The effect of contouring variation and dose in the rectum on toxicity levels for prostate cancer treatment

ARTEMIS BOUZAKI

University of Manchester, School of Physics and Astronomy
and
The Christie Hospital

This project was supervised by Eliana Vasquez Osorio and Alan McWilliam.

May 2023

Abstract

The effect of inter-observer variation in contouring procedures for prostate cancer has led to the automation of the contouring process by deep learning algorithms. This study aims to quantify the contour variations between manual and automatic rectum delineations, while also considering the effects of the planned dose distribution for a dataset of 1,758 patients. The dataset was refined to 1,377 patients after excluding cases where either the prostate or the rectum had been misidentified. A novel method of standardising the rectum length was developed that significantly enhanced the strength of the analysis. A Student's t-test as well as a univariate Cox proportional hazards model were performed to test for significant contour variation regions, but no significance ($p < 0.05$) was found. When the Student's t-test was performed for the dose at the manual rectum, the lower posterior was found to be significant for the side effects of proctitis and bowel urgency. A discrepancy in the results of both CPHM and Student's t analyses was observed between the dose at the manual and automatic rectum contours, suggesting that inconsistency in the contouring process can impact the analysis. The preliminary results of this study suggest that reducing the dose at the lower posterior of the rectum can lead to better toxicity outcomes. Further investigation of the causes of significance and the effect of relevant clinical variables need to be considered in the future. This study is mainly limited by the discrepancy between planned and delivered doses and assumptions made about the structure of the rectum when standardising its length.

Contents

1 Introduction

Prostate cancer is the second most common cancer in men and has incidence rate of nearly 60% in men over 65 years [1]. Radiotherapy is one of the most common types of treatment for prostate cancer and it aims to irradiate cancer cells while sparing organs at risk (OARs), such as the rectum [2]. An important stage in treatment planning is the contouring of the target volume and OARs. Accurate contouring is crucial to plan an optimal dose distribution for the patient. However, the contouring process is inconsistent due to hospital practices and human error, an effect known as Inter-Observer Variation (IOV), which can affect the treatment outcomes [3]. Deep learning algorithms have been developed to automate the contouring process and eliminate IOV [4], [5].
Last semester, a Student's t permutation test was carried out on a sample of 492 patients to investigate the impact of contour variations in the rectum on toxicity levels post-treatment. The deep learning software ADMIRE was used to create automated delineations and was compared to manual delineations made by clinicians. No regions of significance ($p < 0.05$) were found. The dataset included 1,758 patients in total, suggesting limitations in the method followed, which reduced the data sample and potentially affected the analysis [6].
The aim of this study is to quantify the contour variations between the automated and manual rectum contours. Increasing the number of patients used for analysis was one of the main objectives of this study. A new method of standardising the rectum length was introduced which significantly expanded the sample size and allowed for the inclusion of all regions of interest of the rectum in the analysis. The planned dose at the boundaries of the manual rectum contour, as well as the product between the contour variations and the dose were also sampled to investigate their correlation to toxicity levels. The planned dose at the automatic rectum contour was obtained for comparison. The analysis was performed using a Student's t-test as well as Cox Proportional Hazards Model (CPHM).

2 Medical background

2.1 Terminology

Throughout this report, the medical terminology for the different directions and planes is shown in Figure 1. This terminology will be used consistently throughout this study.

2.2 The prostate

The prostate is a walnut-shaped gland of the male reproductive system. Its purpose is to produce fluid for semen. It is separated into different zones with the most important one being the peripheral zone which constitutes over 70% of the glandular prostate. Almost all carcinomas begin at the peripheral zone [8]. The prostate is in close proximity to other organs, including the rectum in the posterior, the bladder and the seminal vesicles. Figure 2 shows the basic anatomy of the prostate and the location of the nearby organs.
Prostate cancer is a type of cancer that develops in the prostate gland and it is most commonly diagnosed in men at the age of 66 years old [10]. It can be asymptomatic at early stages and can require minimal treatment. The exact causes of prostate cancer are not well understood but risk factors include age, family history and genetic mutations. Although prostate cancer is very common, most cases are detected when the cancer is within the prostate and thus the survival rate is high [11], [12].

2.3 Contouring and automation in treatment planning

In order to plan dose delivery, clinicians use CT scans to create delineations of the target volume as well as the organs at risk (OARs). The delineations need to be as accurate as

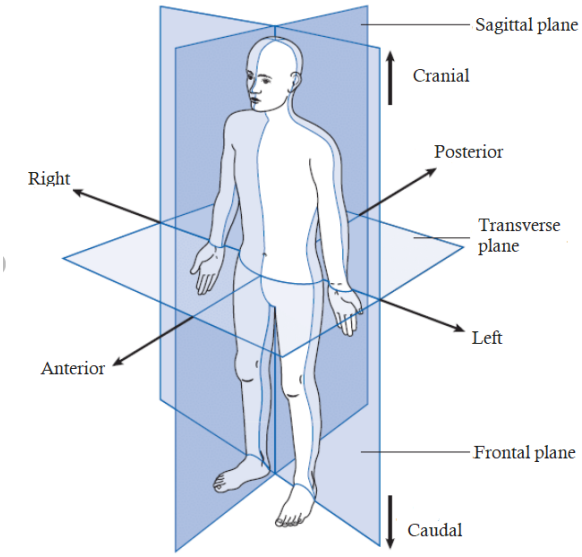


Figure 1: Anatomical planes and directions as used in medicine. The figure was taken and adapted from [7].

.

possible to spare healthy tissue while irradiating the tumour [3].
Manual contouring of the target volume and OARs can vary depending on the clinician and hospital practices. This effect is known as Inter-Observer Variation (IOV). Gual-Arnau *et al.* [13] and Nyholm *et al.* [14] have shown that IOV can be found in contours for both the prostate and the seminal vesicles. IOV in the rectum has also been found in the treatment of both prostate cancer [15] and rectal cancer [16]. Since the target volume and OARs delineations are used for treatment planning, IOV can affect the dose delivered, potentially impacting treatment outcomes [3].
Deep learning algorithms have been developed to tackle the inconsistency of manual delineations by automating the contouring process. Recent studies have demonstrated that deep-learning contouring can achieve a performance of increased accuracy that is comparable to clinicians while reducing the time and effort required [4], [5], [17].
In this study, the deep learning software 'Advance Medical Image Registration Engine' (ADMIRE v1.12) was used to create automatic delineations of the target volume and OARs. The motive of this study was to investigate the potential effect of IOV on the rectum by comparing contours generated by the software to the ones made manually by clinicians.

2.4 Radiotherapy treatment

The aim of radiotherapy is to kill cancer cells with the use of ionising radiation while sparing the surrounding tissue. The patients in this study were treated using external beam radiotherapy (EBRT), which is used commonly for prostate cancer treatment. EBRT delivers high-energy X-rays or proton beams to the prostate from outside the body [2]. Radiation exposure is measured in a unit called gray (Gy), equivalent to one joule of radiation deposited per kilogram of tissue [18].
EBRT kills malignant cells by damaging their DNA, leading to cellular death [19]. Although radiation therapy is primarily aimed at killing tumour cells, it can damage surrounding healthy tissue such as the rectal tissue, leading to toxicity. Various factors contribute to rectal damage, including the volume of the rectum that is being irradiated. One of the most common side effects of rectal damage due to EBRT is proctitis, which can cause rectal bleeding and pain [20], [21]. High levels of radiation at the rectum can also cause bowel problems, such as

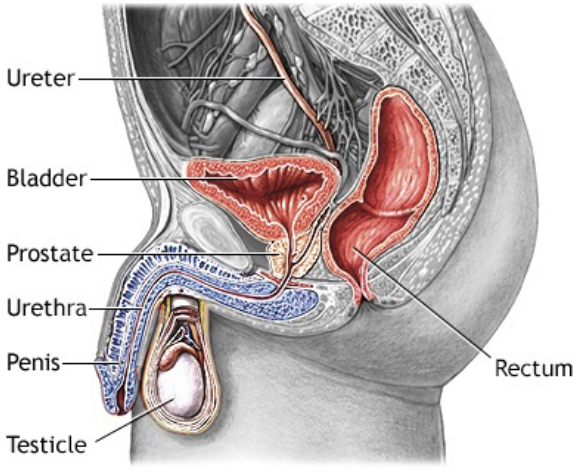


Figure 2: The anatomy of the prostate and the organs in proximity shown in the sagittal plane [9].

bowel urgency [22].

2.5 Toxicity levels - the CTCAE scale

In this study, the toxicity levels for a side effect were measured according to the Common Terminology Criteria of Adverse Events (CTCAE) scale. This scale has grades from 0 to 5 and it measures the severity of a symptom. A grade of 0 indicates that the patient does not display the symptom, whereas a grade of 5 indicates death from the symptom. Grades of 2 and above require medical intervention [23].
For certain side effects, the toxicities were also recorded directly from the patient with no intervention or modification by clinicians. These patient-reported toxicities ranged from 0 to 4 and the grading system was similar to that of the CTCAE scale.

2.6 The dataset

The patient cohort was part of the REQUITE study and consisted of 1,758 patients from 26 hospitals in 8 different countries treated and followed up between April 2014 and March 2017 [24]. For each patient in the dataset, there was a record of different side effects and their clinical and patient-reported toxicity levels. The toxicity level data were obtained before treatment (at 0 months), as well as at 1 month, 12 months, 24 months, and 36 months.
Full access was available to the parameters of each patient's 3D CT scan and planned dose distribution, manual contours made by the oncologist, as well as the automatic contours made by the ADMIRE software. Data about each patient's age were also available. The scans were manipulated using the in-house developed software called Worldmatch, for visualisation and extraction of information.

3 Method

3.1 Data formatting

The data for each patient contained the following DICOM files: CT scans, planned dose distribution, manual delineations and automatic delineations. DICOM files are commonly used in medical imaging to store information [25]. Out of 1,758 patients in the dataset, 210 patients did not have sufficient data due to empty files. For the rest of the files that were

appropriate for analysis, the dose, CT scans and delineations were combined into a single file. These combined files contained both the manual and automatic rectum contours as a series of 2D axial slices. The manual and automatic delineations were different in all dimensions.

3.2 Manual rectum contour identification

Although automatic delineations were consistent in naming, this was not the case for delineations created by oncologists. In order to ensure consistency in the analysis, the manual rectum contour had to be identified and renamed appropriately. In order to reduce computational time, the manual rectum contour was identified by name whenever possible. However, in some cases, the naming could not be predicted due to variations in naming conventions across different regions and institutions. In these cases, a different method was followed.
A mask of the automatic rectum contour was created by setting all the voxel values within that delineation to 255 and voxels outside the delineation to 0. The automatic rectum contour thus appeared white, contrasting the background. This enabled the creation of a distance map, which is a field spanning the entire 3D image. In this field, every voxel contained information about the smallest distance in 3D to the masked delineation. A property of the distance map is that the distance is negative for voxels inside the masked delineation. Therefore, only voxels that are exactly at the boundary of the masked delineation have the value 0 in the distance map field. The mean distance d between each delineation and the automatic rectum mask was found by sampling over all available delineations. It was assumed that the absolute difference between the manual and automatic rectum delineations was small. Therefore, the manual rectum was determined by selecting the delineation with the smallest value of $|d|$.
As a way to validate the distance map, the Sørensen–Dice coefficient was implemented. This coefficient is a statistical tool used to extract the similarity between two samples and is widely used in medical research. In this analysis, the coefficient was used to compare the similarity of the automatic rectum contour with all the other delineations. The Sørensen–Dice coefficient for two delineations is the following [26]

$$SDC(V_{auto}, V) = 2\frac{V_{auto} \cap V}{V_{auto} + V}, \tag{1}$$

where V_{auto} is the volume of the automatic rectum delineation, V is the volume of the delineation being tested and $V_{auto} \cap V$ is the intersection between the two volumes. The Sørensen–Dice coefficient was found for each delineation using Equation 1. The value of ψ was then calculated for each delineation, where

$$\psi = \frac{|d|}{1 + SDC(V_{auto}, V)}. \tag{2}$$

Smaller values of ψ would therefore correspond to increased similarity.
The rectum was misidentified in 68 patients due to other delineations, such as the anus, being very similar. Those patients were removed from the analysis leaving 1,480 patients at this stage.

3.3 Prostate centre of mass

It is important to be able to identify the location of the prostate relative to the rectum in order to normalise rectal geometry. The automatic prostate delineation was used to find the centre of mass of the prostate in each patient. However, there were cases in which the prostate was misidentified due to the nature of the training of the ADMIRE software. Figure 3 shows an example in which the heart was misidentified as the prostate.
To avoid this issue, the centre of mass of the prostate was validated using the planned dose

distribution. It was expected that the centre of the prostate would be located proximate to a high dose point. For this reason, a mask of the dose was made at the areas of 95% of the maximum dose. By comparing the centre of the high-dose mask and the centre of the automatic prostate delineation, it was possible to ascertain whether the prostate had been correctly identified. In particular, in cases where the high dose centre and the prostate centre deviated by more than 2 cm, it was assumed that the ADMIRE software had misidentified the prostate and thus the patient was excluded from the analysis. This procedure eliminated a further 103 patients, resulting in 1,377 patients included in the final analysis.

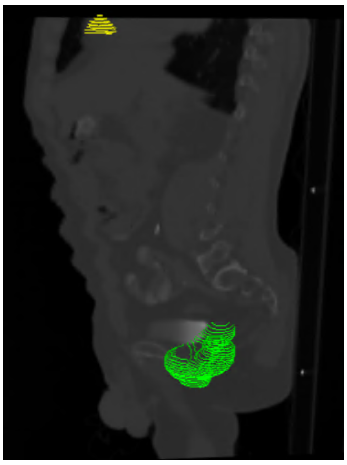


Figure 3: Example scan in the sagittal plane with the manual prostate (green) and automatic prostate (yellow) delineations shown. The two delineations deviate by more than 2 cm since the heart has been misidentified as the prostate by the ADMIRE software. Such cases were not included in the analysis.

3.4 Quantifying contour variations

To quantify contour variations, the manual rectum delineation was cropped according to the automatic rectum delineation, so that the CT scan would only contain the slices that include the automatic rectum. The automatic delineation was chosen as a reference since it is independent of the dose distribution and prostate location and has consistent start and stop positions which depend only on the anatomical structure of the rectum.
At each slice of the craniocaudal axis (y), the centre of mass of the automatic rectum delineation for that slice was identified. An adequate amount of 100 angles was chosen to sample around the transverse plane which is defined by the left-right (x) and posterior-anterior (z) directions. At every angle for a given slice, the radii $r_{auto}(\theta, y)$ and $r_{manual}(\theta, y)$ from the centre of the automatic rectum to the boundary of the automatic and manual rectum delineations were computed. The algorithm was able to identify the edges of a delineation by creating a mask of it. Starting from the automatic rectum centre, the algorithm took steps of one pixel until a pixel value of 0 was encountered, indicating that it had just stepped outside the delineation. The difference in radial distance at that angle would thus be

$$\Delta r(\theta, y) = r_{manual}(\theta, y) - r_{auto}(\theta, y). \tag{3}$$

By repeating this process for each slice of the automatic rectum delineation, maps of varying sizes were generated based on the number of slices of the automatic rectum. These maps contained the radial differences at each angle of every slice. All maps had the same number of columns corresponding to 100 angles around the x-z plane. Since the radial distances were calculated in terms of pixel steps, a scaling function needed to be applied to account for the

fact that a pixel step corresponds to a larger distance at the diagonals. The scaling function had the form $(\sqrt{2}-1)\sin^2(2\theta)+1$, where θ is the sampling angle around the transverse plane. This function peaks at $\sqrt{2}$ at the diagonals and has no effect at the x and z axes.
The dose at the boundary of the manual rectum contour was also sampled for every angle of each slice and saved into maps of the same format as the contour variations. This allowed for the creation of a third type of map, containing the product between the radial differences and the planned dose at the manual rectum. This interaction map allowed for boosting the values of the contour variation in regions where the dose was high and therefore, reducing the impact of large contour variations in regions where the rectum was not significantly irradiated. Lastly, the planned dose at the boundary of the automatic rectum contour was sampled for comparison.
Since the automatic rectum contour was used as a reference for cropping the scan, there were slices that only contained the automatic rectum delineation due to the manual rectum being too small. In these cases, the radial difference was set to a NaN value. This was also applied in cases where the delineations were too curved, resulting in no boundary to be found for either the automatic or the manual rectum delineations at a specific angle. Figure 4 below shows those different cases schematically and how they were treated by the algorithm.

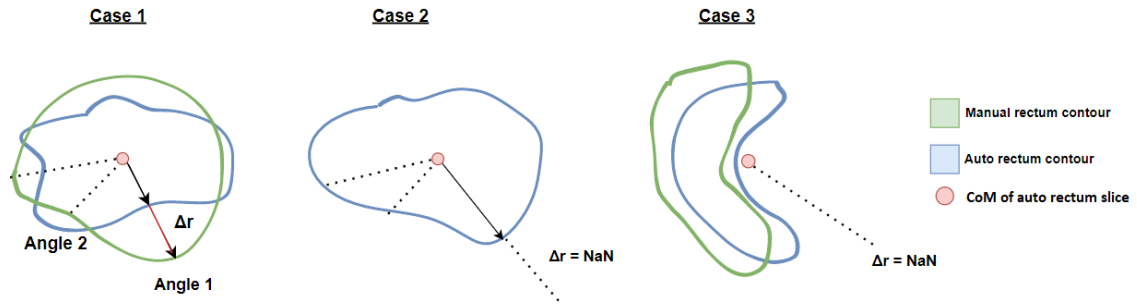


Figure 4: Original diagram showing three different cases that occurred in the sampling process and the value registered for the radial difference in each case. In Case 1 the radial difference is computed normally. Case 2 shows the scenario in which the manual delineation does not exist for a specific slice. Case 3 shows the scenario of extreme curvature, where the centre of mass of the automatic rectum slice lies outside the delineation.

This process allowed unfolding the rectum into 2D surface maps. However, for the analysis to be consistent and useful, all the maps had to be of the same size. This posed an issue since the rectum contours consisted of different amounts of slices for each patient. A solution to this would be to crop the rectum to a predetermined amount of slices, however, Bouzaki and Green [6] found that this would introduce a bias against smaller rectum contours and would result in significantly fewer patients being analysed. The novel method introduced below provides a solution to the problem without having to enforce a limit on the rectum length and narrow the dataset.

3.5 Solving the rectum length problem

To avoid cropping the rectum and excluding patients from the analysis, a coordinate transformation was applied to the maps, allowing all rectum maps to be stretched/compressed to a standard length of 60 slices. The coordinate transformation had the following steps:

- The coordinate of the map corresponding to the centre of mass of the automatic prostate was mapped to 0.

- The slices corresponding to the top half of the rectum from the prostate centre (cranial direction) which had coordinates (prostate centre, a], where a is the map coordinate of the top of the rectum, were transformed to a new map with coordinates (0, 30] using a linear transformation.
- Similarly, the slices corresponding to the bottom half of the rectum from the prostate centre (caudal direction) with coordinates [b, prostate centre), where b is the map coordinate of the bottom of the rectum, were transformed to a new map with coordinates [-30, 0) using another linear transformation. A different transformation was used in each direction due to the fact that the number of slices below and above the prostate was not always symmetrical.

This allowed for consistent mapping of the prostate coordinate at 0 and different linear interpolations between the slices of the top and bottom half of the rectum. This method assumed that the anatomical structure of the rectum can be aligned relative to the prostate and is the same for shorter and longer rectum contours.
This process resulted in a new map of 60 slices in total. The new map was rectilinear since the original uniform map would have a different amount of slices in the caudal and cranial directions from the centre of the prostate, and thus the interpolation would behave differently in each direction. The transformation was also applied in the maps containing the dose data as well as the product between the radial differences and the dose. Figure 5 shows a schematic diagram of the coordinate transformation and the resultant field.

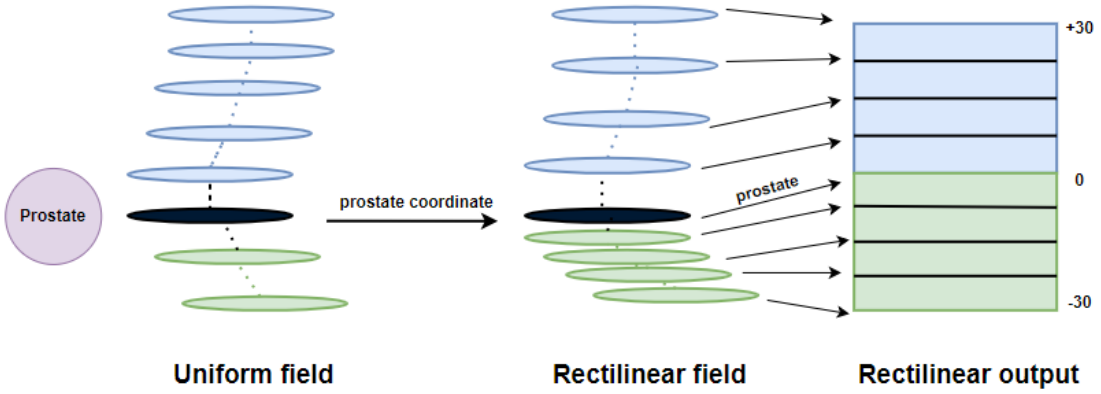


Figure 5: Original schematic diagram showing the coordinate transformation process of a manual rectum contour to a standardised length of 60 slices using linear transformations and interpolation.

The slices containing NaN values in the original uniform map did not contribute to the transformation and therefore their proportions remained the same in the rectilinear map. The resultant field was visualised in the form of 2D images containing the data.

3.6 Dichotomizing the toxicities

The data for the toxicity levels needed to be in a binary format of event cases and non-event cases for the analysis to take place.
Event cases were considered patients with a recorded toxicity of 2 and above for a certain side effect. Subsequently, a toxicity value of 0 or 1 on the CTCAE scale would correspond to a non-event case. The value of 2 on the CTCAE scale was chosen to dichotomize the toxicities due to patients requiring medical intervention at this stage.

4 Statistical background

The statistical analysis performed involves a voxel-by-voxel binary analysis for a given rectal side effect. This can be done using various models however, the primary models used were Student's t permutation test and Cox Proportional Hazards Model (CPHM). Those models were implemented using an in-house developed toolkit called Manchester Toolkit (MTK).
After the toxicity data were dichotomized, the sample of patients had two possible outcomes: non-event (N) corresponding to a toxicity value of 0 and event (E) corresponding to a toxicity value of 1.
The theory outlined below is specific to the analysis carried out for the contour variations. The same analysis was carried out for all types of data collected.

4.1 Hypothesis testing

Hypothesis testing is the process of deciding between two conflicting hypotheses, the null hypothesis H_0 and the alternative hypothesis H_a so that [27]

- H_0 represents the hypothesis that patient toxicity is independent of contour variation
- H_a represents the hypothesis that patient toxicity is dependent on contour variation.

In order to be able to prove H_a substantial evidence is needed to reject H_0. Assuming that the null hypothesis holds, the significance level α can be defined as the threshold probability below which H_0 can be rejected. The significance probability can then be defined as the probability that a test statistic is equal to or more extreme than the observed value. In this analysis, α was chosen to be 0.05 and so $p < 0.05$ is needed to reject the null hypothesis with confidence 100(1 - p)% [28].

4.2 Student's t permutation test

The Student's t-test is based on the Student's t-distribution and it provides a test statistic T to compare two distributions of unknown standard deviation in order to test a hypothesis about the mean [29]. The method described below was taken from Chen *et al* [30]. The same method was used by Bouzaki and Green in 2022 [6].
A new randomised sample with i permutations was created by permuting the data. The Student's t-test statistic was then given by

$$T_{i,k} = \frac{\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}}{\sigma_k}. \tag{4}$$

In Equation 4, $\mu_{\mathrm{E},i,k}$ corresponds to the mean of the event group contour variation for a given sample i and voxel k. Similarly, $\mu_{\mathrm{N},i,k}$ corresponds to the mean of the non-event group contour variation for a given sample i and voxel k. σ_k is the standard deviation of $\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}$ for a given voxel. [30] shows that 1,000 permutations are sufficient for large data sets. In order to calculate σ_k, $\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}$ is evaluated over all permutations. This allows for the calculation of the test statistic. The maximal test statistic for a permutation is defined as $T_{max,i} = max_k(T_{i,k})$ and it represents the most significant statistic of a permutation.
$\tilde{T}_{max}$ is also defined as the maximum test statistic of the non-permuted data. The p-value is then defined as the fraction of permutations below $\tilde{T}_{max}$ or

$$p = \frac{\left| T_{max,i} < \tilde{T}_{max} \right|}{T_{max,i}}. \tag{5}$$

If $p < 0.05$ the null hypothesis can be rejected. For a two-tailed test, the 2.5th and 97.5th

percentiles of $T_{max,i}$ distribution give threshold values $T^*_{min,i}$ and $T^*_{max,i}$ respectively. By comparing the extreme $\tilde{T}_k$ values to the threshold distribution, the confidence level of statistical significance can be found.

4.3 Cox per voxel survival analysis

4.3.1 Survival analysis

Survival analysis is a statistical method for analysing the expected time duration until a dichotomous event occurs. Data in survival analysis are described by either the survival function $S(t)$ or the hazard function $h(t)$. The survival function is the probability that an individual survives from the diagnosis to a specific future time, t. The hazard probability refers to the probability that an individual has an event at time t [31], and is given by

$$h(t) = \lim_{\delta t \to \infty} \left(\frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \right), \tag{6}$$

where T is the time to event. It, therefore, represents the event rate of a patient who has already survived to time t and is related to the survival probability by [32]

$$S(t) = \exp\left\{ - \int_0^t h(t')\, dt' \right\}. \tag{7}$$

A larger hazard function would thus decrease the survival probability.

4.3.2 Cox proportional hazards model

Cox proportional hazards model (CPHM) is a statistical method commonly used in survival analysis to model the relationship between the survival time and one or more predictor covariates. The potential advantage of CPHM analysis as opposed to a univariate model, such as the Student's t-test, is that it takes the effect of other variables into account [33]. Such covariates in the analysis for the rectum can include the age of the patient, rectal volume etc. CPHM models the hazard function as an exponential of an arbitrary baseline hazard $h_0(t)$ when all covariates are null [34]

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^{k} \beta_k x_k \right), \tag{8}$$

where β is the regression coefficient of the covariate x. Cox per voxel survival analysis extends the CPHM to the setting where the predictor variables are images. In this approach, the CPHM model is fitted separately to each voxel, and the resulting per-voxel hazard ratios are used to create a surface map of survival risk [35].

4.3.3 Implementation of Cox per voxel

The implementation of the Cox per voxel algorithm proposed by Green et al. [35] will be utilised. This method uses a standard Newton-Raphson model to produce an estimate of the maximum likelihood of the covariates. The likelihood function is given by [36]

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp\left(\sum_{j=1}^{k} \beta_j x_j \right)}{\sum_{k \in R(t_j)} \exp \sum_{j=1}^{k} \beta_j x_j} \right)^{\delta_j}, \tag{9}$$

where δ_j is an event indicator and is equal to 1 when an event occurs and 0 if no event occurs at t_j. The denominator indicates the fraction of patients who are at risk at time t_j.

From Equation 9 the maximum likelihood function can be defined by the β coefficients that maximise $L(\beta)$ which are denoted by $\hat{\beta}$ [36]. The hazard ratio is then defined as the hazard function of patients in Group 1 relative to patients in Group 2 [37]

$$HR = \frac{h_1(t)}{h_2(t)} = \exp\left(\sum_{j=1}^{k} \hat{\beta}(\mathbf{X}_{i1} - \mathbf{X}_{i2})\right), \tag{10}$$

where $\mathbf{X}_{i1}$, $\mathbf{X}_{i2}$ are the sets of covariates for Groups 1 and 2 respectively [37]. A hazard ratio of greater than 1 suggests that a per-unit increase in the covariate has a detrimental effect on the outcome. A hazard ratio of less than 1 suggests that a per unit increase in the covariate is beneficial to the outcome. It is possible to perform a univariate analysis by excluding other clinical variables and performing the analysis on contour variation alone [35].
The output of this method is a multi-channel image and each channel contains the hazard ratio of the observed data for a given covariate. The statistical significance of the hazard ratio per voxel can be tested using a permutation test of the toxicity values, survival time and other variables with respect to the contour variation distribution. The variables are not permuted with each other. Similar to the Student's t-test, the threshold hazard ratio values per voxel for a permutation can be compared to the observed distribution to obtain the confidence level of statistical significance [35].

5 Results

5.1 Mean contour variation

The average radial difference per voxel was calculated and plotted in a surface map as shown in Figure 6. The map shows some variation in the mean radial differences, however, it is centred around the value $\mu = -0.3$ mm, which is very small. This result is to be expected since the radial differences should not follow a specific trend, and should thus be centred around a value close to 0. There are two clusters in the map for which the absolute radial difference is significantly larger and surpasses 1 mm.

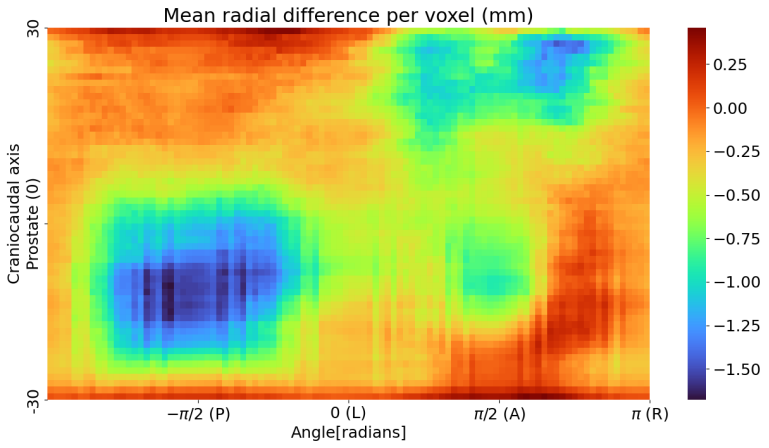


Figure 6: Surface map showing the average radial difference at every pixel for 60 slices and 100 angles. The values of the average radial differences are centred around $\mu = -0.3$ mm. There are two clusters in the lower posterior and upper anterior for which $|\Delta r| \geq 1$ mm.

It was also found that there were 257 patients for which there was at least one voxel with $|\Delta r| > 3$cm. This suggests the existence of contours with very large deviations. The causes and effects of those contours will be investigated in Section 6.5.

5.2 Analysis results

5.2.1 Student's t-test results

When performing a two-tailed Student's t-test analysis on contour variation, no significance was found for any rectal side effect. However, when the same analysis was performed using the dose at the manual rectum, significance was found for proctitis and bowel urgency. The data for proctitis were recorded according to the CTCAE scale whereas the data for bowel urgency were patient-reported. There were 185 event cases for the side effect of proctitis and 857 event cases for bowel urgency. The 95% significance region for both side effects included the lower posterior. For bowel urgency, small regions of the upper and lower anterior as well as the upper posterior demonstrated significance.

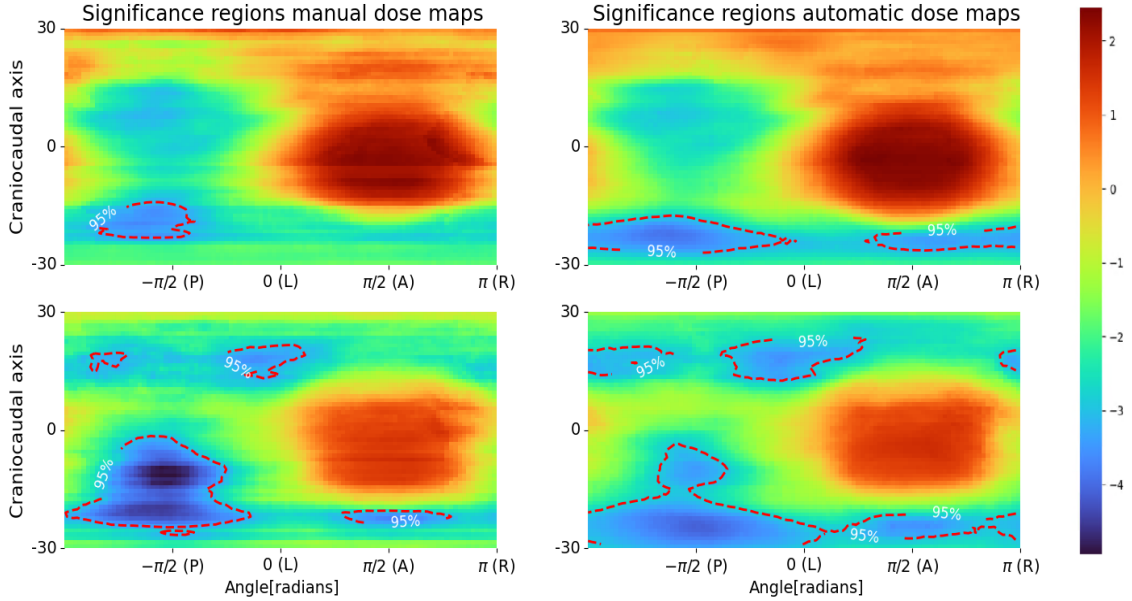


Figure 7: Four surface maps showing the output distribution of $\tilde{T}_k$ values from Student's t-test, with regions of 95% significance outlined. The top two maps illustrate the significant regions for proctitis, comparing the dose at the manual rectum (left) versus the automatic rectum (right). The bottom two maps show the same comparison for bowel urgency. Results suggest that reducing the dose in the significant regions could be beneficial for the outcome.

The same analysis was performed using the dose at the automatic rectum contour. Again, significance was found for the same two side effects. The significance regions produced using the automatic rectum were generally more extended, but still included the lower posterior. Figure 7 shows the regions of significance for proctitis and bowel urgency using the dose at the manual versus the dose at the automatic rectum contour for comparison.
For all the maps shown in Figure 7 the significance region corresponded to the lower tail of the Student's t-test, suggesting that less dose in the region would have a positive impact on toxicity levels for those side effects.

5.2.2 CPHM results

For the same side effects, a univariate CPHM analysis was performed, consisting only of the time to event, the event (0 or 1) and the dose data for every patient. Other covariates that were tested included the age of the patient, the volume of the manual rectum contour and the toxicity values for the side effect before treatment (baseline value). However, due to technical issues with the in-house toolkit, CPHM analysis could not be performed for those covariates properly. Therefore this report will focus on the results of the univariate analysis.

Results from the univariate CPHM analysis did not support the significant regions found in the Student's t-test for the dose at the manual rectum contour. However, when using the dose of the automatic rectum contour, significance was found in the lower posterior for proctitis. The hazard ratio map for proctitis is shown in Figure 8. The values of the hazard ratio at the significant region are larger than 1, suggesting that reducing the dose at the region would be beneficial for the outcome. Specifically, $1.03 \leq HR \leq 1.034$, so a per Gy increase in dose in the region can increase the risk of proctitis by 3-3.4%.
Finally, CPHM was performed using the contour variation and product maps and no significance was found.

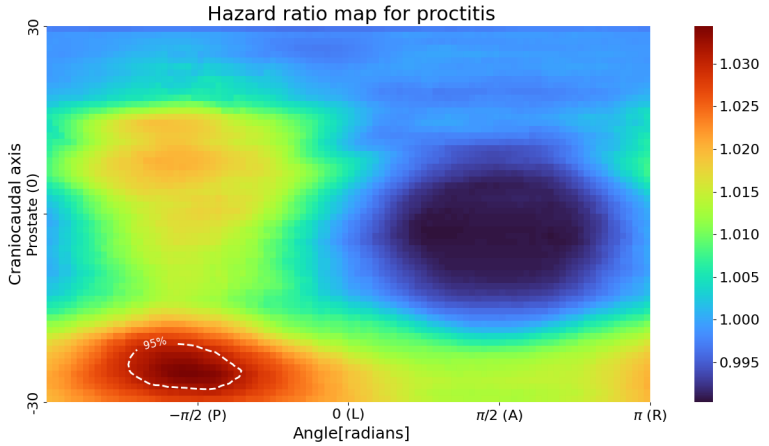


Figure 8: Hazard ratio map from univariate CPHM analysis performed for proctitis using the dose at the automatic rectum. The region of 95% significance is outlined. The map suggests that a per Gy increase in the region can increase risk by 3-3.4%.

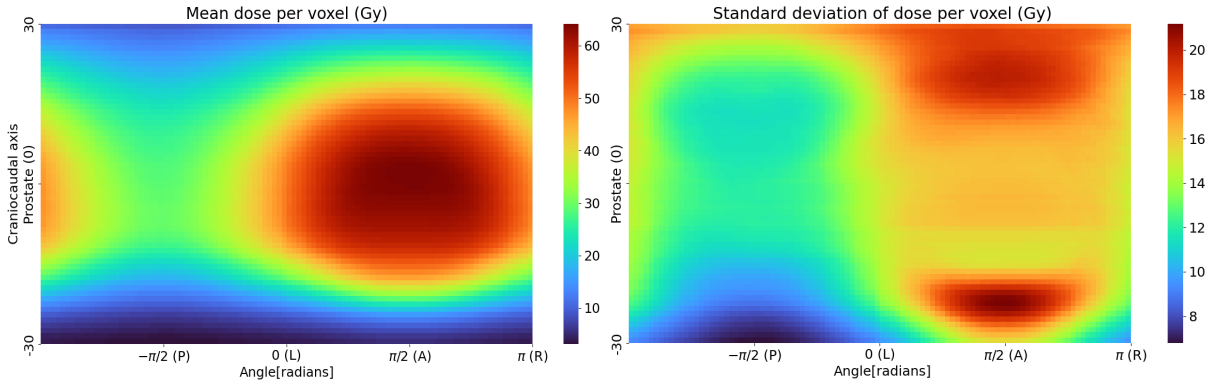
6 Discussion

6.1 Comparison to distance map method

Bouzaki and Green used the method of distance mapping in 2022 [6] to quantify the contour variations between the manual and automatic rectum contours. Distance maps, however, create a field containing information about the 3D distance to a masked object. In the method introduced in this report, the radial differences correspond to the 2D variations between the two contours.
A potential advantage of the method introduced is that it can potentially reduce straightness bias. When using the distance map method, the closest distance in 3D would be chosen as the contour variation between the two contours, which may not be entirely accurate in regions where the rectum is curved, since the distance in the craniocaudal axis may be sampled instead. Another advantage of the 2D variation method is that it is more clinically relevant, as it provides clinicians with an exact region and direction in which contouring variations can impact toxicity levels.
When comparing the output maps for the two methods for a random sample of 15 patients, the overall structure of the maps remained the same, with the same regions being prominent. This acted as a quality assurance that the radial variation maps were implemented correctly. There was a small level of disagreement in the values of the two maps, which was expected.

6.2 Regions of significance

As seen in the average contour variation map of Figure 6, there are certain regions where the average Δr is not close to 0, signifying uncertainty in the contouring variation, perhaps due to increased curvature or the existence of NaN values. One of those regions is the lower posterior which is also a region where significance was found when using the dose distributions. This coincidence poses the question of whether the significant regions are simply an artefact of the Student's t-test.
To test this, a map of the standard deviation of the dose at every voxel of the manual rectum was plotted. Student's t-test is dependent on the standard deviation of the samples and such a test could reveal whether the significance is driven by the existence of outliers. As shown in Figure 9, the regions at which there was a large discrepancy in the planned dose were the upper and lower rectum anterior. The region at which significance was found for proctitis and bowel urgency is located mostly in the rectal posterior, where the dose standard deviation is significantly lower. A low standard deviation indicates that the observed difference between the event and non-event groups is unlikely to have occurred by outliers.
For bowel urgency, there was a small region of significance in the lower anterior which is a region of high standard deviation. This suggested that there may be other underlying reasons or potential outliers causing significance in those regions, rather than the existence of an underlying effect.



(a) (b)

Figure 9: (a) Surface map showing the mean dose at the manual rectum for each voxel (b) Surface map showing the standard deviation of the dose at the manual for each voxel. The highest standard deviation is observed on the upper and lower region of the rectum anterior. The mean dose per voxel and standard deviation per voxel for the dose at the automatic rectum had almost identical structures and values to the maps presented in this figure.

Multiple studies have shown a correlation between the dose delivered at the rectum and late toxicity. [38], [39] performed a dose-volume histogram analysis and showed that rectal complications increase as irradiated volume increases. A recent study by Onjukka *et al.* [40] has shown a correlation between higher doses at the posterior rectal wall and rectal bleeding using a voxel-wise analysis. A very important study by Shelley *et al.* [41] investigated the relationship between voxel-level accumulated dose and rectal toxicity and found regions of significance in the lower posterior area for rectal bleeding and proctitis. The preliminary results of the study presented in this report, therefore, act as a validation of the results found by Shelley *et al.* [41] using a different dataset and methodology.

6.3 Comparison to dose at the automatic rectum and CPHM

There are certain limitations when using the dose at the manual rectum for the analysis. The manual rectum contours are inconsistent and strongly influenced by hospital practices. It was found that more than 80% of the patients in the dataset had a smaller manual than automatic rectum contour in the craniocaudal axis. The method of standardising the rectum into 60 slices assumed that rectal anatomy and structure relative to the prostate are independent of the size of the contour, which may not be entirely accurate for the manual rectum contours. This suggests that the method introduced may be oversampling noise in the data and reducing the accuracy of the mapping, which can affect the results. Furthermore, the analysis may be limited by the existence of NaN values in the maps for the manual rectum, especially in the lower and upper rectum regions, where more voxels with NaN values are expected.
For those reasons, the dose at the automatic rectum contour was also tested for comparison. Using those maps eliminates the uncertainty and oversampling of noise since the software is consistent when contouring the rectal anatomy. It also reduces the appearance of NaN values, resulting in more voxels considered in the analysis.
The results of the Student's t-test indicate that lowering the dose in the rectum lower posterior can lead to better outcomes for both manual and automatic rectal doses. However, the regions of significance for proctitis and bowel urgency using the dose at the automatic rectum are generally larger, suggesting that reducing noise and the appearance of NaN values can lead to slightly different outcomes.
The significance regions found by the Student's t-test based on the dose at the manual rectum contour could not be supported by the results of the univariate CPHM analysis. This could be for a number of reasons. Firstly, the two statistical tests are fundamentally different in nature. Student's t-test is testing the difference in the mean between two independent samples. CPHM, on the other hand, is testing differences in hazard rates between groups and is a time-to-event analysis. It is possible that the differences detected by the Student's t-test were not large enough to affect the hazard ratios. A different explanation would be that the significant regions found with the Student's t-test could be due to other clinical factors not accounted for in the analysis. Interestingly, the significance region for proctitis in the lower posterior was validated when performing CPHM analysis on the dose at the automatic rectum. Therefore, the lack of significant CPHM results for the dose at the manual rectum may be correlated to the existence of NaN values and noise.

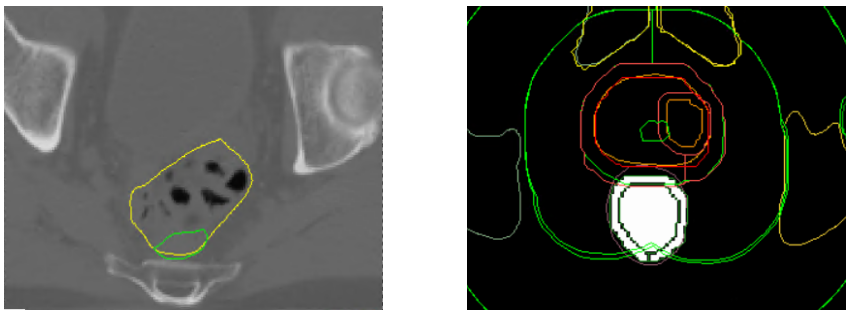
6.4 Rectal volume and motion

Sripadam *et al.* [42] showed that rectal volume can decrease during treatment and that rectal regions corresponding to the prostate base displayed the greatest motion. Van Herk *et al.* [43] has shown that rectal filling is the biggest contributor to prostate motion. Stasi *et al.* [44] found that rectal volume variation between planning and treatment is greater for full rectums. These studies suggest that there is a discrepancy between planned and delivered doses due to rectal motion, which is an important limitation of this study.
This effect can potentially be accounted for in the analysis by including the volume of the rectum as a covariate in the CPHM analysis. Another suggestion would be to normalise the contour variations by dividing each map by the average radius of the rectum for each patient.

6.5 Erroneous patients

There were 257 patients with at least one voxel for which the radial difference was $|\Delta r| \geq$ 3cm. These cases existed due to two primary reasons. Firstly, some of the manual rectum delineations were erroneous due to techniques used by clinicians to create them. This led to large deviations from the automatic contour, especially on the slices of the rectum contours that are most extreme in the craniocaudal axis. Another reason for the large deviations was

that some clinicians contoured the outer and inner rectum walls in a single contour. This made it extremely challenging for the Worldmatch software to accurately mask the contour which affected the sampling of Δr. Examples of these cases are shown in Figure 10.



(a) (b)

Figure 10: (a) shows an example of large discrepancies in the automatic and manual rectum contours. The automatic contour is delineated in yellow and the manual contour is delineated in green. (b) shows an example of a mask of a rectum contour which has the inner and outer rectal walls delineated. The two figures correspond to different patients and the slices shown are in the transverse plane.

A separate analysis was performed with the erroneous patients removed and there was no difference in the final results. There was still significance in the same regions for both side effects and therefore it was decided for the analysis to proceed with the complete patient cohort rather than removing patients.

6.6 Other limitations

The implementation of the linear transformations to standardise the rectum to a map of 60 slices allowed for the inclusion of more patients. This method eliminated the bias towards smaller rectums that would come from simply cropping the rectum to a desired length. Hence, the limitations of the dataset came from patients with no automatic delineations or for whom the ADMIRE algorithm had misidentified the prostate. There were also 68 patients that were excluded due to misidentification of the manual rectum contour. Although this only corresponds to 3.8% of the patients available, this effect can be reduced in the future by explicitly instructing the algorithm to reject contours such as the anus when searching for the manual rectum contours.

7 Future suggestions

7.1 CPHM covariates

This project implemented the Cox per voxel analysis for a single variable, which was either the dose, the distance or the product maps. However, the main advantage of this analysis is that it allows the inclusion of more variables. Due to technical issues with the in-house developed toolkit MTK, it was not possible for this project to perform a more extensive CPHM analysis.
Statistical techniques such as the Akaike Information Criterion (AIC) would be useful to implement in order to determine the best combination of covariates. This will also allow to eliminate covariates that are not as important to the analysis [45].
Some features that were briefly tested were the volume of the rectum, the age of the patient and the toxicity values recorded before the treatment started. Different combinations of the covariates would be required for a concrete analysis.

7.2 Random forest classifiers

As an extension to this project, it was proposed to use Random Forest Classifiers as a more advanced analysis technique. The advantage of random forests compared to permutation tests is that they are non-parametric, meaning that they do not make any assumptions about the data. This method is also robust to outliers and noise [46]. It has been shown that random forest algorithms can perform equally or even better than regression techniques [47], [48]. However, the dataset contains NaN values which posed a problem in the implementation of random forests, as most libraries cannot handle them explicitly. Learning from incomplete data is generally a challenge in machine learning [49]. A common method to deal with NaN or missing values is to completely remove patients that have NaN values, however, that would eliminate more than 80% of patients in this dataset. Imputation techniques such as mean or median imputation are also commonly used and should be considered for the continuation of this type of analysis [50].

8 Conclusion

This study was able to quantify contour variations in the rectum for 1,377 patients. A novel method was developed that standardised the rectum length and structure into 60 slices, which significantly expanded the dataset and included all regions of interest in the analysis. The Student's t-test analysis revealed a region of significance ($p < 0.05$) in the lower posterior when using either the dose at the automatic or manual rectum for proctitis and bowel urgency. A univariate Cox proportional hazards model analysis was also performed. The CPHM analysis could not support the results found from the Student's t-test for the dose at the manual rectum. However, a similar region of significance in the lower posterior was revealed when performing CPHM analysis in the dose at the automatic rectum for proctitis. The discrepancy in results between the manual and automatic rectum suggests that the inconsistency in the contouring process and the existence of NaN values may be affecting the results. No significance was found when using the contour variation data. Overall, the analysis suggests that reducing the dose at the lower posterior of the rectum could lead to decreased toxicities for the two side effects. Further investigation of the causes of significance and the relationships between covariates should be considered in the future for more concrete results. There are a number of limitations to this study with the main being the difference between the planned and delivered doses due to organ motion and the assumption that the geometry of the rectum is independent of contour size.

9 Acknowledgements

I would like to thank my supervisors, Eliana Vasquez Osorio and Alan McWilliam for providing this amazing project and for their continuous guidance and support. I would also like to thank Marcel van Herk without whom it would be impossible to bring the project to such a high level. Lastly, I would like to thank Dylan Green, my partner for the first half of the project and dear friend. His judgement and ideas accelerated this project and his presence certainly made it more fun. This project has been nothing other than inspirational and has certainly opened my eyes to new horizons and ideas. It has been a pleasure meeting the researchers at the RRR group at the Christie hospital and I hope our paths can cross in the future.

References

[1] N. C. Institute, *Seer cancer statistics review, 1975-2013*, 2016. [Online]. Available: `https://seer.cancer.gov/csr/1975_2015/`.

[2] P. J. Hoskin, A. M. Rojas, P. J. Ostler, L. Bryant, and G. J. Lowe, "Randomised trial of external-beam radiotherapy alone or with high-dose-rate brachytherapy for prostate cancer: Mature 12-year results," *Radiother Oncol*, vol. 154, pp. 214–219, Oct. 2020.

[3] C. Njeh, "Tumor delineation: The weakest link in the search for accuracy in radiotherapy.," *Journal of medical physics*, vol. 33, no. 4, 2008.

[4] L. Lin, Q. Dou, Y.-M. Jin, *et al.*, "Deep learning for automated contouring of primary tumor volumes by mri for nasopharyngeal carcinoma," *Radiology*, vol. 291, no. 3, pp. 677–686, 2019.

[5] Z. Liu, X. Liu, H. Guan, *et al.*, "Development and validation of a deep learning algorithm for auto-delineation of clinical target volume and organs at risk in cervical cancer radiotherapy," *Radiotherapy and Oncology*, vol. 153, pp. 172–179, 2020.

[6] A. Bouzaki and D. Green, "The effect of contouring uncertainty in the rectum on clinical outcome for patients treated for prostate cancer," M.S. thesis, The University of Manchester, Jan. 2023.

[7] N. Kh, N. Al-Hayali, J. Chiad, S. Nacy, and O. Hussein, "A review of passive and quasi-passive lower limb exoskeletons for gait rehabilitation," *Journal of Mechanical Engineering Research and Developments*, vol. 44, pp. 436–447, Oct. 2021.

[8] J. E. McNeal, "The zonal anatomy of the prostate," *The Prostate*, vol. 2, no. 1, pp. 35–49, 1981. DOI: `https://doi.org/10.1002/pros.2990020105`.

[9] Atlantic Health System, *Health topics a to z: Prostate cancer*, Website, 2021. [Online]. Available: `https://ssl.adam.com/content.aspx?productid=10&pid=10&gid=000033&site=atlantichealthssl.adam.com&login=ATLA1992`.

[10] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[11] National Cancer Institute, *Surveillance, epidemiology, and end results program (seer)*, Website, 2015. [Online]. Available: `https://seer.cancer.gov/archive/csr/1975_2015/`.

[12] P. Rawla, "Epidemiology of prostate cancer," *World journal of oncology*, vol. 10, no. 2, p. 63, 2019.

[13] X. Gual-Arnau, M. Ibáñez-Gual, F. Lliso, and S. Roldán, "Organ contouring for prostate cancer: Interobserver and internal organ motion variability," *Computerized Medical Imaging and Graphics*, vol. 29, no. 8, pp. 639–647, 2005, ISSN: 0895-6111. DOI: `https://doi.org/10.1016/j.compmedimag.2005.06.002`.

[14] T. Nyholm, J. Jonsson, K. Söderström, *et al.*, "Variability in prostate and seminal vesicle delineations defined on magnetic resonance images, a multi-observer, -center and -sequence study," *Radiation Oncology*, vol. 8, no. 1, p. 126, May 2013.

[15] R. Chicas-Sett, F. Celada-Alvarez, S. Roldan, *et al.*, "Interobserver variability in rectum contouring in high-dose-rate brachytherapy for prostate cancer: A multi-institutional prospective analysis," *Brachytherapy*, vol. 17, no. 1, pp. 208–213, 2018, Treatment Delivery Verification in Brachytherapy: Prospects of Technology Innovation, ISSN: 1538-4721. DOI: `https://doi.org/10.1016/j.brachy.2017.09.015`.

[16] I. White, A. Hunt, T. Bird, *et al.*, "Interobserver variability in target volume delineation for CT/MRI simulation and MRI-guided adaptive radiotherapy in rectal cancer," *Br J Radiol*, vol. 94, no. 1128, p. 20 210 350, Nov. 2021.

[17] D. Jin, D. Guo, J. Ge, X. Ye, and L. Lu, "Towards automated organs at risk and target volumes contouring: Defining precision radiation therapy in the modern era," *Journal of the National Cancer Center*, vol. 2, no. 4, pp. 306–313, 2022, ISSN: 2667-0054. DOI: `https://doi.org/10.1016/j.jncc.2022.09.003`.

[18] C. M. Washington and D. T. Leaver, *Principles and Practice of Radiation Therapy*. Elsevier Health Sciences, 2015, p. 1.

[19] K. M. Prise and J. M. O'Sullivan, "Radiation-induced bystander signalling in cancer therapy," *Nat Rev Cancer*, vol. 9, no. 5, pp. 351–360, Apr. 2009.

[20] C. Fiorino, G. Sanguineti, C. Cozzarini, *et al.*, "Rectal dose-volume constraints in high-dose radiotherapy of localized prostate cancer," *Int J Radiat Oncol Biol Phys*, vol. 57, no. 4, pp. 953–962, Nov. 2003.

[21] M. J. Zelefsky, E. J. Levin, M. Hunt, *et al.*, "Incidence of late rectal and urinary toxicities after three-dimensional conformal radiotherapy and intensity-modulated radiotherapy for localized prostate cancer," *Int J Radiat Oncol Biol Phys*, vol. 70, no. 4, pp. 1124–1129, Mar. 2008.

[22] M. H. Stenmark, A. S. Conlon, S. Johnson, *et al.*, "Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer," *Radiotherapy and Oncology*, vol. 110, no. 2, pp. 291–297, 2014.

[23] N. C. Institute, *Common terminology criteria for adverse events (ctcae)*, 2017. [Online]. Available: `https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_8.5x11.pdf`.

[24] P. Seibold, A. Webb, M. E. Aguado-Barrera, *et al.*, "REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer," *Radiother Oncol*, vol. 138, pp. 59–67, 2019.

[25] P. Mildenberger, M. Eichelberg, and E. Martin, "Introduction to the dicom standard," *European radiology*, vol. 12, pp. 920–927, 2002.

[26] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous dice coefficient: A method for evaluating probabilistic segmentations," *arXiv preprint arXiv:1906.11031*, 2019.

[27] H. N. Yarandi, "Hypothesis testing," *Clin Nurse Spec*, vol. 10, no. 4, pp. 186–188, Jul. 1996.

[28] C. Andrade, "The P value and statistical significance: Misunderstandings, explanations, challenges, and alternatives," *Indian J Psychol Med*, vol. 41, no. 3, pp. 210–215, May 2019.

[29] Student, "The probable error of a mean," *Biometrika*, vol. 6, pp. 1–25, 1908.

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

[31] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, "Survival analysis part i: Basic concepts and first analyses," *Br J Cancer*, vol. 89, no. 2, pp. 232–238, Jul. 2003.

[32] M. Symons and D. Moore, "Hazard rate ratio and prospective epidemiological studies," *Journal of clinical epidemiology*, vol. 55, no. 9, pp. 893–899, 2002.

[33] L. L. JOHNSON and J. H. SHIH, "Chapter 20 - an introduction to survival analysis," in *Principles and Practice of Clinical Research (Second Edition)*, J. I. Gallin and F. P. Ognibene, Eds., Second Edition, Burlington: Academic Press, 2007, pp. 273–282, ISBN: 978-0-12-369440-9. DOI: `https://doi.org/10.1016/B978-012369440-9/50024-4`.

[34] D. Hashim and E. Weiderpass, "Cancer survival and survivorship," in *Encyclopedia of Cancer (Third Edition)*, P. Boffetta and P. Hainaut, Eds., Third Edition, Oxford: Academic Press, 2019, pp. 250–259. DOI: `https://doi.org/10.1016/B978-0-12-801238-3.65102-4`.

[35] A. Green, E. Vasquez Osorio, M. C. Aznar, A. McWilliam, and M. van Herk, "Image based data mining using per-voxel cox regression," *Front. Oncol.*, vol. 10, p. 1178, 2020.

[36] G. Heinze and M. Schemper, "A solution to the problem of monotone likelihood in cox regression," *Biometrics*, vol. 57, no. 1, pp. 114–119, 2001. DOI: `https://doi.org/10.1111/j.0006-341X.2001.00114.x`.

[37] D. G. Kleinbaum, *Survival Analysis A Self-Learning Text, Third Edition* (Statistics for Biology and Health), 3rd ed. 2012. New York, NY: Springer New York, 2012.

[38] A. Jackson, M. W. Skwarchuk, M. J. Zelefsky, *et al.*, "Late rectal bleeding after conformal radiotherapy of prostate cancer (ii): Volume effects and dose–volume histograms," *International Journal of Radiation Oncology*Biology*Physics*, vol. 49, no. 3, pp. 685–698, 2001, ISSN: 0360-3016. DOI: `https://doi.org/10.1016/S0360-3016(00)01414-0`.

[39] E. H. Huang, A. Pollack, L. Levy, *et al.*, "Late rectal toxicity: Dose-volume effects of conformal radiotherapy for prostate cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 54, no. 5, pp. 1314–1321, 2002, ISSN: 0360-3016. DOI: https://doi.org/10.1016/S0360-3016(02)03742-2.

[40] E. Onjukka, C. Fiorino, A. Cicchetti, *et al.*, "Patterns in ano-rectal dose maps and the risk of late toxicity after prostate imrt," *Acta Oncologica*, vol. 58, no. 12, pp. 1757–1764, 2019, PMID: 31298076. DOI: 10.1080/0284186X.2019.1635267.

[41] L. E. Shelley, M. P. Sutcliffe, S. J. Thomas, *et al.*, "Associations between voxel-level accumulated dose and rectal toxicity in prostate radiotherapy," *Physics and Imaging in Radiation Oncology*, vol. 14, pp. 87–94, 2020, ISSN: 2405-6316. DOI: https://doi.org/10.1016/j.phro.2020.05.006.

[42] R. Sripadam, J. Stratford, A. M. Henry, A. Jackson, C. J. Moore, and P. Price, "Rectal motion can reduce ctv coverage and increase rectal dose during prostate radiotherapy: A daily cone-beam ct study," *Radiotherapy and Oncology*, vol. 90, no. 3, pp. 312–317, 2009, ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2008.07.031.

[43] M. van Herk, A. Bruce, A. Guus Kroes, T. Shouman, A. Touw, and J. V. Lebesque, "Quantification of organ motion during conformal radiotherapy of the prostate by three dimensional image registration," *International Journal of Radiation Oncology Biology Physics*, vol. 33, no. 5, pp. 1311–1320, 1995, Implementation of Three Dimensional Conformal Radiotherapy, ISSN: 0360-3016. DOI: https://doi.org/10.1016/0360-3016(95)00116-6.

[44] M. Stasi, F. Munoz, C. Fiorino, *et al.*, "Emptying the rectum before treatment delivery limits the variations of rectal dose–volume parameters during 3dcrt of prostate cancer," *Radiotherapy and Oncology*, vol. 80, no. 3, pp. 363–370, 2006, ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2006.08.007.

[45] E.-J. Wagenmakers and S. Farrell, "AIC model selection using akaike weights," *Psychonomic Bulletin & Review*, vol. 11, no. 1, pp. 192–196, Feb. 2004.

[46] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[47] R. Couronné, P. Probst, and A.-L. Boulesteix, "Random forest versus logistic regression: A large-scale benchmark experiment," *BMC Bioinformatics*, vol. 19, no. 1, p. 270, Jul. 2018.

[48] E. Watanabe, S. Noyama, K. Kiyono, *et al.*, "Comparison among random forest, logistic regression, and existing clinical risk scores for predicting outcomes in patients with atrial fibrillation: A report from the J-RHYTHM registry," *Clin Cardiol*, vol. 44, no. 9, pp. 1305–1315, Jul. 2021.

[49] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[50] J. W. Graham, "Missing data analysis: Making it work in the real world," *Annual review of psychology*, vol. 60, pp. 549–576, 2009.