in Table 4.5.

DVH statistic	Difference (after − before) [Gy]			
	Minimum	Median	Mean	Maximum
D98% [Gy]	0.00	0.04	0.19	3.62
D_{mean} [Gy]	0.00	0.05	0.27	20.30
D2% [Gy]	0.00	0.06	0.32	27.70

Table 4.5: Difference of some DVH statistics before and after registration.

Visual DVH comparison

For a few patients, rectum DVHs were visually compared before and after registration to the CCS. The results are shown in Figure 4.10(a) using the patient's own organ delineations. No large visual differences can be seen, except very minor differences. In Figure 4.10(b), the DVHs using the registered rectums and dose distributions is compared to DVHs for the rectum template and registered dose distributions. The largest differences can be seen in the low-dose region up to $30-35$ Gy, whereas the DVHs nearly align for doses upwards of 50 Gy.

(a) Registered structures and dose vs. native   **(b)** Registered structures vs. template
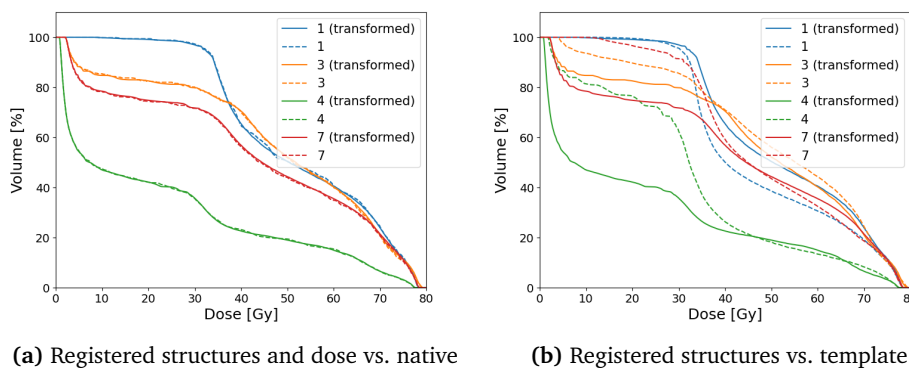
Figure 4.10: DVH comparison before and after registration. **(a)** Structures and dose distributions that are transformed to the CCS (solid lines) and native structures and dose distributions (dashed lines). **(b)** Transformed structures and dose distributions (solid lines) and the template structure and transformed dose distributions (dashed lines).

4.4 Voxel-Based Analysis

4.4.1 Statistical Analysis of Dose-Effect Relationship

Patients with a registration DSC of ≥ 0.5 ($N = 32$) were included in the statistical analysis. The patients were sorted into two groups based on their experienced rectal toxicity rating, measured by the rectal bother score. A voxel-wise t-test was performed between the two groups, and the p-values were plotted over the template CT image to identifying any radiosensitive regions within the rectum. The process was repeated for RBS at 5 ($N = 24$), 12, 18, 24 and 36 months (numbers in parentheses give the number included in each analysis). Results for 12 and 36 months are shown here, while data for 5, 18 and 24 months is found in Appendix D.

Table 4.6: Volume of and mean dose to the subregions of the rectum where dose was significantly different between patients with and without complications. Patients with complications were those with rectal bother score ≥ 2.0. Data is given for 5, 12, 18, 24 and 36 months after inclusion (start of RT), and for significance levels $p < 0.01$ and $p < 0.05$.

Months after RT	RBS ≥ 2.0 Yes	RBS ≥ 2.0 No	p-level	Volume [cm^3]	Mean dose [Gy] With	Mean dose [Gy] Without	Difference [Gy]
5	3	21	0.01	0.35	72.9	55.7	17.2
			0.05	1.30	72.4	58.2	14.2
12	4	24	0.01	0.72	73.8	64.3	9.6
			0.05	5.52	74.3	67.5	6.8
18	4	26	0.01	4.22	75.4	69.6	5.8
			0.05	6.86	74.7	68.8	5.9
24	2	26	0.01	10.50	52.1	44.2	7.9
			0.05	16.92	51.5	45.2	6.3
36	2	22	0.01	8.60	29.1	25.3	3.8
			0.05	12.71	32.8	30.1	2.7
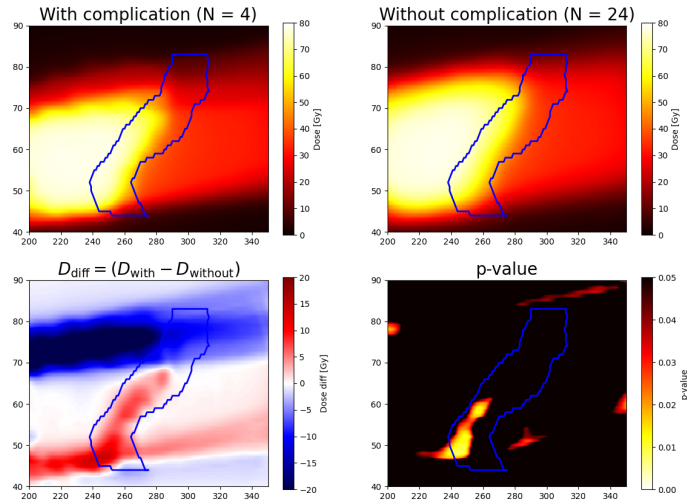
Figure 4.11: *Top:* Mean dose for patients with RBS ≥ 2.0 (left) and RBS < 2.0 (right) at 12 months. *Bottom:* The dose difference (left) and p-values from a voxel-wise t-test (right). $N = 28$ patients were included in the analysis.
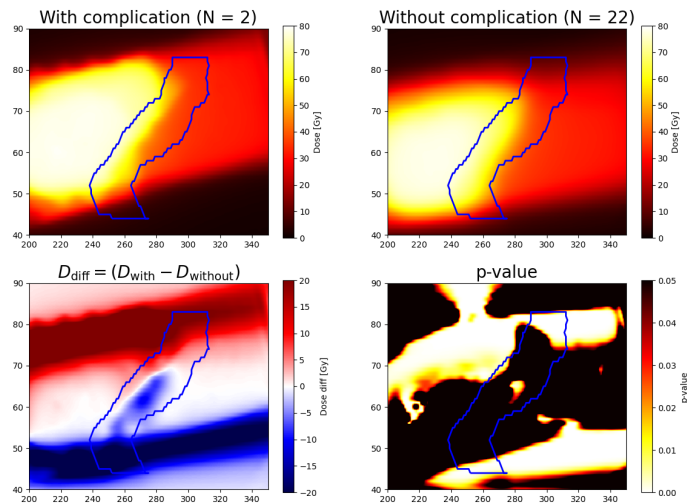
Figure 4.12: *Top:* Mean dose for patients with RBS ≥ 2.0 (left) and RBS < 2.5 (right) at 36 months. *Bottom:* The dose difference (left) and p-values from a voxel-wise t-test (right). $N = 24$ patients were included in the analysis.

5 Discussion

This work has presented various dose-response analyses of rectal complications based on patient-reported outcome. The dose has been quantitatively evaluated for all patients in the RIC study. Comparison of DVHs for patients with and without complications revealed that patients with RBS $\geq$ 2.5 had a significantly larger V2Gy–V30Gy than did patients without complications. For patients in study arm B (daily CBCT verification), the same was true for V5Gy–V65Gy. The parameters for the LKB model for NTCP was fitted to the RIC study data and yielded TD50 = 78.9 Gy, $n = 0.007$ and $m = 0.053$, showing that parameters based on PROMs provide fair agreement with literature values based on clinician-reported outcome.

The second part of the project consisted of a voxel-based analysis based on image registration and per-voxel comparison of dose. Although further improvements in the methodology is required before conclusions can be drawn, the data analysis indicates that a subregion of the rectum located in the inferior, anterior region may be correlated with acute rectal toxicity. However, uncertainty in both the registration step and in the statistical method warrants further investigation.

5.1 DVH Comparison

Comparison between mean DVHs was performed between a variety of groups and parameters. A two-sided t-test was used to compare the DVHs for each dose bin.

5.1.1 DVH Comparison Between Treatment Arms

Irradiated volumes of OARs in the two treatment arms were compared. As seen in Figure 4.2, the patients in arm A had a significantly larger irradiated relative volume than the patients in arm B for all OARs. This could be expected since the increased PTV margins in arm A allow more dose to the surrounding tissue. For arm B, dose is confined to the target volume to a greater degree. The penile bulb DVHs appear to have large variation. This is likely due to the fact that the penile bulb is very small (mean volume of $\approx 4.5\,\text{cm}^3$ in the present study), so that a small variation in the absolute volume receiving a certain dose will lead to large variation in the relative volume.

As expected, the delineated absolute volume of the PTV2 0-70 is significantly larger in arm A (standard margins) than in arm B (reduced margins). The delineated volumes of the other target volumes are not significantly different, which could be expected since

only the margins for the PTV2 differed between the two arms. There is little or no significant differences in the dose received by the target volumes in the two arms. However, when inspecting the PTV1 DVHs and focusing on the $70-78\,$Gy region, it was noted that patients in arm B have slightly lower PTV1 coverage than patients in arm A for doses ranging from $75-78\,$Gy. This is likely due to the fact that smaller PTV margins may leave part of the PTV in the edge of the beam where dose falls off.

Since PTV coverage is according to protocol in both study arms, ($D_{100\%} = 95 - 107\,\%$ of the prescription dose) this would not be taken into account in the treatment planning. If the two arms show any difference in the local recurrence of the cancer, dose to the PTV should be investigated further. A per-voxel analysis could potentially reveal any correlation between reduced PTV margins and recurrence that could be caused by reduced irradiation of microscopic cancer growth around the tumour.

5.1.2   DVH Comparison Between Patients With and Without Complication

Dose-Response and Severity of Outcome

As seen in Figure 4.4(a), when rectal wall mean DVHs are compared for all patients pooled, there seems to be a relationship between elevated volume receiving low-to-medium dose and complication. The difference is significant for doses up to around $30\,$Gy when RBS with cutoff value of 2.5 is considered. These indications contradict literature suggesting the rectum has a serial architecture and is most sensitive to moderate to high doses [11]. This is also the reason why V50Gy or V60Gy often give the dose-volume constraints for the rectum or rectal wall. A possible explanation

The correlation of the outcome with the low-dose region was investigated by considering each of the treatment arms A and B separately (see Figure 4.4(b)-(c)). In arm A, no significant correlation between DVHs and outcome was found for any dose level when the cut-off was 2.5. For a higher cut-off value of RBS ≥ 4.5 the patients without complications receive a slightly higher dose to the rectal wall than the patients with complications (statistically significant for V50Gy–V70Gy), as seen in Figure 4.5. This contradicts the assumption that higher dose to normal tissue increases the complication probability. It is however important to remember that in arm A the patients received only weekly verification. The delivered dose may therefore differ from the planned dose due to inter-fraction variations in treatment position, or in location and filling of the rectum.

In arm B receiving daily CBCT verification, it is likely that the planned dose more accurately represents the actual dose received by the patient. In this treatment arm, the irradiated relative volumes V5Gy–V65Gy are significantly larger for those patient having mild complications (RBS ≥ 2.5). For the higher cut-off value the difference was not significant (see Appendix B.2). Previous studies report that development of side effects is most dependent on the irradiation with high doses [11]. Here, the low-dose irradiation also seems to be correlated with adverse outcome.

In total, with all patients pooled, the low-dose effect dominates, and patients with and without complications have very similar volumes of the middle- and high-dose regions. Modern delivery techniques such as IMRT and VMAT have been shown to reduce the volume of the rectum that is irradiated with low to medium doses, compared to

3D-CRT [53]. These findings are therefore in favour of transition to modern delivery techniques that are able to spare normal tissue compared to 3D-CRT.

The Dose-Response Relationship Depends on the Measured Outcome

It is difficult to investigate the outcome since it varies in many respects; both the time to toxicity, and the type and severity of outcome have to be considered.

The above analysis relied on the use of the composite RBS to measure rectal complications. However, as seen in Figure 4.6 for single-item PROMs, the dose dependency varies with the type of outcome. The use of a composite score may neutralize the effect of an elevated value for any single PROM. Moreover, the dependence of each PROM with the spatial location of dose may vary. The RBS may therefore be suitable for traditional comparison of DVHs or DVH metrics, while PROMs should be investigated separately in VBA to investigate spatial relationships.

In this project, DVHs were only compared based on the toxicity at 36 months after the start of RT, whereas several studies report on cumulative incidence of toxicity (i.e. whether a symptom occurred at any time up to a certain point). Although acute and late effects may correlate with different dosimetric variables, a distinction which is of clinical interest, considering cumulative incidence will likely give an alternative perspective on toxicity. Further, the choice of appropriate cut-off values is not clear. First, it is not possible to know what is the "appropriate" rate of complications in a population of treated patients. In this project, a few different cut-offs were chosen for the DVH comparisons to illustrate exactly this point (see Section 4.1.3 and Appendix B). Ideally, one should investigate the outcome for several different degrees of complication. Second, to ensure reliability of statistical analyses, it was attempted to have a sufficient number of patients in each of the samples that are compared. If the number of patients in a sample is too small the t-test may lose its validity since it assumes the sample values to be normally distributed.

A point which will be further discussed, is that even though an appropriate outcome measure and cut-off value are chosen, the pure dose-volume measures do not say anything about where the high or low doses are deposited. Since relative volume DVHs are common, information about whether the absolute volume of the structure is relevant may also be lost. Efforts have therefore been made to develop methods that take into account the localization of dose [54], such as those in the VBA presented here.

5.1.3 Methods and Sources of Error

For DVH generation, the dose matrix (dimensions (N_x, N_y, N_z)) is used to generate a mask matrix of the same shape and spacing for each structure, which is then used to mask the dose distribution. The spatial resolution of the resulting contour mask therefore matches that of the dose matrix (spacing of 3, 3, and 2 cm in the x-, y- and z-directions, respectively), while the contour is originally delineated on the planning CT image (spacing of 2 cm along the z-axis, and approximately 0.7 to 1.0 cm in the x- and y-directions). When the dose matrix points are determined to be within or outside a contour, some errors are introduced in the volume calculation for the structures since no interpolation or resampling to the smaller spacing is performed. Since the relative

volume DVHs are analyzed, these errors likely influence DVHs for smaller structures such as the rectal mucosa or penile bulb the most, while the larger structures such as the rectum and bladder should be failry accurately represented.

Since dose to the rectum content should not have an impact on the outcome, it was of interest to consider only the dose received to the rectal wall. As mentioned in Section 3.1, the dose constraint for the rectum was defined as a dose to the rectal circumference. A method using the convex hull was developed with the intention of extracting dose-volume data for the rectal wall without contents. The contours outlined by the clinician can be assumed to represent each anatomical region correctly. Therefore, any alterations will have resulted in a less accurate contour.

This method will never decrease the size of the rectum contour, denoted as a set of points Ω, as the convex hull is defined as the smallest convex set containing Ω. Therefore, the rectum can be assumed to maintain its shape if the structure is convex, and in some cases have the contour expand. Since the rectum is located next to the PTV, the contour may have expanded toward the high-dose region. The convex hull approach may thus have lead to increased size of the high-dose region for the rectum, compared to using the original delineation by the clinician.

The convex hull of the rectal mucosa gives an approximation of the outer contour of the rectal mucosa, while the rectal wall was defined as the region between the outer contour of the rectum and the inner contour of the mucosa. Therefore, a 1 mm margin was used to shrink the rectal mucosa convex hull, following Huh et al. [49]. However, the mucosal thicknesses found by Huh et al. vary greatly, so some uncertainty has likely been introduced. Still, shrinking of the contour includes at least part of the mucosal layer in the rectum, which can be assumed to give a better approximation of dose to the rectal wall than using the convex hull directly.

5.2   LKB Model

5.2.1   Interpretation and Comparison to QUANTEC Recommendations

The LKB model parameters TD50, n and m, as defined in Section 2.4.2 were fit to dose distributions to the rectum and PROMs as measured by RBS with a cut-off of 2.5. The parameter value for TD50 $= 78.9 \pm 3.9$ Gy is in good agreement with the QUANTEC recommended value of TD50 $= 76.9 \, (73.7 - 80.1)$ Gy. The TD50 dose fixes the 50 % complication rate to the horizontal axis. Assuming that clinicians accept a rate of complication of around 10-15 %, the tolerance gEUD (volume-weighted average dose) would be below 75 Gy for the RIC study population. This should in most cases be feasible for a prescribed PTV dose of 78 Gy, as it is in this study (no dose plan gEUDs exceeded TD50). It may not be possible if the rectum is small, located very close to, or is overlapping with the PTV.

The m value of 0.053 ± 0.043 is somewhat smaller than the QUANTEC estimate of $m = 0.13 \, (0.10 - 0.17)$, yielding a steeper response curve, as can be seen in the plot in Figure 4.7, when the fitted NTCP curve (orange) and the QUANTEC-recommended curve (gray, dashed line) are compared. The steepness of the response curve relates to the observed variation in response seen for a certain gEUD value. A steep response curve

arises from a narrow distribution of responses, meaning that most patients respond similarly to dose distributions yielding similar gEUDs. This could be attractive for clinicians, as they may be more confident about the response (e.g. the rate of toxicity) to a certain dose plan. However, it is important to note that different dose distributions may give similar gEUDs. This speaks for further investigation of the dose distributions by using spatial metrics and analysis to discover dose patterns that contribute to toxicity.

The n parameter was found to be $n = 0.007$. As described in Section 2.4.2, a small value of the parameter n (closer to zero) indicates that the high-dose region of the rectum is weighted more in the dose-outcome analysis. The low n value suggests a serial architecture of the rectum. From a clinical perspective, that means that even if only a small fraction of the volume is irradiated it may impair function of the organ. For instance, studies have demonstrated relationships between localized dose distributions and acute and late toxicity for the rectum [34, 55].

The fitted value for n is equal to the lower bound of the interval of investigated values ($0.007 - 1.0$). The lower limit of $n = 0.007$ was set due to software limitations; The presence of a factor $1/n$ in the exponent in the calculation of gEUD (Eq. (2.6)) yielded exponent overflow if smaller values of n were allowed. This limitation may have influenced the validity of the n value determined. The observed value for n differs from the QUANTEC value $n = 0.09$ ($0.04 - 0.14$), which may be due to characteristics of the dataset or the specific outcome measured. Some of these aspects will be discussed in the following section.

5.2.2 Influential Factors in NTCP Modelling

When modelling NTCP, there are many factors to be considered, and both the quality and quantity of the data may influence the robustness of the predictions of outcome.

In NTCP modelling of rectal side effects from prostate cancer RT, cohorts of a few hundred up to thousand patients have been used for outcome prediction [55–58]. However, consistency with the QUANTEC parameters for the LKB model have been found in studies using as little as 159 patients where adverse outcome (grade ≥ 2 late rectal bleeding) was observed in only 12 patients (7.5 %) [59]. Therefore, the size of the patient population investigated here should not be a significant limitation.

As mentioned previously, the choice of endpoint will impact the analyses. The choice of time at which the outcome is assessed is believed to influence the LKB parameters. First, late toxicity usually begins to manifest after 3-4 years [11], and may be less apparent in the 36-month follow-up data used for fitting the LKB parameters. Second, several studies report adverse side effects cumulatively over the entire follow-up period (e.g. whether severe rectal bleeding occurs at any time point within a certain period after treatment) [57, 59], whereas in the present analysis of the LKB model only data from the 36-month follow-up is considered.

Many studies investigating the LKB model report on late rectal bleeding as the sole endpoint [56, 59], or use other clinician-reported measures. The choice of endpoint has been demonstrated to produce different LKB parameters or NTCP predictions [60, 61], and these findings are supported by different dose levels corresponding to different PROMs in Figure 4.6. In the RIC study, PROMs are used to assess differences in outcome

between the two treatment arms [48], while few studies have reported LKB parameters based on PROMs. There is ongoing discussion on whether objective or subjective grading of toxicity is more appropriate. Difference between the experience of the patient and the objective scoring of symptoms by a clinician may differ. Although allowing for easier comparison between studies, objective scoring of outcomes does not necessarily give a good indication of the impact that the side effects may have on patients' QoL [2].

The choice of rectal bother score as the outcome measure in this project allows for the evaluation of a range of symptoms. However, the effect of higher toxicity for any single items in the score (while the others remain low) may not result in an observable increase in rectal bother.

Lastly, the dose to the entire rectum (including rectal contents) was calculated and used in the LKB modelling. The QUANTEC review recommends emptying of the rectum prior to planning CT and treatment delivery to avoid errors that may arise with variation in the degree of rectal filling [11]. The studies analyzed in the QUANTEC review have also analyzed different volumes, with some studies reporting dose to the rectum with contents, while other consider only the rectal wall. In the present work the LKB modelling was carried out before the rectal wall data was available, and the dose to the entire rectum was therefore used. Analyzing the rectum plus contents may have influenced the NTCP modelling since dose delivered to the contents, and not to the organ itself, are not necessarily relevant for the clinical outcome.

## 5.3   Image Registration and Voxel-Based Analysis

The pipeline for VBA that has been set up has been demonstrated. The results are given as quantitative metrics for the quality of registration (e.g. DSC and HD for OARs in the CCS), as well as examples of voxel-wise t-testing performed on a subset of the registered patients. This section will discuss aspects of the registration process and quality, as well as attempt to explain the findings presented in Section 4.4.

5.3.1   Evaluation of registration quality

The transformed CT images, dose distributions, and delineation correspond in the CCS, which was verified by visual inspection of the resulting images. The correspondence was also controlled by comparing the DVHs for a few patients before and after registration, which indicated near-perfect correspondence. The calculation of three DVH metrics before and after registration for all patients, shown in Table 4.5, show that the mean and median differences between the metrics are very small, only $0.04 - 0.06\,$Gy and $0.19 - 0.32\,$Gy, respectively. However, the largest differences range from $3.6 - 27.7\,$Gy. All these maximum differences may arise from a single corrupted DVH, caused by a faulty registration.

Notably, during the visual inspection, the patients with low DSC (< 0.1) for both the rectum, bladder and PTV usually had a very poor registration with a large shear component, resulting in the patient being warped beyond recognition. If the shear component of the affine registration is large, volumes may become much smaller after registration,
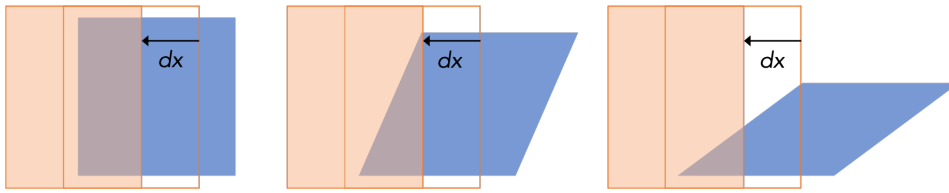
Figure 5.1: Illustration of how shear makes an area smaller by use of a cube. The parallelograms have the same side length as the cube, however the area is smaller with decreasing angle between the bottom and left edges. The orange box illustrates a small displacement dx of an isodose curve, resulting in different relative change in area covered for the different areas.

as illustrated in Figure 5.1. The three parallelograms have the same side length (analogous to the walls of an organ), however the volumes differ. In the left figure, the area of the organ (blue) covered by an imagined isodose curve (orange) is halved when the isodose is displaced by dx. However, in the middle and right figures, around 66 % and 75 % less volume is irradiated by the isodose. Therefore, as illustrated, even a small displacement in the dose distribution or a structure may have an impact on the relative overlap between the structures and dose, and therefore of the DVHs.

Figure 4.9 shows that the entire population of registered rectums cannot be used for voxel based analysis, as a large fraction of rectums extend beyond the rectum template used in the analysis. A DSC cut-off of 0.45 or 0.5 makes a clear difference, and particularly ensures that the majority (> 80 %) of rectums are aligned in the central regions of the rectum template. The figure demonstrates that the largest uncertainties in the registration occur in the superior and inferior parts of the rectum. This may be influenced by variations in the rectum delineations, and could therefore be caused by the properties of the dataset rather than by the registration process. However, inclusion of a local, non-rigid registration step in the registration scheme could improve alignment in the superior/inferior region and reduce these uncertainties.

The registration quality as measured by the DSC and HD was highly variable, ranging from total mismatch between template and registered structures (DSCs of near 0), to what might be described as a decent overlap (DSC of 0.73, 0.81, and 0.82 for the rectum, bladder and prostate, respectively). The PTV generally had the best agreement (highest DSC and lowest HD) between the registered structure and the template. The rectum had the lowest DSC and highest HD of the organs. The small size of the rectum may have contributed to the DSC being smaller than for the bladder and PTV, since the DSC is generally sensitive to differences between small structures [41, 62]. Since the bladder is highly deformable compared to the prostate (which is surrounded by the PTV), a lower DSC for the bladder than the PTV could be expected. The DSC inclusion requirement for the per-voxel statistical analysis was set as low as DSC $\geq$ 0.5, which can be deemed a fair (but not good) agreement [63].

The extent of the rectum in the cranio-caudal direction varied greatly. The extent of the rectum contour was specified in the protocol as the "extension of the segmentation is to the recto-sigmoid junction and in the caudal direction down to and including the anus". Still, the visual inspection revealed that the extent of the contour in the caudal

direction included the anal canal to varying degree. The recto-sigmoid junction can be difficult to determine [11], so variation in the cranial direction was also found. If the delineation procedure has provided varying anatomical regions to be mapped to the template, the similarity might be measured between non-corresponding regions, and will not be accurate even though the registration process works well.

5.3.2   Similarity Metrics Used for Evaluation

It should be noted that the DSC and HD similarity metrics only measure the overall geometric correspondence between figures, and do not measure morphological (shape) similarity or local mapping errors within the structures [41, 62]. When comparing dose distributions the local mapping errors should be minimized in order to ensure an accurate description of spatial dose-response relationships.

Palma et al. [41] state that the DSC and HD are suitable for comparing organ overlap before and after an elastic image registration, where they will measure the improvement of the anatomical overlap. They are, however, not necessarily appropriate for comparing elastic image registration methods when applied on different anatomical regions. The dose-organ overlap measure introduced by Dréan et al. [64] measures the coincidence of both the organ delineations and the dose distributions in the CCS, and may be a helpful tool when assessing the registration quality and comparing registration schemes. In this work, the DSC and HD similarity metrics were used as objective measures to compare both rigid and non-rigid registration schemes for the pelvic region, which was deemed to be a superior method to visual inspection only.

5.3.3   Indications of Radiosensitive Subregions

When performing VBA using PROMs at 5, 12 and 18 months, a region where patients with complications receive a significantly higher dose can be identified in the inferior, anterior part of the rectum. This region is in the proximity of the prostate and subsequently in the high-dose region, a similar region to that which has been found using both dose-surface maps to investigate GI toxicity rated by the RTOG criteria [55]; registration and VBA when considering patients with or without rectal bleeding [36]; and by the method of Dréan et al. [3] for rectal bleeding. All these studies use clinician-evaluated toxicity criteria. A correspondence with results found by using PROMs is interesting from a clinical perspective. Careful consideration of each of the PROMs included in the RBS could further reveal which subregions, if any, each outcome is related to.

The prostate and high-dose are located to the left of the inferior rectum region in the sagittal images. For many of the patients, the registration resulted in the body (and subsequently the rectum) being tilted forward (i.e. counter-clockwise). This may shift the rectum further from the prostate in the inferior region. Since the dose follows the same transform, the high-dose region to the left of the rectum may be shifted closer to the template rectum (which is stationary in the CCS). This may result in the apparent dose to the template being higher than expected. The shifting of dose during registration may occur both for patients with and without complications, and the dose differences are likely not influenced very much if the sample size is sufficiently large. However, as mentioned previously, the sample analyzed here is small, and it cannot be ruled out that

this forward rotation may have contributed to the differences seen in mean dose. Results should be treated with caution and are not adequate for drawing firm conclusions on dose-response relationships.

Calculation of the mean dose to the regions with significance $p < 0.01$ and $p < 0.05$ reveals that the patients with RBS ≥ 2.0 received doses of $72.9 - 75.4$ Gy at these three time points, on average. On the other hand, patients without complications receive on average $14.2 - 17.2$ Gy, $6.8 - 9.6$ Gy and $5.8 - 5.9$ Gy less to the subregions at 5, 12 and 18 months, respectively.

At 24 months, the same significant subregions appears, but is accompanied by a significant dose difference in the superior rectum region. The mean dose of approximately 52 Gy in the significant subregions is slightly lower than for earlier follow-up time points, and a dose excess of $6.3 - 7.9$ Gy is found for patients with RBS ≥ 2.0. This dose difference may be related to artefacts along the beam edges, as discussed above, since only two patients were categorized as having complications.

At 36 months the significant subregions include areas in both the superior and inferior part of the rectum. In the superior region, patients with complications have a dose excess compared to those without complications, whereas the opposite is true for the inferior region. Calculating the dose difference to these regions therefore may be less informative, as the differences cancel out. However, it is worth mentioning that the mean dose to the regions at risk is $29.1 - 32.8$ Gy, lower than for the other follow-up times. A study by Buettner et al. [33], found that low doses between $21 - 27$ Gy were highly correlated with occurrence of loose stools.

5.3.4 Clinical Relevance and Main Issues in VBA

The voxel-based analysis that has been presented here clearly illustrates the possibility of identifying subregions in an organ that are highly correlated with adverse outcome. Despite the uncertainties related to data quality, image registration and statistical analysis, the analysis indicated that elevated dose to the inferior, anterior rectum may be related to worse patient-reported outcome up to 24 months after the RT. A similar region has been identified in studies using both dose-surface maps and VBA to analyze dose differences based on clinician-reported outcome [3, 36, 55].

Lafond et al. [65] have used the subregion correlated with rectal bleeding identified in [3] to perform replanning of the treatment. The planning took into consideration a dose restriction to the subregion, which yielded plans with a significant dose decrease to the rectum and rectal wall, lower V50Gy for the bladder, and maintaining the dose coverage of the PTV. Subsequent NTCP modelling using the LKB model ($n = 0.21$, $m = 0.28$, and TD50 $= 72$ Gy with rectal bleeding as the endpoint [3]) showed a 5 percentage point reduction in NTCP for the subregion, from 22.8 % to 17.6 %.

Although the methodology used by Lafond et al. differs slightly from the VBA method presented here, the possible applications are similar. Any subregion at risk identified in a population of patients may be applied to each patient by registering the patient's CT image to the template, and using the inverse transformation to propagate the contour back to the native CS. Considering a subregion at risk during the treatment planning step has the potential to contribute to lower rate of complications.

Relating to the current project, a major limitation is the image registration step. A large fraction of patients were excluded due to poor registration results. Guiding the process with distance maps can (as discussed above) improve registration quality beyond simple intensity-based registration, which is limited by the poor soft tissue contrast of CT images [36].

5.3.5 Important Choices in the Registration Process

Template selection

The template was selected manually by visual inspection of a subset of the patients ($N = 15$). Although an attempt was made to note the shape, size and rectal and bladder filling of the patients and organs in an objective fashion, several factors such as the axial FOV, the location of bony structures, or the location of the rectum relative to the prostate may have influenced the decision. In particular, the relative location of the rectum and prostate may be important in the analyses following registration, as the high-dose region is centered on the prostate.

Several methods have been employed to attempt to lessen the influence of template selection bias. In elastix (and SimpleElastix), the groupwise registration procedures co-register all patients to an implicit average anatomy by applying the constraint that the mean deformation across the population is zero.

Another method consists of randomly selecting a subset of N patients and selecting the most average of these patients as the template. For each patient as the template, the other $N-1$ patients is registered to the template, and a similarity metric over the registrations is calculated. The process is repeated for all patients (resulting in $N \times (N-1)$ registrations), and the patient showing the greatest similarity to the other patients is selected as the template. This was done by Acosta et al. [36] using affinity propagation clustering [66] and considering the SSD (as defined in Equation (2.7)) after rigid registration.

Registration scheme and transform selection

An issue that was mentioned in Section 2.5.2 is that the transform should be chosen according to each registration problem. For the current work, the region of interest contained two highly deformable organs; the rectum and bladder. A non-rigid transformation would therefore have been suitable to be able to align local variations in the rectum and bladder anatomy. However, the testing and comparison of the four different registration schemes yielded similar rectum DSC for the affine-only and affine and B-spline registrations. Since the DSC for the bladder and PTV were higher for the affine-only scheme, the affine transform was chosen.

During visual inspection it was noted that the apparent alignment did not always reflect in the DSC. Some patients had a good alignment of the centres of mass, but the shape of the registered rectum and the template were so different that the DSC fell below the cutoff of 0.5. Other patients had very good agreement with the template near the PTV, but the shape, size or location of the rest of the rectum was poor. Visual inspection proved to be a necessary procedure to ensure only individuals with adequate

registration were used in the statistical analysis. The cut-off was set to DSC ≥ 0.5 after visual inspection. This resulted in more than two thirds of patients being excluded from the subsequent statistical analysis.

To improve the registration quality, it is necessary to include a non-rigid registration component, in particular in the regions surrounding the rectum and bladder. The default SimpleElastix registration scheme includes both linear (rigid and affine) and elastic (non-rigid) transformations in the registration. Although the registration accuracy generally increases with an increasing number of iterations, this was not the case when comparing the default and adapted default registration schemes (default scheme with an maximum number of iterations set to 512 rather than 256). This indicates that a more guided registration process is needed in order to achieve better anatomical agreement than that when using intensity-based CT-to-CT registration.

Dréan et al. [3, 64] used intensity-based registration together with an anatomical model of the rectum to improve registration accuracy. The model treats the rectum as a hollow tube and consists of rectum centerlines, and Euclidean distance maps and Laplacian scalar fields as structural descriptors. In a cohort of 118 patients, where each of the patients was chosen as the template and the other 117 patients was registered to the patient, a DSC of 0.91 ± 0.03 was achieved for the 118×117 registrations.

Since it is of interest to investigate not just the anatomy of the rectum, but potentially of other organs at risk, the method of Acosta et al. [36] could provide a good starting point. The group developed a framework that uses normalized distance maps for the rectum, bladder and prostate, superimposed on the CT images, to guide the registration process. They report achieving an average DSC for the rectum of 0.75 ± 0.12 (standard deviation), and a HD of $28.17 \, \text{mm} \pm 16$. Comparing these values to the DSC $= 0.39 \pm 0.16$ and HD $= 37.4 \, \text{mm} \pm 22.8$ achieved here, including distance maps for these organs in the registration would likely improve the resulting anatomical match.

Registration mask

Since rectal complications were considered in the current work, a mask focusing the registration on the rectum might have been sufficient with a suitable registration scheme. However, the different mask regions were tested only for a few patients and yielded the best results for the rectum, bladder and PTV mask with an intermediate dilation radius. During testing it was found that including some bone structures in the mask helped guide the global alignment of the patients, while including too many could potentially worsen the alignment (in case of the initial similarity being poor). Therefore, including some bone structures in at least the initial rigid and/or affine transform steps will likely improve the registration performance.

5.3.6 Statistical analysis

The statistical analysis was found to be limited by the small sample size. Poor registration results in particular lead to a large decrease in the number of patients available for analysis, from $N = 115$ to only 32 patients. The number was reduced further by some patients not having all the necessary PROMs available for analysis, and in some of the per-voxel comparisons only 2 patients were in the group with complications. The Python

package that was used for statistical analysis (`scipy.stats`) calculates the t-test statistic from the mean and variance of the two samples. Therefore, at least two patients had to be included in each sample to be able to calculate the sample variance. The cut-off value for rectal bother score had to be set to RBS $\geq$ 2.0 (lower than in the DVH comparison and LKB modelling, where it was RBS $\geq$ 2.5) in order to have two patients in the group. This limitation of the statistical test should be overcome by including a larger population of patients. Then the voxel-based analysis may also be used to investigate dose differences for higher cut-off values (i.e. more severe outcome).

The Student's t-test used for comparison requires that the data is normally distributed, a doubtful assumption when the sample size is as small as it is here. However, a statistical test could not be expected to provide reliable results based on such a small sample. It is therefore necessary to further improve the registration to be able to analyze dose and outcome data in a larger population of patients. Although differences in the organ delineations influence registration, tuning the transform parameters would be the most efficient way to increase the reliability of the statistical analysis, while still using the current dataset.

5.4   Notes on Data Quality

Throughout all dosimetric analyses, it was assumed that the planned dose distribution accurately represents the delivered dose to the patient. However, even with control of positioning using imaging and skin markers, internal anatomical changes such as variations in organ filling will influence the dose delivery. Of greatest importance to the analyses presented here, is the position, shape and size of the rectum, which varies due to inter- or intra-fraction variations in intestinal and rectal filling [11]. In the treatment protocol, the dose constraint for the rectum is that no more than half the circumference in each CT slice should receive a dose above 60 Gy. However, with varying rectal filling the rectal wall might move in and out of the radiation field. No rectal emptying is specified in the treatment protocol, a measure which is often used to attempt to control the degree of rectal filling during each treatment session. This should be specified in a treatment protocol for further studies using image registration.

5.5   Suggestions for Further Work

The RIC study was not designed to consider the types of applications considered in this work. The discussion above leads to some natural points to address in the planning of further studies.

The treatment protocol should attempt to minimize variations between patients. Standardization and adherence to the delineation procedures is of great importance to ensure that corresponding anatomical regions are used for guiding the registration procedure. Furthermore, the protocol could specify rectal and bladder emptying procedures to eliminate some inter-patient and inter-fraction variation in rectal shape and size.

It is also assumed here that the planned dose accurately represents the delivered dose, which is not necessarily the case. The registration framework can potentially be

used to register daily CBCT images to the planning CT in order to track the accumulated dose, which has been shown to provide a better estimate of rectal toxicity than planned dose [67]. However, the registration quality would need to significantly improve in order for this to be feasible.

Several suggestions for improving the registration process have been discussed. The chosen affine transform is likely not sufficiently flexible to align local variations between patients, particularly for the rectum and bladder where the degree of filling may vary. A non-rigid transformation should be implemented, but should be accompanied by the inclusion of a structural descriptor (e.g. distance maps for the rectum, bladder and prostate) to ensure accuracy in the non-rigid registration process. If the general alignment is quite good after only a rigid or affine transformation step, focusing the non-rigid registration on the rectum only could be a good alternative if only rectal complications are of interest.

Although radiosensitive subregions could be identified using the framework in its current state, an NTCP value is not estimated. One motivation for estimating the NTCP values is to use it as a tool in comparison of treatment plans. Implementation of a method for calculating NTCP from a per-voxel assessment would therefore be of relevance from both a clinical and research perspective.

6 Conclusion

In this project a Python programming framework to analyze dose distributions from radiotherapy of prostate cancer was developed and implemented. The framework was used to perform dose-response analyses of rectal side effects based on planned dose distributions and PROMs from the RIC study.

It was found that patients in arm A received higher doses than the patients in arm B to all the investigated OARs, as was expected with larger treatment margins. A correlation between larger irradiated relative volume and rectal toxicity was found for doses up to 30 Gy when all patients were analyzed together, and up to 65 Gy when only patients in arm B were considered. The investigation of various PROMs revealed that different dose-response relationships are found for different outcomes.

The most widely used NTCP model, the LKB model, was implemented to correlate dose distributions with PROMs data. The LKB model parameters were fitted to the RIC study data and showed fair agreement with literature values used clinically. A lower value for the volume-dependence parameter n was found, indicating a serial architecture of the rectum.

CT images for patients in the study arm receiving daily position verification were registered to an anatomical template. The mean Dice coefficient for the rectum was DSC = 0.39. For patients with DSC $\geq$ 0.5 for the rectum, mean dose distributions to patients with and without complications were compared. A subregion in the inferior, anterior rectum received significantly higher dose for patients with complications at 5, 12 and 18 months after the start of RT. Comparisons of toxicity at 24 and 36 months were inconclusive, and more patients need to be analyzed to be able to draw any conclusions.

The dose-response analyses have demonstrated the functionality of the framework that was developed. Analyses based on PROMs provided fair agreement with those based on clinician-reported outcome for most applications. The image registration algorithm should be further improved, as this may lead to more reliable results from the voxel-based analysis. Outcome data from the RIC study at 5 and 10 years after treatment should be analyzed since late toxicity usually manifests within 3-4 years after treatment and these effects may not have been captured in the follow-up data up to 36 months.

With the further improvements suggested here, the voxel-based analysis tools have the potential to identify radiosensitive regions in tissue and contribute to personalized RT for both prostate and other cancer types.

Bibliography

[1] Cancer Registry of Norway, 'Cancer in Norway 2019 - Cancer incidence, mortality, survival and prevalence in Norway,' Cancer Registry of Norway, Oslo, Tech. Rep., 2020.

[2] P. Holch, A. M. Henry, S. Davidson *et al.*, 'Acute and Late Adverse Events Associated With Radical Radiation Therapy Prostate Cancer Treatment: A Systematic Review of Clinician and Patient Toxicity Reporting in Randomized Controlled Trials,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 97, no. 3, pp. 495–510, Mar. 2017.

[3] G. Dréan, O. Acosta, J. D. Ospina *et al.*, 'Identification of a rectal subregion highly predictive of rectal bleeding in prostate cancer IMRT,' *Radiotherapy and Oncology*, vol. 119, no. 3, pp. 388–397, Jun. 2016.

[4] E. Mylona, O. Acosta, T. Lizee *et al.*, 'Voxel-Based Analysis for Identification of Urethrovesical Subregions Predicting Urinary Toxicity After Prostate Cancer Radiation Therapy,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 104, no. 2, pp. 343–354, Jun. 2019.

[5] O. Casares-Magaz, L. P. Muren, V. Moiseenko *et al.*, 'Spatial rectal dose/volume metrics predict patient-reported gastro-intestinal symptoms after radiotherapy for prostate cancer,' *Acta Oncologica*, vol. 56, no. 11, pp. 1507–1513, 2017.

[6] F. M. Khan, J. P. Gibbons and P. W. Sperduto, *Khan's Treatment Planning in Radiation Oncology*, 4th, F. M. Khan, J. P. Gibbons and P. W. Sperduto, Eds. Philadelphia: Wolters Kluwer, 2016.

[7] P. Mayles, A. Nahum and J. C. Rosenwald, *Handbook of Radiotherapy Physics: Theory and Practice*, 1st ed., P. Mayles, A. Nahum and J. C. Rosenwald, Eds. Taylor & Francis Group, 2007.

[8] International Commission on Radiation Units and Measurements, 'The International Commission on Radiation Units and Measurements,' Geneva, Tech. Rep. 1, Apr. 2010.

[9] P. Andreo, M. Evans, J. Hendry *et al.*, *Radiation Oncology Physics: A Handbook for Teachers and Students*, E. Podgorsak, Ed. Vienna: International Atomic Energy Agency, 2005.

[10] B. Emami, J. Lyman, A. Brown *et al.*, 'Tolerance of normal tissue to therapeutic irradiation,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 21, no. 1, pp. 109–122, May 1991.

[11]   J. M. Michalski, H. Gay, A. Jackson, S. L. Tucker and J. O. Deasy, 'Radiation Dose-Volume Effects in Radiation-Induced Rectal Injury,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, no. 3 SUPPL. Pp. 123–129, 2010.

[12]   A. N. Viswanathan, E. D. Yorke, L. B. Marks, P. J. Eifel and W. U. Shipley, 'Radiation Dose-Volume Effects of the Urinary Bladder,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, no. 3 SUPPL. Pp. 116–122, 2010.

[13]   M. Roach, J. Nam, G. Gagliardi, I. El Naqa, J. O. Deasy and L. B. Marks, 'Radiation Dose-Volume Effects and the Penile Bulb,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, no. 3 SUPPL. Pp. 130–134, 2010.

[14]   D. Müssig, 'Re-scanning in scanned ion beam therapy in the presence of organ motion,' Ph.D. dissertation, TU Darmstadt, 2014.

[15]   M. Joiner and A. van der Kogel, *Basic Clinical Radiobiology*, 4th ed., M. Joiner and A. van der Kogel, Eds. London: Hodder Arnold, 2009.

[16]   H. R. Withers, 'The Four R's of Radiotherapy,' in *Advances in Radiation Biology*, vol. 5, 1975, pp. 241–271.

[17]   G. G. Steel, T. McMillan and J. Peacock, 'The 5Rs of Radiobiology,' *International Journal of Radiation Biology*, vol. 56, no. 6, pp. 1045–1048, Jan. 1989.

[18]   H. R. Withers, J. M. Taylor and B. Maciejewski, 'Treatment volume and tissue tolerance,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 14, no. 4, pp. 751–759, Apr. 1988.

[19]   J. Mahmood, A. A. Shamah, T. M. Creed *et al.*, 'Radiation-induced erectile dysfunction: Recent advances and future directions,' *Advances in Radiation Oncology*, vol. 1, no. 3, pp. 161–169, Jul. 2016.

[20]   H. Tøndel, 'Image-Guided Radiotherapy for Prostate Cancer and Side Effects,' Ph.D. dissertation, Norwegian University of Science and Technology, 2019.

[21]   M. Calvert, J. Blazeby, D. G. Altman *et al.*, 'Reporting of Patient-Reported Outcomes in Randomized Trials,' *JAMA*, vol. 309, no. 8, p. 814, Feb. 2013.

[22]   EORTC Quality of Life Group, *EORTC QLQ-C30 (version 3)*, 1995.

[23]   P. Fransson, J.-Å. Lund, J.-E. Damber *et al.*, 'Quality of life in patients with locally advanced prostate cancer given endocrine treatment with or without radiotherapy: 4-year follow-up of SPCG-7/SFUO-3, an open-label, randomised, phase III trial,' *The Lancet Oncology*, vol. 10, no. 4, pp. 370–380, Apr. 2009.

[24]   J. T. Lyman, 'Complication Probability as Assessed from Dose-Volume Histograms,' *Radiation Research*, vol. 104, no. 2, S13–S19, 1985.

[25]   G. J. Kutcher and C. Burman, 'Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 16, no. 6, pp. 1623–1630, Jun. 1989.

[26]   F. Buettner, S. L. Gulliford, S. Webb and M. Partridge, 'Using dose-surface maps to predict radiation-induced rectal bleeding: A neural network approach,' *Physics in Medicine and Biology*, vol. 54, no. 17, pp. 5139–5153, 2009.

[27] N. Yahya, M. A. Ebert, M. J. House *et al.*, 'Modeling Urinary Dysfunction After External Beam Radiation Therapy of the Prostate Using Bladder Dose-Surface Maps: Evidence of Spatially Variable Response of the Bladder Surface,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 97, no. 2, pp. 420–426, Feb. 2017.

[28] S. Monti, G. Palma, V. D'Avino *et al.*, 'Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients,' *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

[29] W. Beasley, M. Thor, A. McWilliam *et al.*, 'Image-based Data Mining to Probe Dosimetric Correlates of Radiation-induced Trismus,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 102, no. 4, pp. 1330–1338, 2018.

[30] M. van Herk, M. Witte, W. Heemsbergen *et al.*, 'Relation between Dose Outside the Prostate and Failure Free Survival in the Dutch Prostate Cancer Trial,' *International Journal of Radiation Oncology*Biology*Physics*, 2008.

[31] M. G. Witte, W. D. Heemsbergen, R. Bohoslavsky *et al.*, 'Relating Dose Outside the Prostate With Freedom From Failure in the Dutch Trial 68 Gy vs. 78 Gy,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 77, no. 1, pp. 131–138, May 2010.

[32] W. D. Heemsbergen, M. S. Hoogeman, G. A. Hart, J. V. Lebesque and P. C. Koper, 'Gastrointestinal toxicity and its relation to dose distributions in the anorectal region of prostate cancer patients treated with radiotherapy,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 61, no. 4, pp. 1011–1018, 2005.

[33] F. Buettner, S. L. Gulliford, S. Webb, M. R. Sydes, D. P. Dearnaley and M. Partridge, '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*, vol. 54, no. 21, pp. 6535–6548, 2009.

[34] R. C. Wortel, M. G. Witte, U. A. van der Heide *et al.*, 'Dose–surface maps identifying local dose–effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer,' *Radiotherapy and Oncology*, vol. 117, no. 3, pp. 515–520, 2015.

[35] C. Chen, M. Witte, W. Heemsbergen and M. v. Herk, 'Multiple comparisons permutation test for image based data mining in radiotherapy,' *Radiation Oncology*, vol. 8, no. 1, p. 293, Dec. 2013.

[36] O. Acosta, G. Dréan, J. D. Ospina *et al.*, 'Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy,' *Physics in Medicine and Biology*, vol. 58, no. 8, pp. 2581–2595, 2013.

[37] N. Dey, W. B. A. Karâa, S. Chakraborty, S. Banerjee, M. A. Salem and A. T. Azar, 'Image mining framework and techniques: a review,' *International Journal of Image Mining*, 2015.

[38] L. J. Isaksson, M. Pepa, M. Zaffaroni *et al.*, 'Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy,' *Frontiers in Oncology*, vol. 10, no. June, 2020.

[39] J. H. Oh, S. Kerns, H. Ostrer, S. N. Powell, B. Rosenstein and J. O. Deasy, 'Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes,' *Scientific Reports*, vol. 7, no. 1, p. 43 381, Mar. 2017.

[40] M. Carrara, E. Massari, A. Cicchetti *et al.*, 'Development of a Ready-to-Use Graphical Tool Based on Artificial Neural Network Classification: Application for the Prediction of Late Fecal Incontinence After Prostate Cancer Radiation Therapy,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 102, no. 5, pp. 1533–1542, Dec. 2018.

[41] G. Palma, S. Monti and L. Cella, 'Voxel-based analysis in radiation oncology: A methodological cookbook,' *Physica Medica*, vol. 69, no. July 2019, pp. 192–204, 2020.

[42] D. Rueckert and J. A. Schnabel, 'Biomedical Image Processing,' in, ser. Biological and Medical Physics, Biomedical Engineering, T. M. Deserno, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. 5. Medical.

[43] A. A. Goshtasby, *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Application*. Hoboken, N.J.: Wiley-Interscience, 2005.

[44] H. F. G. García, A. G. Vega, A. H. Aguirre, J. L. M. Zaleta and C. C. Coello, 'Robust Multiscale Affine 2D-Image Registration through Evolutionary Strategies,' in, 2002, pp. 740–748.

[45] T. Ryen, *Multimodal Volume to Volume Registration between Ultrasound and MRI*, 2006.

[46] S. Klein, J. P. W. Pluim, M. Staring and M. A. Viergever, 'Adaptive Stochastic Gradient Descent Optimisation for Image Registration,' *International Journal of Computer Vision*, vol. 81, no. 3, pp. 227–239, Mar. 2009.

[47] H. J. Johnson, M. M. McCormick and L. Ibáñez, *The ITK Software Guide: Design and Functionality*, 2021.

[48] H. Tøndel, J. Å. Lund, S. Lydersen *et al.*, 'Radiotherapy for prostate cancer – Does daily image guidance with tighter margins improve patient reported outcomes compared to weekly orthogonal verified irradiation? Results from a randomized controlled trial,' *Radiotherapy and Oncology*, vol. 126, no. 2, pp. 229–235, 2018.

[49] C. H. Huh, M. S. Bhutani, E. B. Farfán and W. E. Bolch, 'Individual variations in mucosa and total wall thickness in the stomach and rectum assessed via endoscopic ultrasound,' *Physiological Measurement*, vol. 24, no. 4, 2003.

[50] K. Marstal, F. Berendsen, M. Staring and S. Klein, 'SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration,' in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Jun. 2016, pp. 574–582.

[51] B. C. Lowekamp, D. T. Chen, L. Ibáñez and D. Blezek, 'The Design of SimpleITK,' *Frontiers in Neuroinformatics*, vol. 7, 2013.

[52] S. Klein, M. Staring, K. Murphy, M. Viergever and J. Pluim, 'elastix: A Toolbox for Intensity-Based Medical Image Registration,' *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[53] D. Palma, E. Vollans, K. James *et al.*, 'Volumetric Modulated Arc Therapy for Delivery of Prostate Radiotherapy: Comparison With Intensity-Modulated Radiotherapy and Three-Dimensional Conformal Radiotherapy,' *International Journal of Radiation Oncology*Biology*Physics*, vol. 72, no. 4, pp. 996–1001, Nov. 2008.

[54] M. A. Ebert, S. Gulliford, O. Acosta *et al.*, 'Spatial descriptions of radiotherapy dose: normal tissue complication models and statistical associations,' *Physics in Medicine & Biology*, May 2021.

[55] O. Casares-Magaz, S. Bülow, N. J. Pettersson *et al.*, 'High accumulated doses to the inferior rectum are associated with late gastro-intestinal toxicity in a case-control study of prostate cancer patients treated with radiotherapy,' *Acta Oncologica*, vol. 58, no. 10, pp. 1543–1546, Oct. 2019.

[56] M. Söhn, D. Yan, J. Liang, E. Meldolesi, C. Vargas and M. Alber, '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*, vol. 67, no. 4, pp. 1066–1073, Mar. 2007.

[57] S. L. Gulliford, M. Partridge, M. R. Sydes, S. Webb, P. M. Evans and D. P. Dearnaley, '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*, vol. 102, no. 3, pp. 347–351, 2012.

[58] J. Pedersen, X. Liang, O. Casares-Magaz *et al.*, 'Multivariate normal tissue complication probability models for rectal and bladder morbidity in prostate cancer patients treated with proton therapy,' *Radiotherapy and Oncology*, 2020.

[59] M. Liu, V. Moiseenko, A. Agranovich *et al.*, 'Normal Tissue Complication Probability (NTCP) modeling of late rectal bleeding following external beam radiotherapy for prostate cancer: A Test of the QUANTEC-recommended NTCP model,' *Acta Oncologica*, vol. 49, no. 7, pp. 1040–1044, Oct. 2010.

[60] S. T. Peeters, M. S. Hoogeman, W. D. Heemsbergen, A. A. Hart, P. C. Koper and J. V. Lebesque, '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*, vol. 66, no. 1, pp. 11–19, Sep. 2006.

[61] W. Schaake, A. van der Schaaf, L. V. van Dijk, A. H. Bongaerts, A. C. van den Bergh and J. A. Langendijk, 'Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer patients,' *Radiotherapy and Oncology*, vol. 119, no. 3, pp. 381–387, Jun. 2016.

[62] A. Reinke, M. Eisenmann, M. D. Tizabi *et al.*, 'Common Limitations of Image Processing Metrics: A Picture Story,' Apr. 2021.

[63] A. Akbarzadeh, D. Gutierrez, A. Baskin *et al.*, 'Evaluation of whole-body MR to CT deformable image registration,' *Journal of Applied Clinical Medical Physics*, vol. 14, no. 4, pp. 238–253, Jul. 2013.

[64] G. Dréan, O. Acosta, C. Lafond, A. Simon, R. de Crevoisier and P. Haigron, 'Interindividual registration and dose mapping for voxelwise population analysis of rectal toxicity in prostate cancer radiotherapy,' *Medical Physics*, vol. 43, no. 6, pp. 2721–2730, May 2016.

[65] C. Lafond, A. Barateau, J. N'Guessan *et al.*, 'Planning With Patient-Specific Rectal Sub-Region Constraints Decreases Probability of Toxicity in Prostate Cancer Radiotherapy,' *Frontiers in Oncology*, vol. 10, no. September, 2020.

[66] B. J. Frey and D. Dueck, 'Clustering by Passing Messages Between Data Points,' *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[67] L. E. A. Shelley, J. E. Scaife, M. Romanchikova *et al.*, 'Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy,' *Radiotherapy and Oncology*, vol. 123, no. 3, pp. 466–471, 2017.

A   Questionnaires for Outcome Assessment

The questionnaires QLQ-C30 and QUFW94 that were used for patient-reported outcome are presented here.

A.1   QLQ-C30

■   **EORTC QLQ-C30**   Mnd ☐ ☐ ⓪ ⓵ ⓪⓪⑤ ■

          (Versjon 3.0)

Dato: ☐ ☐ . ☐ ☐ . ☐ ☐ ☐ ☐   Pasientens initialer: ☐ ☐ ☐

Vi er interessert i forhold vedrørende deg og din helse. Vær så vennlig å besvare hvert spørsmål ved å
sette et kryss x i den boksen som best beskriver din tilstand. Det er ingen «riktige» eller «gale» svar.
Alle opplysningene vil bli behandlet konfidensielt.

	Ikke i det hele tatt	Litt	En del	Svært mye
1. Har du vanskeligheter med å utføre anstrengende aktiviteter, slik som å bære en tung handlekurv eller en koffert?	☐	☐	☐	☐
2. Har du vanskeligheter med å gå en lang tur?	☐	☐	☐	☐
3. Har du vanskeligheter med å gå en kort tur utendørs?	☐	☐	☐	☐
4. Er du nødt til å ligge til sengs eller sitte i en stol i løpet av dagen?	☐	☐	☐	☐
5. Trenger du hjelp til å spise, kle på deg, vaske deg eller gå på toalettet?	☐	☐	☐	☐

I løpet av den siste uka:	Ikke i det hele tatt	Litt	En del	Svært mye
6. Har du hatt redusert evne til å arbeide eller utføre andre daglige aktiviteter?	☐	☐	☐	☐
7. Har du hatt redusert evne til å utføre dine hobbyer eller andre fritidsaktiviteter?	☐	☐	☐	☐
8. Har du vært tung i pusten?	☐	☐	☐	☐
9. Har du hatt smerter?	☐	☐	☐	☐
10. Har du hatt behov for å hvile?	☐	☐	☐	☐
11. Har du hatt søvnproblemer?	☐	☐	☐	☐
12. Har du følt deg slapp?	☐	☐	☐	☐
13. Har du hatt dårlig matlyst?	☐	☐	☐	☐
14. Har du vært kvalm?	☐	☐	☐	☐

Bla om til neste side

921

■

Mnd | | 0 | | 1 | 0 | 0 | 5 |

I løpet av den siste uka:

	Ikke i det hele tatt	Litt	En del	Svært mye
15. Har du kastet opp?	☐	☐	☐	☐
16. Har du hatt treg mage?	☐	☐	☐	☐
17. Har du hatt løs mage?	☐	☐	☐	☐
18. Har du følt deg trett?	☐	☐	☐	☐
19. Har smerter påvirket dine daglige aktiviteter?	☐	☐	☐	☐
20. Har du hatt problemer med å konsentrere deg, f.eks. med å lese en avis eller se på TV?	☐	☐	☐	☐
21. Har du følt deg anspent?	☐	☐	☐	☐
22. Har du vært engstelig?	☐	☐	☐	☐
23. Har du følt deg irritabel?	☐	☐	☐	☐
24. Har du følt deg deprimert?	☐	☐	☐	☐
25. Har du hatt problemer med å huske ting?	☐	☐	☐	☐
26. Har din fysiske tilstand eller medisinske behandling påvirket ditt familieliv?	☐	☐	☐	☐
27. Har din fysiske tilstand eller medisinske behandling påvirket dine sosiale aktiviteter?	☐	☐	☐	☐
28. Har din fysiske tilstand eller medisinske behandling gitt deg økonomiske problemer?	☐	☐	☐	☐

Som svar på de neste spørsmålene, sett et kryss i den boksen fra 1 til 7 som best beskriver din tilstand.

29. Hvordan har din helse vært i løpet av den siste uka?

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5   ☐ 6   ☐ 7

Svært dårlig   Helt utmerket

30. Hvordan har livskvaliteten din vært i løpet av den siste uka?

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5   ☐ 6   ☐ 7

Svært dårlig   Helt utmerket

921

A.2 QUFW94

Dato for utfylling ☐☐.☐☐.2 0 ☐☐ Mnd ☐☐ *O* ☐1☐ *O* *O5*

Besvares ved å markere med "X" i den boksen som best beskriver din situasjon DEN SISTE UKEN

31. Begrenser din prostatakreft deg i dine daglige gjøremål?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

32. Har du andre sykdommer som begrenser din daglige aktivitet?

☐ Ja   ☐ Nei

Hvis ja, hvilke sykdommer ....

33. Har du brukt medisiner for å regulere avføringen?

☐ Ja   ☐ Nei

Hvis ja, hvilke medisiner, styrke og hvor mange pr dag ....

34. Hvordan har din evne til å ta initiativ vært?

Meget bra	0	1	2	3	4	5	6	7	8	9	10	Meget dårlig
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

35. Har du hatt plager med urinveiene?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

36. Hvor mange ganger har du på det meste måttet late vannet pr DAG? ☐☐

37. Hvor mange ganger må du opp om natten for å late vannet? ☐☐

38. Har du hatt svie eller smerte når du måtte late vannet?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

39. Måtte du vente lenge før vannlatingen kom i gang?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

12864

Mnd | | 0 | | 1 | 0 | 0 | 5 |

40. Har du hatt urinlekkasje (inkontinens)?

Ikke noe   0   1   2   3   4   5   6   7   8   9   10   Svært mye

□ □ □ □ □ □ □ □ □ □ □

41. Har du brukt bleier (pga urinlekkasje)?

□ Ja   □ Nei   Antall pr dag: | | |

42. Opplevde du mer eller mindre kontinuerlig vannlatingstrang?

Ikke noe   0   1   2   3   4   5   6   7   8   9   10   Svært mye

□ □ □ □ □ □ □ □ □ □ □

43. Har du hatt blod i urinen?

Ikke noe   0   1   2   3   4   5   6   7   8   9   10   Svært mye

□ □ □ □ □ □ □ □ □ □ □

44. Har du brukt kateter?

□ Ja   □ Nei

45. Hvor mye har dine eventuelle urinveisproblemer påvirket din daglige aktivitet?

Ikke noe   0   1   2   3   4   5   6   7   8   9   10   Svært mye

□ □ □ □ □ □ □ □ □ □ □

46. Dersom du har hatt plager med urinveiene, hva har plaget deg mest?

....

47. Har du hatt problemer med avføringen?

Ikke noe   0   1   2   3   4   5   6   7   8   9   10   Svært mye

□ □ □ □ □ □ □ □ □ □ □

48. Hvor mange ganger måtte du på det meste gå på do (avføring) pr døgn? | | |

49. Hvordan var konsistensen på avføringen?

Meget løs   0   1   2   3   4   5   6   7   8   9   10   Meget hard

□ □ □ □ □ □ □ □ □ □ □

12864

Mnd ☐ ☐ ☐ | O | ☐ | 1 | 0 | 0 | 5

50. Hadde du avføringslekkasje?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

51. Medførte ditt avføringsproblem at du måtte planlegge dine toalettbesøk?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

52. Har du hatt problemer med mye luft i magen?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

53. Har du brukt bleie (pga avføringslekkasje)?

☐ Ja   ☐ Nei   Antall/dag ☐ ☐

54. Har du hatt smerter ved avføring?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

55. Har du hatt slim i avføringen?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

56. Har du hatt blod i avføringen?

Ikke | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Svært
noe | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | mye

57. Har du spist spesiell mat for å regulere avføringen?

☐ Ja   ☐ Nei   Hvis ja: ☐ Fettfattig
                                ☐ Fiberrik kost
                                ☐ Mat uten melk
                                ☐ Annet

12864

Mnd [ ] [0]   [1] [0 0 5]

58. Hvor mye har ditt avføringsproblem påvirket din daglige aktivitet?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

59. Dersom du har vært plaget med avføringen, hva har plaget deg mest?

....

....

60. Har du opplevd problemer med ditt seksualliv?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

61. Har du hatt noen partner (gift, samboer)?

☐ Ja   ☐ Nei

62. Har du følt lyst til seksuell aktivitet?

Svært mye	0	1	2	3	4	5	6	7	8	9	10	Ikke noe
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

63a. Har du hatt ereksjon (reisning)?

Svært mye	0	1	2	3	4	5	6	7	8	9	10	Ikke noe
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

63b. Har du fått ereksjon (reisning) uten hjelpemiddel?

Svært mye	0	1	2	3	4	5	6	7	8	9	10	Ikke noe
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

64a. Var ereksjonen (reisningen) tilstrekkelig til å gjennomføre samleie?

☐ Ja   ☐ Nei

64b. Var ereksjonen tilstrekkelig uten hjelpemiddel til å gjennomføre samleie?

☐ Ja   ☐ Nei

12864

Mnd ☐ | O | | 1 | O | O | S |

64c. Har du brukt noe hjelpemiddel for å gjennomføre samleie?

☐ Alltid ☐ Sjelden ☐ Ikke i det hele tatt

☐ Ja ☐ Caverject ☐ Nei, gå til spørsmål 65
        ☐ Bondil
        ☐ Viagra (tablett)
        ☐ Uprima (tablett)

        ☐ Annet _____

64d. Kan du få ereksjon (reisning) med hjelpemiddel?

Svært mye	0	1	2	3	4	5	6	7	8	9	10	Ikke noe
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

64e. Er ereksjonen tilstrekkelig (med hjelpemiddel) til å gjennomføre samleie?

☐ Ja ☐ Nei

65. Hvor mange ganger har du hatt samleie den siste ..

Uken ☐☐   Måneden ☐☐   Året ☐☐   Mer enn et år siden ☐

66. Dersom du har opplevd problemer med seksuallivet, hva plager deg mest?

....

....

67. Har du hatt ømme bryst/brystvorter?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

68. Har du lagt merke til om brystene dine er blitt større?

Ikke noe	0	1	2	3	4	5	6	7	8	9	10	Svært mye
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

69. Synes du noen av spørsmålene er vanskelige å svare på?

Hvis ja, hvilke: ....

....

12864
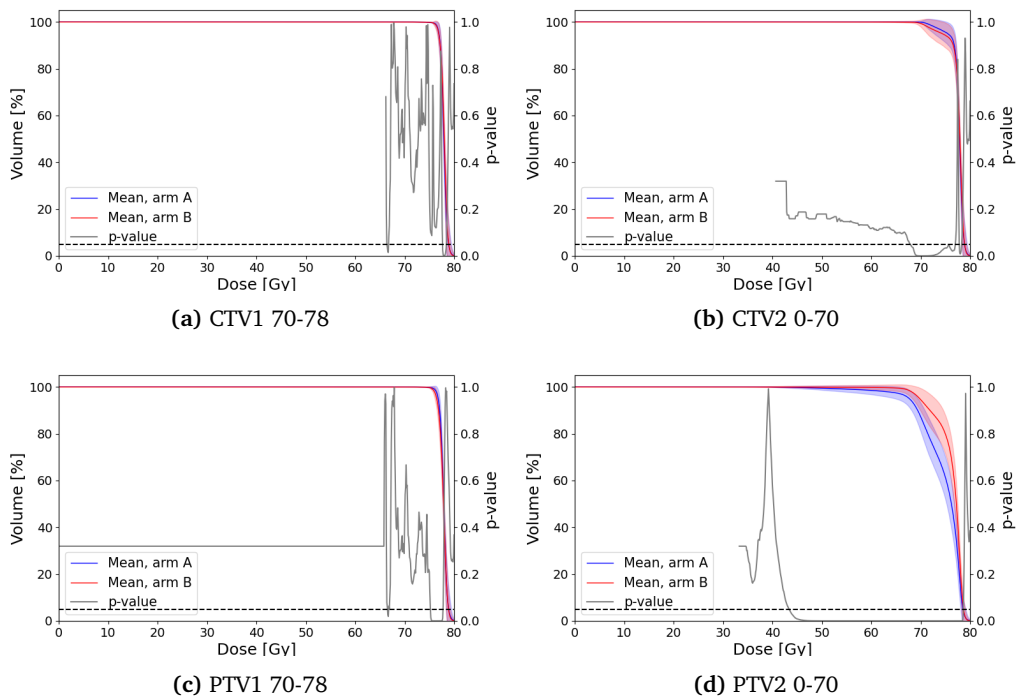
Takk for at du tok deg tid til å svare på spørsmålene!

B DVH Comparison

B.1 DVH Comparison of CTVs and PTVs Between Treatment Arms

Figure B.1: The mean DVHs and standard deviation (shaded areas) for each of the treatment arms are plotted against dose for (a) CTV1 70-78, (b) CTV2 0-70, (c) PTV1 70-78, and (d) PTV2 0-70. The DVHs were compared using a two-sided Student's t-test. The p-values are plotted on the secondary (right) axis, where the dashed black line indicates a significance level of $p < 0.05$.

B.2   DVH Outcome Comparison with Varying Cut-Off

DVHs for the rectal wall were compared as described in Section 3.4.2. The DVHs are presented here for rectal bother score cut-offs of RBS $\geq$ 3.5 and RBS $\geq$ 4.5.
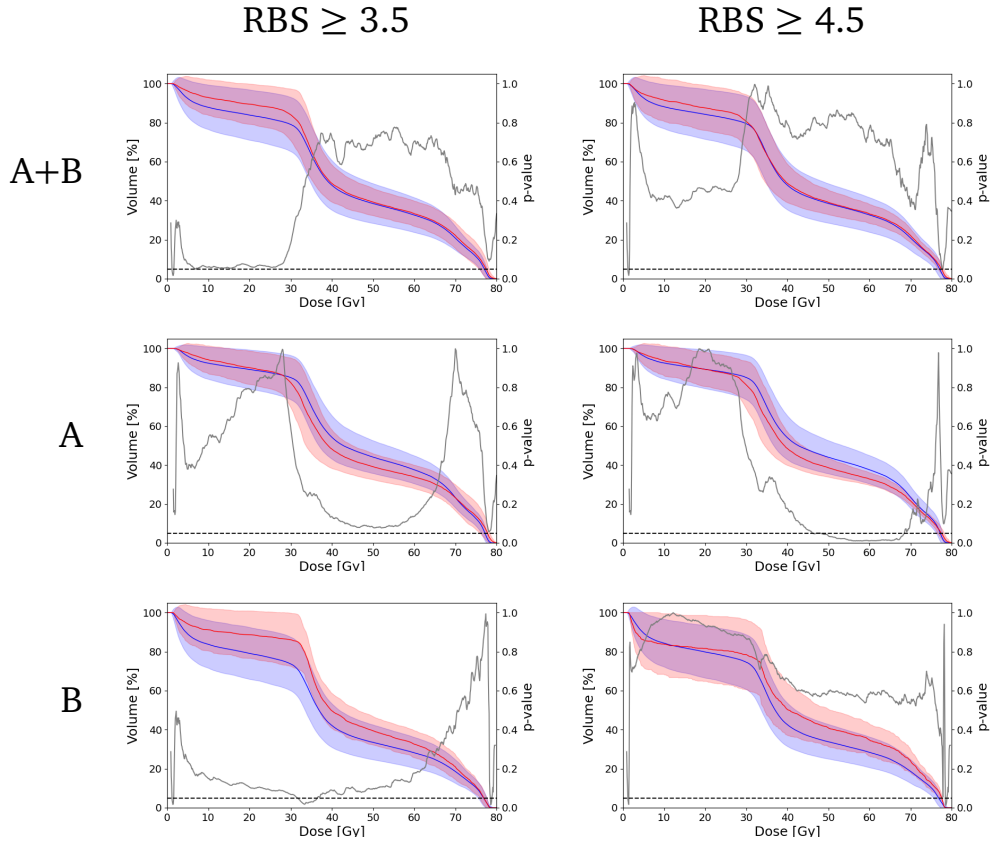
Figure B.2: Mean DVHs and standard deviation (corresponding shaded regions) for the rectal wall are plotted for patients with (red) and without (blue) adverse outcome. The cut-off values are RBS $\geq$ 3.5 (left) and RBS $\geq$ 4.5 (right). DVH comparison was performed for patients in both arms pooled (Fig. (a)-(b)), and arms A and B (Fig. (c)-(d) and Fig. (e)-(f), respectively). The gray lines show p-values from a two-sided t-test, and is plotted on the secondary (right) axis. The dashed black line indicates a significance level of $p < 0.05$.

C   SimpleElastix Parameter Maps

The code written in the project is presented in this appendix. The name of each file is given in the headings. Titles on the form "name1 > name2.py" indicate that the file "name2.py" is in the folder named "name1".

C.1   Default SimpleElastix parameter maps

```
1   ParameterMap 0:
2     (AutomaticParameterEstimation "true")
3     (AutomaticTransformInitialization "true")
4     (CheckNumberOfSamples "true")
5     (DefaultPixelValue 0)
6     (FinalBSplineInterpolationOrder 3)
7     (FixedImagePyramid "FixedSmoothingImagePyramid")
8     (ImageSampler "RandomCoordinate")
9     (Interpolator "LinearInterpolator")
10    (MaximumNumberOfIterations 256)
11    (MaximumNumberOfSamplingAttempts 8)
12    (Metric "AdvancedMattesMutualInformation")
13    (MovingImagePyramid "MovingSmoothingImagePyramid")
14    (NewSamplesEveryIteration "true")
15    (NumberOfResolutions 4)
16    (NumberOfSamplesForExactGradient 4096)
17    (NumberOfSpatialSamples 2048)
18    (Optimizer "AdaptiveStochasticGradientDescent")
19    (Registration "MultiResolutionRegistration")
20    (ResampleInterpolator "FinalBSplineInterpolator")
21    (Resampler "DefaultResampler")
22    (ResultImageFormat "nii")
23    (Transform "TranslationTransform")
24    (WriteIterationInfo "false")
25    (WriteResultImage "true")
26
27  ParameterMap 1:
28    (AutomaticParameterEstimation "true")
29    (AutomaticScalesEstimation "true")
30    (CheckNumberOfSamples "true")
31    (DefaultPixelValue 0)
32    (FinalBSplineInterpolationOrder 3)
33    (FixedImagePyramid "FixedSmoothingImagePyramid")
34    (ImageSampler "RandomCoordinate")
35    (Interpolator "LinearInterpolator")
36    (MaximumNumberOfIterations 256)
37    (MaximumNumberOfSamplingAttempts 8)
38    (Metric "AdvancedMattesMutualInformation")
39    (MovingImagePyramid "MovingSmoothingImagePyramid")
40    (NewSamplesEveryIteration "true")
41    (NumberOfResolutions 4)
42    (NumberOfSamplesForExactGradient 4096)
43    (NumberOfSpatialSamples 2048)
```

```
44      (Optimizer "AdaptiveStochasticGradientDescent")
45      (Registration "MultiResolutionRegistration")
46      (ResampleInterpolator "FinalBSplineInterpolator")
47      (Resampler "DefaultResampler")
48      (ResultImageFormat "nii")
49      (Transform "AffineTransform")
50      (WriteIterationInfo "false")
51      (WriteResultImage "true")
52
53   ParameterMap 2:
54      (AutomaticParameterEstimation "true")
55      (CheckNumberOfSamples "true")
56      (DefaultPixelValue 0)
57      (FinalBSplineInterpolationOrder 3)
58      (FinalGridSpacingInPhysicalUnits 10)
59      (FixedImagePyramid "FixedSmoothingImagePyramid")
60      (GridSpacingSchedule 2.80322 1.9881 1.41 1)
61      (ImageSampler "RandomCoordinate")
62      (Interpolator "LinearInterpolator")
63      (MaximumNumberOfIterations 256)
64      (MaximumNumberOfSamplingAttempts 8)
65      (Metric "AdvancedMattesMutualInformation" "TransformBendingEnergyPenalty")
66      (Metric0Weight 1)
67      (Metric1Weight 1)
68      (MovingImagePyramid "MovingSmoothingImagePyramid")
69      (NewSamplesEveryIteration "true")
70      (NumberOfResolutions 4)
71      (NumberOfSamplesForExactGradient 4096)
72      (NumberOfSpatialSamples 2048)
73      (Optimizer "AdaptiveStochasticGradientDescent")
74      (Registration "MultiMetricMultiResolutionRegistration")
75      (ResampleInterpolator "FinalBSplineInterpolator")
76      (Resampler "DefaultResampler")
77      (ResultImageFormat "nii")
78      (Transform "BSplineTransform")
79      (WriteIterationInfo "false")
80      (WriteResultImage "true")
```

C.2   Default SimpleElastix with increased maximum number of iterations

This parameter set is mostly equal to the previous parameter map in C.1. However, the maximum number of iterations in each transform and resolution is set to 512 (rather than the default 256) in the script by the call

```
1      elastixImageFilter.SetParameter("MaximumNumberOfIterations", "512")
```

The MaximumNumberOfIterations parameter is set for all initialized parameter maps in the elastixImageFilter.

C.3 Affine Transform

The following parameter map was used for the affine transform.

```
1   //
2   // Created by: Ingrid Tveten
3   // Created on: 2021-04-01
4   //
5   // What: Affine transform parameter file for MSc project
6   // How: 3D CT-to-CT registration using intensity-based metrics
7   //
8
9   // ================ Main settings ================
10  (FixedInternalImagePixelType "float")
11  (MovingInternalImagePixelType "float")
12  (FixedImageDimension 3)
13  (MovingImageDimension 3)
14  (UseDirectionCosines "true")
15
16  // =============== Main Components ===============
17  (Transform "AffineTransform")
18
19  (Registration "MultiResolutionRegistration")
20  (Optimizer "AdaptiveStochasticGradientDescent")
21  (Interpolator "BSplineInterpolator")
22  (Metric "AdvancedMattesMutualInformation")
23  (ResampleInterpolator "FinalBSplineInterpolator")
24  (Resampler "DefaultResampler")
25
26  // =========== Registration parameters ============
27  (NumberOfResolutions 4)
28  (ImagePyramidSchedule 8 8 8  4 4 4  2 2 2  1 1 1 )
29  (FixedImagePyramid "FixedRecursiveImagePyramid")
30  (MovingImagePyramid "MovingRecursiveImagePyramid")
31
32  // =========== Transform parameters ============
33  (HowToCombineTransforms "Compose")
34  (AutomaticScalesEstimation "true")
35  (AutomaticTransformInitialization "true")
36
37  // ============= Similarity measure =============
38  (NumberOfHistogramBins 32)
39  (ErodeMask "false")
40
41  // ================= Optimizer =================
42  (MaximumNumberOfIterations 500)
43
44  // =============== Image sampling ===============
45  (ImageSampler "Random")
46  (NumberOfSpatialSamples 2048)
47  (NewSamplesEveryIteration "true")
48
49  // ========== Interpolation and Re-sampling ==========
50  (BSplineInterpolationOrder 1)
51  (FinalBSplineInterpolationOrder 3)
52  (DefaultPixelValue 0)
53  (WriteResultImage "true")
54  (ResultImagePixelType "short")
55  (ResultImageFormat "nii")
```

C.4 Affine and B-Spline Transform

For the affine and B-spline transform, the following B-spline parameter map was added to the affine transform parameter map in the previous section (C.3). The two transforms were applied consecutively to the moving image.

```
1   //
2   // Created by: Ingrid Tveten
3   // Created on: 2021-04-01
4   //
5   // What: B-spline parameter file for MSc project
6   // How: 3D CT-to-CT registration using intensity-based metrics
7   //
8
9   // =============== Main settings ================
10  (FixedInternalImagePixelType "float")
11  (MovingInternalImagePixelType "float")
12  (FixedImageDimension 3)
13  (MovingImageDimension 3)
14  (UseDirectionCosines "true")
15
16  // =============== Main Components ===============
17  (Transform "BSplineTransform")
18
19  (Registration "MultiResolutionRegistration")
20  (Optimizer "AdaptiveStochasticGradientDescent")
21  (Interpolator "BSplineInterpolator")
22  (Metric "AdvancedMattesMutualInformation")
23  (ResampleInterpolator "FinalBSplineInterpolator")
24  (Resampler "DefaultResampler")
25
26  // =========== Registration parameters ============
27  (NumberOfResolutions 4)
28  (ImagePyramidSchedule 8 8 8  4 4 4  2 2 2  1 1 1 )
29  (FixedImagePyramid "FixedRecursiveImagePyramid")
30  (MovingImagePyramid "MovingRecursiveImagePyramid")
31
32  // =========== Transform parameters ============
33  (HowToCombineTransforms "Compose")
34  (AutomaticScalesEstimation "true")
35  (AutomaticTransformInitialization "true")
36
37  // ============= Similarity measure =============
38  (NumberOfHistogramBins 32)
39  (ErodeMask "false")
40
41  // ================ Optimizer =================
42  (MaximumNumberOfIterations 500)
43
44  // ============== Image sampling ===============
45  (ImageSampler "Random")
46  (NumberOfSpatialSamples 2048)
47  (NewSamplesEveryIteration "true")
48
49  // ========== Interpolation and Re-sampling ==========
50  (BSplineInterpolationOrder 1)
51  (FinalBSplineInterpolationOrder 3)
52  (DefaultPixelValue 0)
53  (WriteResultImage "true")
54  (ResultImagePixelType "short")
55  (ResultImageFormat "nii")
```

D Results from Voxel-Based Analysis

Results from the VBA are given here for PROMs (rectal bother score) at 5, 18 and 24 months after inclusion. The cut-off for RBS was set to 2.0, and dose to patients with and without complications was compared per voxel using a two-sided Student's t-test.
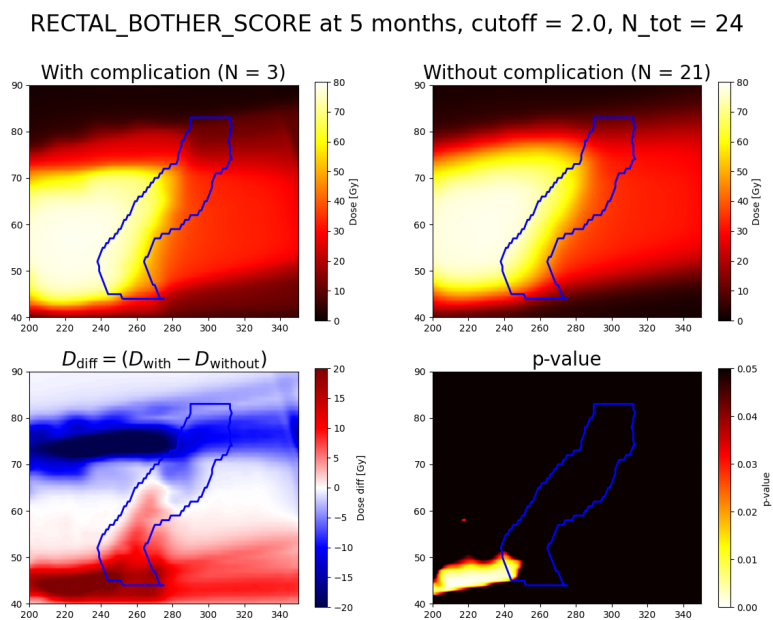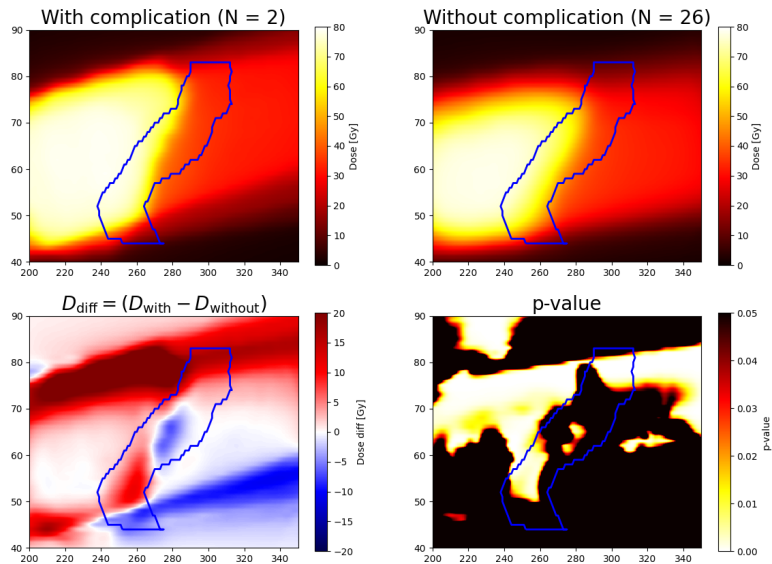

Figure D.1: *Top:* Mean dose for patients with RBS ≥ 2.5 (left) and RBS < 2.5 (right) at 5 months. *Bottom:* The dose difference (left) and p-values from a voxel-wise t-test (right). $N = 24$ patients were included in the analysis.
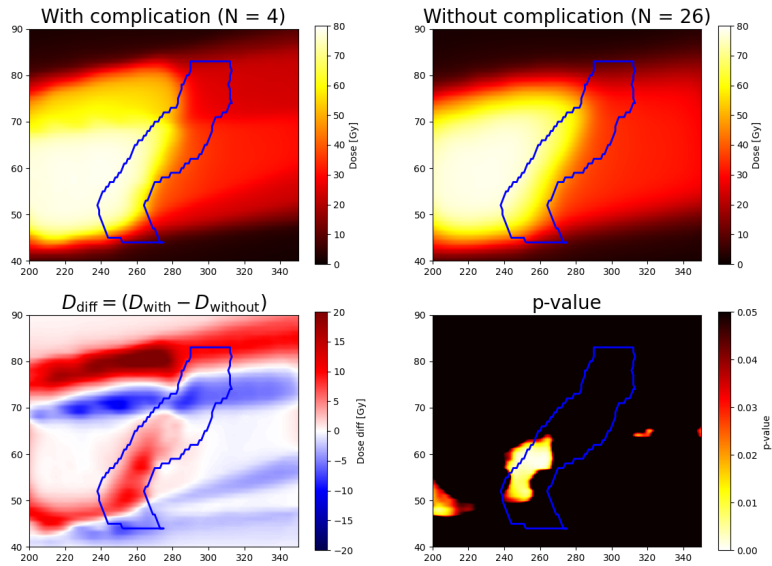
Figure D.2: *Top:* Mean dose for patients with RBS ≥ 2.5 (left) and RBS < 2.5 (right) at 18 months. *Bottom:* The dose difference (left) and p-values from a voxel-wise *t*-test (right). *N* = 30 patients were included in the analysis.

Figure D.3: *Top:* Mean dose for patients with RBS ≥ 2.5 (left) and RBS < 2.5 (right) at 24 months. *Bottom:* The dose difference (left) and p-values from a voxel-wise *t*-test (right). *N* = 28 patients were included in the analysis.

E   NACP Abstract

The abstract "Quantitative dose metrics and normal tissue complication probability modeling in prostate cancer" was submitted to the Nordic Association for Clinical Physics (NACP) digital symposium 2020/2021, and was presented as a poster.

NACP 2020 Abstract

Quantitative dose metrics and normal tissue complication probability modeling in prostate cancer

With recent years' advances in prostate cancer radiotherapy (RT), patient survival is improved. However, for some patients the treatment results in side-effects that may reduce quality of life. Better methods for optimizing RT dose delivery and modelling normal tissue complications are needed to predict outcomes and further optimize and personalize the treatment. The purpose of the current project was to create a programming framework for quantifying dose metrics and plan quality in delivered RT plans and for modelling normal tissue complication probability (NTCP).

The RIC study (a **R**andomized trial of daily CT vs standard **I**mage-guided RT in prostate **C**ancer) sought to investigate whether reduced safety margins in prostate cancer RT could lead to a reduction in side-effects. The study material included delivered dose plans from 245 patients from both study arms (A: standard margins and weekly orthogonal verification, B: reduced margins and daily CT verification) and associated patient-reported outcome measures (PROMs) from 36 months of follow-up.

A programming framework was developed in Python. First, the program was used to calculate a range of quantitative dose metrics, such as three-dimensional dose-volume histogram (DVH) characteristics, to evaluate whether delivered radiation dose to the planning target volume (PTV) and organs at risk (OAR) was according to the study protocol. Second, the Lyman-Kutcher-Burman (LKB) NTCP model was implemented and its parameters were fitted to dose data and PROMs for the OARs.

The basic functionality of the program was demonstrated through calculation of quantitative dose metrics. Although evaluation of the DVH characteristics for the PTV showed that all except one plan were in agreement with the treatment protocol, other metrics such as the $V_{60Gy} \leq 50\%$ constraint given in the protocol was shown to be exceeded by 5.5 % of dose plans for the rectum and by 7.5 % of dose plans for the rectal mucosa. Fitting of the LKB parameters to the RIC study data yielded a tolerance dose for 50 % complication rate $TD50$=77.0 Gy, n=0.007, and m=0.016. Analysis of dose metrics and NTCP modelling parameters for other OARs (bladder, sexual function) is ongoing and will be presented at the meeting.

A framework for quantitative evaluation of dose metrics and NTCP modelling of prostate cancer was developed. The framework identified dose plans deviating from the treatment protocol. The LKB model parameters showed fair agreement with literature values used clinically, values based on clinical assessment and not PROMs. The low value of the volume-dependent parameter n may indicate a serial architecture of the rectum and increased sensitivity to high doses. The implemented framework provides a starting point for developing more advanced NTCP models that also take spatial dose localization into account. Such a model may overcome the limitation in the organ based NTCP models that provide the basis for the dose limits used in the clinic today.

Figure 1:
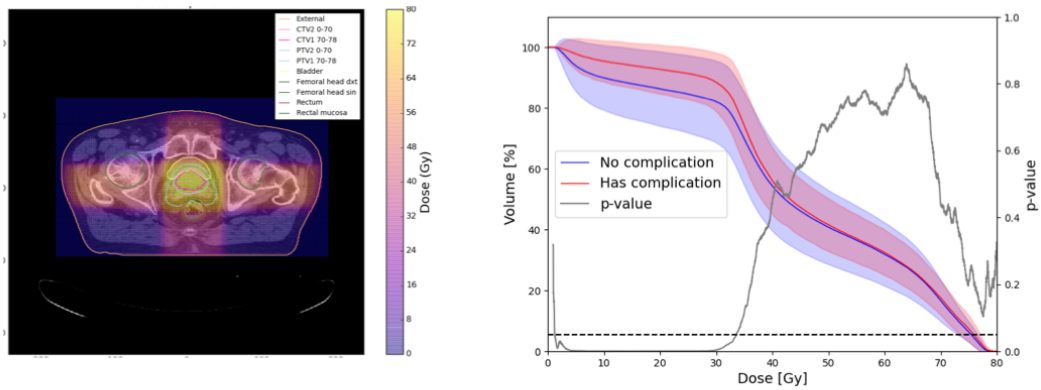
Figure 1: *Left:* A sample dose plan for a patient is shown to the left. The dose level is indicated by the colour bar, and the dose is overlaid a CT image of the patient where structures are outlined. *Right:* The mean DVH for the rectum (plus contents) is shown for patients with (red) and without (blue) complications, measured by a composite score of five items from the prostate cancer-specific questionnaire QUFW94. The standard deviation for each of the two groups is indicated by the red/blue shaded area. The p-value indicates whether samples are significantly different ($p < 0.05$, indicated by the dashed line), as tested by Student's t-test.

Table 1:

Table 1: Optimized parameters for the LKB model for the rectum. The outcome was measured by a composite score of five items from the prostate cancer-specific questionnaire QUFW94. Of the patients, 8.3 % were classified as having complications. The optimal value $\pm$ estimated standard deviation is given. Recommended values from the QUANTEC review are provided for reference.

Parameter	Best-fit parameters	QUANTEC recommendation
TD50 [Gy]	77.0 ± 1.5	76.9
n $[-]$	0.007 ± 0.005	0.09
m $[-]$	0.016 ± 0.006	0.13

Ingrid Elisabeth Tveten

Dosimetric Predictors of Toxicity in Prostate Cancer Radiotherapy

NTNU
Norwegian University of
Science and Technology