

Assessment of delivered rectal dose in
prostate cancer radiotherapy

A thesis submitted for the degree of Doctor in Philosophy

Ashley Marigold Anne d'Aquino

Department of Radiotherapy and Imaging

The Institute of Cancer Research and The Royal Marsden

University of London

Declaration

The work presented in this thesis is entirely my own, except where otherwise stated.

Signed: AAV Date: 01/02/21

Acknowledgements

I am very grateful to many people for their support and assistance in the completion of this thesis. I would like to dedicate it in loving memory to my father Alan.

First and foremost, I cannot fully express my gratitude to the unwavering support of my supervisors Professor Dearnaley, Dr Harris, Dr Gulliford and Dr McNair. Their encouragement and dedication has always been an inspiration. I am also extremely grateful to Professor Jeff Bamber, and my mentors Sarah Armstrong and Dr Cynthia Eccles for their guidance and calming advice. Special thanks must also go to the Radiotherapy Physics department, especially Dr Alex Dunlop, Dr Dualta McQuaid and Dr Ian Hanson for always making the time to patiently discuss my work with me. I would also like to thank my colleagues in Radiotherapy, especially the Superintendents whose support was instrumental in the completion of this thesis.

I would like to take this opportunity to thank my parents, Caroline and Peter, and my siblings James and Philippa and her partner Michael, who have always been there for me. I am grateful for their tireless encouragement and belief in my ability to complete what felt like a “never-ending story”. I am grateful to Steve, who joined me late in this journey, but whose support and love I couldn’t have crossed the finish line without. To my friends and urban family, thank you for cheering me on!

I gratefully acknowledge Prostate Cancer UK for funding my PhD studentship, and for their supporters who continue to raise much needed funds for research to improve prostate cancer treatments and care. Finally, I am very thankful to the patients whose data was used in this study. Their contribution is invaluable in improving care for future patients.

Abstract

Prostate radiotherapy gives excellent long-term disease control; however morbidity from treatment can negatively impact quality of life. To minimise the risk of toxicity a treatment plan is created, which meets dose volume constraints for organs at risk (OAR) such as the rectum. These constraints have been informed by dose response studies which have attempted to establish a relationship between dosimetric parameters and the incidence of toxicity. The position and shape of the surrounding OAR vary over the course of treatment, and as a consequence, the treatment plan may not represent the actual delivered dose. Precise knowledge of the relationship between dose to the OAR and toxicity, essential for improving the radiotherapeutic index, may be limited if derived using the treatment plan. The aim of this PhD was to determine if the accumulated dose could be used to more accurately predict toxicity than using the planned dose.

Delivered dose to the rectum was calculated using daily cone beam computed tomography (CBCT) scans to describe the patient anatomy at each treatment fraction in 86 patients. Deformable image registration was used to establish spatial correspondence between the CBCTs, enabling the daily dose to be accumulated. Dose was also accumulated using dose surface maps which described the spatial distribution of dose to the rectal wall. Differences between planned and accumulated dose were compared and correlated with toxicity data, collected using established toxicity scoring measures.

This study quantified uncertainties in the dose accumulation methodology including the effect of missing data. It was found that the delivered dose to the rectum accumulated using a subset of CBCTs gave a comparable result to the dose accumulated using a CBCT from all fractions. Comparison of both planned and accumulated rectal dose-volume parameters and spatial metrics showed that planned dose is a good approximation of accumulated rectal dose when state-of-the-art daily image guidance is used to correct for geometric uncertainty in prostate position. The small differences between planned and accumulated dose meant neither was found to be a better predictor of toxicity. Although not statistically significant, the results suggest that dose surface metrics may have had a stronger association with late toxicity than dose-volume metrics.

Table of Contents

Chapter 1	<i>Introduction and Background</i>	19
1.1	Introduction	19
1.2	Radiotherapy	19
1.2.1	Radiotherapy treatment planning	20
1.2.2	Prostate radiotherapy	21
1.2.3	Toxicity.....	22
1.2.4	Dose volume constraints.....	24
1.3	Organ motion and correction strategies	28
1.3.1	Bowel preparation	29
1.3.2	Image guided radiotherapy.....	30
1.4	Studies calculating delivered fractional dose to the rectum	32
1.5	Dose accumulation	33
1.5.1	Studies accumulating dose to the rectum	34
1.5.2	Deformable image registration	35
1.5.3	Dose surface maps	38
1.6	Aim and organisation of thesis	40
1.6.1	Aim.....	40
1.6.2	Organisation of thesis	41
1.6.3	Summary of Chapters	41
Chapter 2	<i>Methodology</i>	44
2.1	Introduction	44
2.2	The Dose Escalation to Intra-prostatic tumour Nodules in localisEd prostate cancer (DELINEATE) trial	44
2.2.1	DELINEATE inclusion and exclusion criteria	45
2.2.2	Radiotherapy planning and treatment	46
2.2.3	Treatment verification using CBCT.....	49
2.2.4	Toxicity data collected	49
2.2.5	PhD patient population and governance	51
2.2.6	Intra-prostatic tumour nodules	53
2.3	Dose accumulation	53
2.3.1	Calculation of daily delivered dose	54
2.3.2	Delineation of the rectum and regions of interest	58
2.3.3	Deformable image registration	61
2.3.4	Dose surface maps	64

2.4	Methods for exploring the dose response	67
2.4.1	Univariate analyses	67
2.4.2	Receiver operating characteristic curves	67
2.4.3	Atlas of complication incidence	68
2.5	Statistical analysis.....	69
Chapter 3	<i>Evaluation of observer variation in delineation of rectal contours</i>	71
3.1	Introduction.....	71
3.2	Aims	74
3.3	Methods and Materials.....	75
3.3.1	Patient data.....	75
3.3.2	Segmentation.....	75
3.3.3	Measuring contour variation and evaluating automatic segmentation models.....	76
3.3.4	Measuring the effect of contour variation on deformable image registration and dose accumulation.....	76
3.4	Statistical analysis.....	79
3.4.1	Analysis of contour variation and automatic segmentation models	79
3.4.2	Analysis of contour variation on dose accumulation	79
3.5	Results.....	80
3.5.1	Comparison of manual contours.....	80
3.5.2	Comparison of rectal contours generated by automatic segmentation models	83
3.5.3	The effect of interobserver variation on deformable image registration and calculation of dose	84
3.6	Discussion.....	86
3.7	Limitations of this study.....	91
3.8	Conclusion	91
Chapter 4	<i>The effect of missing data on the calculation of accumulated dose</i>	92
4.1	Background and Introduction.....	92
4.2	Aim	94
4.3	Methodology	94
4.3.1	Summary of missing imaging data	94
4.3.2	Simulation of missing inferior rectum on CBCTs.....	95

4.3.3	Evaluation of the difference in accumulated dose calculated using full and reduced imaging datasets	97
4.4	Statistical Analysis	99
4.4.1	Summary of missing imaging data	99
4.4.2	Simulation of the missing inferior rectum on CBCTs.....	100
4.4.3	Evaluation of the difference in accumulated dose calculated using full and reduced imaging datasets	100
4.5	Results.....	101
4.5.1	Summary of missing imaging data	101
4.5.2	Evaluation on the effect of simulated inferior rectum on delivered dose.....	103
4.5.3	The effect of missing CBCTs on accumulated dose.....	106
4.6	Discussion.....	110
4.6.1	Missing images.....	110
4.6.2	Assessment of correction strategy for missing rectal images.....	112
4.6.3	Assessment of impact of incomplete data sets on dose accumulation	113
4.7	Limitations of this study.....	115
4.8	Conclusion	115
Chapter 5	<i>Difference in rectal dose-volume between planned and delivered dose</i>	116
5.1	Introduction.....	116
5.2	Aims	120
5.3	Methods and materials	121
5.3.1	Rectal volume variation and comparison of different enema regimes in the DELINEATE trial	122
5.3.2	Dose-volume constraints	122
5.3.3	Difference between planned and accumulated dosimetry for combined cohorts....	122
5.4	Statistical analysis.....	123
5.4.1	Rectal volume variation and comparison of different enema regimes in the DELINEATE trial	123
5.4.2	Dose-volume constraints	123
5.4.3	Difference between planned and accumulated dose for combined cohorts	124
5.5	Results.....	124
5.5.1	Rectal volume variation and comparison of different enema regimes in the DELINEATE trial	125
5.5.2	Dose constraint analysis	129

5.5.3	Difference between planned and accumulated dose for combined cohorts	134
5.6	Discussion	137
5.6.1	Rectal volume variation and comparison of different enema regimes in the DELINEATE trial	137
5.6.2	Dose-volume constraints	139
5.6.3	Difference between planned and accumulated dose for combined cohorts	141
5.7	Conclusion	142
Chapter 6	<i>Association between planned and accumulated dose-volume with toxicity</i>	143
6.1	Introduction.....	143
6.2	Aims	143
6.3	Methods and materials	143
6.4	Statistical analysis.....	143
6.4.1	Patient and clinical data.....	143
6.4.2	Association of dose and toxicity	144
6.5	Results.....	145
6.5.1	Patient and clinical data.....	145
6.5.2	Association of dose and toxicity	149
6.6	Discussion	159
6.7	Conclusion	165
Chapter 7	<i>Association between planned and accumulated spatial characteristics with toxicity.....</i>	166
7.1	Introduction.....	166
7.2	Aim	171
7.3	Methods and materials	171
7.4	Statistical Analysis	173
7.4.1	Clinical characteristics and incidence of toxicity.....	173
7.4.2	Difference between planned and accumulated spatial characteristics	173
7.4.3	Association of spatial characteristics and toxicity.....	173
7.5	Results.....	174
7.5.1	Clinical characteristics and incidence of toxicity.....	174
7.5.2	Difference between planned and accumulated spatial characteristics	179

7.5.3	Association of spatial metrics and toxicity.....	181
7.6	Discussion	189
7.6.1	Difference between planned and accumulated spatial characteristics	189
7.6.2	Association of spatial metrics and toxicity.....	190
7.6.3	Spatial dose metrics as predictors of toxicity.	191
7.7	Conclusion	193
Chapter 8	<i>Discussion and final conclusion</i>	<i>194</i>
8.1	Geometric uncertainties in the dose accumulation methodology.....	194
8.1.1	Delineation of the rectum.....	194
8.1.2	Missing data.....	196
8.2	Difference between planned and delivered dose	197
8.3	Association of dose and toxicity.....	199
8.4	Summary of future work and recommendations.....	201
8.5	Final conclusion	202
	References	203
	Appendices	217

List of tables

TABLE 2.1 TABLE OF THE PLANNING TARGET VOLUME (PTV) AND ASSOCIATED CLINICAL OR GROSS TUMOUR VOLUME, PTV MARGIN AND DOSE PRESCRIBED TO EACH VOLUME FOR BOTH THE STANDARD AND HYPOFRACTIONATED COHORTS.....	48
TABLE 2.2 THE NORMAL TISSUE DOSE-VOLUME CONSTRAINTS USED FOR EACH COHORT IN THE DELINEATE TRIAL	49
TABLE 2.3 DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PATIENTS INCLUDED IN THIS STUDY.	52
TABLE 4.1 THE MEDIAN (IQR) FOR THE DSC AND MDA VALES FOR EACH PATIENT, AND FOR ALL PATIENTS COMBINED.....	104
TABLE 5.1 CLINICAL CHARACTERISTICS AND DEMOGRAPHICS OF THE PATIENTS ANALYSED (N=86).	125
TABLE 5.2 MEDIAN (IQR) PLANNING RECTAL VOLUME AND THE MEDIAN (IQR) RECTAL VOLUME ON A WEEK BY WEEK BASIS FOR EACH COHORT.....	127
TABLE 5.3 THE NUMBER OF FRACTIONS WHERE THE ACCUMULATED RECTAL DOSE-VOLUME EXCEEDED THE REQUIRED DVC AND OPTIMAL DVC.	131
TABLE 5.4 TABLE SHOWING THE NUMBER OF FRACTIONS WHERE THE DAILY RECTAL DOSE-VOLUME EXCEEDED THE REQUIRED DVC AND THE OPTIMAL DVC.	131
TABLE 6.1 THE FREQUENCY OF RECORDED PEAK TOXICITY (%) (N = 86) IS GIVEN FOR ALL ENDPOINTS FOR BOTH GRADE 1 ⁺ AND GRADE 2 ⁺ TOXICITY.	148
TABLE 6.2 TABLE PRESENTING THOSE ENDPOINTS AND DOSE LEVELS WHICH WERE STATISTICALLY SIGNIFICANT (PRIOR TO CORRECTION FOR MULTIPLE TESTING) USING ROC ANALYSIS FOR BOTH PLANNED AND ACCUMULATED DOSE-VOLUME.	155
TABLE 7.1 TABLE OF STUDIES WHICH HAVE REPORTED STRONG CORRELATIONS AT GIVEN DOSES BETWEEN SPATIAL METRICS WITH TOXICITY.	170
TABLE 7.2 COMBINED CLINICAL CHARACTERISTICS FOR THE 80 PATIENTS ANALYSED.....	175
TABLE 7.3 TABLE OF TOXICITY (%) (N = 80) IS GIVEN FOR ALL ENDPOINTS FOR BOTH GRADE 1 ⁺ AND GRADE 2 ⁺ TOXICITY.....	178
TABLE 7.4 TABLE OF THE P VALUES CALCULATED USING MWU TO TEST FOR SIGNIFICANCE BETWEEN EACH OF THE SPATIAL CHARACTERISTICS AND GRADE 1 ⁺ TOXICITY FOR BOTH PLANNED AND ACCUMULATED DSM (N=80).....	183
TABLE 7.5 TABLE OF THE P VALUES CALCULATED USING MWU TO TEST FOR SIGNIFICANCE BETWEEN EACH OF THE SPATIAL CHARACTERISTICS AND GRADE 2 ⁺ TOXICITY FOR BOTH PLANNED AND ACCUMULATED DSM (N=80).....	184
TABLE 7.6 THE AUC, P VALUE AND 95 % CONFIDENCE INTERVALS ARE PRESENTED FOR GRADE 1 ⁺ TOXICITY ENDPOINTS FOR PIXEL COUNT BOTH PLANNED AND ACCUMULATED DOSE.....	185

TABLE 7.7 THE AUC, P VALUE AND 95 % CONFIDENCE INTERVALS ARE PRESENTED FOR TOXICITY ENDPOINTS FOR LONGITUDINAL EXTENT FOR BOTH PLANNED AND ACCUMULATED DOSE. .	186
TABLE 7.8 THE AUC, P VALUE AND 95 % CONFIDENCE INTERVALS ARE PRESENTED FOR GRADE 2+ TOXICITY ENDPOINTS FOR PIXEL COUNT BOTH PLANNED AND ACCUMULATED DOSE.	187
TABLE 7.9 THE AUC, P VALUE AND 95 % CONFIDENCE INTERVALS ARE PRESENTED FOR GRADE 2+ ENDPOINTS FOR LONGITUDINAL EXTENT FOR BOTH PLANNED AND ACCUMULATED DOSE. .	188
TABLE 7.10 THE AUC, P VALUE AND 95 % CONFIDENCE INTERVALS ARE PRESENTED FOR GRADE 2+ ENDPOINTS FOR LATERAL EXTENT FOR BOTH PLANNED AND ACCUMULATED DOSE.	188

List of figures

FIGURE 1.1 HISTOGRAM OF MAXIMAL TOXICITY GRADE IN RT01 TRIAL.....	26
FIGURE 1.2 AXIAL SLICE OF PELVIS AT CT AND THE FIRST TWO FRACTIONS OF TREATMENT .	27
FIGURE 1.3 HISTOGRAM OF PLANNED, DELIVERED AND CUMULATED BLADDER DVHS.	33
FIGURE 1.4 SCHEMATIC SHOWING THE DOSE TO THE SURFACE OF THE RECTUM WHICH IS DEFINED BY THE RECTAL CONTOUR.....	38
FIGURE 1.5 SCHEMATIC REPRESENTATION OF RECTAL FILLING CAUSING DEFORMATION OF THE RECTUM.	39
FIGURE 1.6 SUMMARY OF AIM AND OBJECTIVES ADDRESSED WITHIN THIS THESIS	41
FIGURE 2.1 SCHEMATIC OF DELINEATE COHORTS.....	45
FIGURE 2.2 OVERVIEW OF THE DOSE ACCUMULATION METHODOLOGY USING DEFORMABLE IMAGE REGISTRATION (DIR) AND DOSE SURFACE MAPS (DSM).....	54
FIGURE 2.3 ILLUSTRATION OF A CBCT GREY LEVEL VALUES TO MASS DENSITY CORRECTION.....	56
FIGURE 2.4 SAGITTAL IMAGE OF PELVIS SHOWING CT AND CBCT FIELD OF VIEW.	57
FIGURE 2.5 AXIAL SLICE OF PELVIS CBCT. THE WINDOWING ON THE CBCTS HAVE BEEN ADJUSTED TO HELP DEFINE THE CBCT FOV.....	58
FIGURE 2.6 SAGITTAL SLICE OF PELVIS. THE IMAGE SHOWS HOW THE RECTUM (YELLOW) EXTENDS PAST THE CBCT FOV.	59
FIGURE 2.7 SAGITTAL SLICE OF A FUSED CT AND CBCT PELVIS, SHOWING A RECTAL VOLUME BEFORE AND AFTER THE CBCT HAS BEEN DEFORMED TO THE CT.....	62
FIGURE 2.8 SCHEMATIC SHOWING THE PROCESS OF ACCUMULATING DOSE USING DIR.....	63
FIGURE 2.9 SCHEMATIC OF A SET OF BINARY IMAGES FROM A DSM.	65
FIGURE 2.10 SCHEMATIC SHOWING THE PROCESS OF ACCUMULATING DOSE USING DSM. THE INITIAL PROCESS (DENOTED BY GREY BOXES) IS ALSO USED IN DIR BASED DOSE ACCUMULATION.	66
FIGURE 2.11 ACI SHOWING THE RELATIONSHIP BETWEEN THE DOSE-VOLUME TO THE PENILE BULB AND ERECTILE POTENCY (COURTESY OF MURRAY ET AL., [18]).....	69
FIGURE 3.1 SCHEMATIC OF THE METHODOLOGY USED TO ACHIEVE EACH OBJECTIVE.....	78
FIGURE 3.2 BOXPLOT OF RECTAL VOLUME DELINEATED BY EACH OBSERVER.....	80
FIGURE 3.3 BOXPLOT OF DICE SIMILARITY COEFFICIENT (DSC) BETWEEN CONTOURS DELINEATED BY PAIRS OF OBSERVERS.	81
FIGURE 3.4 BOXPLOT OF MEAN DISTANCE TO AGREEMENT (CM) BETWEEN CONTOURS DELINEATED BY PAIRS OF OBSERVERS.	82
FIGURE 3.5 DICE SIMILARITY COEFFICIENT (DSC) FROM PAIRWISE COMPARISONS BETWEEN ADAS MANUALLY DELINEATED RECTAL CONTOURS AND EXPERT CONTOURS, AND CONTOURS AUTOMATICALLY GENERATED AND EXPERT CONTOURS.	83

FIGURE 3.6 BOXPLOT OF MEAN DISTANCE TO AGREEMENT (MDA) MEASUREMENTS FROM PAIRWISE COMPARISON BETWEEN EXPERTS AND ADA (ADA), MANUALLY DELINEATED EXPERT CONTOURS AND CONTOURS AUTOMATICALLY GENERATED CONTOUR. RECTAL CONTOURS WERE AUTOMATICALLY GENERATED USING ATLAS (ATLAS), MODEL BASED SEGMENTATION (MBS) AND DEFORMABLE IMAGE REGISTRATION (DIR).....	84
FIGURE 3.7 DICE SIMILARITY COEFFICIENT (DSC) FOR EACH CBCT FROM PAIRWISE COMPARISON BETWEEN RECTAL VOLUME SEGMENTED ON THE CT SCAN AND THE DEFORMED RECTAL VOLUME CREATED WHEN EACH OBSERVER CONTOUR WAS USED TO DEFORM THE CBCT.....	85
FIGURE 3.8 MEAN DISTANCE TO AGREEMENT (MDA) FOR EACH CBCT FROM PAIRWISE COMPARISON BETWEEN RECTAL VOLUME SEGMENTED FROM THE CT SCAN AND THE DEFORMED RECTAL VOLUME CREATED WHEN EACH OBSERVER CONTOUR WAS USED TO DEFORM THE CBCT.....	85
FIGURE 3.9 BOXPLOT OF PERCENTAGE DIFFERENCE IN DOSE-VOLUME BETWEEN EXP1 AND REMAINING OBSERVERS.	86
FIGURE 4.1 AN AXIAL SLICE OF A CBCT OF THE PELVIS WHICH WAS EXCLUDED DUE TO POOR IMAGE QUALITY AS A RESULT OF INCOMPLETE ACQUISITION.	93
FIGURE 4.2 SAGITTAL CBCT SLICE OF PELVIS WITH A FOV (DEFINED BY THE RED LINES) THAT DID NOT FULLY ENCOMPASS THE FULL RECTAL LENGTH.....	95
FIGURE 4.3 SAGITTAL SLICE OF A PLANNING CT IMAGE REGISTERED TO ONE OF THE PATIENT'S CBCT IMAGES.....	96
FIGURE 4.4 SCHEMATIC SHOWING THE NUMBER AND WEIGHTINGS OF DELIVERED DOSES USED TO CALCULATE THE ACCUMULATED DOSE FOR FULL AND REDUCED DATASETS USING DIR. [10]..	98
FIGURE 4.5 STACKED HISTOGRAM SHOWING FOR EACH PATIENT THE PERCENT OF CBCTS MISSING BY CAUSE. EACH PATIENT HAS AT LEAST 50% OF THE TREATMENT CBCTS	102
FIGURE 4.6 BOXPLOT OF THE ORIGINAL AND SIMULATED RECTAL VOLUMES FOR EACH OF THE 5 PATIENTS ANALYSED.	103
FIGURE 4.7 COMPARISON OF THE ACCUMULATED DVHS FOR THE PLANNED, ORIGINAL AND SIMULATED RECTAL VOLUMES.	105
FIGURE 4.8 BOXPLOT SHOWING THE DIFFERENCE IN RECTAL VOLUME BETWEEN THE PLANNED AND ACCUMULATED DOSE USING A FULL DATASET (LABELLED PLAN), AND THE ACCUMULATED DOSE USING THE FULL DATASET AND REDUCED IMAGING DATASETS FOR 5 PATIENTS.	106
FIGURE 4.9 PLOTS FOR PATIENT 10 PROVIDED AS AN EXAMPLE WHICH SHOWS THE MEAN PERCENTAGE DIFFERENCE IN PIXEL COUNT, LONGITUDINAL EXTENT AND LATERAL EXTENT.	108
FIGURE 4.10 BOXPLOT OF THE DIFFERENCE BETWEEN THE LATERAL EXTENT OF ACCUMULATED DSM AND DATASETS WHICH HAS BEEN REDUCED BY 10 DSM (MINUS 10) FOR ALL PATIENTS.....	110
FIGURE 5.1 MEDIAN (MARKERS) AND IQR (SHOWN BY THE ERROR BARS) OF THE RECTAL VOLUME DURING THE COURSE OF TREATMENT.....	126
FIGURE 5.2 SCATTER PLOT ILLUSTRATING THE CORRELATION BETWEEN THE PLANNED AND MEDIAN TREATMENT RECTAL VOLUME. THE LINE OF BEST FIT, PEARSON CORRELATION COEFFICIENT, R, AND P VALUE ARE GIVEN.	128

FIGURE 5.3 SCATTER PLOT ILLUSTRATING THE CORRELATION BETWEEN THE PLANNED RECTAL VOLUME AND IQR OF THE TREATMENT RECTAL VOLUME.	128
FIGURE 5.4 THE AVERAGE PLANNED DVH (GREY) AND AVERAGED ACCUMULATED DVH (GREEN) FOR THE STANDARD COHORT. THE STANDARD DEVIATIONS IS GIVEN AS AN ERROR BAR. THE REQUIRED CONSTRAINTS (BLUE), AND OPTIMAL DOSE-VOLUME CONSTRAINTS (RED) ARE ALSO PLOTTED.	130
FIGURE 5.5 THE AVERAGE PLANNED (GREY) AND AVERAGE ACCUMULATED DVHS (GREEN) FOR PATIENTS IN THE HYPOFRACTIONATED COHORT (NO EQD2 CORRECTION).	130
FIGURE 5.6 A BOXPLOT OF THE DIFFERENCE IN RECTAL VOLUME BETWEEN THE OPTIMAL PLANNING CONSTRAINT AND THE PLANNED ACCUMULATED AND AVERAGED DVH FOR STANDARD COHORT IS SHOWN.	132
FIGURE 5.7 DIFFERENCE IN RECTAL DOSE BETWEEN THE OPTIMAL CONSTRAINT VOLUME AND THE PLANNED, ACCUMULATED AND AVERAGED DVH FOR THE HYPOFRACTIONATED COHORT. ..	133
FIGURE 5.8 BOXPLOT OF THE DIFFERENCE IN RECTAL VOLUME (%) BETWEEN PLANNED AND ACCUMULATED DOSE FOR 10 GY DOSE INTERVALS.	134
FIGURE 5.9 SCATTER PLOTS ILLUSTRATING THE CORRELATION BETWEEN PLANNED AND ACCUMULATED RECTAL VOLUME AT 10 GY INTERVALS.	136
FIGURE 6.1 HISTOGRAM OF PATIENTS EXCLUDED AT BASELINE FOR EACH ENDPOINT.	146
FIGURE 6.2 THE INCIDENCE OF PEAK TOXICITY FOR EACH ENDPOINT WHEN THE COHORTS ARE COMBINED (N=86).	147
FIGURE 6.3 THE AVERAGE PLANNED (GREY) AND AVERAGE ACCUMULATED (RED) DVH FOR PATIENTS EXPRESSING GRADE 1+PROCTITIS (RTOG) IS PLOTTED. THE STANDARD DEVIATION IS GIVEN AS AN ERROR BAR.	150
FIGURE 6.4 THE AVERAGE PLANNED (GREY) AND ACCUMULATED (RED) DVH FOR PATIENTS WHO DID NOT EXPRESS GRADE 1+ PROCTITIS (RTOG).	150
FIGURE 6.5 THE AVERAGE PLANNED (GREY) AND ACCUMULATED (RED) DVH FOR PATIENTS WHO EXPRESSED GRADE 2+ BOWEL FREQUENCY (RTOG).	151
FIGURE 6.6 THE AVERAGED PLANNED (GREY) AND ACCUMULATED (RED) DVH FOR PATIENTS WHO DID NOT EXPRESS GRADE 2+ BOWEL FREQUENCY (RTOG).	151
FIGURE 6.7 THE AVERAGED PLANNED (GREY) AND ACCUMULATED (RED) DVH FOR PATIENTS WHO EXPRESSED GRADE 2+ DIARRHOEA (CTCAE).	152
FIGURE 6.8 THE AVERAGED PLANNED (GREY) AND ACCUMULATED (RED) DVH GIVEN FOR PATIENTS THAT DID NOT EXPRESS GRADE 2+ DIARRHOEA.	152
FIGURE 6.9 COMPARISON OF PLANNED AND ACCUMULATED CURVES FOR ROC CURVES WHICH HAD $AUC \geq .06$ AND SIGNIFICANT P VALUES (UNCORRECTED FOR MULTIPLE TESTING).	154
FIGURE 6.10 ACI FOR RTOG GRADE 1+ PROCTITIS FOR PLANNED (A.) AND ACCUMULATED (B.).	156
FIGURE 6.11 ACI FOR RMH GRADE 2+ BOWEL FREQUENCY FOR PLANNED (A.) AND ACCUMULATED (B.).	157
FIGURE 6.12 ACI FOR CTCAE GRADE 2+ DIARRHOEA FOR PLANNED (A.) AND ACCUMULATED (B.)...	158

FIGURE 6.13 AVERAGE PLANNING DVH FOR ALL PATIENTS INCLUDED IN RTO1 (74 GY/3 7#) , CHHIP (74 GY/37 # ARM), AND DELINEATE (EQD2 CORRECTED) PATIENTS ANALYSED IN THIS STUDY (HYPOFRACTIONATED ARM EQD2 CORRECTED).	164
FIGURE 7.1 FIGURE SHOWING THE AVERAGE DSM FOR PATIENTS WITH RTOG GRADE 2+ TOXICITY (A.) AND NO TOXICITY (B.). THE MAPS WERE GENERATED FROM 49 PATIENTS TREATED WITH IMRT TO A DOSE OF 81 GY.	168
FIGURE 7.2 HISTOGRAM OF PATIENTS EXCLUDED AT BASELINE FOR EACH ENDPOINT.....	176
FIGURE 7.3 HISTOGRAM SHOWING THE INCIDENCE OF PEAK TOXICITY AT TWO YEARS FOR THE PATIENTS INCLUDED IN THIS DSM STUDY WITH COHORTS COMBINED.	177
FIGURE 7.4 BOXPLOT OF THE DIFFERENCE (%) BETWEEN THE PLANNED PIXEL COUNT AND THE ACCUMULATED PIXEL COUNT AT 10 GY INTERVALS (10 GY TO 70 GY) FOR THE COMBINED COHORTS (N=80).	179
FIGURE 7.5 BOXPLOT OF THE DIFFERENCE IN THE LONGITUDINAL EXTENT OF THE PLANNED DSM AND THE ACCUMULATED DSM AT 10 GY INTERVALS (10-70 GY) FOR COMBINED COHORTS (N=80).....	180
FIGURE 7.6 BOXPLOT OF THE DIFFERENCE IN THE LATERAL EXTENT OF THE PLANNED DSM AND ACCUMULATED DSM AT 10 GY INTERVALS (10 TO70 GY) FOR COMBINED COHORTS (N=80). 181	

Abbreviations

3D CRT – Three dimensional conformal radiotherapy

AAPM – The American Association of Physicists in Medicine

ANACONDA – Anatomically constrained deformation algorithm

BRFS – Biochemical relapse free survival

CBCT – Cone beam computer tomography

CHHiP - The Conventional or Hypofractionated High-dose Intensity- modulated Radiotherapy for Prostate Cancer

CT – Computed tomography

CTCAE – Common Terminology Criteria for Adverse Events

CTV – Clinical target volume

DELINEATE - Dose EscalaLation to Intra-prostatic tumour Nodules in localised prostATE cancer

DIR – Deformable image registration

DRE – Digital rectal exam

DSC – Dice similarity coefficient

DSH – Dose surface histograms

DSM – Dose surface maps

DVC – Dose volume constraints

DVF – Displacement vector fields

DVH – Dose volume histograms (unless stated these are cumulative)

DWH – Dose wall histograms

EBRT – External beam radiotherapy

EPIC-26 – Extended Prostate Cancer Index Composite-26

FLAME – Focal Lesion Ablative Microboost in Prostate Cancer

FOV Field of view

GI – Gastrointestinal

GTV – Gross tumour volume

GU - Genitourinary

HIFU – High-intensity focused ultrasound

IBDQ – Inflammatory Bowel Disease Questionnaire

IGRT – Image guided radiotherapy

IMRT – Intensity modulated radiotherapy

IMRT – Intensity modulated radiotherapy

IPL – Intra-prostatic lesion

LENT SOMA – Late Effect on Normal Tissue: Subjective, Objective, Management and Analytical scale

MDA – Mean distance to agreement

MRC RT01 – Medical Research Council RT01 trial

MRI – Magnetic resonance imaging

MVCT – Megavoltage computer tomography

MWU – Mann Whitney U

NCCN – National Comprehensive Cancer Network

OAR – Organs at risk

PRO – Patient reported putcomes

ProtecT - Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer

PSA – Prostate-specific antigen

PTV – Planning target volume

QoL – Quality of life

QUANTEC – Quantitive Analyses of Normal Tissue effects in Clinic

RMH – Royal Marsden NHS Trust

RADAR – Randomised Androgen Deprivation and Radiotherapy trial

SBRT – Stereotactic Body Radiotherapy

SIB – Simulated integrated boost

SVs – Seminal vesicles

TPS – Treatment planning system

XVI – X-ray volumetric imaging

Chapter 1 Introduction and Background

1.1 Introduction

Prostate cancer is the most common cancer in the UK, with more than 47,500 men diagnosed each year [1]. One in eight men will be diagnosed with prostate cancer in their lifetime, with over 400,000 men living with and after prostate cancer in the UK [1, 2]. The incidence of prostate cancer in the UK has increased by 44% since the 1990s [2]. The introduction of prostate-specific antigen (PSA) screening has resulted in men being diagnosed earlier [3]. As a result more men are being diagnosed with localised prostate cancer and are being treated with curative intent [4]. Localised prostate cancer is treated based on risk group, with men stratified based on clinical staging using pre-treatment PSA, digital rectal exam (DRE), Gleason score on biopsy, and the number of involved biopsy cores [3, 5]. Men are stratified into low, intermediate (intermediate and unfavourable high risk) and high risk (high risk and very high risk) [6]. Treatment options include but are not limited to one or a combination of watchful waiting, active surveillance, radical prostatectomy, radiation (brachytherapy or external beam therapy), androgen deprivation therapy, high-intensity focused ultrasound (HIFU) and cryotherapy [1, 3, 5, 7]. Currently 30% of men diagnosed with prostate cancer will have radiotherapy as part of their primary treatment [2].

1.2 Radiotherapy

External beam radiotherapy (EBRT) with androgen deprivation therapy is an established treatment for men with intermediate or high risk disease [8], and it is estimated that almost 16,000 men a year will receive radical radiotherapy in the UK [2]. Advances in radiotherapy treatment planning and delivery techniques have resulted in better local control, and a decrease in toxicity [6, 9-11]. Further dose escalation may improve biochemical relapse free survival (BRFS) [12-14]. However, the total radiation dose delivered to the prostate is constrained by the radiation tolerances of the surrounding organs at risk (OAR), such as the bladder, rectum and penile bulb [6, 15, 16]. Erectile dysfunction is one of the most common reported toxicities related to radiotherapy [17], and is associated with dose to the penile bulb [18, 19]. Dose to the bladder can result in complaints of nocturia, frequency, urgency

and hesitancy [19, 20]. Dose to the rectum can include rectal bleeding, urgency, bowel frequency, rectal pain and faecal incontinence [21-24].

The work in this thesis focuses on better quantifying the delivered dose to the rectum using daily treatment images. Association of delivered dose with toxicity may improve our understanding of the dose response relationship of the rectum, enabling greater dose escalation to the prostate which may improve BRFS.

1.2.1 Radiotherapy treatment planning

Prior to the commencement of radiotherapy, careful planning is required in order to optimise the treatment. A computed tomography (CT) scan of the patient provides the position of the target volume and the surrounding structures, which allows a treatment plan to be created with the aim of delivering a tumoricidal dose to the target volume, whilst minimising dose to the normal tissues. The target volume consists of the tumour and a margin of tissue to account for uncertainties in planning and delivery, and includes:

(i) The gross tumour volume (GTV), defined as “the gross demonstrable extent and location of the tumour” [25].

(ii) The clinical target volume (CTV) encompasses the GTV with a margin of tissue to include subclinical disease.

(iii) The planning target volume (PTV), an expansion of the CTV with a margin to account for uncertainties in patient positioning and organ motion [25, 26].

Depending on the tumour location, the PTV margin may partially encompass surrounding critical structures such as the rectum. The size of the margins used, and subsequent volume of normal tissue included, is influenced by factors such as organ motion, reproducibility in patient positioning, treatment technique and verification protocol. The planning CT scan also provides electron density values which are used to calculate the dose [27]. One of the key methods to limit dose to the rectum is to optimise the treatment plan using dose volume constraints. Dose volume constraints (DVC) play pivotal role in reducing the risk of rectal toxicity by ensuring that dose to the rectum remains within prescribed limits [28].

These constraints have been informed by dose response studies, which have modelled the relationship between the dose received by OAR, and the severity and duration of the toxicity experienced by the patient. Dose parameters used in these studies are often derived from the treatment planning CT in the form of Dose Volume Histograms (DVH). An acknowledged limitation of dose response studies is that the dose volume parameters are derived from the planning CT only, and therefore do not account for variation in the shape and volume of the rectum during the treatment course [29, 30]. Rectal deformation during treatment affects the dose delivered, with some studies showing large differences in delivered rectal dose from planned as a result [31, 32]. In order to improve the therapeutic ratio and safely escalate dose to the prostate, a better understanding of the relationship between dose and GI toxicity is required. It is possible that this can be achieved by calculating the delivered dose to the rectum using treatment images to calculate the daily delivered and accumulate the total delivered dose. By associating the accumulated dose with toxicity, improved motion inclusive dose constraints can be derived to further optimise planning.

1.2.2 Prostate radiotherapy

Studies suggest that giving a higher dose of radiation to the prostate could increase BRFS [3][33-35]. However, despite technical advances in treatment delivery, the rectum is one of the principal dose limiting structures [12, 36, 37]. Increasing dose to the prostate may lead to increased toxicity which is especially of concern for patients with long term survival [38]. One solution is to deliver a focal boost to the dominant lesion within the prostate. Local recurrence usually occurs at the site of the dominant lesion [39-42]. It is hypothesised that delivering a standard dose to the prostate, while simultaneously delivering a dose escalated boost to the dominant lesion, may achieve greater tumour control without increasing toxicity [39, 43, 44]. This targeted approach is currently being tested in a phase II trial, Dose Escalation to Intra-prostatic tumour Nodules in Localised ProstATE cancer (DELINEATE) (Rec No 11/L0/0510) at the Royal Marsden Hospital (RMH). The aim of the trial is to assess whether the toxicity experienced by the patients is within safe limits [42]. Recent publication of the preliminary 1 year toxicity data of the first 105 patients in the DELINEATE trial showed low incidence Grade 2 or worse gastrointestinal (GI) toxicity was 3.6% in the standard fractionation cohort and 8% in hypofractionated cohort [42]. A similar trial, FLAME (Focal Lesion Ablative Microboost in Prostate Cancer), showed that there was no significant difference in GI toxicity between the standard dose arm of 77 Gy/35

(late cumulative GI toxicity of 11.1%) and the arm with the same prescription plus a concomitant focal boost of 95 Gy (10.2%) [43].

Conventionally, radiotherapy has been delivered in fractions of 1.8 – 2.0 Gy to doses of 74.0 – 79.2 Gy over a period of 7 to 8 weeks [45]. There is evidence that prostate cancer has a lower alpha-beta ratio than previously thought, suggesting an increased sensitivity to fraction size [38, 46-48]. This suggests that higher daily doses of radiation can be given with a reduced number of treatment fractions over shorter time periods with similar outcomes. Recent publication of randomised trials evaluating moderate hypofractionation (≥ 3 Gy) have strengthened the evidence base for this [38, 49-51]. The largest of these trials is the CHHiP trial (Conventional or Hypofractionated High-dose intensity- modulated radiotherapy for Prostate cancer), which is the largest randomised trial to have compared conventional and hypofractionated dose prescriptions. Results showed that a prescription of 60 Gy/20 # was non-inferior to 74 Gy/37 # in biochemical or clinical failure-free rate, with no significant differences in RTOG Grade ≥ 2 toxicity (13.7% in the 74 Gy/37 # arm and 11.9% in the 60 Gy/20 # arm) [52]. The hypofractionation prescription of 60 Gy/20 # is now considered the standard of care for localised prostate cancer at RMH and internationally, and is one of the dose prescriptions used within the DELINEATE trial.

1.2.3 Toxicity

Radiotherapy to the prostate can result in genitourinary (GU) toxicity, GI toxicity and sexual dysfunction. With over 400,000 men a year living with and after prostate cancer in the UK, it is important to minimise the effect of toxicity on Quality of Life (QoL). A systematic review by Lardas et al., compared the incidence of toxicity across four different treatment interventions which were EBRT, brachytherapy, radical prostatectomy and active surveillance [53]. The authors reported that EBRT had a more pronounced negative impact on bowel function compared to surgery but a lesser impact on bladder function and erectile dysfunction than prostatectomy [53]. Subsequently the key UK randomised controlled trial ProtecT (Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer), which compared surgery, radiotherapy and active monitoring confirmed these findings demonstrating again that radiotherapy had less impact on urinary incontinence and erectile dysfunction than surgery with a small increase in rectal bleeding [54, 55]. The

rectum is a dose limiting structure, and lowering GI toxicity can influence the dose which can be safely delivered to the prostate [21].

Acute GI toxicity is usually defined as symptoms which occur up to three months post radiotherapy, and include diarrhoea, urgency and mucus discharge. Late side effects are defined as those that are observed at six months post radiotherapy or later, and include rectal bleeding, urgency, bowel frequency, rectal pain and faecal incontinence [21-24]. Previously, because it is a more objective measure, rectal bleeding was routinely used by studies as the main GI toxicity endpoint [30, 56]. However, other side effects also impact on patient QoL, and as a result it is recommended that scoring systems should be used which capture different symptoms and reflect the total impact on QoL [9, 21, 23]. Scoring systems such as The Common Terminology Criteria for Adverse Events (CTCAE) [57], and Late Effect on Normal Tissue: Subjective, Objective, Management, and Analytical scale (LENT SOMA) [58] are now commonly utilised allowing comparison of efficacy and adverse events across trials. These systems are quantitative and reflect different symptoms experienced by the patient [22]. However, although scoring systems such as LENT SOMA better reflect the effect of toxicity on QoL, the use of patient reported outcomes (PRO) is also recommended [29]. PRO have been shown to detect side effects more reliably than clinician reported outcomes [59]. Lower GI radiation induced toxicity can present as symptoms which are similar to other bowel disorders. As a result QoL measurements tools used by gastroenterologists specialising in pelvic radiotherapy side effects are often used. These include modified Inflammatory Bowel Disease Questionnaire (IBDQ), which is a QoL measurement tool [60] and the Vaizey questionnaire [61]. The Extended Prostate Cancer Index Composite-26 (EPIC-26) is another common PRO tool which also measures adverse effects such as urinary incontinence, urinary irritation/obstruction, bowel, sexual and vitality/hormonal function [62].

Interpreting the incidence of radiation induced sequelae is complicated. Pre-existing conditions, such as irritable bowel syndrome and haemorrhoids, are reportedly associated with complication risk [9]. Patient factors such as advanced age, smoking, abdominal surgery and diabetes have also been reported to increase risk of toxicity [63-67]. Studies have also shown that different pathophysiology may arise from specific anorectal subsites [30, 68]. For instance, faecal leakage has been reported as being more strongly correlated with dose to the anal sphincter than the anorectal wall

[69]. Analysis of data from the Randomised Androgen Deprivation and Radiotherapy (RADAR) trial showed the impact of dose on different anatomical regions of the rectum by creating DVH for the anorectum, rectum and anal canal. Association with toxicity showed that for stool frequency and urgency, significant dose-volume effects are dominant in the anorectum and anal canal rather than the superior rectum [68]. A similar study by Wilkins et al, associated the planning dose volume parameters for the anorectum, rectum and anal canal with reported toxicity. The study reported different strengths of association between the dose volume parameters for distinct subsites and different symptoms. For instance, the anal canal had a weaker relationship for dose volume and faecal incontinence than the rectum or anorectum [70]. This suggests that in order to more clearly understand the mechanism of injury it may be important to analyse the spatial distributions of dose in different anatomical regions of the organ.

1.2.4 Dose volume constraints

Radiation induced GI toxicity cannot be completely avoided, however by keeping the dose to the rectum within acceptable limits, it is possible to safely escalate dose to the prostate. These limits (or constraints) have been defined by dose response studies which have evaluated the risk of GI toxicity as function of the dose distribution. By associating toxicity with dosimetry, a dose volume cutpoint can be calculated which discriminates between patients who exhibit toxicity from those who do not. From these data, prognostic dosimetric constraints can be derived which can be used to optimise treatment planning, and reduce the risk of toxicity [68, 71, 72].

Dose volume parameters are usually summarised as DVH [68]. DVH are used to extract data from the plan, graphically summarising 3-Dimensional (3D) dosimetry into a 2-Dimensional (2D) plot, by describing the amount of dose received by a specified volume [71, 73]. In the clinical setting the most common rectal dose constraints are related to percentage volume (V_d) receiving more than a threshold dose, d . In 2010, the Quantitative Analyses of Normal Tissue Effects in Clinic (QUANTEC) built on the previous seminal Emani paper [9, 74], and published a collection of recommended dose volume constraints for a number of sites including the rectum. The recommended dose constraints for reducing rectal toxicity are $V_{50}<50\%$, $V_{60}<35\%$, $V_{65}<25\%$, $V_{70}<20\%$, and $V_{75}<15\%$ [9]. The studies from which these constraints were derived used 3D Conformal Radiotherapy (3D CRT). Subsequent use of IMRT, characterised by steep dose gradients and the ability to deliver more conformal and convex dose

distributions, will have resulted in the ability to achieve tighter dose constraints. Daily use of IGRT has also enabled the margin between the CTV and PTV to be reduced, decreasing the percentage of rectum in the high dose region without compromising the coverage of the prostate [10]

It should also be noted that it is only since 2006 that researchers have begun to report on endpoints other than rectal bleeding [30]. Subsequently, endpoints such as stool frequency and faecal incontinence have been investigated. Although occurring less frequently (in approximately less than 5% of patients [75]), they are often of a chronic nature which can negatively impact a patient's quality of life [30]. The pathophysiology of the rectal toxicity is complex, with several factors affecting the continence mechanism. These include injury to the rectal mucosal, impairment of rectal capacity, anal sphincter dysfunction and stool frequency [76, 77]. As a result different rectal morbidity may be the result of dose to different subsites of the rectum, such as the anal wall and the pelvic floor muscles [68, 78, 79]. In addition research has shown that different endpoints are associated with different doses, further complicating the derivation and application of dose volume constraints. For instance, bowel frequency is associated with low to mid dose range and faecal incontinence is associated with a low dose range [70]. Correlation between rectal injury and dosimetry can also be confounded by patient factors such as discussed previously in Section 1.2.3.

Nevertheless Fiorino et al., states that at present DVHs are, to date, the most reliable method of predicting toxicities such as late rectal bleeding [65]. This is supported in a study by Gulliford et al., where rectal dose volume constraints were retrospectively applied to the Medical Research Council (MRC) RT01 trial data to investigate the association with different toxicity endpoints [28, 80]. No rectal dose constraints were used in the RT01 trial other than the dose to the rectum should not exceed the randomised prescription dose [13]. The study observed that the number of patients experiencing Grade 2 toxicity increased as the number of retrospectively applied dose constraints was exceeded (see Figure 1.1).

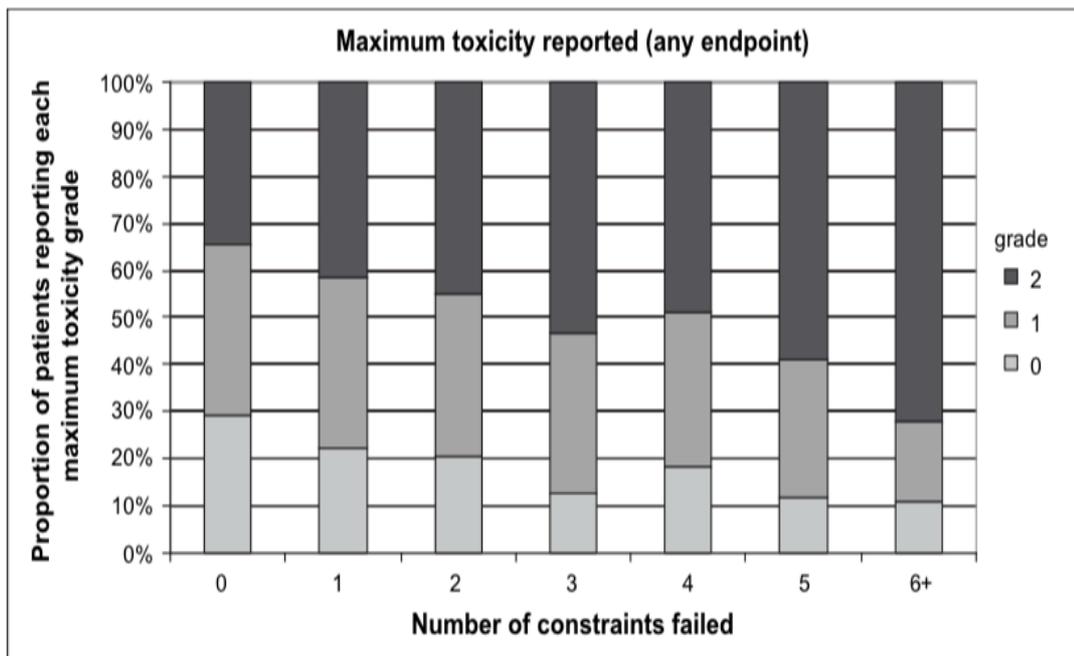


Figure 1.1 Histogram of maximal toxicity grade in RT01 trial.

The histogram shows a summary of the maximal toxicity grade reported (any endpoint) relative to the number of constraints failed from the RT01 trial. The only rectal dose constraint used in the RT01 trial was that the dose to the rectum should not exceed the dose prescribed to the prostate. Constraints were retrospectively applied to the data. Reproduced with kind permission by Dr Gulliford et al, 2010 [28].

Gulliford et al. also observed that when all dose constraints were met in a group of patients, a third still reported toxicity as Grade 1 or Grade 2. A possible reason for the failure of this dose volume data to consistently predict toxicity is that the DVH is based on a snapshot of the patient’s anatomy at the time of the planning CT scan. The CT does not reflect the motion and deformation of the pelvic organs over the course of treatment [31]. Rectal volume changes during treatment therefore reduce the reliability of any predictive models based on planned DVH estimates (see Figure 1.2) [71, 81]. The dose constraints can only be as accurate as the data used in the model [82].

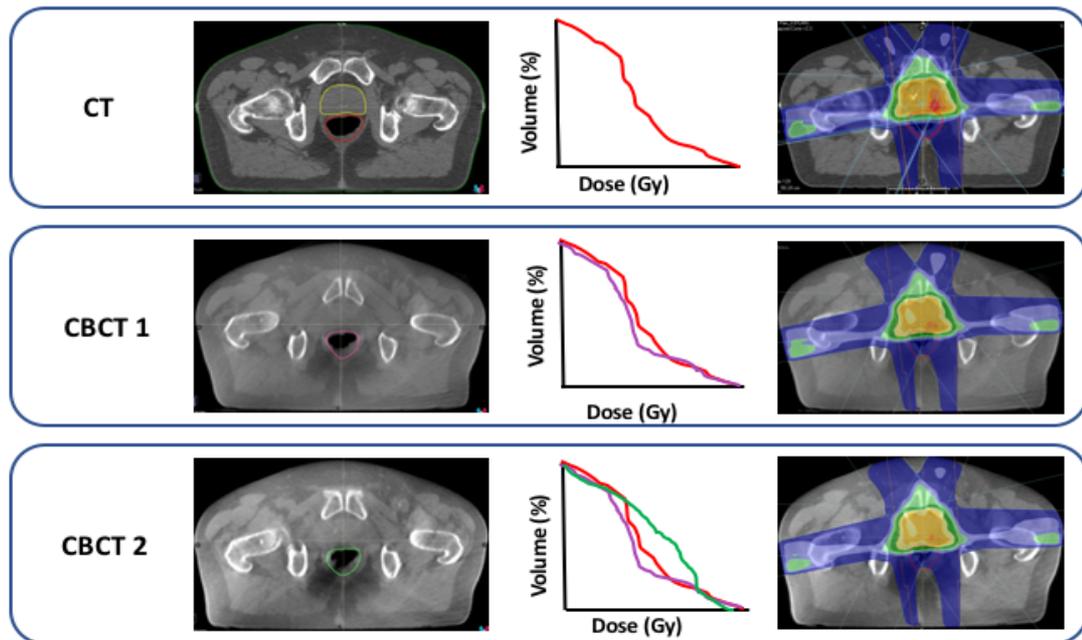


Figure 1.2 Axial slice of pelvis at CT and the first two fractions of treatment . The images show the variation in rectal volume and shape at different time points, and the subsequent variation in calculated DVH and dose distribution as a consequence of rectal deformation is also given.

The accuracy of DVH also depends on the accuracy with which the boundary of the structure was defined [73, 83]. For solid organs the DVH defines the volume encompassed by the outer contour of the organ. For hollow organs such as the rectum the area of interest is not the contents of the organ but the dose to the rectal wall [84]. Dose Surface Histograms (DSH) and Dose Wall Histograms (DWH) have both been proposed as an alternative to DVH for use in hollow structures such as the rectum and bladder [85, 86]. Dose wall histograms create a histogram based on the volume of the rectal wall which is defined by contouring both the external and internal rectal wall. The rectal wall is thin with some models specifying between 2 and 3 mm [85, 86]. Usually an automatic model is used to contract the outer delineated wall by the stated thickness to reduce uncertainty in delineation of the inner wall. A limitation of this method however is that it does not account for further thinning of the rectal wall when it is distended. This can result in thinning of the rectal wall to 1mm [87], which means that the rectal wall volume may not be conserved. Dose surface histograms are computed in a similar way to DVH, relying on the delineation of the outer rectal wall. They are used to describe the dosimetry to the surface of the rectum by binning surface elements rather than volume elements [88].

Another possible reason for the inability of DVHs to consistently predict morbidity is that they do not provide spatial information on how the dose is distributed within the organ, instead assuming every region of normal tissue is equally important [89]. Both methods for calculating dose to the rectal wall, also lack spatial information. Dose surface maps (DSM) provide spatial information about the distribution of the dose as well as specifying dose to the rectal wall surface. DSM summarise the 3D spatial dose distribution as a 2D map enabling the extent of the distribution to be analysed [90]. There are a small number of studies which have shown correlation between spatial metrics derived from DSM and toxicity [91-93]. Buettner et al., [94] created rectal DSM using the planning scan and investigated the association of the spatial distribution of dose with a number of toxicity endpoints and found there was a correlation for some endpoints. For instance, rectal bleeding correlated strongly with the lateral extent of the dose distribution, whereas loose stools correlated most strongly with longitudinal extent. Similarly, Onjukka et al., reported the lateral extent (59 Gy-71 Gy) was associated with rectal bleeding, and faecal incontinence was associated with lateral extent (5-49 Gy), and longitudinal extent (60 Gy) [95].

1.3 Organ motion and correction strategies

Improved tumour control and reduced toxicity relies on accurate delivery of the treatment plan created at pre-treatment stage. Geometric uncertainty such as setup errors and organ motion can affect the accuracy of treatment delivery [26, 96]. Changes in the dimension of the rectum affects the position of prostate, mainly in the anterior-posterior and the superior-inferior directions [97-99]. Changes which occur between treatments are defined as interfraction motion, with studies reporting interfraction motion of the prostate to be greatest in the anterior-posterior direction with motion of a maximum of 1 cm [99]. Intrafraction motion describes the deviation of the target during the treatment and is primarily a result of rectal peristalsis and filling [96, 100]. Different types of prostate intrafraction motion have been described as a result of rectal filling [101], and these fall broadly into two categories [102]. The first type is non-resolving with a slow drift of the prostate predominantly in the posterior direction. The second type of motion is sudden and transient, usually in the superior and anterior direction [102]. Intrafraction motion has been shown to be common, and can be larger than 1 cm in some cases although more commonly less than 5 mm [100, 103, 104].

A study by de Crevoisier et al., underlines the significance of rectal variation in prostate radiotherapy [87]. The study measured the diameter of the rectum on the planning CT and found that patients with a distended rectum at planning had a poorer 5 year biochemical relapse free survival compared to those patients with an empty rectum at planning (62% vs 94%). The authors proposed that as treatment progressed the rectum became smaller and the prostate partially dropped out of the high dose region, resulting in the prostate being under dosed [87]. This was supported in a study by Sripadam et al., who evaluated rectal variation during treatment by using daily Cone Beam Computer Tomography (CBCTs) to measure the rectal volume in 15 patients [105]. The study observed that the rectal volume decreased over the course of treatment for 13 patients. However, they also reported that patients who had a small rectal volume at planning, had a small proportion of treatments where the rectum was larger than planned, resulting in a greater volume of rectum in the high dose region. There are two key strategies for addressing the effect of rectal variation on the prostate. The first is to introduce consistency in rectal volume through the use of bowel preparation or a device such as a rectal balloon. The second is to correct for positional changes using Image Guided Radiotherapy (IGRT).

1.3.1 Bowel preparation

Interventional strategies are commonly employed with the aim of producing an empty rectum at planning which can be consistently replicated during treatment. Common strategies include the use of laxatives [106, 107], enemas [108] and dietary interventions such as a high fibre diet [108, 109]. A systematic review by McNair et al., evaluating different interventional strategies found that no single interventional strategy reviewed was superior or preferable, to another, in controlling rectal dimensions and associated prostate motion [97]. However, this may be because studies often investigated multiple strategies concomitantly. Alternatively, a large rectal volume can be reproduced through the use of endorectal balloons which may increase the distance between prostate and the posterior and lateral rectal wall [110]. The device ensures consistency and may reduce intrafraction motion [111, 112]. Another method is to introduce a hydrogel spacer between the prostate and the rectal wall, moving the rectal wall further from the high dose region of the prostate [113, 114]. Many of these interventions are invasive and may also incur an additional cost.

1.3.2 Image guided radiotherapy

IGRT has been instrumental in improving treatment precision [115]. It has been defined as any imaging at the pre-treatment or treatment stage which results in an action which will improve or verify the precision of radiotherapy [116]. Most often it is now used to describe techniques which image the target and normal tissues immediately prior to or during radiotherapy treatments, and which permit correction of positional changes of the target as a result of setup errors and organ motion. Consequently, the PTV margin, which is used to partly compensate for organ motion can be reduced. Previously the PTV margin for the prostate was in the order of 10 to 15 mm which resulted in a larger percentage of the rectal volume in the high dose region [117, 118]. Image guided IMRT has resulted in steeper dose gradients and optimisation of PTV margins, reducing the volume of rectum in the high dose region [119, 120]. For example the DELINEATE trial uses a simultaneous integrated boost technique with posterior margins of 6 mm/3 mm /0 mm for the three planning target volumes (see section 2.2.2 Table 2.1) [42]. However tighter margins may mean that treatment accuracy is more sensitive to setup errors and organ deformation, which consist of both systematic and random errors.

All errors have a systematic component which can be introduced at any point of the treatment pathway, but usually occur at the planning stage, for instance due to differences in setup procedures between simulation and treatment or structure delineation. Systematic errors are similar in direction and magnitude, and shift the dose envelope away from the CTV [116]. Without correction, the error will affect all treatments uniformly, and may result in a partial underdose of the target volume and overdose of surrounding critical structures. Random errors are unpredictable in direction and magnitude, and blur the dose envelope around the CTV. Random errors are caused by occurrences such as changes in patient position, patient setup errors or organ motion [116]. Measurement of the displacement of the target volume from planning to treatment, is typically based on rigid registration of images (usually a planning image and treatment image). The deviation of the target volume from the planned position can generally be corrected by applying couch shifts [119]. However, this can only correct translational displacements; it cannot correct for organ deformation. Online imaging approaches aim to correct for daily target displacement by correcting for both random and systematic translational errors [119]. Offline

approaches do not image every day but rather correct systematic errors by calculating corrective offsets for days when imaging is not performed [96, 119].

Although easy to identify on imaging, bone is a poor surrogate for prostate position since soft tissues can move independently from the bony anatomy. Set up using bony anatomy is however considerably better than skin marks alone [121]. To improve the visualisation of the prostate, gold fiducial markers are commonly implanted into the prostate [122]. Typically, two markers are placed at the posterior base of the prostate and one marker at the apex [96]. There is a wide range of imaging modalities available in clinical practice to measure interfraction variation including orthogonal megavoltage (MV) or kilovoltage (kV) images. Volumetric imaging can be acquired using CT on rails [123], ultrasound, and CBCT using kV or MV x-ray energies. Less commonly used are IGRT modalities used to measure intrafraction motion such as Calypso (Calypso Medical, USA) which uses electromagnetic transponders implanted into the prostate to track prostate position [101]. Alternatively 4-Dimensional (4D) monitoring using ultrasound can also be used which provides real time evaluation of the motion [124].

Currently the gold standard for prostate IGRT at RMH is a daily online imaging protocol, correcting for daily displacement. The CHHiP IGRT and DELINEATE trials use daily CBCT with implanted gold fiducial markers. CBCT uses a kV x-ray source and flat panel detector mounted at 90° to the gantry head. It offers volumetric imaging enabling the quantification of prostate motion, and the visualisation of interfraction soft tissue changes such as rectal filling. The image quality of CBCTs is affected by the imaging geometry, the wide beam of radiation increases scatter resulting in poor soft tissue contrast and artefacts. They do not produce true Hounsfield Units (HU) which prevent direct calculation of dose using the CBCT [125]. Strategies have been developed which enable direct dose calculation using the CBCT [126, 127], including more recently bulk density correction. This method is fast and can be easily incorporated into the clinical workflow [27, 125]. As a result, DVH of the target volume and OAR can be calculated for a given treatment. However, to calculate the total dose delivered over a course of treatment, dose volume parameters from multiple fractions should not be simply summed or averaged [128]. This is because the dose distributions within the organ will vary from fraction to fraction. Rather, dose should be accumulated by establishing spatial correspondence between images [128, 129].

1.4 Studies calculating delivered fractional dose to the rectum

There have been a small number of studies which have used volumetric treatment images to evaluate the delivered fractional dose to the rectum (i.e. the dose per treatment fraction). These studies show a large variation in the fractional rectal dose during treatment as a result of deformation, causing some authors to conclude that the planned DVH is not a good representation of delivered dose [32, 130-132]. Variation in rectal dose occurs despite online correction for geometric inaccuracy of the prostate [115, 132]. This is because, although image guidance enables improvement in treatment delivery by enabling compensation of target displacement, it does not handle organ deformation [128]. A study by Chen et al., used linear regression analysis to measure the correlation between rectal filling and dose. Unlike the bladder, where the mean dose increased as the bladder volume decreased, the author reported no correlation between volume and dose for the rectum. Chen et al., concluded that changes in rectal dose were not only due to volume changes but also shape, most notably as a result of gas bubbles [31].

Studies analysing the difference between planned and delivered dose report differences in rectal dose between 40 to 70 Gy [31, 32, 130]. Although studies reported that some patients had greater rectal doses at treatment, it was often reported that these patients also had a proportion of treatments where the rectal dose was lower than planned [31, 32, 130]. For instance, Chen et al., calculated the delivered dose for 28 patients for 4 treatments using CT on rails. In 28% of fractions, analysed the rectal DVH was greater than planned, and did not meet constraints of $V_{40} < 35\%$, and 27% of fractions did not meet $V_{65} < 17\%$. However, for 70 % of fractions analysed the rectal dose was lower than planned [31].

The quality of data for these studies was limited with some studies using only a small number of daily images to measure delivered dose [31, 32, 132]. Rectal volume changes are random and individual to the patient, and by only imaging once or twice a week the true extent of volume changes cannot be analysed [32, 131, 133]. Studies which used daily images to calculate the fractional dose showed a greater range of rectal doses, giving a better description of dose over the whole course of treatment [130, 134]. Another limitation is that rather than accumulating dose, some studies have instead summed or averaged the daily doses [31]. As previously discussed dose volume parameters from multiple fractions should not be summed as there is no spatial

correspondence between images. This is illustrated in study by Rigaud et al, who used a numerical phantom to create an accumulated bladder DVH ground truth (here they refer to accumulated as cumulated). The accumulated DVH was compared to simulated daily doses which were then summed (Fig 1.3) [129]. This study showed that summing or averaging the daily dose resulted in overestimation in the actual delivered dose.

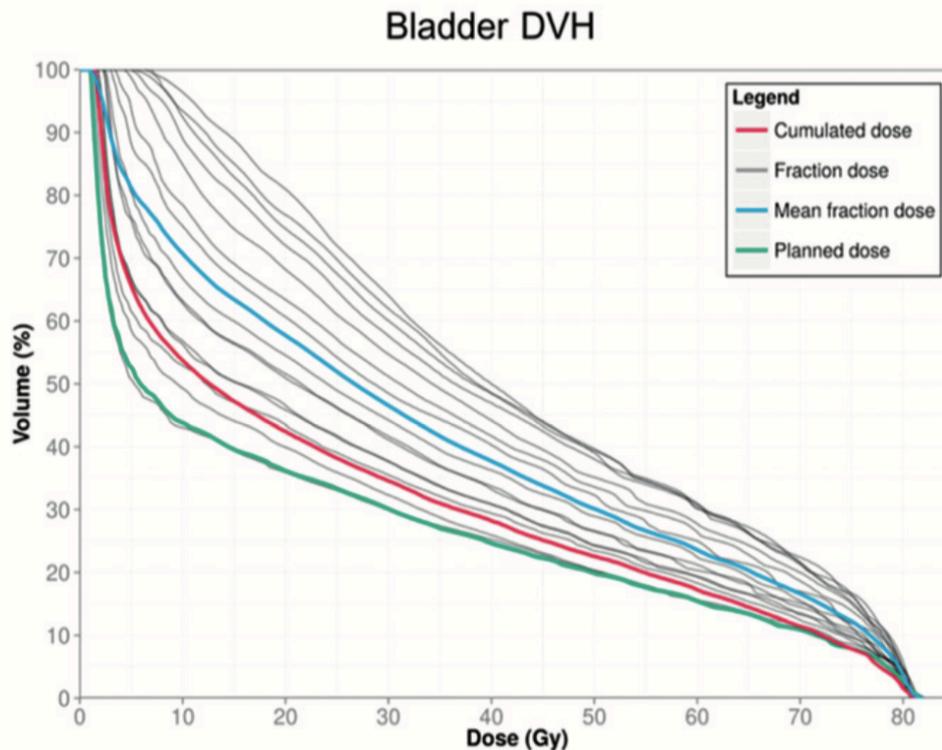


Figure 1.3 Histogram of planned, delivered and cumulated bladder DVHs. The histogram illustrates why dose should not be summed or estimated using the fractional dose in a deforming structure. The DVHs were created using a numerical phantom so the ground truth is known. The cumulated DVH (red) is the ground truth and shows that fraction or mean dose can overestimate dose volume difference [129].

1.5 Dose accumulation

Dose accumulation can be achieved by measuring differences in patient anatomy between planning and treatment by tracking changes in anatomy using treatment images. These differences can occur gradually, such as those due to weight loss or tumour shrinkage, or the changes can be rapid and transient such as those due to rectal filling [134, 135]. Measuring interfraction variation of deforming structures, and accumulating the dose, has the potential to quantify not only differences between planned and delivered dose, but to support treatment adaptation and measure treatment

response [128, 129]. Jaffray et al, describes the key elements required in order to accumulate dose [134]. Firstly, a description of the patient's anatomy at the time of treatment is required. This is provided by volumetric images, which describe the position and shape of the organs at treatment. Secondly, a method for calculating dose at those time points is required so that the daily delivered dose can be calculated. Finally, a method to establish spatial correspondence between images is required in order to accumulate dose on a voxel by voxel basis. The most commonly reported method of accumulation is Deformable Image Registration (DIR). However, DSM has also been used to accumulate dose in organs which aren't solid such as the bladder and rectum [91, 92].

1.5.1 Studies accumulating dose to the rectum

The number of dose accumulation studies is increasing and has been used in a number of sites including cervix, lung and head and neck [133, 136, 137]. However, the use in prostate radiotherapy is limited, with only a small number of studies accumulating dose to the rectum using the principles defined by Jaffray et al., [134]. In addition to evaluating planned and accumulated dose these studies have used accumulation to evaluate frequency of imaging [138], the effect of rectal balloons and spacers on toxicity [110, 139] and as a means of evaluating margin reduction [140]. One of the most promising developments is the use of dose accumulation to better understand dose response, and optimise planning and treatment delivery.

The VoxTox research group is the only group to have associated accumulated dose with reported toxicity for the rectum during prostate radiotherapy to date [92, 93]. The group accumulated dose using DSM for 109 patients receiving prostate radiotherapy using TomoTherapy (Accuray, USA) and daily megavoltage CT (MVCT). The results showed that accumulated dosimetric parameters extracted from the DSM had a stronger correlation with rectal bleeding and proctitis than with planned dose. A limitation of this study however was that the toxicity was retrospectively acquired with no baseline data. Also, an acknowledged limitation of the MVCT images used was the small field of view, which gave only partial converge of the rectal length. To overcome this problem the VoxTox group used the CT volume as a substitute for the missing data. However, this method masks the variation in the inferior and superior direction of the rectum.

1.5.2 Deformable image registration

Image registration is the process of establishing spatial correspondence between images taken at different time points [141]. Rigid registration generally only permits geometric transformation between two images using rotation or translations, and thus it is unable to account for deformation of tissue [142]. DIR establishes geometric correspondence by deforming or distorting the image grid (e.g. shrinking or stretching) [129, 143]. The resulting transformation can be used to map dosimetric or other information (e.g. contours) between images on a voxel by voxel basis. For example, it is used for multi-modal image fusion [129], atlas based segmentation [144], propagation of contours from one image to another [145], as well as for dose tracking and accumulation [138]. There are three key steps required to accumulate dose using DIR and treatment images. The first step is to calculate the daily dose using the treatment image. If the treatment image lacks true Hounsfield units, such as in the case of a CBCT, a correction must first be applied to the image to enable the dose calculation (as discussed in section 1.3.2). The image being transformed (moving image) is then deformed to the reference image (which remains static). The resulting deformation is used to deform the daily dose to the reference image. The reference image is usually the planning CT, which provides a common coordinate system to which multiple treatment images are deformed. Deformed doses can be summed to accumulate the dose.

The DIR process can require a number of iterations to obtain the optimal correspondence. There are four key components to deformable image registration algorithms. Firstly, *similarity measures* are used to measure how well the moving and reference image match [141]. A *transformation model* is used to define how the features in the moving image can be deformed relative to the reference image. The model also interpolates between features when there is no useable image information [141]. The transformation is usually presented as displacement vector fields (DVF), which are used to describe the direction and magnitude of the voxel movement in the moving image to the corresponding fixed image. Unlike the rigid registration, voxels can move with a different displacement vector than that of its neighbour [135]. It is the DVF which are used to deform the dose or propagate contours from one image to another. Transformation models fall into either parametric or non-parametric methods [129]. During DIR, the rigidity (or elasticity) of the moving image is absent, and this can result in deformed anatomy which is not physically meaningful. An example of

this is tearing or folding of anatomy to achieve correspondence [135]. A *regularisation* term can be employed to restrict the transformation and ensure that the deformation is reasonable. Finally the *optimiser* refers to the process used to vary the parameters of the transformation model in order to improve similarity between the images [141].

The choice of similarity measure is important as it guides the transformation and measures how well one image is matched to another [135]. The most commonly used measures are feature (or geometry) based methods, intensity based method, and more recently the hybrid method which combines intensity and feature [135]. Feature based models use geometric characteristics, such as points, curves or contours to guide the transformation. Often these have to be previously extracted through use of manual delineation, which can be time consuming and subject to uncertainty associated with delineation. Feature based models are better when registering multimodality imaging (e.g. Magnetic resonance imaging (MRI) and CT) however, because of the focus on enhancing the similarity between features, distortions in other areas of the image can occur [135, 143]. A type of feature based model is biomechanical deformable registration. This particular type of model uses features to guide the registration. However, the model also incorporates the mechanical properties of the tissues contoured [135]. It is commonly used where structures slide against each other, such as in the case of the lung and rib cage [135]. Intensity based models do not require prior extraction of features but rather use the greyscale information from the images to determine correspondence [141]. Intensities are matched voxel-wise which means that the image is usually matched globally (using all voxels). However, this does not guarantee that the deformation will be anatomically reasonable, especially in areas of low image contrast [146]. A hybrid algorithm utilises both feature and intensity models making it a good choice for deforming images of the pelvis. A study compared the accuracy of an intensity and a hybrid algorithm for DIR of the rectum, and showed greater concordance with the reference image when the hybrid algorithm was used [147].

Caution is recommended when using DIR as it can be subject to limitations which can introduce uncertainties that affect the validity of the registration. The accuracy of the DIR can be affected by the quality of the images used in the registration. The size of the reference image and the moving image can vary as typically is the case in the registration of the CT and CBCT [142]. The field of view of the planning CT scan can

often be longer than the CBCT. This can lead to unrealistic distortion of the anatomy in the cranio-caudal dimension as the moving image is stretched to match the reference image [142]. Another potential source of uncertainty when using clinical images, especially of the pelvis, is correspondence ambiguity. This can occur when the registration is unable to establish one-to-one physical correspondence between images. For instance, there may be a large bubble of rectal gas in one image, which is not present in the other [128]. Poor image quality (either due to artefacts, or low-dose imaging) can also lead to a difference in the fidelity of the anatomy between the images being registered, and subsequently the algorithm may struggle to establish correspondence. [135]. Differences in tissue contrast between the images can also be a source of error. Registering CT to CBCT can be an example of this due to differences in scatter and beam hardening [135].

Careful consideration of the sources of uncertainties, and methods of verification, are essential to minimise errors which can potentially influence clinical decisions, especially when accumulating dose [142, 148]. Verification of each image registration using DIR is recommended by The American Association of Physicists in Medicine (AAPM) [142]. The difficulty in verifying the DIR is that the ground truth is unknown which makes it difficult to determine the accuracy of the registration, However, deformed registrations should be evaluated to ensure they are reasonable. [135, 142]. Propagation of contours using DIR can be visually checked by the observer and edited as required. However, dose accumulation is more susceptible to uncertainties in DIR process. Small errors in registration can lead to large dose differences in areas of steep dose gradients [142, 148]. Brock et al., gives an example of a mean geometrical error of 3.85 mm for the carotid bifurcation which resulted in a mean accumulated dose error of 0.69 Gy. As with rigid image registration, visual inspection of the deformed registration is recommended with commercial systems providing tools for this to be achieved [142]. Nevertheless, because of the complexity of the registration quantitative methods of verification should also be employed.

One such method is to compare the similarity in shape between the deformed contour of a structure to the contour on the reference image. Most common metrics used to evaluate the similarity include Dice similarity coefficient (DSC) which is an overlap metric, and mean distance to agreement (MDA), which is a measure of the mean distance between contours [128]. A limitation of this method is that it only evaluates

the registration at the boundary of the ROI, not the entire volume [135]. However, it can be used easily to evaluate the accuracy of a large number of registrations [135]. The accuracy of the registration can also be measured by evaluating the alignment of corresponding points identified on each image to be registered [135, 142]. These points can either be anatomical, such as a vessel bifurcation, or implanted, such as fiducial markers in the prostate [135]. A limitation of this method is that the evaluation is not valid away from the defined points and in particular, within the rectum there is a lack of stable anatomy.

1.5.3 Dose surface maps

Dose surface maps (DSM) are generated by calculating the dose to the surface of the rectum. The rectum is then virtually unfolded by slicing along the posterior axis to project a 2D representation of the dose distribution to the rectal surface (see Fig. 1.4). This enables the distribution of dose at the surface of the rectum to be described using spatial metrics such as lateral or longitudinal extent [110, 149].

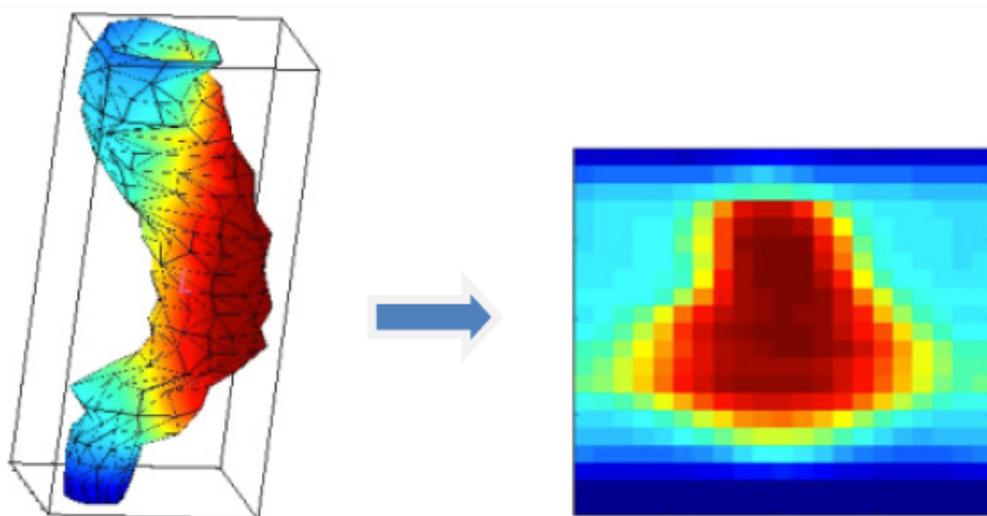


Figure 1.4 Schematic showing the dose to the surface of the rectum which is defined by the rectal contour. The contour is then virtually unfolded to create DSM (courtesy of Florian Buettner et al, 2009) [94]. The colour reflects the difference in dose across the rectal surface. The areas receiving the greatest dose are coloured red and are posterior to the prostate. The areas receiving the least dose are blue.

DSM can be used to analyse toxicity and organ motion relative to the radiation field, with studies showing that shape of delivered dose may be a better predictor of toxicity than DVH alone [84, 150]. Buettner et al., associated reported toxicity with planning DSM for 388 patients from the MRC RTO1 trial. The study showed that there was a

stronger correlation between the shape of the dose distribution and some outcomes than there was for DSH. Rectal bleeding correlated strongly with the lateral extent of the dose distribution between doses of 39 Gy and 61 Gy. And loose stools correlated strongly with the longitudinal extent of the dose distribution between 21 Gy and 33 Gy [149].

DSM can also be used to accumulate dose independent of DIR. Murray et al., proposed a method to accumulate dose by creating daily DSM using daily CBCTs [91]. The rectum was manually delineated on each CBCT and bulk density corrections applied to the CBCT so the dose to the rectum could be calculated. Daily DSM were created and then summed to create an accumulated DSM. However, due to variations in rectal height it is first necessary to normalise the DSM using interpolation [91]. It is assumed that the length of the rectum remains constant, but the height of the rectum visualised in the CBCT on different days changes because of rectal deformation (see Fig. 1.5).

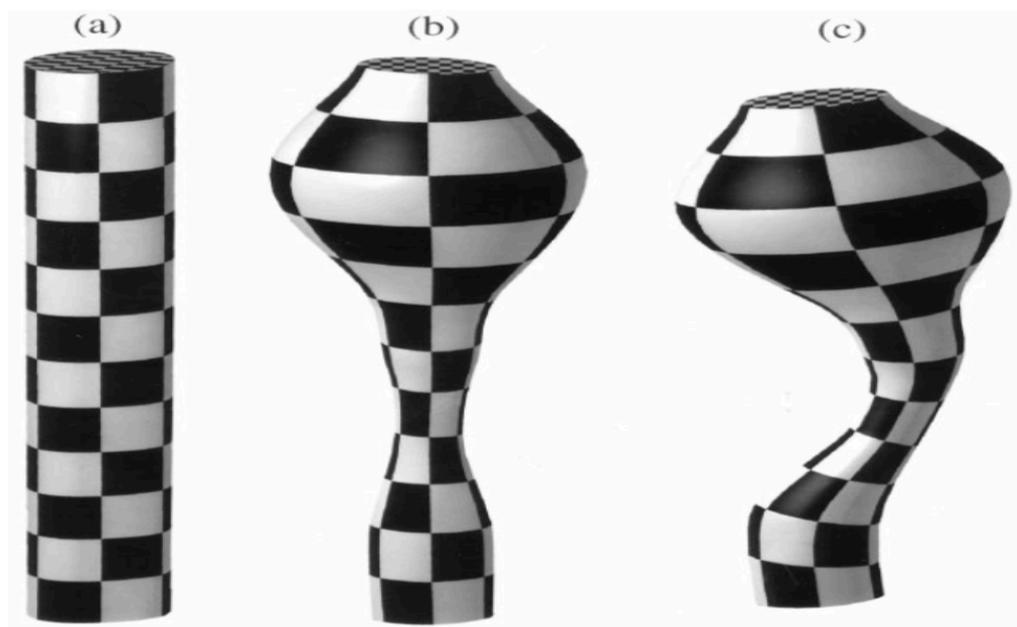


Figure 1.5 Schematic representation of rectal filling causing deformation of the rectum. The schematic describes deformation of the rectum caused by rectal filling which results in perceived difference in rectal length. (a) fixed volume of rectum which is principally cylindrical. (b) the rectum stretches due to filling however the height remains constant. (c) the rectum curves due deformation of the rectum and surrounding structures. The length remains constant but the height of the rectum is altered (Courtesy of Meijer et al., [86])

The uncertainties associated with DSM are similar to many of those related with DIR. Namely, the uncertainty associated with delineation of the organ, and the calculation of dose on the treatment image. Accumulating dose using DSM however negates the need to first deform the dose which also is subject to uncertainties (discussed Section 1.5.2.)

1.6 Aim and organisation of thesis

To improve local control through dose escalation safely, it is important to accurately quantify the dose delivered to the OAR such as the rectum, and develop a better understanding of the dose response relationship of these organs. Current rectal dose volume constraints are based on dose volume data from the planning scan only which may explain why some patients still experience toxicity despite meeting dose constraints at planning [28]. This thesis will evaluate the dose response of the rectum. Daily CBCTs will provide a description of the patient's anatomy at the time of treatment. The delivered dose to the rectum will be accumulated using two methods, DIR and DSM. The accumulated dose will be compared to the planned dose to measure the effect of rectal deformation on delivered dose. The dose volume parameters and spatial characteristics from the planned and accumulated DVHs and DSM will be associated with reported toxicity data. Comparison of the strength of the associations will determine whether delivered dose is a better predictor of toxicity than planned dose, and will aid the evaluation of current dose constraints.

1.6.1 Aim

The primary aim of this thesis is to determine whether accumulated dose is a better predictor of toxicity than the parameters derived from the planning CT scan alone.

1.6.2 Organisation of thesis

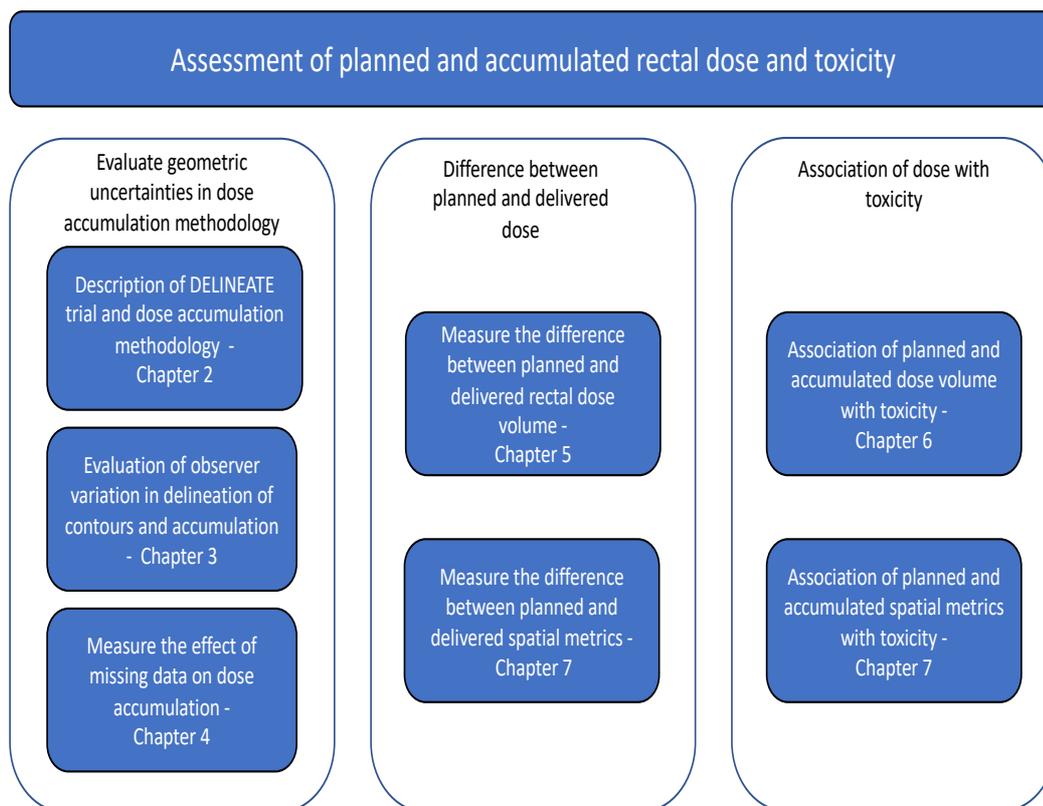


Figure 1.6 Summary of aim and objectives addressed within this thesis

A brief summary of each chapter is given below.

1.6.3 Summary of Chapters

1.6.3.1 Chapter 2

This chapter describes the methodology used to accumulate dose using DIR and DSM. A brief overview of the DELINEATE trial is given, including the scoring systems used to evaluate toxicity. The workflow for accumulating dose using DIR and DSM is defined. This includes importing the CBCTs and assigning mass density correction so that the daily dose can be calculated on the CBCT. The criteria used to delineate the rectal volume is also described. The DIR and DSM process is also detailed.

1.6.3.2 Chapter 3

This chapter analyses the uncertainty associated with manual delineation, which has been identified as one of the greatest sources of uncertainty in radiotherapy. The rectum is delineated on CBCTs which have poorer soft tissue contrast than CT. Thus, particular attention was paid to training, and subsequently validating the author's delineations by comparing them to experts' delineations. Three automatic segmentation methods were also evaluated to determine whether one could reduce the burden of manual delineation. This chapter also begins to explore the effect of interobserver variation on the uncertainty in DIR image registration, and subsequent dose calculation.

1.6.3.3 Chapter 4

As with many clinical studies, some patients did not have a complete dataset. There were two main sources of missing data identified in this study. Firstly, the rectal length was not always encompassed by the field of view (FOV) on the CBCT. This chapter describes the method used to simulate and validate the missing rectal length so that the full rectal volume could be deformed to the CT. Secondly, not all CBCTs were available or suitable for dose accumulation. The methodology used to measure the effect of missing imaging data on accumulated dose by comparing dose accumulated using a complete and artificially reduced imaging datasets is described.

1.6.3.4 Chapter 5 and Chapter 6

The difference in planned and accumulated dose volume parameters and subsequent association with dose is split over two chapters, Chapter 5 and Chapter 6. In Chapter 5 the delivered rectal dose is accumulated using DIR. In order to measure the variation in delivered dose, the difference between the rectal volume over treatment is quantified, and the difference between planned and accumulated DVH is analysed using descriptive and quantitative analysis. In Chapter 6 the toxicity experienced by the patient cohort is described. The planned and accumulated dose volume parameters are associated with treatment related side-effects to determine if one or other is a better predictor of toxicity.

1.6.3.5 Chapter 7

This chapter describes the spatial characteristics of dose accumulated using DSM. The toxicity experienced by the patient cohort is described. The difference between accumulated and delivered dose is described using spatial metrics such as pixel count, longitudinal and lateral extent of the dose distribution. Each of the spatial metrics analysed is associated with toxicity to determine whether accumulated dose described using spatial metrics is a better predictor of toxicity than the planned dose.

1.6.3.6 Chapter 8

Findings are summarised and clinical relevance discussed. Suggestions for future research directions are outlined.

Chapter 2 Methodology

2.1 Introduction

The aim of this PhD was to investigate if accumulated dose to the rectum is a better predictor of rectal toxicity than planned rectal dose. In order to accumulate dose, a description of the patient's anatomy (usually provided by a 3-dimensional image) at the time of treatment is required, as well as a method of establishing spatial correspondence between images so that the dose calculated at those time points can be summed [134]. In addition to accumulation of dose, this study also requires toxicity data which captures a range of adverse GI events that may be associated with different dose-volume parameters. Patient data from the DELINEATE trial (ISRCTN04483921) was identified as a source of the information required to meet the aims and objectives of this PhD since patients had daily 3D imaging data, as well as comprehensive baseline and follow-up toxicity data. This chapter describes the DELINEATE trial, the toxicity data collected as part of that trial, and the methodology used to accumulate dose.

2.2 The Dose EscaLation to Intra-prostatic tumour Nodules in localisEd prostATE cancer (DELINEATE) trial

DELINEATE is a single centre, Phase II dose escalation trial, evaluating the toxicity and feasibility of delivering a standard dose to the whole prostate, whilst delivering a simultaneous integrated boost (SIB) to escalate dose to intraprostatic lesions (IPL) [42]. The aim was to maintain current levels of late toxicity, despite delivering a higher total dose. The IPL were identified using multiparametric MRI. Patients were treated with IMRT or VMAT, and a daily on-line image protocol using CBCTs and implanted intraprostatic gold fiducial markers.

The trial opened in July 2011, with initial recruitment to the standard fraction cohort of 74 Gy/37 fractions (#) with a SIB of 82 Gy. In July 2012 the conventional versus hypofractionated high dose intensity modulated radiotherapy for prostate cancer trial (CHHiP) reported two-year safety results which showed that hypofractionated radiotherapy was equally well tolerated as standard radiotherapy [151]. Subsequently, in July 2013, the DELINEATE trial began recruitment to a moderate hypofractionated cohort of 60 Gy/20 # with a SIB of 67 Gy. The trial remains open and has begun

recruiting to a focal boost stereotactic body radiotherapy (SBRT) in 5 #, as well as a cohort for patients requiring pelvic node irradiation (treated with standard fractionation of 37 #). This PhD used patient data from the first two cohorts (i.e. standard and hypofractionated cohorts) (see Figure 2.1).

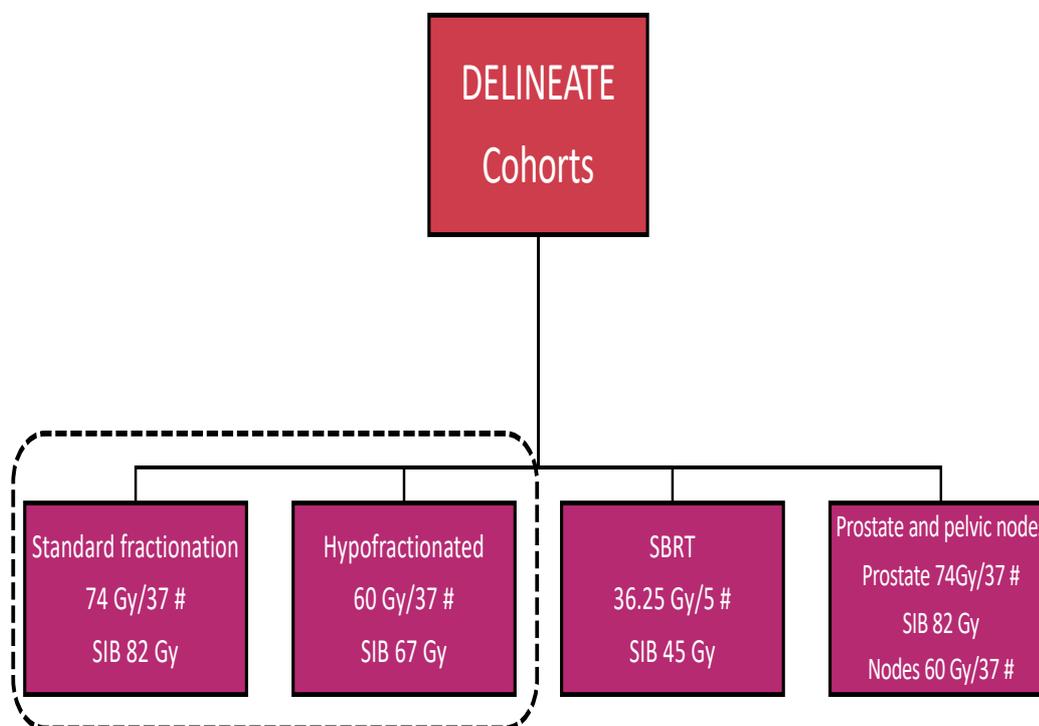


Figure 2.1 Schematic of DELINEATE cohorts. The dashed box encompasses the cohorts and fractionation schedules used in this study.

2.2.1 DELINEATE inclusion and exclusion criteria

The inclusion and exclusion criteria for the DELINEATE trial is presented below.

Inclusion Criteria

- Age > 18 years
- Histologically confirmed adenocarcinoma of the prostate
- National Collaborative Cancer Network (NCCN) risk groups of high or intermediate localised prostate cancer
- Normal blood count Haemoglobin (Hb) >11 g/dl, white blood cell count (WBC) > 4000/mm, platelets > 100,000/mm

- WHO Performance status 0 or 1
- Life expectancy of 10 years or more
- Written informed consent
- Patients must be prepared to attend follow-up

Exclusion Criteria

- Prior radiotherapy to the prostate or pelvis
- Bilateral hip replacement
- Prior hormone therapy
- Radical prostatectomy
- Prostate specific antigen (PSA) 50 or over
- NCCN Favourable risk group
- Evidence of seminal vesicle invasion, nodal or metastatic disease
- Any previous invasive cancer in the past 5 years, with the exception of non-melanoma skin cancer
- Patients with medical contraindication to MRI scanning
- Patients allergic to the contrast agent gadopentetate dimeglumine (Gd-DTPA)
- Co-morbid conditions likely to impact on toxicity of radiotherapy e.g. clinically significant inflammatory bowel disease, recent lower abdominal surgery
- Anticoagulants or antiplatelet medications that cannot be stopped temporarily for fiducial implant
- Increased risk of toxicity from implanted fiducial markers, e.g. valvular heart disease, severe toxicity, (infection/bleeding) from diagnostic transrectal ultrasound (TRUS) biopsy, anti-coagulated patients

2.2.2 Radiotherapy planning and treatment

Within 10 days of enrolment to the trial multiparametric MRI scans were performed to determine whether the IPL(s) was suitable for a boost. If the IPL was either not easily identifiable or alternatively considered too large to boost ($\geq 50\%$ of prostate volume) or more than 3 IPLs were identified, the patient would be advised to not take part in the trial but to receive prostate radiotherapy as per standard of care, i.e., off-trial. The IPL(s) was delineated on the MRI by an experienced radiologist.

Patients were prescribed androgen deprivation therapy after the first MRI. Prior to starting radiotherapy, patients should have responded to the hormone therapy with a stable or falling PSA < 4ng/ml. Hormone therapy was continued for the duration of radiotherapy. At least 5 days prior to the radiotherapy treatment planning CT scan, 3 gold fiducial markers were inserted into the prostate under transrectal ultrasound guidance.

For the planning CT scan patients were positioned with knee and ankle supports using the Combifix device (Civco, US), and CT images of 1.5 mm slice thickness were acquired (Brilliance Big Bore, Philips, Netherlands). The Royal Marsden Hospital (RMH) department bladder and rectal preparation protocol was followed which instructed patients to use a microenema (Micolette® or Relaxit) 90 mins before the CT planning scan. Sixty minutes before the CT scan, they were asked to void their rectum and bladder, and drink 350 ml of water. Patients were asked to have a full bladder at treatment, and to use enemas for the first 1 to 2 weeks of treatment.

An anatomical MR using T1 and T2 weighted and diffusion weighted sequences was acquired on the same day as the planning CT. This was registered with the CT scan in the treatment planning system (TPS), Pinnacle (Elekta AB, Stockholm, Sweden), using the fiducial markers. The registered MRI, with reference to the radiologist delineated IPL(s) on the first MRI, was used to define the prostate and the IPL(s). This was necessary as the IPL(s) could no longer be reliably identified on MRI following regression after the initial hormone therapy. The treatment was planned and delivered using a simultaneous integrated boost technique and 5 field IMRT or VMAT. The target volumes and prescriptions for both the standard and hypofractionated cohort are given in Table 2.1

Planning Target Volume (PTV)	Clinical Target (CTV) or Gross Target Volume (GTV)	PTV Margin		Standard Fraction Prescription (37#)	Hypofractionated Prescription (20#)
		Anterior, Right, Left, Superior and Inferior	Posterior		
PTV1	Prostate and seminal vesicles (SVs)	6 mm	6mm	59 Gy	48 Gy
PTV2	Prostate and base of SVs	6 mm	3 mm	71 Gy	57.6 Gy
PTV3	Prostate and base of SVs	3 mm	0 mm	74 Gy	60 Gy
PTV4	Intraprostatic lesion	2 mm	2 mm	82 Gy	67 Gy

Table 2.1 Table of the Planning target volume (PTV) and associated clinical or gross tumour volume, PTV margin and dose prescribed to each volume for both the standard and hypofractionated cohorts.

In addition, OARs were manually contoured including rectum, bladder, urethra, anus, bowel, femoral heads and penile bulb. The rectum was outlined to include the full circumference of the rectum, including contents. Inferiorly the rectum was defined as extending to the anal verge given as the level of ischial tuberosity. The superior border was the rectosigmoid junction, defined as the level at which there was an anterior inflection of the bowel. A table of required and optimal normal tissue dose-volume constraints for all OAR can be found in Table 2.2. For the treatment plan to be approved for treatment, it was necessary that the required normal tissue dose-volume constraints be met. Optimal dose-volume constraints are more stringent and further limit the dose to the OAR. However, it was not necessary for the treatment plan to meet the optimal dose-volume constraints for it to be approved.

Organ at Risk	Standard Fractionation			Hypofractionation		
	Volume Constraint	Optimal	Mandatory	Volume Constraint	Optimal	Mandatory
Anus	Mean Dose	30Gy	-	Mean Dose	24.3 Gy	-
	Volume to 35 Gy	< 60%	-	Volume to 28.38 Gy	< 60 %	-
	Volume to 40 Gy	< 40%	-	Volume to 32.43 Gy	< 40 %	-
	Volume to 60 Gy	< 0.01%	-	Volume to 48.65 Gy	< 0.01%	-
Rectum	Volume to 30 Gy	< 80%	-	Volume to 24.32 Gy	< 80%	-
	Volume to 40 Gy	< 65%	-	Volume to 32.43 Gy	< 65%	-
	Volume to 50 Gy	< 50%	< 60%	Volume to 40.54 Gy	< 50%	< 60%
	Volume to 60 Gy	< 35%	< 50%	Volume to 48.65 Gy	< 30%	< 50%
	Volume to 65 Gy	-	< 30%	Volume to 52.70 Gy	-	< 30%
	Volume to 70 Gy	-	< 15%	Volume to 56.76 Gy	-	< 15%
	Volume to 75 Gy	< 3%	< 5%	Volume to 60.81 Gy	< 3%	< 5%
	Dose to 2%	-	< 76 Gy	Dose to 2%	-	61.6 Gy
Bladder	Volume to 50 Gy	< 50%	-	Volume to 40.54 Gy	< 50%	-
	Volume to 60 Gy	< 25%	-	Volume to 48.65 Gy	< 25%	-
	Volume to 65 Gy	-	< 50%	Volume to 52.70 Gy	-	< 50%
	Volume to 70 Gy	< 5%	< 35%	Volume to 56.76 Gy	< 5%	< 35%
	Volume to 75 Gy	< 3%	< 25%	Volume to 60.81 Gy	< 3%	< 25%
	Volume to 80 Gy	< 0.02%	< 15%	Volume to 64.86 Gy	< 0.2%	< 15%
Urethra	Dose to 2%	< 77Gy	< 83 Gy	Dose to 2%	< 62.4Gy	< 67.3Gy
Bowel	Volume to 45 Gy	< 78 cc	< 158 cc	Volume to 36.49 Gy	< 78 cc	< 158 cc
	Volume to 50 Gy	< 17 cc	< 110 cc	Volume to 40.54 Gy	< 17 cc	< 110 cc
	Volume to 55 Gy	<14 cc	< 28 cc	Volume to 44.59 Gy	<14 cc	< 28 cc
	Volume to 60 Gy	< 0.5 cc	< 6 cc	Volume to 48.65 Gy	< 0.5 cc	< 6 cc
	Volume to 65 Gy	-	< 0.01 cc	Volume to 52.70 Gy	-	< 0.01 cc
Femoral Joint	Volume to 50 Gy	< 10%	< 50%	Volume to 40.54 Gy	< 10%	< 50%
Penile Bulb	Dose to 50%	< 27 Gy	-	Dose to 50%	< 21.8 Gy	-

Table 2.2 The normal tissue dose-volume constraints used for each cohort in the DELINEATE trial. Both optimal and required constraints are included.

2.2.3 Treatment verification using CBCT

A daily on-line imaging protocol was used to correct for interfraction prostate motion. CBCT images were matched to the CT planning scan using dual registration (XVI version 4.5, Elekta Synergy, Elekta AB, Stockholm, Sweden), which was rigid registration of bony anatomy followed by rigid registration of fiducial markers. Treatment was given only if the fiducial markers in the planning CT and CBCT were at least partially overlapping after registration, and if the prostate was encompassed by the PTV contour. If rectal distension resulted in a large proportion of the rectum being pushed into the PTV, the patient was removed from the couch and asked to try to expel any rectal gas.

2.2.4 Toxicity data collected

It is important when analysing the relationship between late GI toxicities and dose that appropriate, validated scoring systems are employed that capture the extent and the

variety of symptoms associated with adverse effects of treatment [28, 30, 70, 152]. It is important that details of pre-existing conditions which may increase complication risk are collected to help interpret the incidence of radiation induced sequelae [9]. Baseline data should likewise be collected to help distinguish between adverse events associated with rectal injury or pre-existing conditions which may present with similar symptoms [9]. The duration of follow up should also be considered as GI toxicity can occur up to a few years after treatment. Patient reported outcomes (PRO) have been shown to detect side effects more reliably than clinician reported outcomes [153, 154]. The use of patient reported toxicity questionnaires are also recommended by QUANTEC [9] as a means of measuring the impact of side effects on quality of life (QoL). The scoring system and collection protocol of the DELINEATE trial meets these specifications.

Clinical characteristics and toxicity data were prospectively collected as part of the DELINEATE trial. Scoring systems for clinician reported outcomes (toxicity) (CRO) used in the DELINEATE trial included the Royal Marsden Hospital (RMH) scoring system [80] which measures bowel frequency and rectal bleeding. The Gulliford rectal scores [56] is a combination of seven questions drawn from Late Effects of Normal Tissue (LENT SOMA) [58] and University of California Prostate Cancer Index (UCLA-PCI) [153]. Although UCLA-PCI is a patient reported scoring system, 3 questions regarding GI toxicity were also asked by clinicians. The questions were chosen to reflect endpoints that are clinically relevant and impact on patient QoL. Scoring systems which were also used as part of DELINEATE were the Radiation Therapy Oncology Group (RTOG) [155] and the Common Terminology Criteria for Adverse Events (CTCAE) [57]. Baseline CRO was collected at registration and at 4 weeks prior to the start of radiotherapy (referred to as baseline toxicity). Late toxicity data was collected at 6, 12, 18, and 24 months, and then annually. Acute toxicity was measured using RTOG and CTCAE v4.0 at weeks 1 to 8, 10, 12 and 18 from start of radiotherapy. PRO were measured using Vaizey [61], a Modified Inflammatory Bowel Disease Questionnaire (IBDQ_B) [60] and the Expanded Prostate Cancer Index Composite-26 (EPIC-26) [62]. Questionnaires were distributed at baseline, week 18 and at 6, 12, 18 and 24 months from the start of radiotherapy. However, due to initial low patient response, PROs were not analysed in this thesis.

A copy of the CRO can be found in Appendix 1. Analysis and results of the toxicity data are given in Chapter 5 and Chapter 7.

2.2.5 PhD patient population and governance

Approval for use of the DELINEATE trial data was granted by the Trial Management Committee in January 2015. The Trial Management Committee considered that this PhD meets one of the secondary objectives of the DELINEATE trial, which is to assess the patients received dosimetry to prostate and OAR, therefore no additional ethics approval was required.

At the start of this study (PhD) the first 106 patients treated in the DELINEATE trial were identified for inclusion in this study. Using the initial patients from both cohorts ensured that there would be two-year toxicity data for the analyses during the timescale of the PhD. These patients were treated between 13th July 2011, and 16th January 2015. Of the 106 patients, 56 and 50 patients were in the standard and hypofractionated cohorts, respectively. One patient in the standard fractionation cohort withdrew from the DELINEATE trial (patient choice), and was excluded from analysis in this study. Median (range) follow-up for the remaining 105 patients was 28.9 (31.7) months for the standard fractionation cohort, and 28.3 (25.1) months for the hypofractionated cohort at the time of data lock (defined as time of data provided by trial statistician for analysis as part of this PhD).

The demographics and clinical characteristic of the 105 patients included in this study are given below in Table 2.3.

	Standard Fractionation (n=55)	Hypofractionation (n=50)
<u>Median Age at registration (years)</u>	70 (57-80)	71.5 (61-79)
<u>T Stage</u>		
T1	20 (36 %)	24 (48 %)
T2a	6 (11 %)	14 (28 %)
T2b	18 (33 %)	0 (0 %)
T2c	5 (9 %)	2 (4 %)
T3a	6 (11 %)	10 (20 %)
<u>Combined Gleason Score</u>		
6	13 (25 %)	13 (26 %)
7	40 (73 %)	34 (68 %)
8	2 (4 %)	3 (6 %)
<u>NCCN Risk Group</u>		
Intermediate	40 (73 %)	30 (60 %)
High	15 (27 %)	20 (40 %)
<u>Diabetes</u>		
Yes	2 (4 %)	15 (30 %)
No	53 (96 %)	35 (70 %)
<u>Hypertension</u>		
Yes	21 (38 %)	27 (54 %)
No	34 (62 %)	23 (46 %)
<u>Inflammatory bowel or diverticular disease</u>		
Yes	2 (4 %)	8 (16 %)
No	53 (98 %)	42 (84 %)
<u>Pelvic surgery</u>		
Yes	4 (7 %)	12 (24 %)
No	51 (93 %)	38 (76 %)
<u>Symptomatic Haemorrhoids in the last 12 months</u>		
Yes	11 (20 %)	2 (4 %)
No	44 (80 %)	48 (96 %)
Unknown		
<u>Previous transurethral resection of the prostate</u>		
Yes	5 (9 %)	4 (8 %)
No	50 (91 %)	46 (92 %)
<u>Statins</u>		
Yes	20 (36 %)	24 (48 %)
No	35 (64 %)	26 (52 %)
<u>Current Smoker</u>		
Yes	5 (9 %)	4 (8 %)
No	50 (91 %)	46 (92 %)

Table 2.3 Demographic and clinical characteristics of patients included in this study. There were 55 patients in standard fractionation and 50 in hypofractionated cohort.

In order to have a larger sample size when associating toxicity with planned and accumulated dose-volume data, the dose arms were combined, converting the hypofractionated dose to an equivalent dose of 2 Gy/# (EQD2) on a pixelwise basis to simulate the standard fractionated prescription. This was calculated by exporting the hypofractionated dose-volume data to MATLAB (Mathworks Inc., Natick, MA, USA) where it was converted using the Wither's formula given below (script written by Dr Gulliford) [156].) An alpha-beta ratio, α/β , a measure of radiation fraction sensitivity, of 3 Gy was used for the rectum [9], D was the total dose and d was the dose per fraction

$$EQD2 = D \frac{(d + \frac{\alpha}{\beta})}{(2 + \frac{\alpha}{\beta})}$$

2.2.6 Intra-prostatic tumour nodules

The position and location of the high dose boost regions in the prostate is important in terms of rectal toxicity. Each treatment plan has been reviewed in RayStation. For the DELINEATE trial, the gross tumour volume (GTV) was defined as the intra-prostatic tumour nodule plus a margin of 2 mm. Each GTV was considered as a single intra-prostatic tumour nodule. If there were multiple GTVs per patient, then all GTVs were included in this analysis and the number were recorded. Intra-prostatic nodules in close proximity may have been outlined using a single contour and therefore, using the GTV as a surrogate for the number of nodules, may not accurately reflect the true number of intra-prostatic nodules boosted within the patient population. To describe the location of the GTVs, the prostate was divided into the base, mid-gland and apex. Each section was further divided into four quadrants providing and anterior/posterior and right/left. The position of the GTV was determined by visually assessing which section the majority of the GTV was located.

Two of the eighty six patients were excluded from the analysis as the GTV extended outside of the prostate to include the seminal vesicles. Of the remaining 84 patients the median volume (IQR) of the prostate was 35.9 cm³ (21.6 cm³). In 59 (70 %) patients a single nodule was boosted, two nodules in 20 (24 %) patients, and three nodules in 5 (6 %). The median (IQR) of all nodules was 2.4 cm³ (3.98 cm³). There were a total of 22 (19 %) nodules boosted in the right anterior aspect of the prostate, 11 (10 %) in the left anterior, 37 (32 %) in the right posterior and 44 (39 %) in the left posterior. Of the 114 nodules boosted 19 (17 %) were found in the base of the prostate, 36 (32 %) in the apex and 59 (41 %) in the mid-gland.

2.3 Dose accumulation

Two distinct methods were used in this thesis to accumulate the total dose received by the rectum over the course of treatment. The first method accumulated the dose by establishing spatial correspondence between the daily CBCTs and the original planning CT using deformable image registration (DIR). This required regions of interest (ROIs) to be defined, in this case, the rectum, to help guide the DIR. The deformation information was then used to deform the daily dose, permitting dose from multiple fractions to be summed. The second method created 2D dose surface maps (DSM) of the rectum using the daily CBCTs. The maps were normalised to the same

length, which permitted multiple maps to be summed together to create a DSM of the total delivered dose. RayStation TPS (RaySearch Laboratories AB, Stockholm, Sweden) has an established workflow for DIR and dose accumulation [19]. However, the DSM were created in MATLAB using an in-house script created by Dr McQuaid and Dr Gulliford, which extracted dose data from RayStation using IPython (www.ipython.org) [20]. The schematic in Figure 2.2 gives an overview of the dose accumulation process.

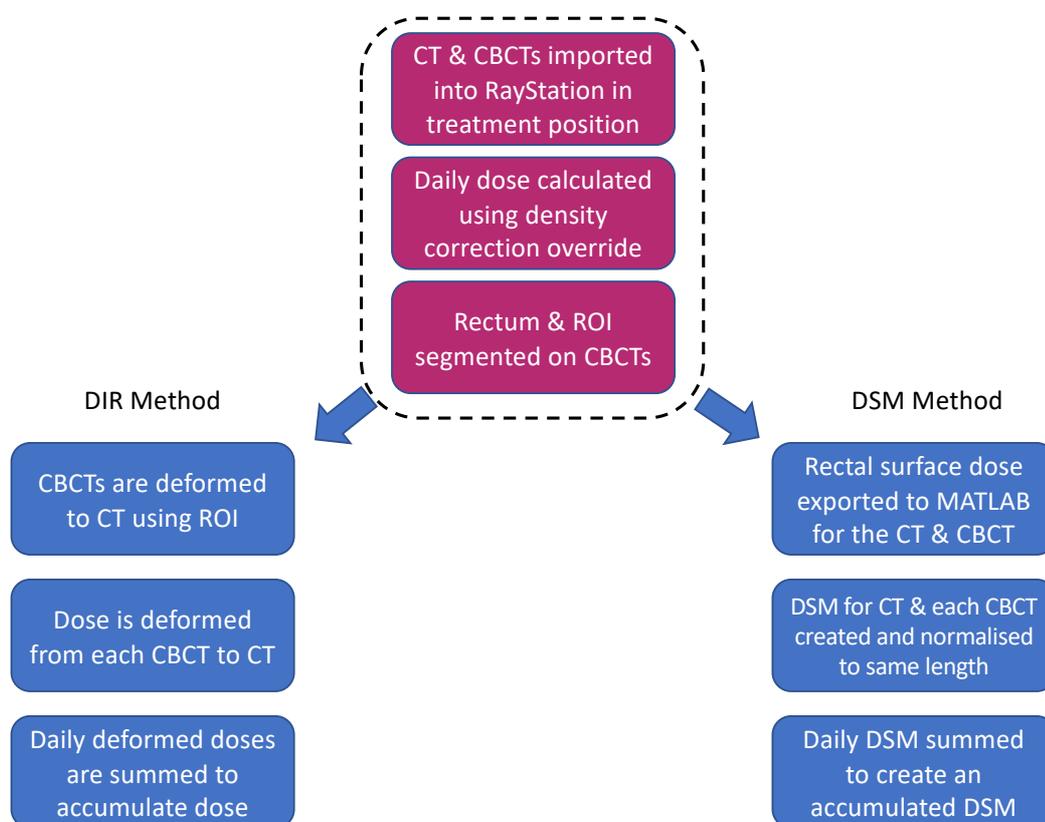


Figure 2.2 Overview of the dose accumulation methodology using deformable image registration (DIR) and dose surface maps (DSM). The schematic gives a broad overview of the key steps involved in accumulating dose using either DIR or DSM. The initial steps (highlighted by the dashed box) are common to both dose accumulation methods.

2.3.1 Calculation of daily delivered dose

Both methods of dose accumulation require, in the first instance, the calculation of the daily delivered dose using the CBCT, and the delineation of the rectum. This was performed using RayStation. To calculate the daily delivered dose using RayStation

and CBCTs, two steps were required [157]. Firstly, the CT voxel Hounsfield units (HU) and CBCT voxel grey level values were assigned mass densities which are required for the dose calculation. Secondly, the external contour of the patient was defined on both the CT and CBCT.

2.3.1.1 Assignment of CT and CBCT voxel mass density

The planning CT, with the associated treatment plan, structure sets (3D contours of the target volumes and OAR) and dose distributions were imported into RayStation from Pinnacle TPS (Philips N.V., Eindhoven, Netherlands). Daily CBCTs were exported to RayStation from the X-ray Volumetric Imaging (XVI) system (Elekta AB, Stockholm, Sweden) with the set-up correction applied (i.e., the patient is represented in the position they would have been for treatment). RayStation contains a library of HU to density conversion tables that was created for each of the department CT scanners as part of the TPS commissioning. The appropriate CT density table was manually assigned to the planning CT scan to convert each voxel HU to mass density.

CBCT voxel grey scale level values do not correspond to HU and no conversion tables were available to convert grey scale level values to mass density [27, 125]. RayStation uses a stepwise mass density correction method to create a density table for each CBCT (see Figure 2.3). An automatic algorithm uses grey level thresholds to bin the voxels of the image into one of six types of tissue, with a corresponding mass density. The types of tissue are air (0.00121 g/cm^3), lung (0.26 g/cm^3), adipose (0.95 g/cm^3), connective tissue (1.05 g/cm^3), cartilage/bone (1.60 g/cm^3) and high density tissue (3.00 g/cm^3) [158]. Although the values of tissue densities used are fixed, how they are proportioned is individual to the patient. For average-sized patients (anterior-posterior distance $< 25 \text{ cm}$), the automatically-run algorithm produced satisfactory threshold values, but for larger-sized patients (anterior-posterior distance $> 25 \text{ cm}$), the amount of adipose tissue relative to muscle tissue sometimes required manual adjustment [125, 157]. The mass density generated using RayStation's correction method was used to calculate the dose on CBCTs. Dunlop et al., compared the dose to the rectum, calculated using CT and mass density corrected CBCT and reported a mean difference of 0.7 % (range: -1.3 % to 0.6 %) [125]. Similar results are reported by Chen et al., with a mean PTV dose-volume difference between CBCT-based and CT-based plans in the pelvis of 0.2 % (SD 1.3 %) [31].

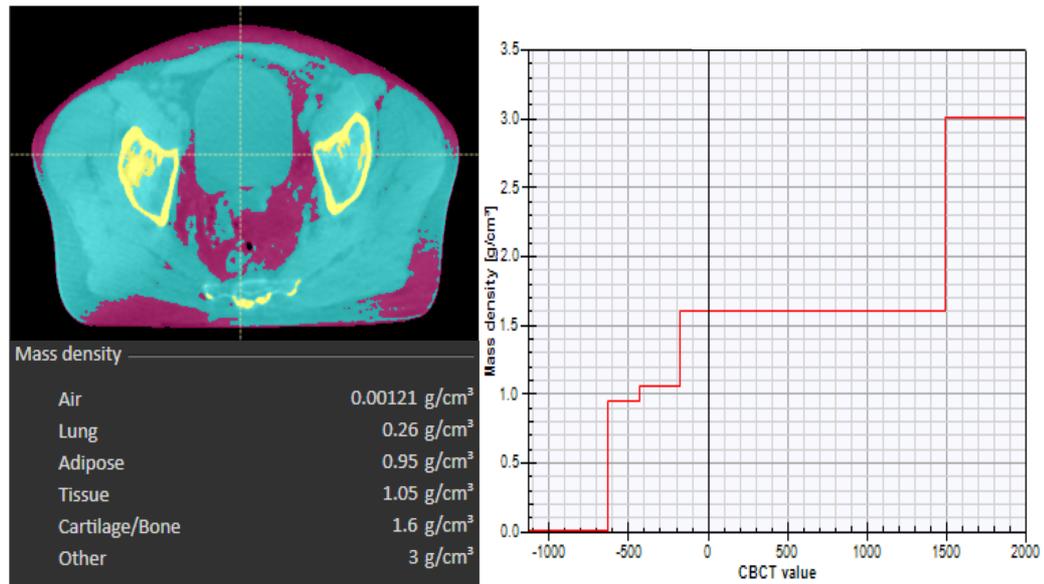


Figure 2.3 Illustration of a CBCT grey level values to mass density correction. The top left image shows an axial slice of the pelvis. Each voxel of the image is assigned one of six types of tissue. The different tissues are colour coded: black = air, pink = adipose tissue, yellow = bone and light blue = connective tissue (lung and high density (labelled as other) tissues are not present in this image). Each of the six tissue types has a corresponding mass density, which is shown in the bottom left image. An automatic algorithm (RS_auto) extracts the density thresholds and creates a CBCT value to mass density conversion table (shown on the right) [125, 158].

2.3.1.2 Defining the external contour

An external body contour was created using an automatic thresholding technique, which used the image data to find the patient's surface. The body contour was edited to remove small discontinuities in the contour, and those contours that did not appear to accurately represent the external surface of the patient. Due to the limited field of view (FOV) of the CBCT, the external CBCT contour was shorter in length than the CT external contour (Figure 2.4.). Without addressing the discrepancy in length between the two images, the accuracy of the DIR would be affected as the algorithm would distort the shorter CBCT in the superior-inferior direction to match the longer CT [142]. To overcome this the RayStation user manual recommended that the external contour on the CBCT is converted into an model based segmentation (MBS) mesh [157]. This is used in biomechanical deformable registration where ROIs are converted into triangular meshes. The external MBS from the CBCT was used to define an external MBS of the same magnitude on the CT. Defining the external MBS on both images ensured that the same superior-inferior length of image was used when deforming the CBCTs. In addition, the external MBS can be selected as a ROI to be used when deforming the CBCT, and ensure the deformation is anatomically realistic in the superior-inferior direction.

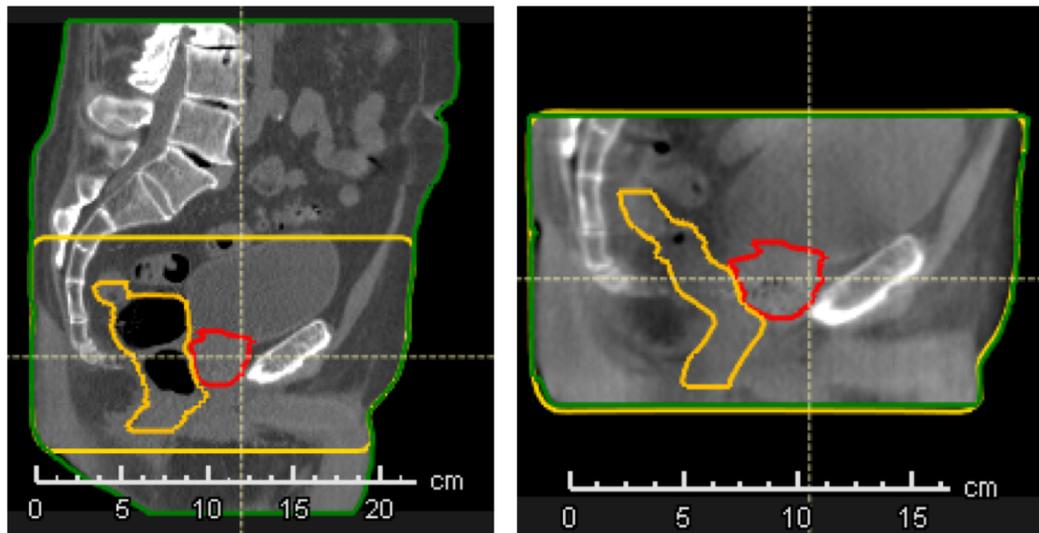


Figure 2.4 Sagittal image of pelvis showing CT and CBCT field of view. Images showing the external CT body contour (green) and the external MBS (yellow), the prostate (red) and the rectum (orange). The left sagittal image is the planning CT, with a greater volume of the patient encompassed by the longer FOV. The right image is a sagittal CBCT slice of the same patient showing the shorter external contour defined by the smaller FOV of the CBCT. The external MBS of the CBCT was used to define the external MBS on the CT. This ensured the same length of the image was used when deforming the CBCT to the CT.

In some instances, the CBCT FOV did not fully encompass the circumference of the patient which resulted in ‘missing tissue’ (Figure 2.5.). Calculating the dose with missing tissue (defined by the external contour) would lead to dosimetric errors. In these instances, the missing tissue was simulated by using the external CT contour as a proxy. When this occurred the simulated tissue outside of the CBCT FOV was assigned the density of water (1.00 g/cm^3). In the study by Dunlop et al., they reported the mean difference in dose calculated using water density when compared to the planning CT for the rectum was 0.20 % (range: -1.10 % to 1.10 %) [125].

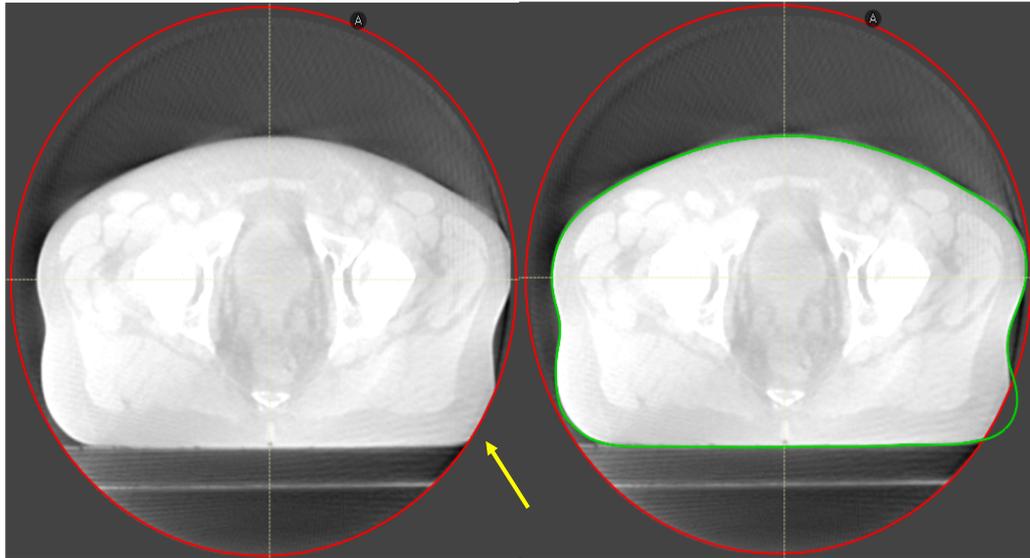


Figure 2.5 Axial slice of pelvis CBCT. The windowing on the CBCTs have been adjusted to help define the CBCT FOV. On the left, the axial slice shows the CBCTs limited FOV (red circle) does not fully encompass the patient resulting in missing tissues (highlighted by yellow arrow). Dose calculation without the true patient volume would lead to inaccuracies. To overcome this the CT contour was used to estimate the body contour in this region at the time of treatment, as shown on the right (green outline). The missing tissues, the extent of which were defined by the CT external contour and the FOV of the CBCT, were assigned water density.

2.3.2 Delineation of the rectum and regions of interest

After receiving training from two experts (Consultant Radiation Oncologists) the rectum was manually contoured on all CBCTs. The rectal contour on the CBCTs was defined using the criteria specified in the DELINEATE trial protocol. The rectum defined on the patient's planning scan was used as a visual reference, as previous studies have shown that differences between observers defining the length of the rectum can lead to dosimetric uncertainty when calculating the rectal dose [159-161]. In order to minimise dose uncertainties when comparing the accumulated dose to the planned dose, particular care was taken in contouring and defining the extent of the rectum across all images. This ensured that the rectum was defined as the level of recto-sigmoid junction to the ischial tuberosities on the planning CT, and was replicated on all CBCTs e.g. defining the superior rectum.

There were instances when the superior-inferior (longitudinal) FOV of the CBCT scan was not extensive enough to encompass the most inferior part of the rectum (Figure 2.6.). This was because the department CBCT imaging protocol required that the smallest collimator was chosen to minimise concomitant dose to normal tissues. The primary purpose of daily CBCTs was to correct for prostate motion, and therefore the

inclusion of the full rectal length was not a consideration. To enable the same rectal length to be defined on the CBCT as on the CT, the missing rectal volume was simulated. The validation of the technique used to simulate the missing rectal volume is discussed further in Chapter 4.

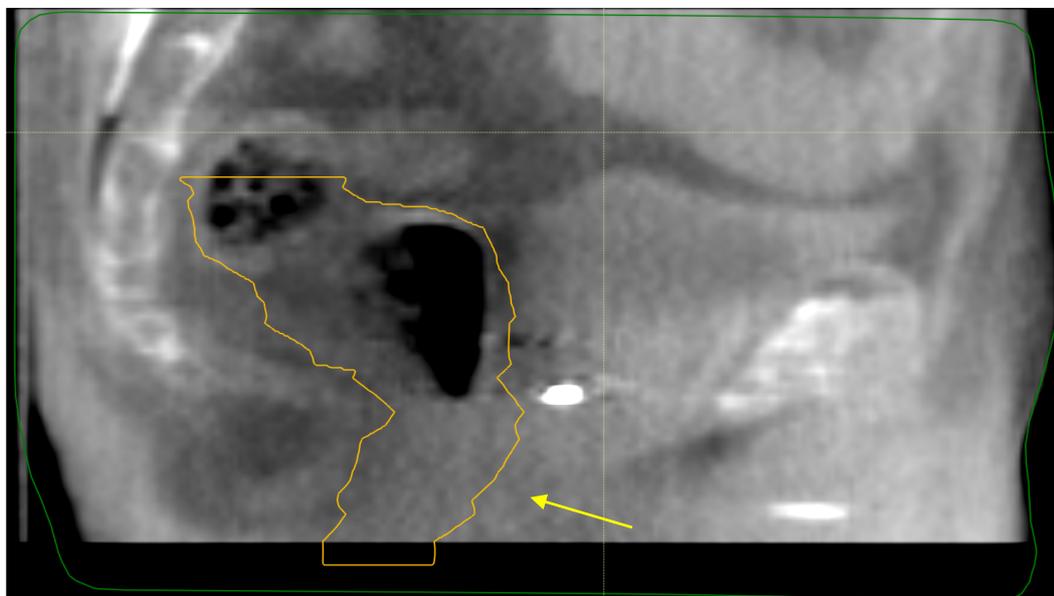


Figure 2.6 Sagittal slice of pelvis. The image shows how the rectum (yellow) extends past the CBCT FOV. The missing rectal volume was simulated as described in Chapter 4.

In addition to the rectum and the external MBS (discussed in Section 2.3.1.2), the prostate was also defined on the CBCTs, and used as a controlling ROI for DIR. The apex and the base can be overestimated when contouring the prostate on CT scans [96]. Examination of the prostate and base of SVs in repeat CT scans by Bostel et al., showed little variation in prostate volume (relative volume difference between the planning examination and treatment scan $0.3 \pm 10.9\%$) [138]. The soft tissue contrast of the CBCT was found not to be sufficient to enable definition of the longitudinal extent of the prostate, and therefore the CT defined prostate ROI was copied to the CBCTs. Once the rectal contour was defined, the prostate ROI (except for the longitudinal extent) was adjusted by editing the shape so that it did not overlap with the rectal ROI. The bladder could not be delineated, again because the limited FOV of the CBCT meant that the bladder volume was not fully encompassed on the majority of the CBCTs.

Automatic segmentation models (ASM) may be of benefit in reducing the time consuming nature of manual segmentation [144]. However, prior to use ASM

performance needs to be evaluated. An ASM for rectal delineation is described and evaluated in Chapter 3.

2.3.2.1 Evaluation of observer contouring agreement

Interobserver variation has been shown to be greater than intraobserver variation and one of the sources of uncertainty in DIR [142]. Dice similarity coefficient (DSC) and mean distance to agreement (MDA) are established metrics for determining the agreement between contours that have been outlined by multiple observers (interobserver agreement), or outlined by one observer multiple times (intraobserver agreement) [135]. These metrics can also be used to assess the performance of computer algorithms that can be used to automatically contour tissues, by comparing the automatically generated contour with that of experts, which is discussed further in Section 3.3.2 [135, 162].

DSC is an overlap metric which is defined as two times the intersection of contour A and contour B, divided by the sum of volume A and B and is shown in the formula below [135].

$$DSC = \frac{2(A \cap B)}{(A + B)}$$

If there is perfect overlap of A and B the DSC is 1, and if there is no overlap then the value is 0. MDA is calculated by finding the mean Euclidean distance between every point on the surface of A, and its nearest point on surface B, with perfect agreement of 0 mm [142, 163]. The author could not find a consensus threshold value given in the literature as to what would show good agreement, as this can be influenced by factors such as image modality and the structure being analysed. However a review of the literature showed that $DSC \geq 0.7$ [147, 164, 165] and a MDA of ≤ 3 mm are typically considered to show good agreement between contours [142].

In this study, for planned and accumulated dose-volume data to be comparable, it was important that the rectal contour outlined on the CBCT encompassed the same extent of the rectum outlined on the CT. To measure agreement between contours outlined by the author and the observer who outlined the rectum for the purposes of planning

the treatment, the rectum was recontoured on the planning CT. A pairwise comparison was then performed between the original rectal contour and the author delineated contour using DSC and MDA (script written by Dualta McQuaid) [135, 142, 163]. The comparison showed a good level of agreement with a median DCE score of 0.91 (range 0.87 to 0.94) and a median MDA of 0.09 mm (range 0.06 mm to 0.15 mm). The ability of the author to contour the rectal volume on CBCTs was also assessed by comparison to expert contours, and is discussed further in Chapter 3.

2.3.3 Deformable image registration

RayStation's ANatomically CONstrained Deformation Algorithm (ANACONDA) was used to deform the daily CBCT images (moving image) to the planning CT (reference image). ANACONDA is a hybrid solution, which uses a combination of pixel (or voxel) intensity and anatomic information to deform the image [146]. The anatomical information is provided by the contoured ROIs (controlling ROIs). ANACONDA is based on a mathematical formulation where the registration problem is formulated as a non-linear optimisation problem [146]. When contours are used as the controlling ROI, a penalty term is added to the optimisation problem in order to deform the selected structure(s) in the moving image to the structure(s) in the reference image [166]. A grid regularisation term is designed to keep the deformed image grid smooth and invertible. The similarity term is measured using the correlation coefficient [146, 166]. ANACONDA is particularly useful in anatomical areas such as the pelvis where soft tissue contrast is poor and intensity based information alone may not result in anatomically realistic deformation [146]. Or when images from different imaging modalities, and subsequent different greyscale information, are being deformed [143].

To deform the CBCT, the rectal, prostate and external MBS contours were used as the controlling ROIs, with a deformation resolution grid of 0.25 cm/voxel. This was the smallest resolution grid available in RayStation at the time of calculation. The DIR produced displacement vector fields (DVF) specifying the direction and magnitude of the deformation. Figure 2.7 shows a rectal volume from a CBCT before and after deformation to the planning CT, and also the DVF describing the deformation. To accumulate the dose to the rectum, the DVF (which were used to deform the image) were used to deform the dose grid calculated on the CBCTs and map it to the reference CT. The deformed doses were then summed to give a total accumulated dose.

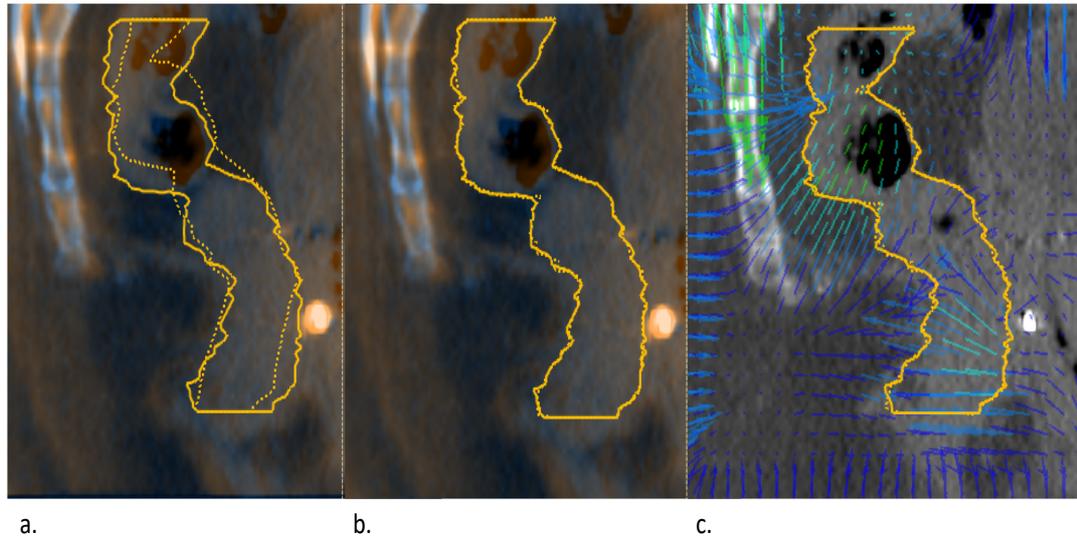


Figure 2.7 Sagittal slice of a fused CT and CBCT pelvis, showing a rectal volume before and after the CBCT has been deformed to the CT. Image a. shows the reference contour (CT rectal volume) with a solid yellow line and the undeformed moving rectal contour (CBCT) with a yellow dotted line. Image b. shows the moving rectal contour after deformation where it matches the reference contour. Image c. shows the deformation vector fields (DVF), the arrows indicating the direction and magnitude of the deformation.

To evaluate the performance of the DIR a comparison was performed between the deformed rectum and the reference rectum (CT image) using DCE and MDA. Only those deformations with a $DCE \geq 0.7$ and a $MDA \leq 3$ mm were considered for dose accumulation. If a full dataset of CBCTs was not available, daily doses were weighted to account for missing images. The effect of missing images on dose accumulation is described in Chapter 4. The planned and accumulated dose-volume data was exported for analysis. A schematic of the DIR process for dose accumulation is given in Figure 2.8.

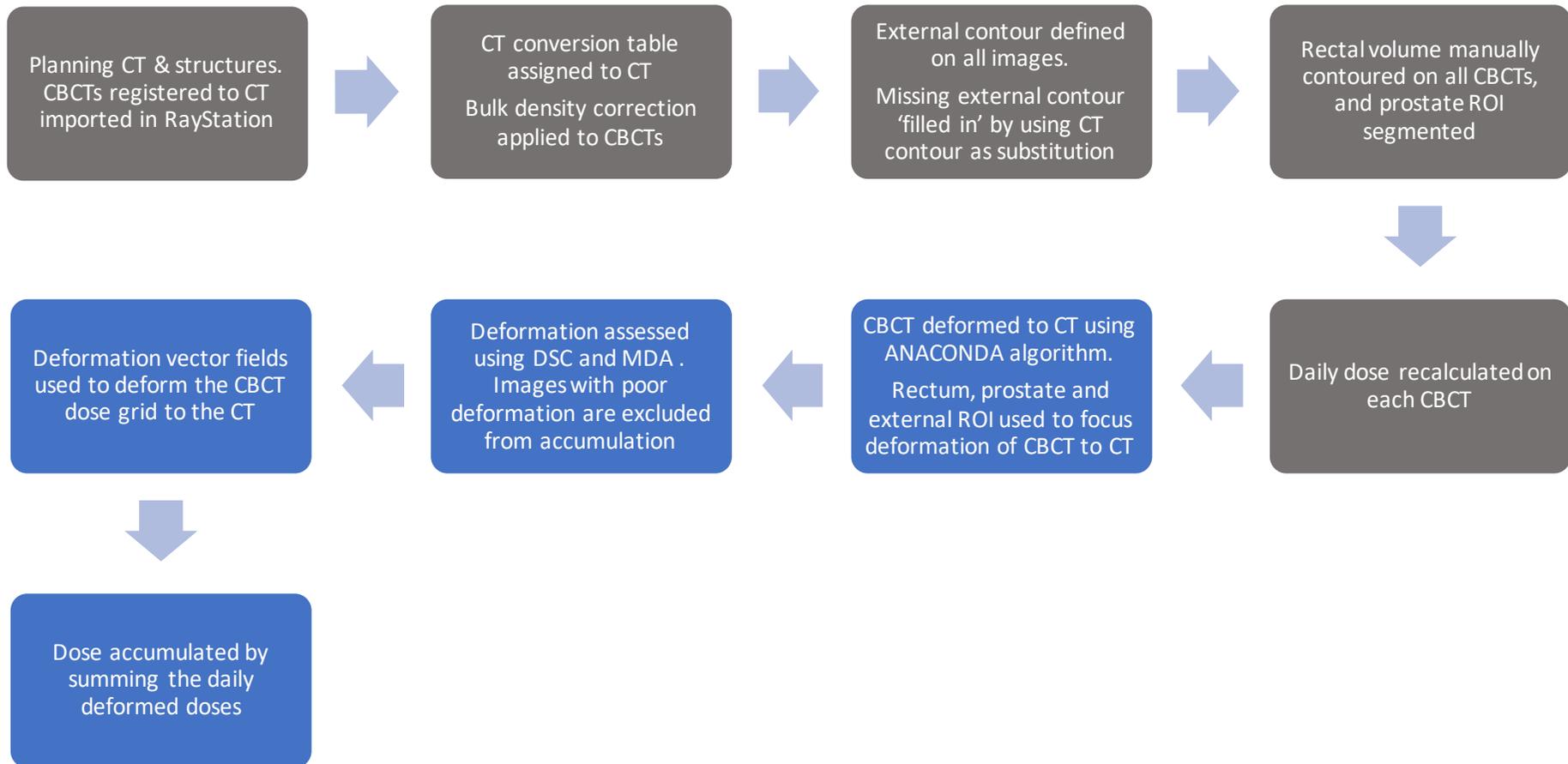


Figure 2.8 Schematic showing the process of accumulating dose using DIR. The initial process (denoted by grey boxes) is also used to create DSM. Abbreviations are ROI = to region of interest, DSC = Dice similarity coefficient and MDA = mean distance to agreement.

2.3.4 Dose surface maps

Dose surface maps (DSM) can be used to report the dose to the surface of the rectal wall, and are an established method for analysing radiation-induced toxicity [90, 93, 167]. However, RayStation did not have the functionality to generate DSM. Instead dose data was exported from RayStation using IPython, and DSM were generated in MATLAB using in-house software created by Dr. McQuaid and Dr. Gulliford. DSM were generated for the contoured rectum on the planning CT, and for each of the undeformed CBCTs. The undeformed CBCT was used to provide a method of accumulation independent from DIR, and possible uncertainties associated with DIR methodology. Unpublished work undertaken by Murray as part of her doctoral thesis compared rectal DSM generated using DIR and summation of CBCTs [168]. The study reported similarity in spatial dose information across all doses.

Different algorithms to generate the rectal DSM exist [150, 169]. However this study uses methodology developed by Buettner et al., and Murray et al., which has been previously reported [91, 94]. A binary mask of the rectum is first created in RayStation by sampling the dose at 2° intervals around the circumference of each axial slice of the rectum, providing 180 equidistant points. Longitudinally, the cylindrical array was sliced at 0.3 cm intervals. This interval was chosen as it has been previously reported [94, 149]. The cylindrical array was then exported to iPython using in-house software (written by Dr Dualta McQuaid). To create the 2D map of the rectal dose, the cylindrical array was virtually unfolded by cutting each slice at its most posterior location. This produced a 2D array of $180 \times n$ slices, where n is the number of axial slices (see Figure 1.4 in Section 1.5.3)

As the rectal length varied between treatments, the DSM of the CBCTs were resampled in the lateral and longitudinal extent to create maps of 21×21 pixels in order to facilitate accumulation [94]. To accumulate the DSM, the absolute dose was summed on a pixel by pixel basis. For patients treated in the hypofractionated cohort, the DSMs were converted to equivalent dose in 2 Gy/# (EQD2) using the Wither's formula, and as described in section 2.2.5. An interpolated 21×21 pixel DSM was also created for the CT rectal contour. A dose difference map was created by subtracting the accumulated DSM from the planned DSM.

Metrics to describe the spatial distribution of the dose were extracted using the method developed by Buettner et al., [90]. Using Matlab, DSM were converted into a series of binary images by thresholding the maps at 2 Gy intervals between 5 Gy and 59 Gy (software created by Dr Buettner and adapted by Dr Gulliford). Pixels receiving dose equal to or above a particular dose were assigned a value of 1, and pixels below this dose were assigned a value of 0. In order to extract features from the stacked binary images, all 1-valued pixels were grouped into clusters (a set of connected regions). The anterior rectum lies adjacent to the prostate and usually receives the highest dose. This explains why the binary maps of the rectum commonly resulted in a radial dose distribution extending from the centre. Consequently, features were extracted from the largest cluster of pixels present (see Figure 2.9 a.). The metrics used to analyse the spatial distribution were the maximum longitudinal and lateral extent at each dose. This was calculated by counting the number of pixels at the greatest extent which is illustrated in Figure 2.9 b. Dose surface histograms were also calculated by counting the number of pixels with a value of 1 at a given dose threshold. Schematic of the DSM accumulation process is shown in Figure 2.10.

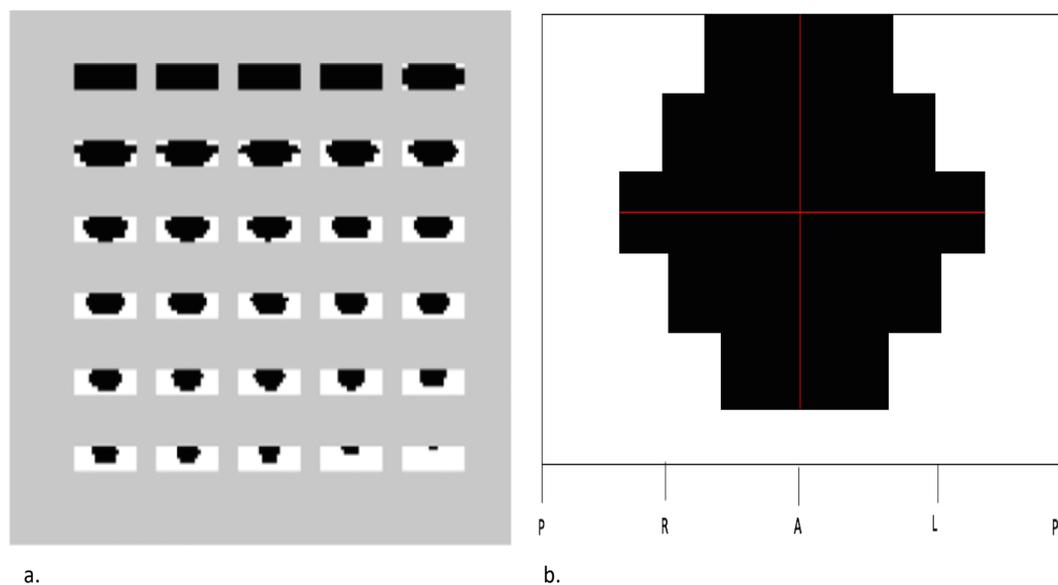


Figure 2.9 Schematic of a set of binary images from a DSM. Figure (a.) shows binary maps corresponding to different dose levels showing the cluster of pixels with a value of 1 (black), and pixels with a value of 0 (white). Figure (b.) shows an example of a binary image with the pixels receiving equal to or greater than a given dose shown in black (1). The lateral and longitudinal extent of the cluster are quantified by measuring the number of pixels at the greatest extent.

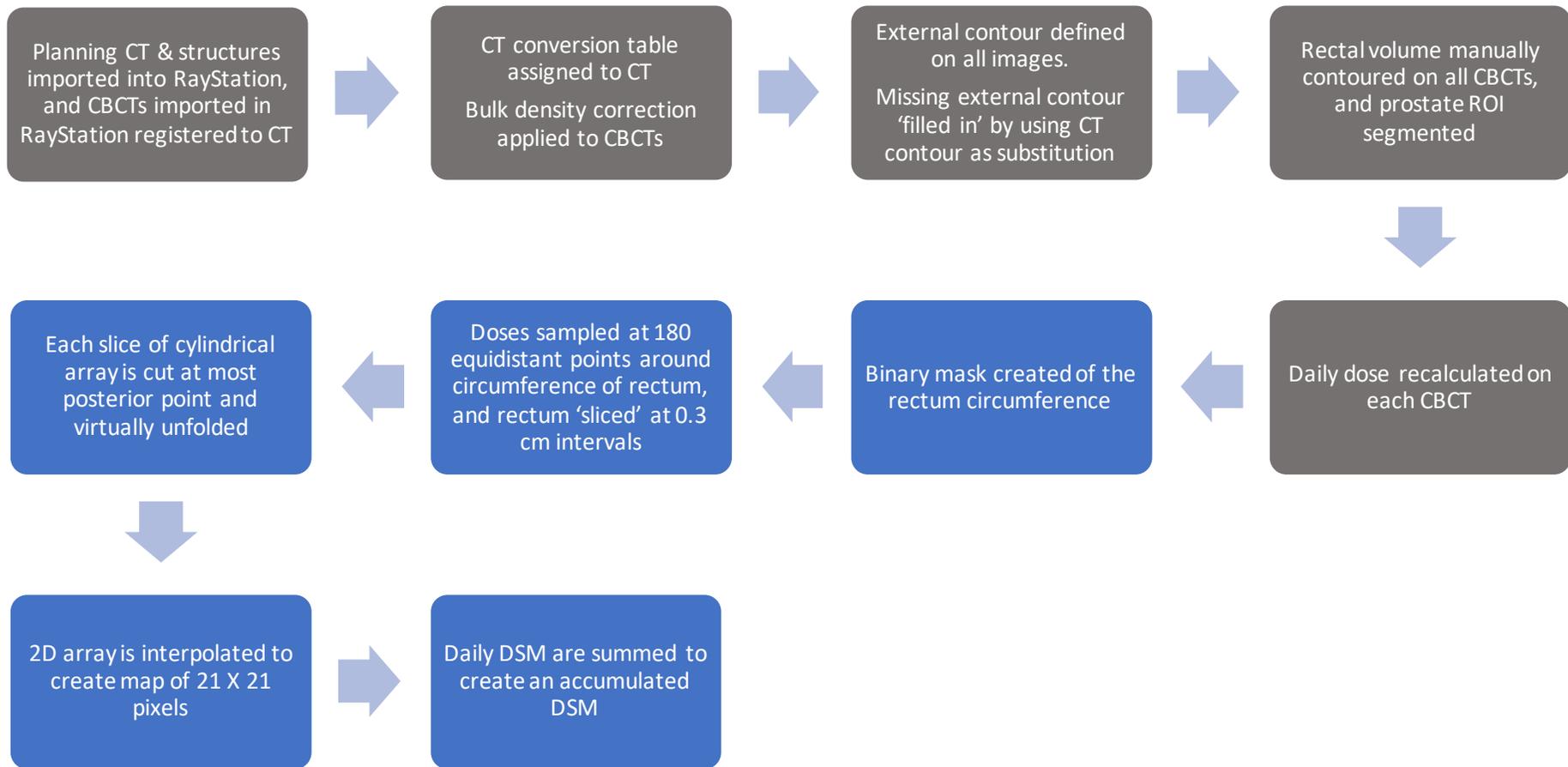


Figure 2.10 Schematic showing the process of accumulating dose using DSM. The initial process (denoted by grey boxes) is also used in DIR based dose accumulation.

2.4 Methods for exploring the dose response

Dose-volume constraints are widely used in radiotherapy to optimise the planning process and reduce the incidence of toxicity [9]. They are evidence-based limits which specify the amount of dose a specific volume of the OAR can receive. In 1991, in the landmark Emami and Lyman paper [74], investigators pooled their judgement and experience in describing whole and partial organ tolerance doses. In 2010, the QUANTEC papers were published which summarised the available reports of association between dose-volume parameters and normal tissue adverse effects [9]. Reported in QUANTEC are dose-volume constraints which define a threshold volume or dose which is likely to limit the risk of toxicity and is used to guide treatment plan optimisation.

To explore the relationship between dose from planned and accumulated rectal DVH and DSM for the data from DELINEATE, univariate analyses was used as well as receiver operating characteristics curves (ROC) and atlas of complication incidence (ACI) [29].

2.4.1 Univariate analyses

Univariate analysis was performed using a Mann Whitney U test to test for a statistically significant difference in rectal volume between patients with and without toxicity, for both the planned and accumulated rectal volume at specific doses of patients that did or did not experience toxicity.

2.4.2 Receiver operating characteristic curves

Receiver operating characteristic curves were used to derive dose constraints by discriminating between patients with and without a given outcome [29, 70, 170, 171]. For a given dose level, thresholds (cutpoints) in the percentage of rectal volume between patients were evaluated to predict presence or absence of a toxicity endpoint. ROC curves were generated by plotting sensitivity as a function of 1 – specificity for all possible cutpoints [172]. The dose levels were analysed at 10 Gy intervals from 10 Gy to 70 Gy. The outcome was the toxicity endpoint of interest (e.g. 1 = Grade 1⁺ diarrhoea, 0 = no diarrhoea experienced). Only ROC with area under the curve (AUC) values greater than 0.6, with the lower 95% confidence interval limit greater than 0.5 were considered as being statistically significant [56, 70, 93]. The Youden index was

used to determine the volume constraint which best discriminates between volumes predicting for toxicity [173].

2.4.3 Atlas of complication incidence

Atlas of complication incidence [174] were used to qualitatively assess the relationship between dosimetry and toxicity [18, 170, 174, 175]. The horizontal and vertical axes of the atlas are the same as for a DVH. The plot is divided into a grid of boxes with each one representing dose in 10 Gy, and a range of 5% of the relative volume of rectum. DVHs are plotted through the grid and a numeric fraction is given for each grid box. Each atlas was based on data for a specific toxicity endpoint and the results are represented as a colour-coded fraction in each box. The denominator is the total number of patients whose DVH passes through each box. The numerator is the number of patients whose DVH passed through the box and experience the reported endpoint. This methodology has been used for visual association of dose-volume parameters with trial outcomes [18, 170, 174, 175]. An example of an ACI taken from Murray et al., is shown below in Figure 2.11 for illustrative purposes. The ACI shows the relationship between dose-volume to the penile bulb and erectile function in the IGRT and reduced margin arm of the CHHiP trial [18].

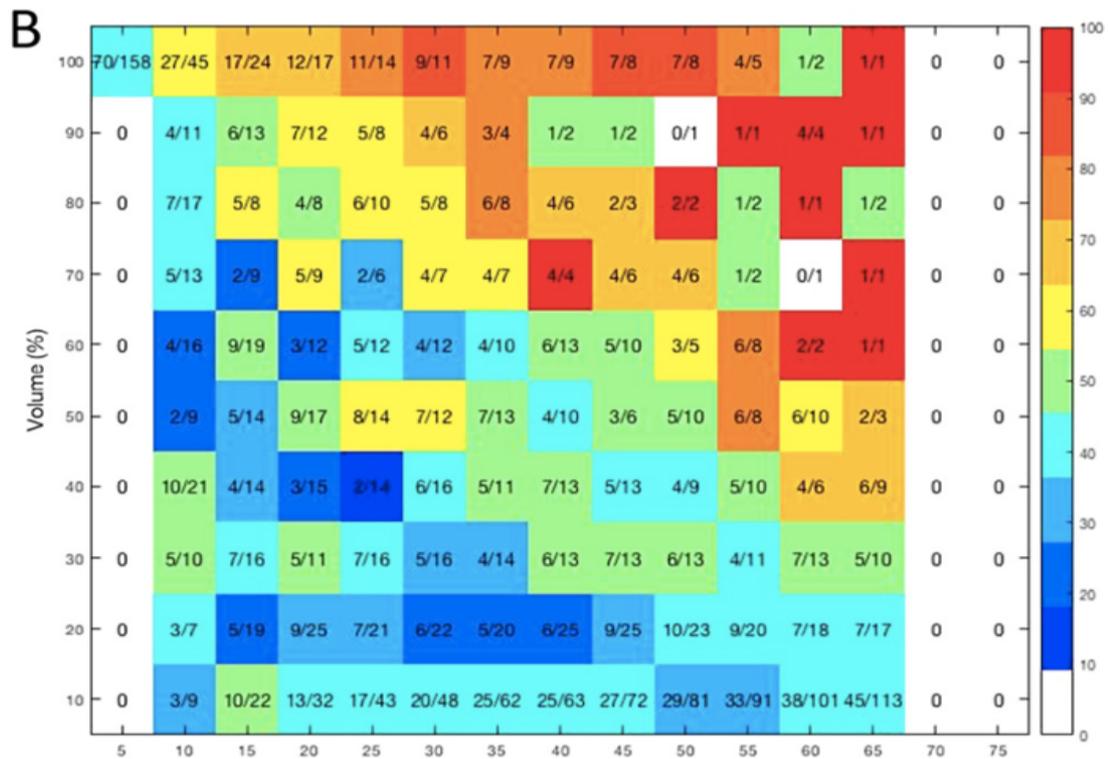


Figure 2.11 ACI showing the relationship between the dose-volume to the penile bulb and erectile potency (courtesy of Murray et al., [18]). The colour of the box is determined by the fractional incidence e.g. orange – red represents 70-100%.

Combining these methodologies, the data from DELINEATE was analysed and reported in Chapter 6 and Chapter 7.

2.5 Statistical analysis

Statistical analysis undertaken in this thesis was performed in SPSS v.25 (IBM Inc., Armonk, NY, USA). Prior to analysis all data was tested for normality using the Shapiro Wilk's test and found to be non-parametric [176], the following non-parametric statistical tests were subsequently used.

As discussed in section 2.4.1 a Mann-Whitney U test is a rank based test which can be used to determine if there is a statistically significant difference between two independent variables. However, where more than two independent variables were compared a Kruskal Wallis H test was used. A Wilcoxon Signed Rank test was used to measure the difference between two dependent variables, and a Friedman test was used for 3 or more dependent variables.

Where multiple testing was performed e.g. association of a defined toxicity endpoint and Vd, an adjustment was applied to the alpha level to reduce the risk of a type 1 error [177]. A Bonferroni correction was applied as part of the statistical test undertaken in SPSS for both the Kruskal Wallis H test and the Friedman test. For all other tests (including MWU and ROC curves), a Holm Bonferroni correction of the alpha level was calculated by the author [178]. Both the uncorrected and corrected p value are presented for each test.

Chapter 3 Evaluation of observer variation in delineation of rectal contours

3.1 Introduction

Organs at risk (OAR) and target volumes are contoured on the planning CT scan as part of the treatment planning process. The contours are used to calculate dose delivered to those structures, and to optimise the dose distribution [179]. Identification of the position and shape of an organ on an image typically relies on the manual delineation of the organ boundary by an observer [144]. This process is subjective, and as such *inter-* and *intra-* observer variation in the delineation of structures is acknowledged in the literature as a source of geometric uncertainty in the delivery of radiotherapy [119, 142]. This is because observer variation can potentially influence tumour control and normal tissue toxicity [180, 181]. In order to calculate the delivered dose to the rectum in this PhD, a crucial step is the accurate delineation of the rectum on the CBCT. The rectal contour is used not only to calculate the daily dose received by the rectum, but as a controlling region of interest (ROI) used for deformable image registration of the CBCT to the CT (see section 2.3.3). Manual delineation of the prostate and OAR on the CBCT presents a greater challenge than on conventional CT due to reduced soft tissue contrast, which may lead to greater uncertainty in delineation and observer variation [122]. This chapter aims to evaluate the uncertainty associated with the delineations of the rectal volume on the CBCTs used in this PhD, and to evaluate whether interobserver variation has an effect on deformable image registration (DIR) and subsequent dose accumulation.

Vinod et al., undertook a systematic review of the uncertainties in volume delineation, and reported that only 21% (25/119) of the studies reviewed evaluated the impact of delineation variation on dosimetry at planning. Of those, only ten studies evaluated the impact of contour variation on the dosimetry of OAR [180]. Only a small number of those studies assessed the impact of observer variation on the planned dose distribution for prostate radiotherapy and the dose to the rectum [159, 182, 183]. Relevant to this work, Livsey et al., reported no significant difference in rectal dose-volumes described by DVHs (V90, V85, V80, V70 or V50) as a result of contour variation between 5 clinicians asked to delineate the rectum on the planning CT [182]. However, some studies did report significant differences in rectal dose-volume

parameters [159, 183]. This difference was commonly attributed to disparities in how observers defined the cranial and caudal extent of the rectum on the planning CT [159, 183]. Further work, however, has shown that training, and the provision of clear contouring guidelines can reduce the variation between observers within and across institutions [184-187].

Observer variation can also differ between anatomical structures. For example, the boundary between two adjacent soft tissue structures can be difficult to discern, and in these cases observer variation has been reported to increase [188-190]. This difficulty in defining organ boundaries can be more challenging if the quality of the image is poor. For example, in the case of CBCT which suffers from poorer image contrast and spatial resolution than CT due to X-ray scatter within the patient that is detected by the CBCT imager [191]. A study by White et al., reported that 5 observers had difficulty agreeing on the location of the peri-prostatic interface when delineating the prostate on CBCTs [191]. However, the largest variations between observers were seen in defining the superior-inferior extent of the prostate (maximum superior deviation 9.7 mm). A similar study was undertaken by Weiss et al., in which 5 observers repeatedly delineated the prostate, rectum and bladder on the CT and CBCTs of 4 prostate cancer patients [188]. The study found that *inter*- and *intra*-observer variation was greater on the CBCT than the CT for all structures except the rectum. The study reported no intermodality difference in patient averaged centre of mass (COV) between the CT and CBCT for the rectum ($p < 0.37$), which suggests that the accuracy of manual delineation of the rectum on CBCT is comparable to delineation of the rectum on a CT.

There is no common consensus on the metrics used to measure observer variation, or the minimum number of observers or images required [162]. Commonly reported metrics include volume comparisons (volume measurements), measures of overlap (e.g., concordance index or Dice similarity coefficient (DSC)), or distance measures (e.g., mean surface distance or Hausdorff distance) [180]. Using a combination of metrics is desirable as no one metric can fully characterise the variation. For instance, two contours may have equal volume and shape, but may not encompass the same region of tissue. Or two contours may have similar shape but their volumes may differ [162]. DSC and MDA may be combined to compare the similarity in the shape and dimensions of contours [135], and both were employed in this study to evaluate

variation in observer delineation of the rectum (see section 2.3.2.1 for description of DSC and MDA).

Manual delineation remains the standard method of organ delineation [192] for a large number of anatomical structures. However, this can be very time consuming, which has stimulated significant research into automated segmentation [144, 164, 193]. To transition to using multiple CBCTs for dose accumulation and treatment adaptation, automatic segmentation models (ASM), which can quickly and accurately (with respect to a manually defined gold standard) delineate structures, are required [128]. Three automatic segmentation methods were available in RayStation for use in this study and were evaluated for the automated segmentation of the rectum.

The first automatic segmentation method is an Atlas based segmentation model, which uses *a priori* knowledge of the size and shape of the structures to be segmented. An Atlas is a library of reference images on which structures have been manually delineated and may contain a single image or multiple images from one or more patients. Segmentation of corresponding structures on a new image is obtained by finding the optimal transformation between the images in the Atlas and the new image [179]. The algorithm will search the Atlas for an image which best matches the image to be segmented, and the contours are then propagated to the image using DIR [194]. The performance of the Atlas based segmentation is influenced by the precision of the delineated volumes on the reference image, as well as the image quality and the similarity of the anatomy in the new image, and the performance of the DIR algorithm [144].

The second automatic segmentation model is a shape based model segmentation (MBS), which also uses *a priori* knowledge. RayStation provides a catalogue of model organ shapes for different sites based on manual contours delineated by experts (depending on the anatomical site, between 10 and 50 contour sets are used to create the model) [158]. The desired organ shapes are manually selected by the user from the catalogue. RayStation then automatically ‘drops’ the organs in approximate positions using a registration algorithm [195]. The algorithm then adapts the model by deforming the shape to match the patient’s anatomy, which results in a contour [158, 195].

The third method uses the rectal contour delineated on the CT as *a priori* knowledge. The deformation between CBCT and the CT is calculated using ANACONDA to generate a 3D deformation vector field (described in detail in Section 2.3.3). By reversing the direction of the vectors, and applying the reversed vector field to the CT, the CT rectal contour is deformed to the same shape of the rectum imaged on the CBCT. Segmentation accuracy is affected by the accuracy and precision of the initial contour on the CT scan, as well as the image quality. The performance of the DIR may also be subject to interobserver contour variation when deforming one image to another.

3.2 Aims

The aims of this chapter were to evaluate the consistency of rectal contouring by comparing manual contours of different observers, to evaluate manual and automated approaches, and to find approaches to minimise inconsistencies in rectal contouring.

The work described in the chapter had the following objectives:

- 1) To quantify the interobserver variation (IOV) between experts' manual contours of the rectum delineated on CBCT. This will provide a benchmark against which to measure the author's contours. Intraobserver variation will also be measured to quantify the repeatability of the author's contours.
- 2) To evaluate the performance of three ASMs (Atlas, MBS and contour propagation using DIR) by comparing automatically segmented contours of the rectum to the expert manually delineated contours.
- 3) To evaluate the effect of contour variation on the performance of the deformable image registration and the subsequent variation in accumulated dose.

3.3 Methods and Materials

3.3.1 Patient data

The first five patients treated with hypofractionated radiotherapy were chosen from the study population. For each patient, five CBCT datasets were randomly selected using a random number generator (Excel, Microsoft Corporation, Redmond, WA, USA). A density correction was performed enabling dose to be calculated directly on the CBCT (described in Chapter 2.3.1).

3.3.2 Segmentation

3.3.2.1 Manual delineation:

In addition to the author (ADA), three experts were asked delineate the rectum on CBCTs (Exp1, Exp2 and Exp3). All three experts were Consultant Radiation Oncologists at the RMH, specialising in uro-oncology. Observers delineated the rectum on each of the 25 CBCTs using the DELINEATE trial delineation criteria described in Section 2.2.2. To measure intraobserver variability, the rectum was delineated a second time four weeks later by ADA.

3.3.2.2 Automatic Segmentation:

To evaluate the accuracy of the ASMs available in RayStation, rectal contours were automatically segmented on the same subset of patient data. The methodology was as follows:

- 1) Atlas based segmentation: The manually delineated rectum of five patients from the hypofractionated cohort were used to build the Atlas (these five patients were not those included in the IOV study). The patients selected had an image for all 20 fractions. The author manually delineated the rectum on all 20 CBCTs and the planning CT using the DELINEATE trial delineation criteria. In total 5 CTs and 100 CBCTs were used to create the Atlas (in total 105 images), which was subsequently used to automatically segment the rectum on the CBCTs in the IOV study.
- 2) Model based segmentation (MBS): For each CBCT, the author manually selected the rectal shape from the male pelvis model provided in RayStation.

The same model was used for all 5 patients. The application superimposed the rectal shape onto the image, and then deformed the rectal shape automatically using the statistical model shape. The adaption of the model to match the patients' anatomy on the CBCT utilises a combination of greyscale gradients and shape statistics [157].

- 3) Contour propagation using DIR: The rectal contour from the relevant CT scan was propagated to the CBCT using the deformable image registration algorithm ANACONDA. First, the CBCT was deformed to the CT. No controlling ROI was used other than the external contour (this is described in Section 2.). The deformation vector fields which describe the deformation of the CBCT to the CT were reversed in order to propagate the CT rectal contour to the CBCT.

3.3.3 Measuring contour variation and evaluating automatic segmentation models

The volume of the rectum, DSC and mean distance to agreement (MDA) were used to measure the variation between contours. To measure the interobserver variation between the three experts, a pairwise comparison was used i.e., the contours from Exp1 and Exp2, Exp1 and Exp3, and Exp2 and Exp3 were compared. The author's contours were then compared to each expert in turn. The intraobserver variation of the author was measured by pairwise comparison of repeat contours delineated four weeks apart. Pairwise analysis using DSC and MDA was also used to evaluate the performance of the automatically segmented rectal volumes by comparing them to the expert contours.

3.3.4 Measuring the effect of contour variation on deformable image registration and dose accumulation

To measure differences in the DIR due to variation in delineation of the controlling ROI (i.e. rectal contour), each CBCT was deformed to the CT using each observer's contour as the controlling ROI in turn. Specifically, the differences were measured by determining the conformity of the deformed CBCT rectal contour to the CT rectal contour, and the difference in accumulated rectal dose. For each CBCT, this provided four deformed images, one for each observer contour. The difference in the performance of the deformable image registration was analysed using pairwise

comparison between the original CT contour and each deformed contour (one for each observer) using DSC and MDA to measure similarity. DIR was performed using ANACONDA, as described in section 2.3.3.

For each deformed CBCT (one for each observer) the daily dose was mapped to the CT using the deformation vector field (as described in section 2.3.3). Dose-volume histograms (DVHs) were created for the rectum for each of the four deformed dose distributions. For subsequent work in this thesis, if the deformed rectal contour had poor agreement with the CT rectal contour, the corresponding CBCT was excluded from accumulation. This method was used to ensure only DIR which were anatomically reasonable, and had a good agreement with the reference image, were included when accumulating dose. This approach was also adopted in the current analysis. Therefore, if a deformed rectal contour had a DSC < 0.7 or MDA > 0.3 cm, the deformed dose was excluded from analysis of observer variation (see section 2.3.2.1). To enable comparisons, the same CBCT was excluded for all observers. A schematic of the methodology used for each objective is given below in Figure 3.1. As the ground truth was not known, a reference contour from which to measure the other contours against was selected [162]. Exp1 had the greatest experience and therefore their contours were used as the gold standard for comparative purposes. The remaining observers delineated volumes were subtracted from Exp1 to calculate the difference in volume at 10 Gy intervals (10 to 60 Gy).

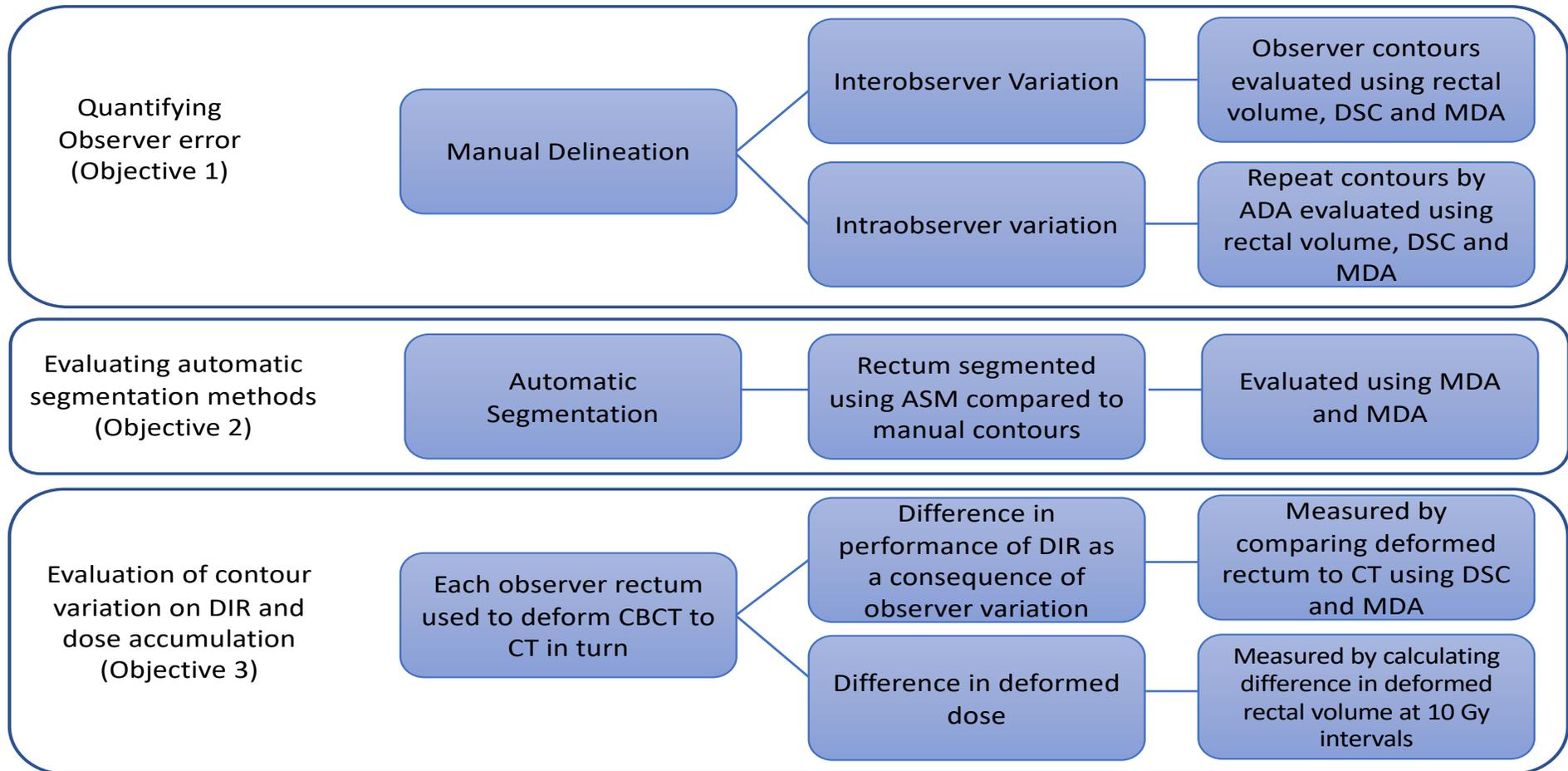


Figure 3.1 Schematic of the methodology used to achieve each objective.

3.4 Statistical analysis

3.4.1 Analysis of contour variation and automatic segmentation models

To quantify *inter-* and *intra-* observer variation, the difference in the median volumes defined by the observer delineated contours were calculated and tested for statistical significance using the Kruskal Wallis H test. The variation between expert contours was measured using DSC and MDA, and by calculating the mean pairwise agreement, i.e., Exp1 compared to Exp2, Exp1 compared to Exp3 and Exp2 compared to Exp3. This was used as a benchmark to evaluate the accuracy of ADA's delineation. This was assessed by measuring the mean pairwise agreement between ADA's contour and each expert contour in turn using DSC and MDA, i.e., ADA to Exp1, ADA to Exp2 and ADA to Exp3. A Mann Whitney U test was used to determine if there was a statistically significant difference in the median DSC and MDA between the two groups.

The variation in expert contours was also used as a benchmark to assess the performance of the three ASM methods (Atlas, MBS and contour propagation using DIR). The contours generated from each of the three ASM tested were compared to each of the experts in turn using pairwise comparison of MDA and DSC. Finally, a Wilcoxon rank sum test was used to determine if there was a statistically significant difference in the DSC and MDA between the pairwise comparisons of Exp1, Exp2 & Exp3 to DIR, and Exp1, Exp2 & Exp3 to ADA.

3.4.2 Analysis of contour variation on dose accumulation

Pairwise agreement of DSC and MDA was also used to measure the variation in the rectum after the CBCT was deformed using each observer contour as the controlling ROI. To measure the variation in dosimetry between the deformed rectums, a boxplot of the difference in rectal volume between observers was generated, with the volumes of observers subtracted from Exp1. A Friedman test with post hoc analysis was used to test for statistical difference in the rectal dose-volume between observers at specific doses.

3.5 Results

All data presented below were found to be non-parametric ($p > 0.05$).

3.5.1 Comparison of manual contours

3.5.1.1 Volume of delineated rectum

The volumes encompassed by the delineated contours are shown for each observer in a boxplot in Figure 3.2.

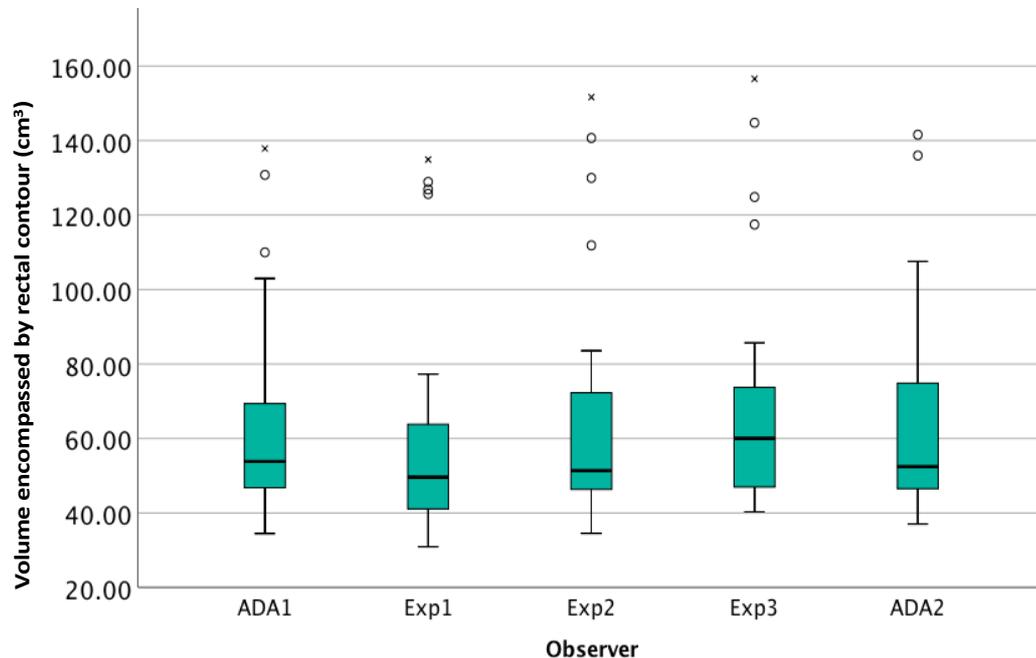


Figure 3.2 Boxplot of rectal volume delineated by each observer. Boxplot of the volume encompassed by the rectal contour delineated by each observer, including both first and second attempts by ADA (ADA1 = 1st attempt and ADA2 = 2nd attempt). Exp1, Exp2 and Exp3 refer to the experts. The box represents the interquartile range (IQR) and the median is denoted by the central bold black line. The whiskers indicate the lowest and highest values which are no greater or less than 1.5 times the IQR. Outliers which are between 1.5 to 3 times the IQR are denoted by circles. Extreme outliers with value greater than 3 times the IQR are denoted by crosses.

There was no statistically significant difference in the volume of the delineated rectal contour between expert observers ($p = 0.22$, Kruskal Wallis test), and no statistically significant difference between the volume of the rectal contour delineated by ADA when compared to expert observers ($p = 0.34$). A Wilcoxon rank sum test showed no statistically significant difference in the volume of the rectal contour between ADA's first and second contoured volumes ($p = 0.76$).

3.5.1.2 Similarity metrics

Boxplots of DSC and MDA for the pairwise comparison between ADA contours and each of the experts, and between experts only is shown below in Figure 3.3 and Figure 3.4.

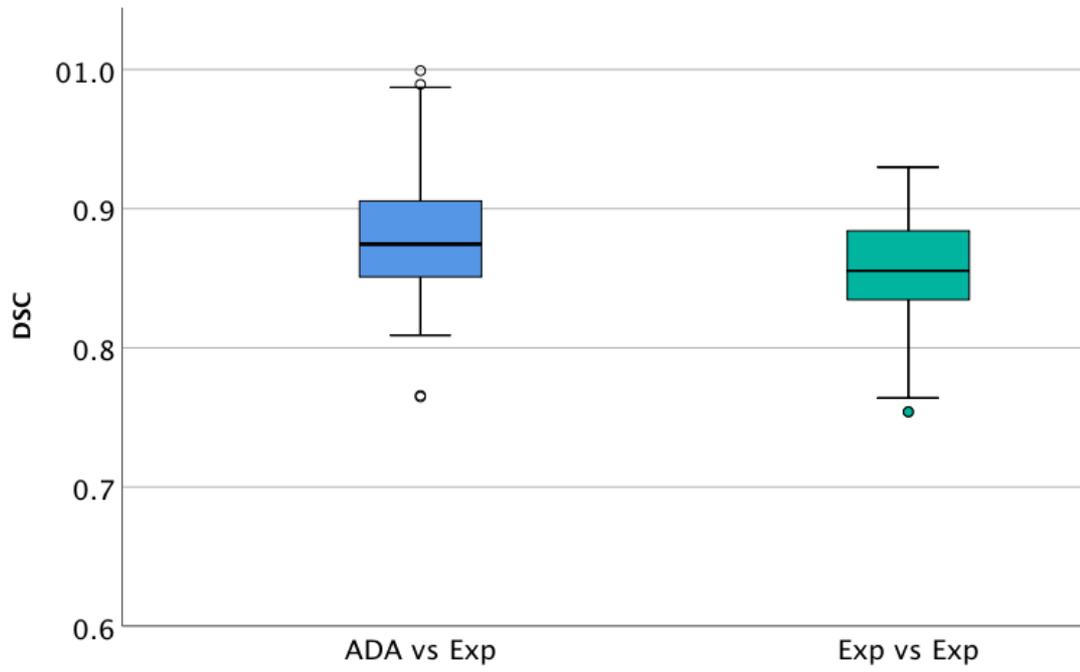


Figure 3.3 Boxplot of Dice similarity coefficient (DSC) between contours delineated by pairs of observers. ADA vs Exp are DSC values obtained when contours delineated by ADA1 were compared to contours delineated by each Expert for each CBCT image. Exp vs Exp are DSC values obtained when contours delineated by each expert were compared to contours delineated by each other expert for each CBCT image.

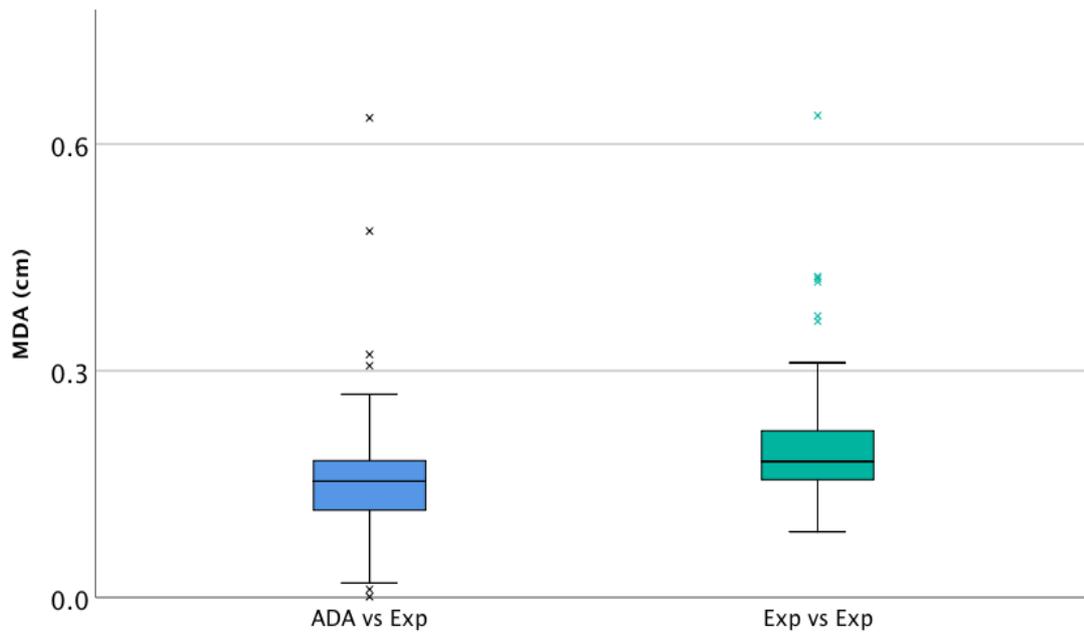


Figure 3.4 Boxplot of mean distance to agreement (cm) between contours delineated by pairs of observers. ADA vs Exp are MDA values obtained when contours delineated by ADA1 were compared to contours delineated by each expert for each CBCT image. Exp vs Exp are MDA values obtained when contours delineated by each expert were compared to contours delineated by each other expert for each CBCT image.

The median and interquartile range (IQR) of the DSC between contours delineated by ADA and experts was 0.87 (0.05), and between experts only was 0.86 (0.05). The median (IQR) MDA between contours delineated by ADA and Experts was 0.15 (0.07) cm, and between experts only was 0.18 (0.08). There were statistically significant differences between the median DSC between ADA vs. Experts, and Experts vs. Experts ($p < 0.05$), for both DSC indicating that on average ADA's contours had greater agreement with the expert contours, than the experts had between their contours.

ADA intraobserver variation was measured by pairwise comparison between contours from repeated delineation. The median (IQR) DSC was 0.90 (0.02), and MDA was 0.11 (0.02). Intraobserver variation was statistically significantly smaller than interobserver variation for both DSC and MDA (Kruskal-Wallis test $p < 0.005$).

3.5.2 Comparison of rectal contours generated by automatic segmentation models

Boxplot of the pairwise comparison between experts and ASM using DSC and MDA are shown below in Figures 3.5. and Figure 3.6.

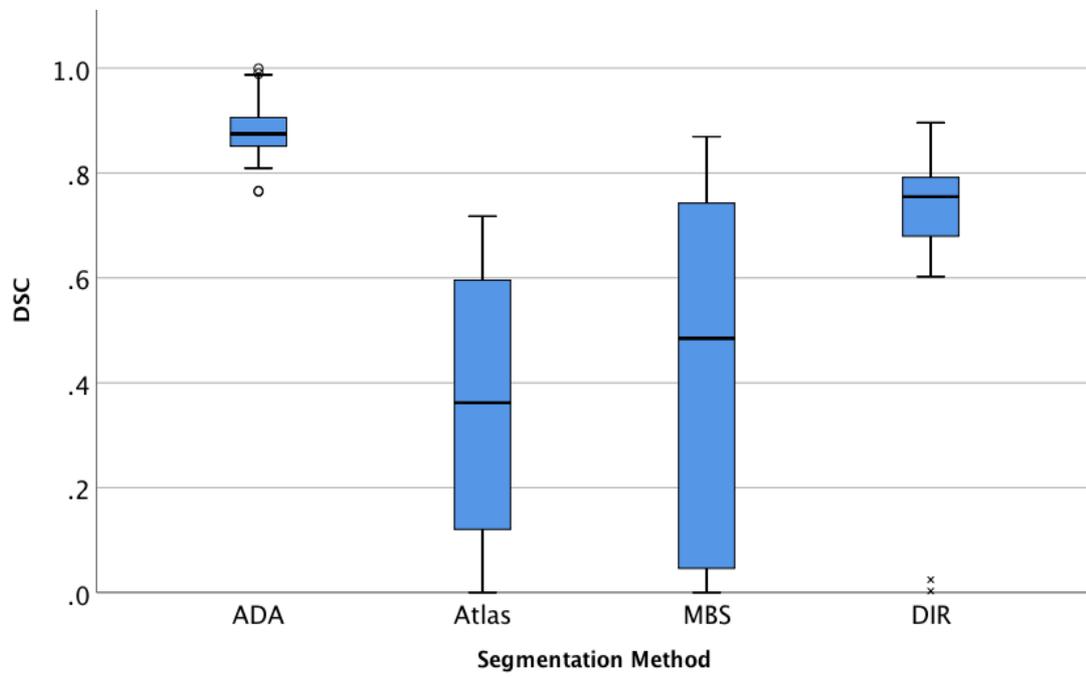


Figure 3.5 Dice similarity coefficient (DSC) from pairwise comparisons between ADAs manually delineated rectal contours and expert contours, and contours automatically generated and expert contours. Rectal contours were automatically generated using Atlas (Atlas), model based segmentation (MBS) and contour propagation using deformable image registration (DIR)

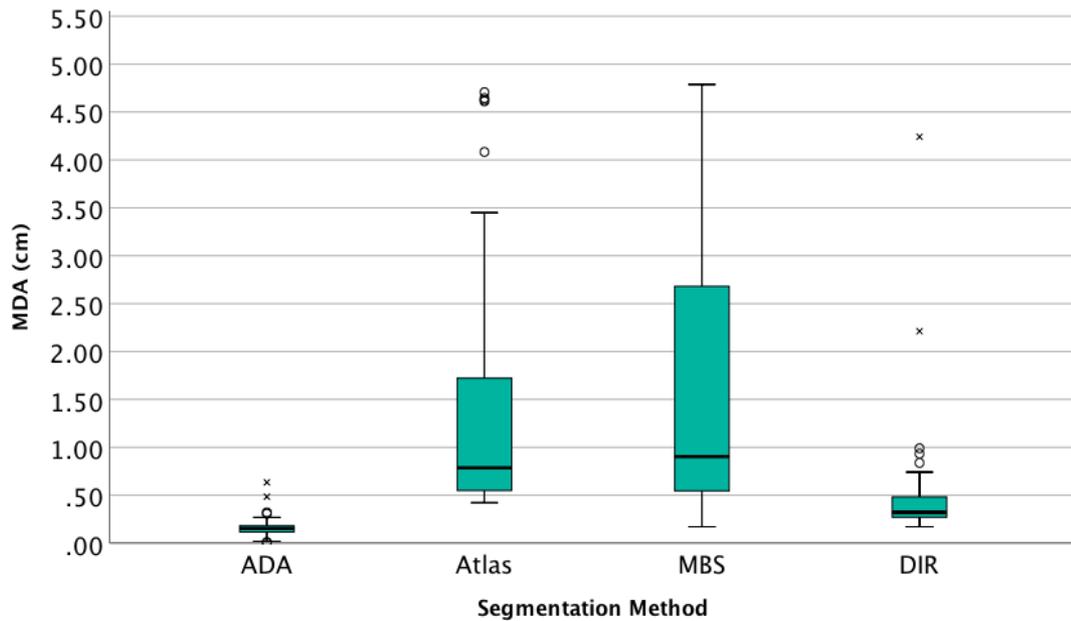


Figure 3.6 Boxplot of mean distance to agreement (MDA) measurements from pairwise comparison between experts and ADA (ADA), manually delineated expert contours and contours automatically generated contour. Rectal contours were automatically generated using Atlas (Atlas), model based segmentation (MBS) and deformable image registration (DIR)

The median (IQR) DSC for comparisons between Experts and ADA was 0.87 (0.05), Experts and Atlas was 0.36 (0.47), Experts and MBS was 0.48 (0.70) and Experts and DIR was 0.75 (0.11). The median (IQR) MDA for comparisons between Experts and ADA were 0.15 (0.07), Experts and Atlas was 0.71(1.10), Experts and MBS was 0.69 (2.21), Experts and DIR was 0.32 (0.21),

Visual inspection of the box plots and descriptive statistics showed that ADA contours had better agreement (greater DSC and less MDA) with experts' contours than Atlas and MBS methods. ADA contours had statistically significantly better agreement with experts compared to contours generated using DIR for both DSC and MDA ($p < 0.05$, Wilcoxon Signed Rank test).

3.5.3 The effect of interobserver variation on deformable image registration and calculation of dose

3.5.3.1 Similarity Metrics

DSC and MDA values measured between the deformed rectum and the CT for each observer are given in Figures 3.7 and 3.8.

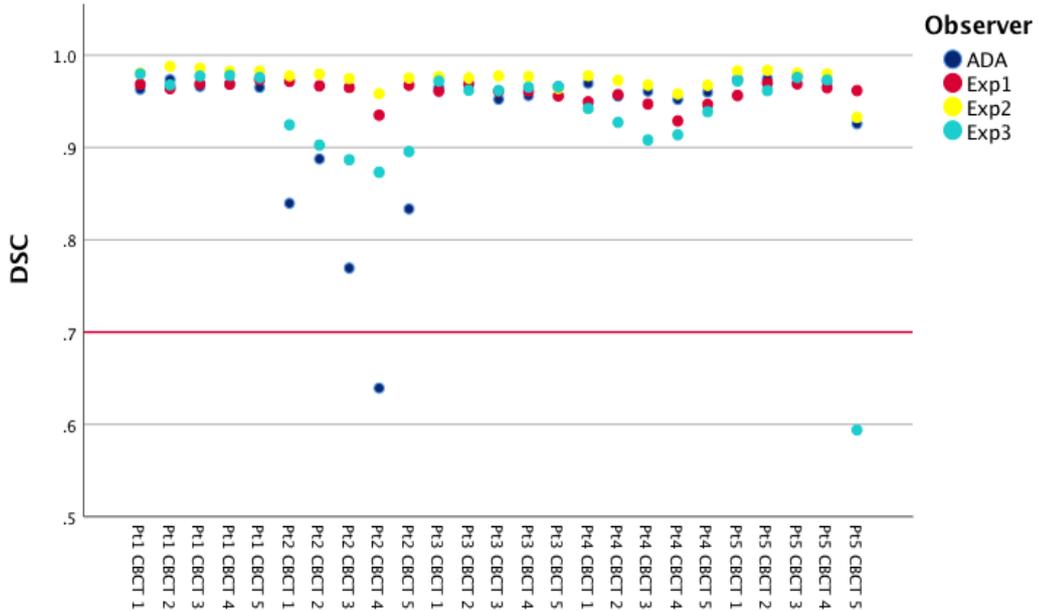


Figure 3.7 Dice Similarity Coefficient (DSC) for each CBCT from pairwise comparison between rectal volume segmented on the CT scan and the deformed rectal volume created when each observer contour was used to deform the CBCT. DSC measurements above 0.7 indicate good agreement between contours (red line).

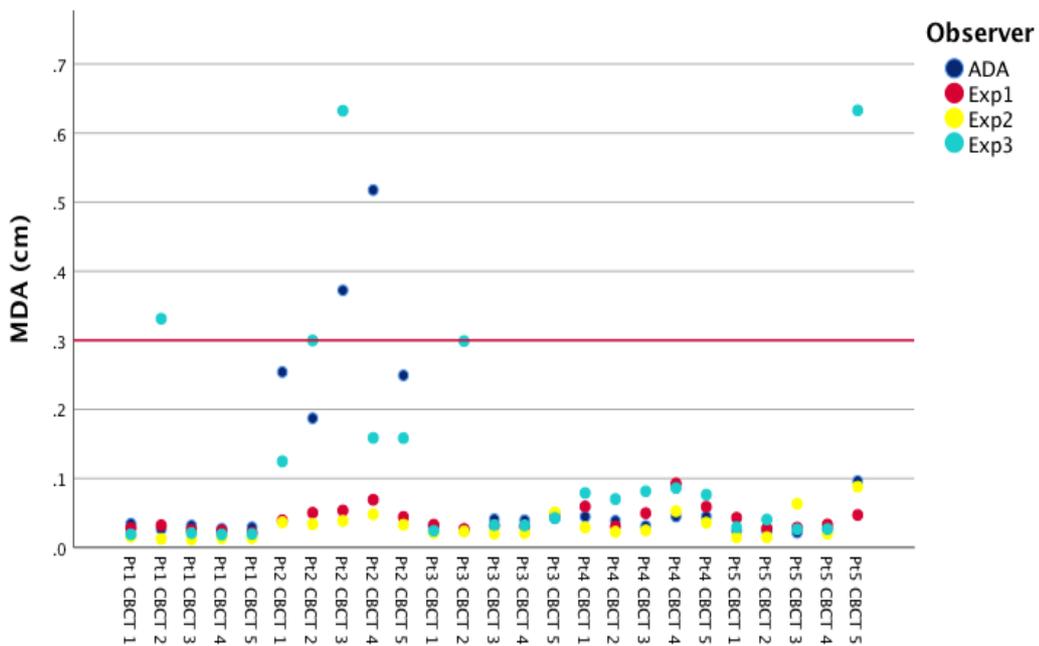
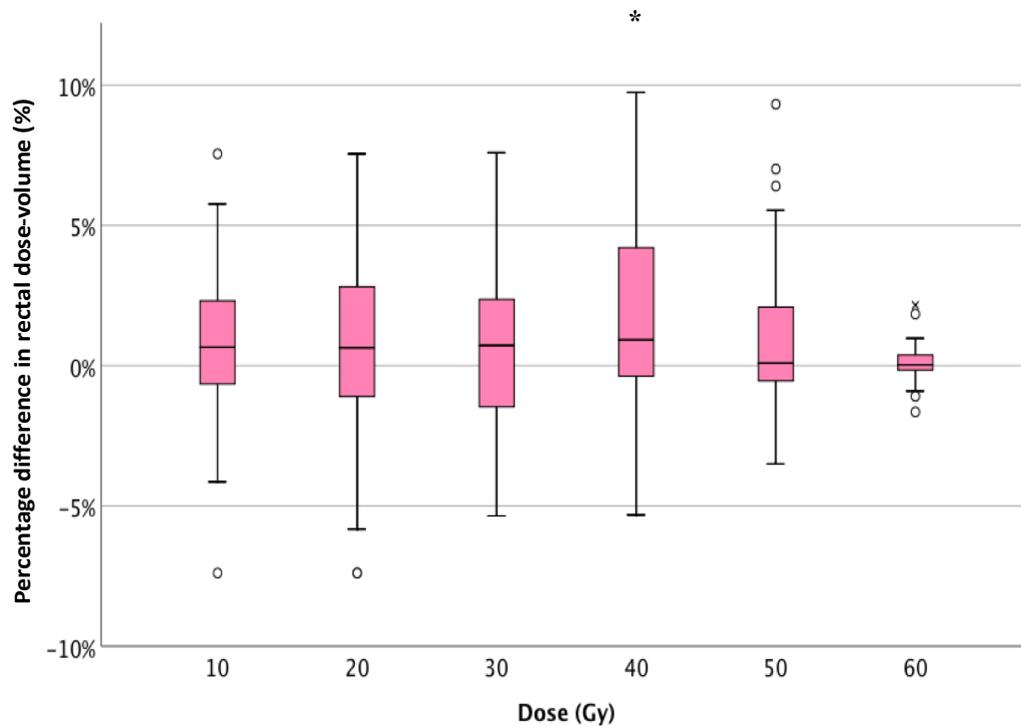


Figure 3.8 Mean distance to agreement (MDA) for each CBCT from pairwise comparison between rectal volume segmented from the CT scan and the deformed rectal volume created when each observer contour was used to deform the CBCT. $MDA \leq 0.3$ (red line) is seen to show good levels of agreement between contours.

3.5.3.2 Effect of interobserver variation on delivered rectal dose

A boxplot of the difference in rectal dose-volume between Exp1 (defined as gold standard for comparative purposes) and ADA, Exp2 and Exp3 is given below.



*Figure 3.9 Boxplot of percentage difference in dose-volume between Exp1 and remaining observers. The difference in rectal dose-volume between Exp1 and ADA, Exp2 and Exp3 was analysed at 10 Gy dose intervals from 10 Gy to 60 Gy. Statistically significant differences in the volume between Exp1 (gold standard) and observers (prior to correction for multiple testing) are highlighted with *.*

The difference in dose-volume at 10 Gy intervals between observers was tested using a Friedman's test with post hoc analysis. There was no statistically significant difference in the rectal dose-volume between the observers at 10 Gy ($p = 0.06$), 20 Gy ($p = 0.15$), 30 Gy ($p = 0.53$), 50 Gy ($p = 0.08$) and 60 Gy ($p = 0.49$). A statistically significant difference was found at 40 Gy ($p = 0.009$). Post hoc analysis revealed a statistically significant difference between the volume of rectum at 40 Gy between Expert 1 and ADA ($p = 0.005$ with Bonferroni correction).

3.6 Discussion

The author (ADA) had no previous experience of manual delineation of any structures. The aim in this chapter was therefore to evaluate the quality of the delineations used

in this PhD by comparing them with expert delineations of the rectum. In addition, there was concern that the poorer soft tissue contrast of the CBCTs would result in greater observer variation when delineating the rectum. The results showed a good level of agreement in nearly all the rectal contours delineated by experts suggesting that image quality of the CBCTs did not increase observer variation between experts. This is in keeping with the study by Weiss et al., who reported observer variation for the rectum to be a similar magnitude on CBCT as for CT [188]. Most importantly, there was also a good level of agreement between the rectal contours delineated by ADA and the experts. Comparison of the rectal volume showed no statistically significant difference between observers ($p = 0.22$). Visual analysis of boxplots, and the median and IQR of DSC and MDA, showed that there was actually better agreement between ADA and the experts than between the experts only. When the two groups were compared using Man Whitney U, there was a statistically significant difference ($p < 0.05$). This may be explained because two of the experts (Exp1 and Exp2) trained the author. Exp3 may have different outlining practices to the other two expert observers, and by extension ADA. There was some evidence of this, Exp3's contours tended to differ more from Exp1 and Exp2, than Exp1 and Exp2's contours did with each other (see Figure 3.7 and 3.8). Intraobserver variation measured by repeat delineation of the rectum by ADA was shown to be smaller than interobserver variation between experts. This is expected from the literature [159, 162] and provides confidence in the author's ability to consistently contour the rectum on the CBCT. However, it should be acknowledged that intra-observer variation was only measured using two time points. Although there are studies in the literature which also use two time points to measure intra-observer variation [159, 196], reproducibility of the contours would be more reliably measured had a third time point been used.

Although overall there was a good level of agreement between observers, there were five specific images where there was poor agreement (defined as $DSC < 0.7$ and $MDA > 0.3$ cm) between one or more observers. There was a good level of agreement between Exp1 and Exp2 for all images. There were two images where there was poor agreement between ADA and the experts (patient 2, images 3 and 4). There was poor agreement between Exp3 and the other Expert observers for patient 1 (image 2), patient 2 (image 2 and 3), and patient 3 (image 2). Visual analysis of these images showed that poor agreement was due to variation in the delineation of the superior length of the rectum. Review of the literature showed that variation in defining the

length of the rectum is a common source of uncertainty which can result in variation in dose-volume parameters [99, 159, 183, 188]. As a result of this study, ADA received further training in defining the superior extent of the rectum.

Manual delineation of structures is still considered to be the gold standard in both research and clinical practice. However it is time consuming [179]. Methods of automatically segmenting structures have the potential to reduce observer variation as well as the clinical burden of delineation [193]. This PhD requires the delineation of the rectum on over two thousand CBCTs. In order to reduce the time required to delineate the rectum, three ASM available in RayStation were evaluated. The contours generated using the three ASM were compared to the experts' contours using pairwise agreement. Visual inspection showed that there was poor agreement between the contours automatically segmented and those delineated by the experts. The MBS method and Atlas based segmentation model performed the worst; contours propagated using DIR gave the greatest agreement with experts.

A small study by Delpon et al., compared different commercially available ASM in contouring the prostate bed and OAR on ten CT scans [197]. The software systems compared were Workflow Box (Mirada Medical Ltd, Oxford, UK), MIM Maestro (MIM Software Inc., Beachwood, OH, USA), SPICE (Philips AB, Eindhoven, Netherlands), ABAS (Elekta AB, Stockholm, Sweden) and the Atlas based segmentation model from RayStation. The ASMs performed well for high contrast structures such as femoral heads, but the algorithms struggled with segmenting structures in areas where there was poor soft tissue contrast such as the rectum. The study reported that RayStation's Atlas model had mean (SD) DSC scores for the rectum of 0.51 ± 0.17 . This result is similar to this study, which reported mean and IQR 0.36 (0.47) when the Atlas model was compared to the experts' delineations. An acknowledged possible limitation of the Delpon et al. study is that only 10 images were used to create the Atlas library when RayStation recommends using at least 20 images [158]. For this PhD, the library was created using the CT scans and CBCTs for 5 patients (hypofractionated cohort), which created a library using 105 images. Although this is a greater number of images than recommended by RayStation, the images came from a small number of patients. The Atlas model works by creating a library of clinical scenarios. The best matching Atlas is rigidly registered to the CBCT, and the structures are deformably registered. Using only a small subset of patients for

the Atlas may have resulted in limited variation in shape and volume of the rectum, resulting in limited options for the Atlas. Finally, the accuracy of the manually delineated contours used to create the Atlas library could also affect the integrity of the contours generated by the Atlas. It is unclear why the MBS performed badly. It may be that because the MBS model was generated using CT scans [158] it was affected by the poorer soft tissue contrast on the CBCTs.

The most promising method of automatic segmentation was the use of DIR to propagate the contours from the planning CT to the CBCT. When the propagated contours were compared to the experts, the median (IQR) DSC scores were 0.75 (0.11) and MDA 0.32 (0.21). However, the agreement of the authors contours with the experts, was better both in terms of average and dispersion of DSC (0.87(0.05)) and MDA (0.15(0.07)). A study by Gardner et al., reported a 10 % decrease in agreement of DSC when propagating rectal contour from CBCT to subsequent fractions using DIR [198]. Thor et al., found that CBCT image quality affected the propagation of contours using DIR, with mean DSC for rectum 0.73 (range 0.34 to 0.87) [199]. Propagation using DIR provides a good starting place but all the contours required further editing which is not unusual [198]. Visually checking and editing the rectal contours was found to be almost as time consuming as manual delineation. It has been previously shown that accumulated dose accuracy is influenced by contour accuracy [148]. Manual delineation by the author was shown to be superior to propagation of contours by DIR, which may reduce the introduction of further uncertainty in dose calculation. The author therefore decided to manually delineate all rectal volumes for this PhD study.

Validation of the deformable registration on a voxel by voxel basis is beyond the scope of this PhD. However, the effect of contour variation on the deformed CBCT rectal contour was shown to be small. The CBCT rectal contours delineated by Exp1 and Exp2 showed an excellent level of agreement with CT after deformation by DIR for all images with DSC > 0.9 and MDA < 0.1 cm. For the author, the deformed rectal contours on two images (patient 2, CBCTs 3 and 4) had poor agreement with the CT. Exp3 had three CBCT rectal contours which showed poor agreement with the CT (patient 2 CBCTs 3 and 4, and patient 5 CBCT 5). These were the same images which had shown poor agreement with other expert observers i.e., when used for the interobserver variation study. The poor agreement was likely to be as a result of

variation in defining the superior extent of the rectum which was observed in these images. This incorrect delineation of the superior extent of the rectum is likely to have resulted in poor performance of the DIR as it struggled with correspondence ambiguity. This can occur when the registration is unable to establish one to one physical correspondence between images due to large anatomical differences between the images [128]. Or it may simply be because the rectum delineated on the CBCT was incorrect. Wen et al., also reported large distortions in the superior extent of the rectum when the CBCTs were deformed in five patients [140]. It was suggested that the distortions were a result of poor CBCT image contrast. No details regarding the length of the rectal contour on the CBCT compared to that of the CT were given. Difficulties in defining the rectal length could have led to large differences resulting in poor registration as shown in this study.

Poor performance of the DIR can lead to inaccuracies when accumulating the dose [142]. It is difficult to validate DIR because there is no comprehensive ground truth. It is therefore recommended that users should not only visually inspect the resulting transformation for errors but also employ quantitative methods of verification such as DSC and MDA [142]. Images which have been incorrectly deformed, as determined by contour conformality, were excluded from the above analysis and will be for all subsequent dose analysis. The differences in dosimetry between deformed rectal contours was analysed by calculating the percentage of rectal volume at given doses (10 Gy intervals). Friedman's test with post hoc analysis showed no statistically significant difference except for a small statistically significant difference (maximum 9.13%) at 40 Gy between the ADA and Exp1. The author could not determine any obvious reason for this difference. A study by Nassef et al., measured the effect of poor registration on dose accumulation in the rectum [148]. A numerical phantom was used to simulate deforming organs in 24 patients. The simulated images were deformed using contour based DIR and showed differences of mean (SD) rectal volume of 0.7 ± 0.3 % for V70. The study concluded that small registration errors in high dose region can result in large differences in dose. The high dose region for the rectum is at the prostate rectal interface. However, this has been shown to be an area where there is least contour variation between observers [181]. Visual inspection of the deformed contours showed good agreement between the experts at the prostate rectal interface in this study.

3.7 Limitations of this study

The primary purpose of this initial study was to evaluate whether the author could precisely contour the rectum on the CBCT prior to delineating the rectum on the population used in the PhD. As a result only 5 patients were chosen from the PhD population for the inter-observer study. The author acknowledges that this is a small sample size, and the inter and intra-observer study would have benefitted from a larger number of patients.

This study evaluated the effect of contour variation on deformable image registration and dose calculation. This was because the contour is used to deform the image, and the resulting deformation is used to deform the dose. A poor registration can therefore introduce uncertainty into calculation of dose. This study did not look at the effect of observer variation on DSM, nor could the author find any reference to this in the literature. DSM are generated by sampling the dose along the circumference of the rectum. It was assumed that the DSM would therefore be influenced by the delineation of the rectum. As the author's contour delineation is comparable to the experts and there were very small differences in the dose-volumes no further analysis was performed.

3.8 Conclusion

This study has shown that the author's delineation of the rectum on CBCT was comparable to expert delineations. The ASM available in RayStation did not result in anatomically rigorous segmentation of the rectum. Manual delineation, despite the time-consuming nature, will continue to be used for delineation of the rectum on CBCT for this PhD. Future work could measure the effect of observer variation on dose surface maps, as well as the effect of contour variation of deformable image registration and subsequent dose accumulation.

Chapter 4 The effect of missing data on the calculation of accumulated dose

4.1 Background and Introduction

In order to accumulate dose accurately, a description of the patient's anatomy at the time of treatment is required, allowing the dose at each time point to be calculated [129, 134]. In this PhD, daily CBCTs from patients treated in the DELINEATE study were used to provide a description of the patient anatomy at the time of treatment. The CBCTs needed to be of sufficient quality to visualise the rectum for delineation, and for accurate deformable image registration to the planning CT (see section 2.3.3). In the early part of the project, it was appreciated that some CBCT images were not of ideal image quality, or had a small field of view (FOV) limiting the length of the rectum imaged. Reasons for sub-optimal quality were collated, methods to mitigate deficiencies in FOV were explored, and the impact of missing images on the robustness of analysis of dose accumulation studied. Ideally, the imaging FOV should also encompass all of the anatomy of interest i.e., the inclusion of the entire length of the rectum. The importance of accurate delineation of the rectum to facilitate calculation of rectal dose is discussed in Chapter 3.

Assessment of the image quality degradation of the CBCTs can be summarised, but not limited to, streaking artefacts, low soft tissue contrast in patients with a large volume separation, and artefacts arising from bowel gas [200]. These factors have previously been reported to affect image quality of CBCTs [188]. Another cause of poor image quality in this study was the partial acquisition of the CBCT, i.e., when the CBCT was acquired with an incomplete 360° arc. With large volumes such as a pelvis, the amorphous silicon panel needs to be offset, so that half the CBCT is acquired in the first 180°, and the second half of the CBCT in the last 180°. Interruption to the 360° arc results in only a partial volume being acquired, which made visualisation of the rectum difficult, and would have resulted in poor performance of the deformable image registration algorithm. An example of a CBCT with partial acquisition is shown below in Figure 4.1.

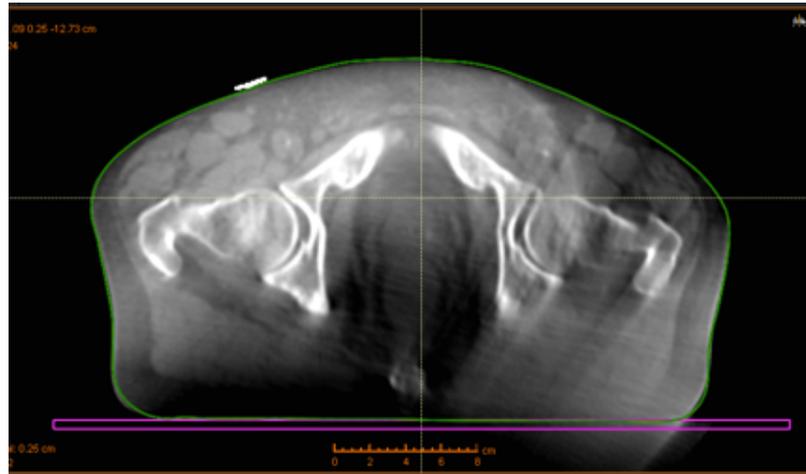


Figure 4.1 An axial slice of a CBCT of the pelvis which was excluded due to poor image quality as a result of incomplete acquisition. It can be observed that it is difficult to define anatomy in the posterior half of the image.

Another example of missing data are the CBCTs that were excluded from dose accumulation after deformable image registration of the deformation was poor (i.e. the deformed CBCT rectal contour and the CT rectal contour had Dice similarity coefficient (DSC) < 0.7 and mean distance to agreement (MDA) $> 0.3\text{cm}$ (described in more detail in section 2.3.2.1). This is because the deformation vector fields, which quantify the magnitude and direction of how the image is deformed, are used to map the dose from the CBCT to the CT. Poor deformation would result in dose being mapped imperfectly leading to inaccurate dose accumulation [135].

A potential source of correspondence ambiguity (where there is a lack of one to one physical correspondence between two images) in the imaging dataset is the ‘missing’ inferior rectal length on some of the CBCTs due to a small field of view. If not addressed, the deformable image algorithm would attempt to stretch the reduced rectal length on the CBCT inferiorly to match the CT rectal length. This would lead to incorrect and poor deformations, resulting in inaccurate dose accumulation. Although the dose surface map (DSM) methodology would not result in stretching of the rectum, it would result in a partial DSM. Partial DSMs were reported by the VoxTox group who created rectal DSM using MVCT from patients treated with Tomotherapy [92, 93]. Due to the limited field of view of the MVCT, the superior and inferior extent of the rectum was missing from their DSM. To enable accumulation of the DSM the VoxTox group simulated the missing rectal length using the planned DSM as a surrogate. As this PhD uses both DIR as well as DSM this method was not feasible. Instead it was decided to simulate the missing inferior rectal length on each image to

enable the CBCT to be deformed. An assumption was made based on the literature which reported less variation in the rectal volume and shape in the inferior part of the rectum than the middle and superior parts of the rectum due to the surrounding musculature and the sacrum [150, 201]. This chapter evaluates the types of missing data, describes the compensation strategies used, and evaluates the effects of those strategies on the dose accumulation methodology.

4.2 Aim

The aim of this chapter was to determine the effect of missing data on dose accumulation to the rectum. The work described in this chapter had three objectives:

1. To determine the number and cause of missing images in the study imaging dataset. This was used to identify the most common cause of missing imaging data and develop strategies to compensate for the missing data.
2. To measure the accuracy of a method used to simulate ‘missing’ inferior rectum when the CBCT field of view did not fully encompass the rectum (described in 2.3.2). The method used to simulate the missing rectal volume is described along with the effect of simulating the rectum on the absolute rectal volume, the performance of the DIR and, most importantly, the accuracy of the estimated accumulated dose.
3. To determine the degree to which dose accumulated is incorrectly estimated when using a partial imaging set. Secondly, to determine if the accumulated dose using a partial image set remains a better estimate of delivered dose than the dose calculated using the planning CT scan only. This was achieved by comparing full and artificially reduced datasets using DVH and DSM.

4.3 Methodology

4.3.1 Summary of missing imaging data

The number and cause of unusable CBCTs was recorded for each patient in order to determine the greatest cause of missing imaging data. Common causes were identified and a summary was created for all patients.

4.3.2 Simulation of missing inferior rectum on CBCTs

The rectum was contoured on each CBCT as per the DELINEATE contouring criteria (as described in 2.2.2). When a CBCT was found to have a limited FOV, resulting in partial coverage of the rectal length, the missing rectal volume was simulated by copying the rectal contour from the most inferior CBCT slice (Figure 4.2.). This gave a simulated rectal contour that extended outside of the patient's anatomy captured within the CBCT FOV. In order to enable calculation of dose to the whole rectum, an additional patient volume was simulated by extending the external model based segmentation (MBS) mesh which defines the external contour of the patient. The simulated voxels in the additional tissue were assigned the mass density of water (1.00 g/cm^3). The method for this is described further in section 2.3.

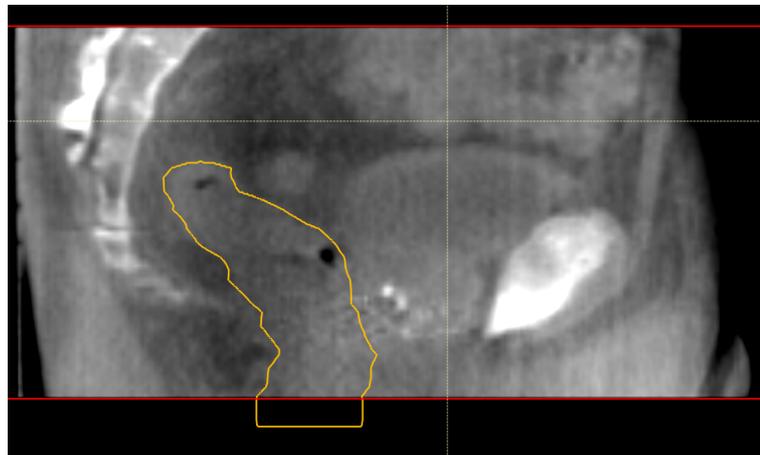


Figure 4.2 Sagittal CBCT slice of pelvis with a FOV (defined by the red lines) that did not fully encompass the full rectal length. The inferior rectal contour has been simulated by copying the last contoured slice of visible rectum giving a rectal contour that extends outside of the CBCT.

The number of additional slices required to simulate the missing contour was determined by using measurements from bony landmarks such as the symphysis pubis, sacrum or coccyx. The bony anatomy had to be visible using the fused CT and CBCT images (Figure 4.3.). On the CT, the 'ruler' measuring tool was used to measure from the bony landmark to the inferior extent of the rectum. This was repeated on the CBCT. The difference in rectal length between the two images was calculated and divided by 1.5 mm, which was the slice thickness used for both the CT and CBCT. The inferior rectal contour was then simulated by copying the last rectal slice on the CBCT inferiorly by the difference calculated. The fused images were then visually checked to ensure the simulation seemed anatomically reasonable (i.e., the inferior

extent of the simulated rectal length did not extend past the inferior rectal contour on the CT by more than 1 or 2 slices). Copying the last slice inferiorly resulted in a rectal contour that simulated the total rectal length. Note that the simulated length was straight, and therefore did not represent the actual rectal shape accurately, i.e., it did not flex posteriorly at the inferior aspect (Figure 4.3.).

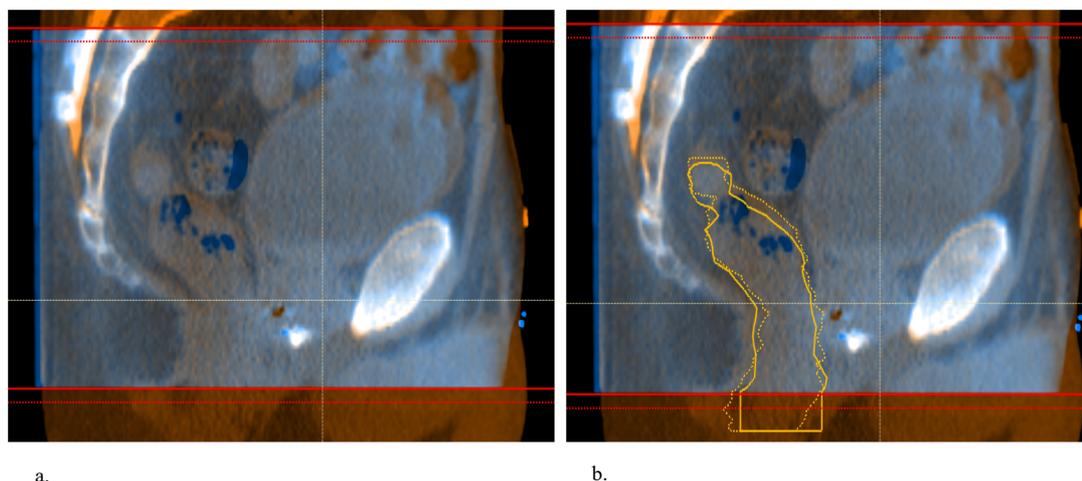


Figure 4.3 Sagittal slice of a planning CT image registered to one of the patient’s CBCT images. The red line in Fig (a.) shows the limit of the CBCT FOV. Pelvic bony anatomy such as the symphysis pubis is visible on the fused image in Fig (a.) which can be used to determine the extent of the missing rectum, and the number of slices which need to be simulated. The dotted yellow line in Fig (b.) is the rectal contour on the CT and the solid line is the simulated rectal contour on the CBCT. Comparison of the rectal contour visually shows differences in rectal shape but not length.

4.3.2.1 Evaluation of the effect of simulated inferior rectum contour on dose-volume histograms and the performance of deformable image registration.

To evaluate the effect of simulating the missing inferior rectum on accumulated dose, the rectal length for 5 patients were artificially reduced, and the ‘missing’ rectal volume simulated. Five patients were chosen from the hypofractionated cohort who had daily CBCT scans that encompassed the full length of the rectum, and a complete imaging dataset (20 CBCTs). The rectal volume was contoured on all 100 CBCTs as per the DELINEATE contouring criteria (see section 2.2.2).

To artificially reduce the rectal length, copies of the original rectal contour were created on each image. The most inferior extent of the rectal contour was then deleted and the missing volume simulated as already described in section 2.3.2. Two rectal

lengths, 1.0 cm and 1.5 cm, were deleted and then simulated. These volumes were chosen based on preliminary work which showed that in this study population the missing volume was usually between a minimum of 1.0 cm and a maximum of 1.5 cm. As the CBCT slice thickness was 0.15 cm, multiples of 7 and 10 slices were used to simulate the missing rectal contour, resulting in lengths of 1.05 cm and 1.5 cm, respectively.

To evaluate the effect of the simulated rectal length on the performance of DIR, each CBCT was deformed to the CT using the original and the simulated rectal contours in turn as the controlling ROI. This resulted in 3 deformed rectal volumes. A pairwise comparison was performed between the CT rectal contour and each of the deformed rectal contours (the deformed original contour and each of the deformed simulated contours) using DSC and MDA. The performance of the DIR using simulated rectal contours was deemed good if the DSC and MDA were within acceptable limits as defined by the literature (i.e. $DSC \geq 0.7$ of $MDA \leq 0.3$ cm).

4.3.3 Evaluation of the difference in accumulated dose calculated using full and reduced imaging datasets

To evaluate the effect of missing images on the calculation of accumulated dose, a full imaging dataset (i.e., when a treatment image is available for every fraction) was artificially reduced. To simulate an incomplete dataset, a limited number of CBCTs (each CBCT in turn, then CBCTs for 3, 5 and 10 fractions) were chosen using a random number generator (Microsoft Excel) and removed. A pairwise comparison between the planned dose and the accumulated dose using the full and artificially reduced dataset was performed. Accumulated DVHs were created for the full and reduced imaging datasets for five patients, and these were used to analyse the difference in rectal volume receiving specific doses between datasets. A second, larger study consisting of 15 patients used DSM to analyse the difference in spatial metrics when DSM are summed using a reduced dataset. A larger study was made possible as the process of removing and summing the DSM was automated using MATLAB.

4.3.3.1 Dose-volume histograms

Five patients from the hypofractionated cohort with complete imaging datasets and full rectal lengths were randomly chosen. The rectum was contoured on each image

and the daily delivered dose was calculated using the CBCT as described in Section 2.3.1 Each patient’s images were deformably registered to the CT using ANACONDA, with the rectum as the controlling ROI. The delivered dose was accumulated using the full imaging dataset by summing all deformed daily doses. The dose was then accumulated using the artificially reduced datasets. The dose per fraction was weighted according to the number of remaining fractions, and the deformed doses were summed (Figure 4.4.). Dose-volume histograms of the dose accumulated using full and reduced datasets were compared.

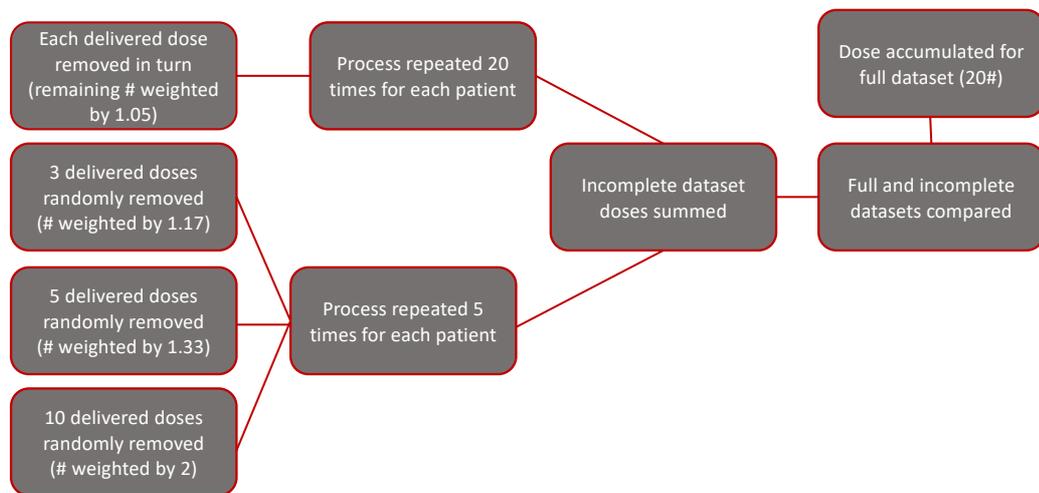


Figure 4.4 Schematic showing the number and weightings of delivered doses used to calculate the accumulated dose for full and reduced datasets using DIR. [10]

4.3.3.2 Dose surface maps

This study included 15 patients. Nine patients with full imaging datasets from the hypofractionated were used, with a further 6 patients from the standard prescription cohort. Standard cohort patients were randomly chosen and their plans were recalculated using the hypofractionated prescription. The imaging datasets were simulated by randomly selecting 20 CBCTs from the original imaging dataset using a random number generator (Microsoft Excel).

The methodology built on previous work by Murry [91] and Buettner [149] by using DSM to analyse the spatial characteristics of delivered dose. The longitudinal and lateral extent of dose was calculated as well as the percentage of pixels receiving at

least a specific dose, referred to in this thesis as pixel count. To compare the effect that an incomplete imaging dataset would have on the spatial characteristics, a comparison between a DSM of the plan, the DSM generated using a full imaging dataset, and the DSM generated using a reduced imaging dataset was performed. Reduced datasets were generated by randomly removing either 3, 5 or 10 DSMs. Artificially reducing the dataset by randomly removing the DSM, and subsequently weighting the dataset, was performed in MATLAB. Automation of the process enabled 1000 permutations to be performed for each reduction of 3, 5 or 10 DSMs.

Evaluation of the differences in spatial metrics between DSM was based on the method used by Buettner et al [149]. This is discussed more fully in Section 7.3.5 but the characteristics analysed are briefly given below.

- *Pixel count* - the number of pixels with value 1 at or below a given dose were calculated as a percentage of the total number of pixels for the entire image (180 x length). For each patient, the rectal length was normalised to the planning length. Note that the rectal length was not normalised across patients and therefore no interpatient comparisons were made.
- *Longitudinal extent* - was the maximum number of pixels with a value of 1 across the longitudinal extent of the largest cluster at a given dose.
- *Lateral extent* - was the maximum number of 1-valued pixels across the lateral extent of the largest cluster at a given dose.

To calculate the difference between full and reduced datasets, the percentage difference between metrics was calculated at intervals of 2 Gy from doses of 5 Gy to 59 Gy.

4.4 Statistical Analysis

4.4.1 Summary of missing imaging data

Qualitative analysis was used to describe and summarise the number and cause of the missing CBCTs.

4.4.2 Simulation of the missing inferior rectum on CBCTs

The volume for the original CBCT rectal contour, and each of the simulated contours (1.05 cm and 1.5 cm) was calculated for each fraction. A Kruskal Wallis H test was used to test for a statistically significant difference between the original and simulated rectal volumes for each patient. Pairwise comparison between the deformed original and simulated rectal contours with the CT rectal contour was performed using DSC and MDA, and was tested for statistical significance using a Kruskal Wallis H test. A comparison between the mean rectal dose accumulated by deforming the CBCTs using the original and simulated rectal contours was also performed. Accumulated DVHs were created for each volume to allow comparison using visual inspection. A Wilcoxon rank sum test was used to test for statistically significant differences in rectal volume receiving a specific dose x (or dose-volume, V_x) calculated using the original contour and each of the simulated contour.

4.4.3 Evaluation of the difference in accumulated dose calculated using full and reduced imaging datasets

4.4.3.1 Analysis using dose-volume histograms

Dose volume histograms were calculated for the plan, the full and each of the reduced imaging datasets (where 1, 3, 5 or 10 CBCTs had been removed). The difference between dose-volume accumulated using the full imaging dataset and the plan was calculated at each of the specified doses (10 Gy intervals between 10 to 50 Gy). The difference in rectal volume between dose accumulated using the full imaging datasets, and each of the reduced imaging datasets was also calculated. A boxplot of the difference in dose-volume at 10 Gy dose intervals was created, and paired analyses using Wilcoxon rank sum test was used to test for a statistical significance in dose-volume between the full and the reduced imaging datasets where 10 CBCTs had been excluded. This was chosen as the worst case scenario as all patients in the study had at least half the imaging dataset.

4.4.3.2 Analysis using Dose Surface Maps

Descriptive statistics were used to evaluate the difference in spatial metrics between complete and reduced datasets. Wilcoxon rank sum test was used to test for statistical significance between the spatial characteristics of each reduced imaging dataset

(where 3, 5 and 10 DSM had been removed) for 2 Gy intervals between 5 Gy and 59 Gy.

4.5 Results

4.5.1 Summary of missing imaging data

Only 13 of the 86 patients included in the study had complete imaging datasets, all of which were in the hypofractionated cohort (reasons for use of 86 patients instead of 105 are discussed in Section 5.5). In total, 11.2% of CBCTs (279 of 2502 CBCTs) were not useable or missing. A stacked histogram (Figure 4.5.) shows the percentage and cause of images missing from each patient's dataset. The cause and proportion of missing images varied from patient to patient. The loss of imaging data within a patient's dataset was typically the result of a combination of factors rather than a single cause. The most common cause was incomplete acquisition of the CBCT, which accounted for 42 % (116 of 279) of all missing CBCTs. CBCTs where the rectum had been poorly deformed by DIR were excluded from accumulation. As described in Section 2.3.3, each deformable image registration was evaluated, and if the rectum was found to be poorly deformed (defined as $DSC < 0.7$ and $MDA > 0.3$ cm) when compared to the CT, the image was excluded from accumulation. This accounted for the 19 % (54) of the imaging data excluded. Artefacts accounted for 17 % (48) of CBCTs excluded, and a further 12 % (32) of CBCTs could not be found in the imaging archive. The remaining 10 % (29) of images excluded were for various reasons including missing superior rectum (which was not possible to simulate), or where the author could not confidently contour the rectum due to poor image quality.

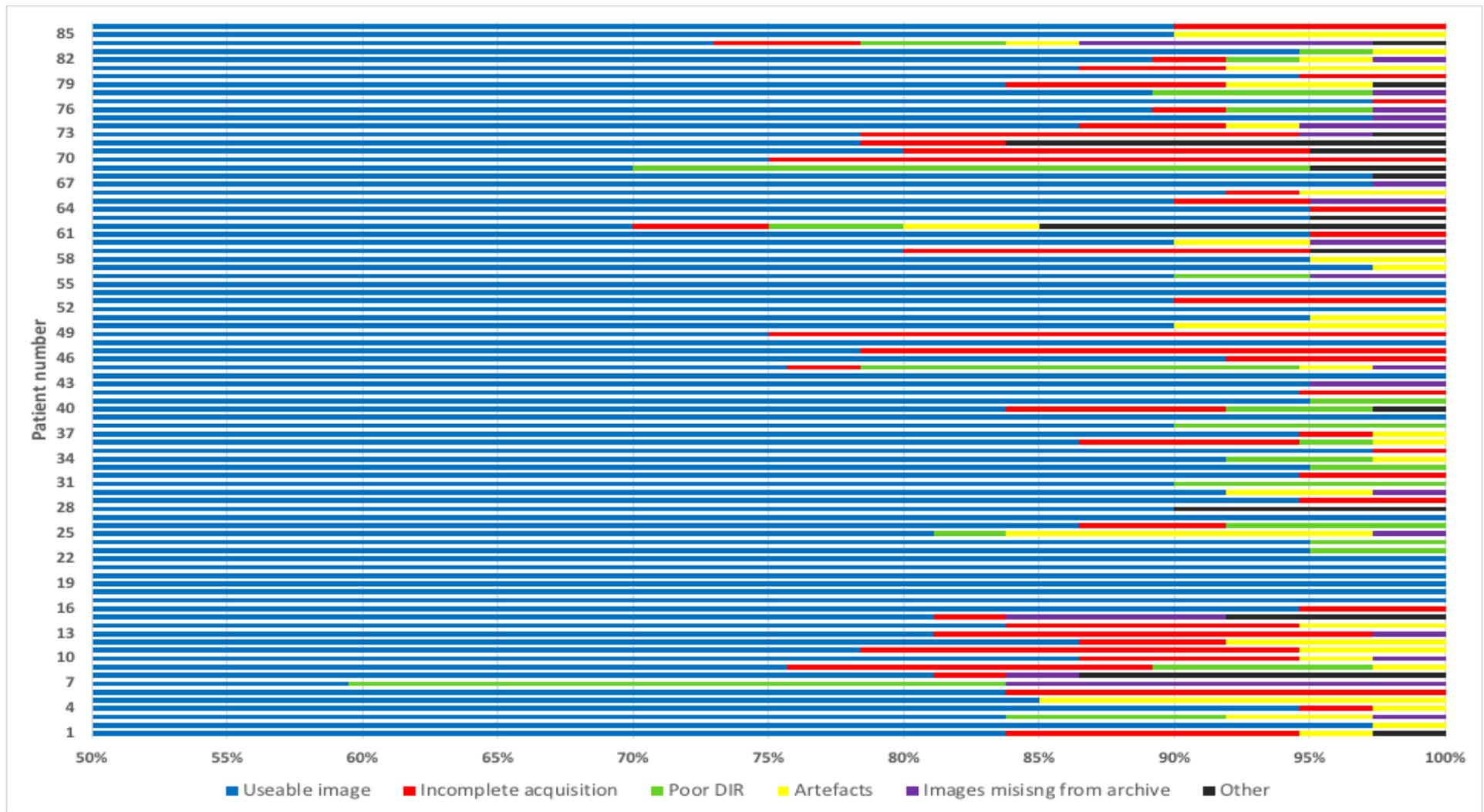


Figure 4.5 Stacked histogram showing for each patient the percent of CBCTs missing by cause. Each patient has at least 50% of the treatment CBCTs

4.5.2 Evaluation on the effect of simulated inferior rectum on delivered dose

Of the 86 patients included in the dose accumulation study, 20% (17) patients required simulation of the missing inferior extent of the rectum.

4.5.2.1 Variance in simulated rectal volume

There was no statistically significant difference between the mean original and simulated rectal volumes generated for each of the five patients when analysed using a Kruskal Wallis test. A boxplot of the original and simulated rectal volumes for all CBCTs for each patient are shown below in Figure .4.6.

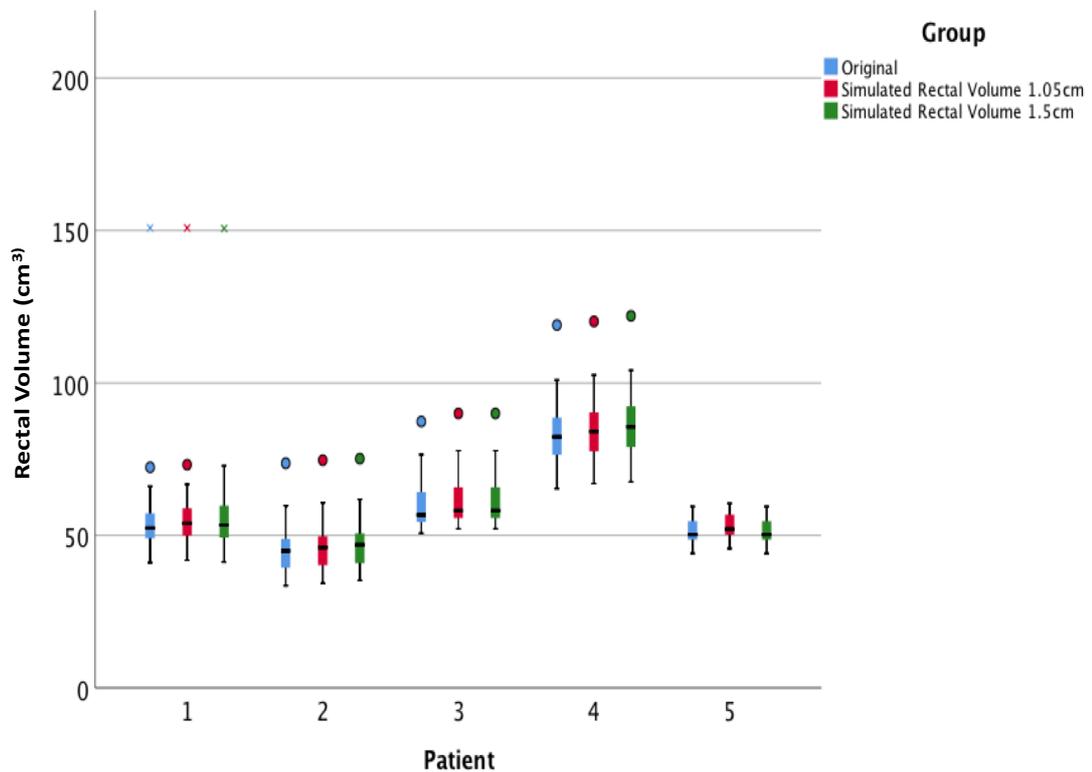


Figure 4.6 Boxplot of the original and simulated rectal volumes for each of the 5 patients analysed. The box represents the interquartile range (IQR), the median is denoted by the black line. The whiskers extending from the box indicate the lowest and highest values which are no greater than 1.5 times of the IQR. Outliers which are greater than 1.5 to 3 times the IQR are denoted by circles. Extreme values which are greater than 3 times the IQR are denoted by crosses.

4.5.2.2 The effect of rectal volume changes on performance of DIR

DSC and MDA were used to measure the similarity between the rectal volume contoured on the CT and the deformed rectal volumes delineated on CBCT, where the original and simulated CBCT delineated rectal contours were used in turn as the

controlling ROIs to focus the deformation. The median and inter-quartile range (IQR) for the DSC and MDA values for each deformed volume are shown in Table 4.1.

	DSC values			MDA values (mm)		
	Original Rectal Volume	Simulated Rectal Volume 1.05cm	Simulated Rectal Volume 1.5cm	Original Rectal Volume	Simulated Rectal Volume 1.05cm	Simulated Rectal Volume 1.5cm
Patient 1	0.97 (0.01)	0.94 (0.01)	0.94 (0.01)	0.32 (0.06)	0.65 (0.13)	0.87 (0.27)
Patient 2	0.96 (0.01)	0.95 (0.02)	0.94 (0.01)	0.33 (0.13)	0.52 (0.15)	0.64 (0.17)
Patient 3	0.98 (0.01)	0.98 (0.01)	0.96 (0.01)	0.23 (0.15)	0.23 (0.14)	0.51 (0.20)
Patient 4	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.21 (0.04)	0.20 (0.03)	0.20 (0.04)
Patient 5	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.16 (0.02)	0.16 (0.02)	0.16 (0.02)
All patients	0.98 (0.02)	0.98 (0.03)	0.96 (0.04)	0.26 (0.14)	0.26 (0.36)	0.49 (0.51)

Table 4.1 The median (IQR) for the DSC and MDA vales for each patient, and for all patients combined. A DSC of 1 indicates a perfect overlap between the two contours. A MDA value of 0 indicates no difference between the two contours.

A Kruskal Wallis test was used to test for differences between median DSC and MDA measurements for all patients in each of the three groups. Pairwise comparison with adjusted p values using Bonferroni correction for multiple testing showed no statistically significant difference in DSC and MDA between the original and simulated 1.05 cm rectal volumes. There was a small but statistically significant difference in the values between the original and simulated 1.5 cm length (DSC, $p = 0.002$; MDA, $p < 0.001$). Note, MDA values are in units of millimetres and therefore differences are very small.

4.5.2.3 Accuracy of dose accumulation using simulated rectal volumes

For each patient, the dose to the rectum was accumulated using the deformed original volume and each of the simulated rectal volumes (1.05 cm and 1.5 cm). A DVH of the rectum from the plan, and the accumulated DVH using the deformed original and simulated rectal volumes was created for each patient (Figure 4.7.). Visual inspection of the accumulated DVH using the original and simulated rectal contours showed small differences in the DVHs at lower doses only. A Wilcoxon rank sum showed no statistically significant difference between dose-volumes (V10, V20, V30, V40, V50 and V60) accumulated using the original contour, and each of the simulated contours.

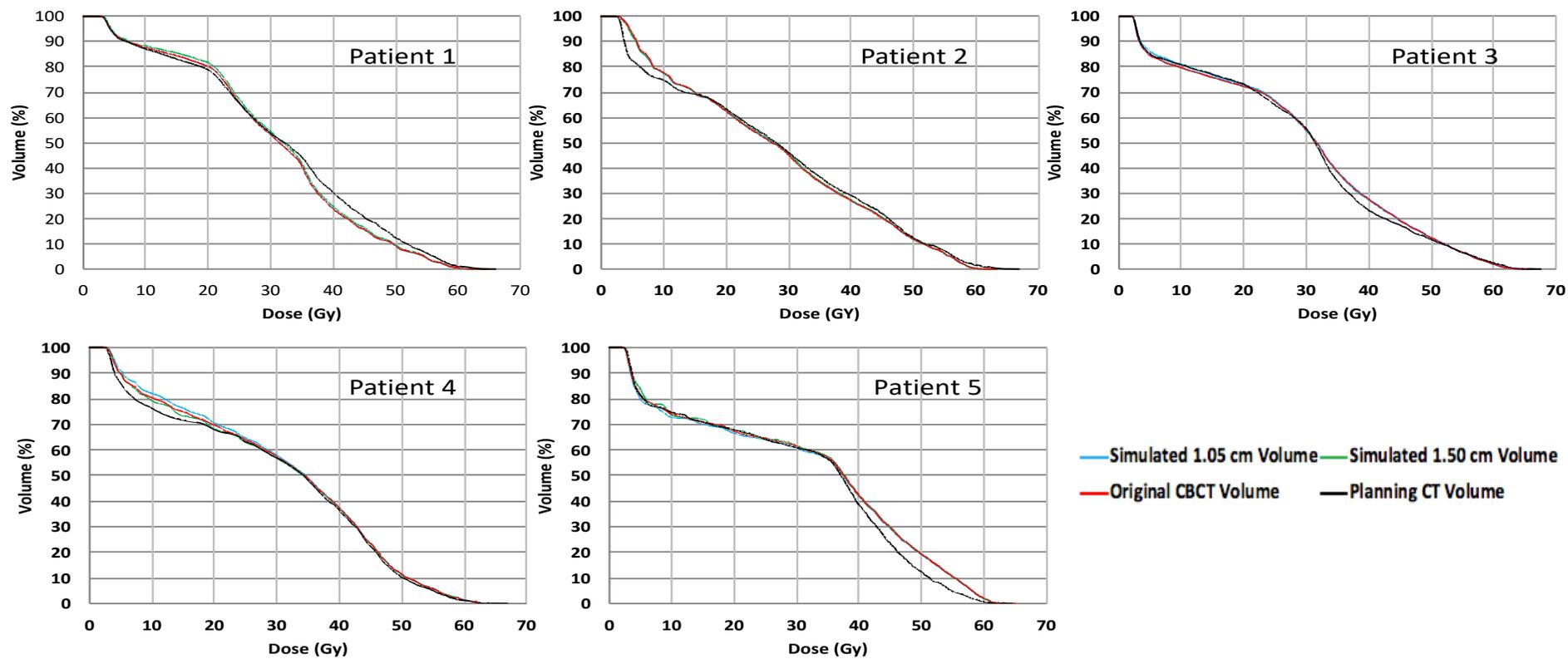


Figure 4.7 Comparison of the accumulated DVHs for the planned, original and simulated rectal volumes. The accumulated DVHs using the original and simulated rectal length as controlling ROI, have been plotted for each patient. Visual inspection shows the variation between dose accumulated using the original rectal contour and the simulated rectal contour is small. For each patient the difference between the original and simulated rectal contours is smaller than the difference between the planning and accumulated DVHs.

4.5.3 The effect of missing CBCTs on accumulated dose

4.5.3.1 Difference in rectal volume between full and reduced imaging datasets

The difference in rectal dose-volume between the accumulated and planned datasets was calculated at 10 Gy intervals between V10 and V50. This was repeated for accumulated DVHs using full and reduced imaging datasets (see Figure 4.8.).

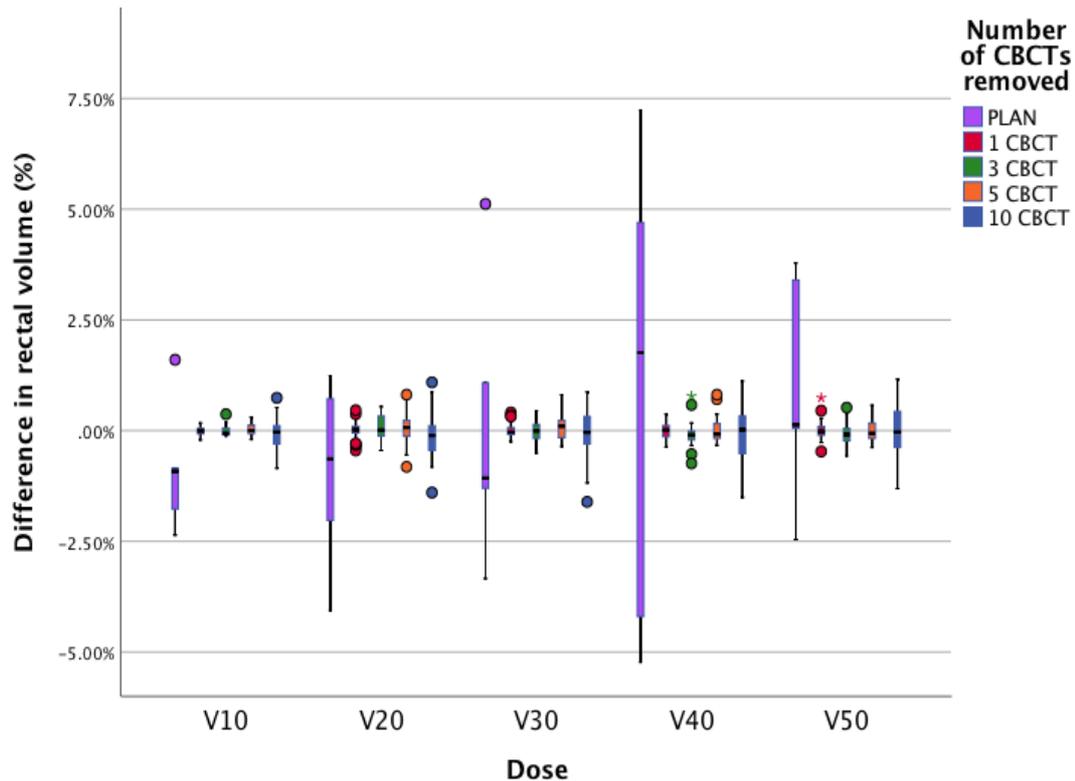


Figure 4.8 Boxplot showing the difference in rectal volume between the planned and accumulated dose using a full dataset (labelled Plan), and the accumulated dose using the full dataset and reduced imaging datasets for 5 patients. The boxplot shows the increase in the difference of the rectal volume between datasets as more CBCTs are removed. The label 1 CBCT refers to the difference in between the full and reduced datasets where one CBCT has been removed. The nomenclature is continued for the reduced imaging datasets where 3, 5 and 10 CBCTs are removed.

A Wilcoxon rank sum test showed no statistically significant difference in rectal volume at 10 Gy intervals between the full and reduced imaging dataset, where 1, 3, 5 or 10 CBCTs had been removed.

4.5.3.2 Effect of missing images on the spatial characteristics of delivered dose

For each patient the difference in the spatial metrics between dose accumulated using a full imaging dataset and the plan, and dose accumulated using full and reduced datasets was analysed by calculating the percentage difference in pixel count,

longitudinal and the lateral extent. The data was assumed to be normally distributed due to the number of permutations (1000) [176]. A plot of the mean difference between the full and reduced datasets was created for each metric to enable visual inspection. To provide an example, a plot for one patient is shown in Figure 4.9

For all patients, visual inspection of the plots shows that the difference between full and reduced datasets increases as more DSM are removed from the dataset. This holds true across all metrics. For pixel count and lateral extent, the difference between full and reduced datasets is smaller than the difference between full and planned dataset. On average, the mean longitudinal extent is smaller than the difference between accumulated and planned DSM. However, there were a small number of instances where the standard deviation is equal to that difference. This occurs at different dose levels and more frequently as more CBCTs are removed. This suggests that the longitudinal extent is more sensitive to removal of CBCTs than the lateral extent or pixel count.

Each of the spatial metrics from the accumulated DSM were compared to the DSM which had been generated using the reduced datasets (3, 5 or 10 DSM excluded). There were statistically significant differences across all metrics for a small number of doses. For pixel count, significance was noted at high doses of 55 and 61 Gy. There were statistically significant differences at 55 Gy ($p = 0.02$) and at 61 Gy ($p = 0.04$) when 3 DSM were removed. There were statistically significant differences when 5 DSM were removed at 55 Gy ($p = 0.04$) and at 61 Gy ($p = 0.02$). Finally, when 10 DSM are removed at 61 Gy ($p = 0.01$).

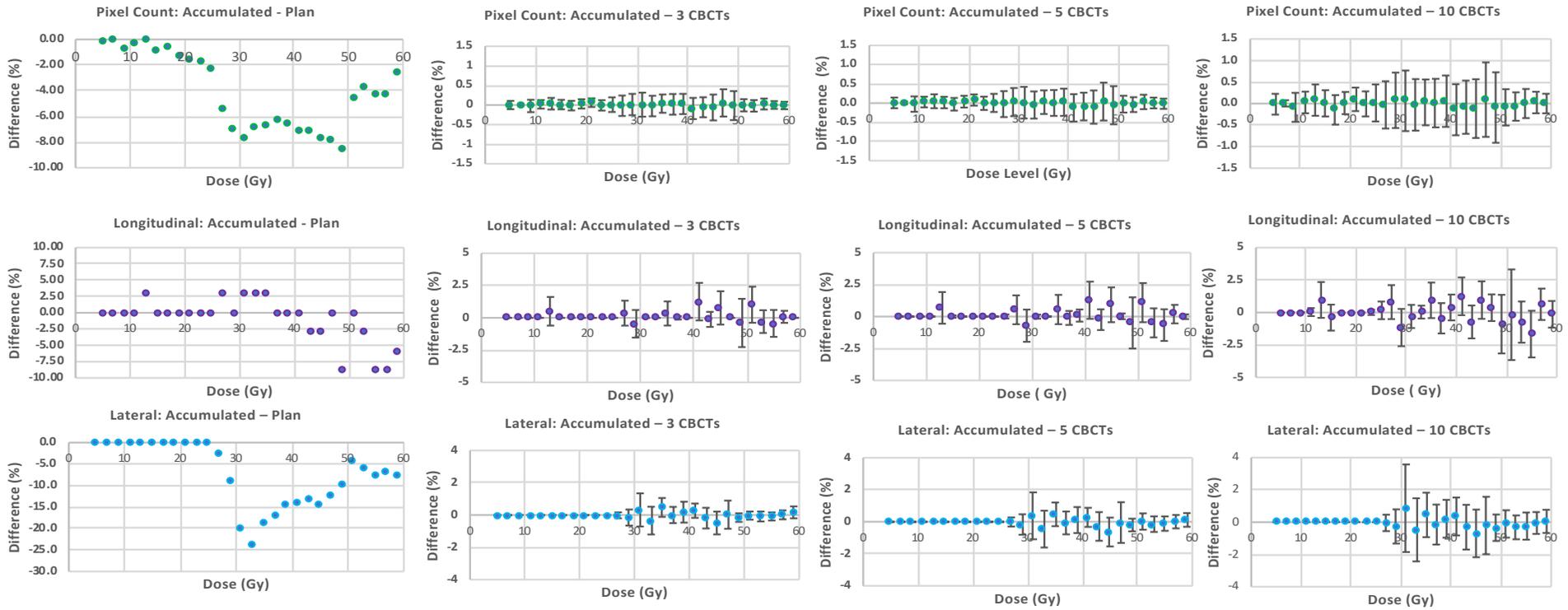


Figure 4.9 Plots for Patient 10 provided as an example which shows the mean percentage difference in pixel count, longitudinal extent and lateral extent. The plot shows the mean difference in DSM accumulated using a full imaging dataset and the plan, and full imaging datasets and the reduced imaging datasets where 3, 5 and 10 DSMs have been removed. The mean difference for each metric was calculated at 2 Gy interval from 5 Gy to 59 Gy. Accumulated – Plan refers to the mean difference between accumulated DSM and Plan DSM. Accumulated – 3 CBCTs refers to the average difference between accumulated DSM and the reduced DSM where 3 DSM were randomly removed a 1000 times. The error bars represent the standard deviation across all fractions. The nomenclature continued for datasets where 5 and then 10 DSM were randomly removed

In the longitudinal extent, a statistically significant difference was noted across a range of doses, and tended to increase as more DSM were removed. When 3 DSM were removed, significant differences were noted at 5 Gy ($p < 0.01$) and 17 Gy ($p = 0.01$). When 5 DSM were removed statistically significant differences were noted at 5 Gy ($p = 0.01$), 17 Gy ($p = 0.01$) and 47 Gy ($p = 0.02$). When 10 DSM were removed statistically significant differences were noted at 17 Gy ($p = 0.02$) and 47 Gy ($p = 0.03$).

When assessing the lateral extent, the greater the number of DSM removed, the more statistically significant differences were noted between datasets. Doses of 31 Gy and greater were analysed as there was no difference between accumulated and reduced datasets in doses lower than this. When 3 DSM were removed, a statistically significant difference was found at 43 Gy ($p = 0.01$). When 5 DSM, are removed statistically significant difference was noted at 43 Gy ($p = 0.01$) and at 55 Gy ($p = 0.02$). When 10 DSM were removed, significant differences were noted at 43 Gy ($p < 0.01$), 51 Gy ($p = 0.03$), 55 Gy ($p < 0.005$) and at 57 Gy ($p = 0.04$).

Although the Wilcoxon rank sum test found significant differences at certain doses when reduced datasets were used, the difference between the datasets were less than 2 %. This is smaller than the difference between the spatial metrics accumulated using a full dataset and the plan. A boxplot of the difference in lateral extent between the doses for which significant differences were noted when the dataset was reduced by 10 DSM and the accumulated dataset, and the difference between the accumulated datasets and the plan is shown in Figure 4.10.

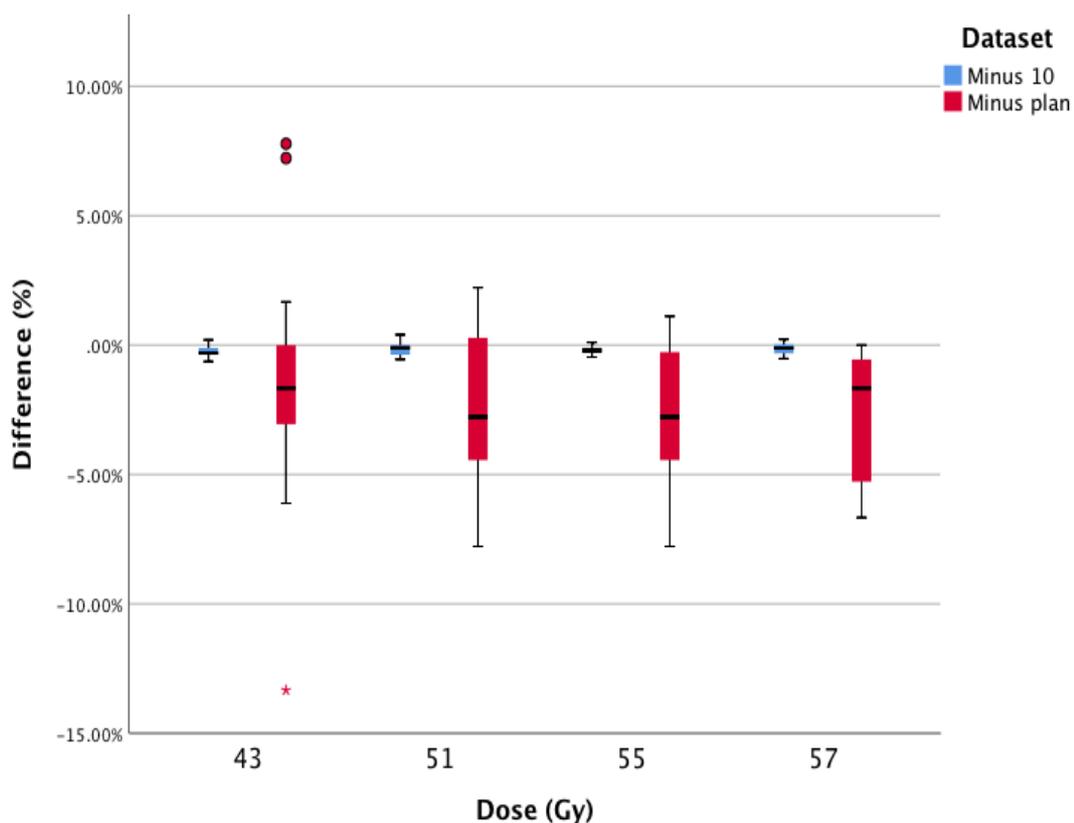


Figure 4.10 Boxplot of the difference between the lateral extent of accumulated DSM and datasets which has been reduced by 10 DSM (Minus 10) for all patients. The difference in lateral extent between accumulated DSM minus the Plan DSM is also plotted. Doses chosen are those which showed a statistically significant difference when a Wilcoxon rank sum test was used to test between the lateral extent of the accumulated DSM using a full dataset and a reduced dataset where 10 DSMs had been removed.

4.6 Discussion

4.6.1 Missing images

Seventy three out of 86 patients had incomplete imaging datasets in this study. It was determined that the greatest source of missing images was due to incomplete acquisition of the CBCT (42 % of all missing images). After discussion with the department engineers, it was found that the interruption to the CBCT acquisition was due to a software fault, which has since been fixed. However, other causes of image exclusion such as poor image quality are less easily rectified. Poor image quality can be a contributing factor to poor deformable registration [142]. Repeat imaging of the patient to acquire an image of better quality may not always be justified. Under the Ionising Radiation (Medical Exposure) Regulations (IRMER) [202], operators are

required to keep concomitant doses to the patient as low as reasonably practicable (ALARP). If the treatment image provides enough detail to correct for set-up error then additional exposure by repeating the image may not be warranted, especially if the repeat image is still likely to be affected by artefacts such as those caused by bowel gas. However, if in the future dose accumulation can be demonstrated to optimise treatment delivery, then there may be a benefit in increasing the dose to improve images (patient dependent), or the justification for repeat images [116, 119]. As we move from using CBCTs for image guidance to dose guidance, further discussion as a profession will be required, and guidelines should be issued building on the work of Society and College of Radiographers [116].

CBCTs were also excluded due to poor performance of the deformable image registration algorithm 19% (54). Poor image registration can be a result of poor image quality, or difficulty establishing spatial correspondence due to large differences in anatomy as a consequence of rectal gas [128]. Validation and verification of the deformable registration process is strongly recommended in the literature due to the potential inaccuracy when accumulating dose [134, 135, 148, 203, 204]. Validation refers to the overall process to ensure that accurate image registration can be performed on a consistent basis [142]. Validation for deformable image registration uses methods which look for similarity in anatomy or contours [204], or use physical [205] or digital phantoms [148]. ANACONDA has been validated for the thorax, pelvis and head neck using published CT and CBCT datasets and similarity metrics, and has been shown to perform well with CT to CBCT registration [146, 166]. Verification is the process of confirming the accuracy of the registration is acceptable for the intended use [142]. In the current study, images were verified by measuring the similarity between the reference contour (CT) and the deformed contour using DSC and MDA (see section 2.3.2.1). An alternative method of verification uses corresponding points such as stable anatomy or markers identified by observers [135, 203]. In the lung, this could be the carina, or fiducial markers could be used in the prostate [135]. This method was not found to be suitable for the rectum because of a lack of stable anatomy due to the deforming nature of the organ [135, 148]. Using DSC and MDA was a fast and effective method of verifying a large number of images (2300 images), and could be used to assess the performance of DIR when evaluating uncertainty from simulation of missing rectum.

4.6.2 Assessment of correction strategy for missing rectal images

A methodology to simulate the missing inferior rectum was developed and validated. The effect of simulating the missing inferior rectum by copying the last contoured inferior slice has been shown to be small when compared to the original length for missing rectal lengths of up to 1.5 cm. Differences in absolute rectal volume between the original and simulated rectal length of 1.05 cm and 1.5 cm were shown not to be statistically significant. The difference in rectal shape and volume of the simulated rectal volumes did not negatively affect the performance of the DIR. Although the Kruskal Wallis test showed statistically significant differences in the DSC and MDA between the original and simulated 1.5 cm volume when deformed, both the DSC and MDA measurement were within those stated in the literature as clinically acceptable. Most importantly, when the dose was accumulated using the original and simulated rectal contours in turn, there were no statistically significant difference in dose-volume at intervals of 10 Gy. This means that dose accumulation using a simulated inferior rectal length of up to 1.5cm will give an accumulated dose similar to that if the CBCT had included the full rectal length. The positive results observed from simulating the missing rectum in this manner may be because the anorectum is such a small proportion of the total rectum, and is in close proximity to the edge of the radiation field where dose is low.

It was difficult to evaluate the approach used in the PhD of simulating the inferior rectum. There were no comparative studies in the literature which have measured the uncertainty in dose accumulation which arises from the simulation. The most similar study is by the VoxTox group who simulated missing rectal extent on the DSM by substituting missing pixels with those from the planning DSM [92, 93]. This approach was justified by the authors based on previous work they had undertaken where they quantified the variation in rectal position during treatment [201]. This showed that the greatest variation in rectal circumference was seen in the superior third of the rectum (attached to the rectosigmoid junction). However, they did not perform a comparative study as performed in this PhD to measure the difference between artificially simulated rectal length and the original rectal length. In keeping with findings of VoxTox, Chong et al., also reported greater differences in the random error for lateral movement on the superior third of the rectum than the other two-thirds of the rectum, based on weekly CBCT scans for 16 patients' receiving chemoradiation for rectal cancer [206]. Hoogeman et al., created dose surface maps of the rectum using repeat

CT scans, to analyse the displacement of the rectum between scans [150]. They reported that the largest variation in the position of the rectal wall was in the anterior aspect of the rectum, 6 cm cranial to the anus. There was less variation near the anus and at the posterior side of the rectum, which they postulated was possibly due the musculature surrounding the anus, and the proximity of the sacrum. Although not directly comparable, these studies support the decision to simulate the inferior rectum but not the superior rectum.

In future, prospective dose accumulation studies should ideally ensure that the CBCT collimator is a sufficient length that the FOV encompasses the organ of interest. Although the method developed in this PhD is suitable for simulation of the missing inferior rectum (up to 1.5cm) on the CBCT for the purposes of this PhD and results in small differences in the deformed contours, the process is time consuming, and does not lend itself to simulation of the superior rectum.

4.6.3 Assessment of impact of incomplete data sets on dose accumulation

This chapter has shown that accumulation of dose using a full imaging dataset is not always feasible, especially when using a retrospective dataset. Never the less, an estimation of the delivered dose can be calculated using a partial dataset, although the certainty in the calculated dose increases as the number of CBCTs being accumulated increases. However, even accumulating the dose with only half the imaging dataset, the difference between full and the reduced imaging dataset is smaller than the difference between the accumulated and planned dose. This suggests that even if only half the imaging data is available, using a subset of images will provide a more accurate estimate of the actual delivered dose than using the plan alone. This may be because the rectum deforms within a pattern individual to each patient. Thus, using a small number of images which describe the variation individual to the patient will give a better estimate of the delivered dose than a single snapshot at CT. A similar trend was also observed for the difference in spatial metrics which describe the DSM (pixel count, longitudinal and lateral extent) when estimated using full and reduced datasets. Even when half the images were removed (10 DSMs) the difference between full and reduced datasets was small (for example less than 1 % for the lateral extent).

In this PhD the maximum number of images removed was 10, which is half the imaging dataset (hypofractionated cohort). Further evaluation could be performed by

removing additional images. However, the motivation for the study was to ensure that the results presented on dose accumulation were robust when some CBCT images are unusable. All imaging datasets within this study had at least 50% of images available for accumulation. A study by Nassef et al., is the only other study which has evaluated the uncertainty of missing images on accumulated dose, but with only using 3 patients [148]. For each of the 3 patients, the dose difference between accumulated and an increasing number of images (1 CBCT to 35 CBCTs) was evaluated using Dmean and V70. Mean and standard deviations of the difference between using 8 CBCT and 35 CBCTs were, for the rectum wall: 0.03 ± 0.60 Gy for the Dmean, and 0.37 ± 1.6 % for the V70. The conclusion was that accumulating dose with weekly images (7#) would still give an accumulated dose closer to the actual delivered dose than the plan.

Although the work undertaken as part of this PhD supports the findings by Nassef et al., future work could evaluate dose accumulation using an offline no action level protocol which other centres may prefer to use (i.e. image first 3#, correct for systematic error and then image weekly which may result in as few as 6 images for a hypofractionated schedule). This was considered. However, the data to perform this analysis was not easily obtainable, and would have required further simulation e.g., applying systematic corrections retrospectively. Again, as the primary motivation for this study was achieved, and many centres are moving to daily online corrections strategy, it was decided not to pursue this at the time.

However, this work highlights the need for further evaluation and discussion around the implementation of future dose accumulation. As demonstrated in this chapter images are not always suitable for accumulation. Therefore, if accumulating dose in a clinical setting, or in subsequent studies, should a policy of perfection be pursued i.e. as many images as possible, will repeat images be required if for example the image acquisition is interrupted? The additional dose may not be justified if dose can be accumulated using a subset of images as shown in this study. Future work should also consider analysing the minimum number of images required to give a good estimate of delivered dose i.e. the first 5 fractions. This would enable the identification of those patients who would benefit from replanning or adaption to be identified.

Finally, different cancer sites or OAR will require further evaluation and the number of images required will be influenced by how much the organs deform. Delineation of

structures is also a resource issue as demonstrated in Section 3.5.3, auto segmentation of the rectum on the CBCT was poor when compared to manual delineation. If a subset of dataset is required to accumulate dose, this will essentially reduce the workload. Dose accumulation when treating with extreme hypofractionation will also need consideration, it may be necessary to accumulate dose using an image for each treatment because of the large dose per fraction. The author is unable to address these questions as they are outside the scope of this thesis, and will require further analysis and discussion on a larger scale.

4.7 Limitations of this study

Measuring the effect of simulating the rectum, and accumulating the dose using a partial dataset, was undertaken early in the PhD when the author began to delineate the rectum on the initial patients. Not all patient's imaging datasets had been exported to RayStation from the archive database at this time. As a consequence only a small number of patients were identified as having a full rectal length which was needed for simulating the rectum study, and why a 'full' imaging dataset (20#) was simulated using the standard fraction patients (37#) for the missing images study. Due to the time intensive nature of exporting the images to RayStation, and preparing the images for delineation, it was decided to undertake a study with the patients which were available.

4.8 Conclusion

This work has established that an estimate of the delivered dose can be calculated using a partial imaging dataset, and that the certainty of the delivered dose calculated increases proportionally with imaging data accumulated. This not only supports the use of incomplete imaging datasets in this PhD but will also assist future dose accumulation studies. It will be of benefit to support radiographers' choice when determining whether additional images may be justified in clinical dose accumulation studies. This chapter has also developed and validated a method for simulating the missing inferior rectum on the CBCT to enable dose to be accumulated.

Chapter 5 Difference in rectal dose-volume between planned and delivered dose

5.1 Introduction

Rectal toxicity can adversely affect the quality of life (QOL) of men who have received prostate radiotherapy [207, 208]. Limiting the dose to the rectum to within defined thresholds is an established method of reducing the severity and incidence of toxicity [9, 28]. These thresholds, or dose-volume constraints (DVC), have been derived from a number of studies which have established associations between clinical and dosimetric parameters, and acute and late GI toxicities [28, 30, 65, 70]. As previously discussed in Section 1.2.4, Gulliford et al. analysed data from the MRC RTO1 trial showing statistically significant association between dose-volume and rectal bleeding as well as proctitis, loose stools and rectal urgency across a range of doses (30 to 70 Gy). Most notably the study found that the number of patients experiencing Grade ≥ 2 GI toxicity increased as the number of DVC exceeded increased [28]. Fonteyne et al., evaluated the impact of different DVC on rectal toxicity. Within the study, 637 patients were treated in one of three groups with increasingly stringent DVC. For Group 1 (193 patients), rectal DVC were $V_{50} < 100\%$, $V_{60} < 60\%$, $V_{65} < 50\%$ and $V_{70} < 30\%$. For Group 2 (212 patients) the DVC were $V_{50} < 60\%$, $V_{60} < 50\%$ and $V_{70} < 30\%$. For Group 3 (232 patients), initial DVC to limit Grade ≥ 2 was $V_{40} < 84\%$, $V_{50} < 69\%$, $V_{60} < 59\%$ and $V_{65} < 48\%$. If, in Group 3, the initial planning DVC were easily met, the plan was reoptimized using a second set of constraints ($V_{40} < 64\%$, $V_{50} < 46\%$, $V_{60} < 35\%$ and $V_{65} < 34\%$), which aimed to lower the rectal dose further and prevent Grade ≥ 1 toxicity. The authors reported that the risk of developing late Grade ≥ 2 rectal toxicity decreased from 16 % to 7 %, and then 5 % as increasingly tighter DVC were applied ($p < 0.001$) [209].

Historically, late rectal bleeding was often the only rectal related endpoint which was analysed, partly due to the objective nature of the endpoint as well as its prevalence [9, 90]. However, other GI endpoints such as bowel frequency are chronic conditions which can reduce a patient's quality of life [207, 208]. There have been a number of studies which have analysed the correlations of patient and treatment related parameters with acute and late toxicity in order to optimise treatment planning [28, 65, 70, 210].

Despite the number of studies evaluating the dose response of the rectum, the application of published DVC has limitations [9, 28, 211]. A study by Faria et al., investigated the incidence of late rectal toxicity in 71 patients treated with 66 Gy/22 # [212]. The aim of the study was to explore the relationship between the number of patient radiotherapy plans violating different rectal DVC to determine which DVC would have better predicted rectal toxicity. The patients were divided into one of three groups with differing rectal DVC that were obtained from three previously conducted randomised trials [213-215]. Despite using DVC from published trials Faria et al., were unable to find a correlation between the DVC and toxicity. Difficulties in applying DVC across centres may exist due to institutional differences, such as how the rectum is defined, with some institutions defining the rectum including content [52], while others analyse dose to the rectal wall only [92]. There is also variation in the length contoured with some outlining the rectum from the anal verge to the recto-sigmoid junction [52], while others define a portion of the rectum above and below the prostate [115]. These variations result in differences in DVHs making consensus across institutions difficult [30]. Other differences may be related to differences in treatment techniques, prescription, and imaging protocol. For example, with limited image guidance, greater planning target volume margins are required which may mean that DVCs are more likely to be breached. However, large multi-institutional randomised trials have also shown that interinstitutional variations can be reduced through use of guidelines and, where appropriate, training [13, 52, 159]. Furthermore, analysis of DVH and toxicity can be confounded by non-dose related factors such as irritable bowel syndrome, haemorrhoids, advanced age, smoking, abdominal surgery and diabetes, which have been shown to be associated with complication risk [9, 63-67].

In addition to institutional differences, there are uncertainties associated with the data used to generate the DVC. Currently, DVC are derived using the dose-volume parameters from the planning CT scan only, and do not account for variations in delivered dose due to changes in the shape and volume of the rectum during the treatment course [29, 30]. There have been several studies which have used a range of treatment imaging modalities including CT on rails [216], CBCT [32, 130] and megavoltage CT (MVCT) [92] to determine the shape and position of the rectum during treatment, and used this information to calculate the delivered dose to the rectum, and measure the difference from the planned dose. All of these studies

concluded that despite online imaging verification and correction of patient position, the anatomical changes in the rectum affect the dose delivered with the conclusion that the planned dose is not a true representation of delivered dose.

A study reported by Kupelian et al., was one of the first to calculate the daily dose in 10 patients (78 Gy/39 #) using MVCT [115]. The study found that daily targeting of the prostate using intraprostatic fiducial markers ensured the prostate received the prescribed dose. However, despite the online imaging correction strategy, the rectal volume receiving 2 Gy varied from day to day and for a number of treatments. The average (\pm SD) of the absolute rectal volume receiving 2 Gy was 7.0 (\pm 8.0 cc) (range: 0.1 to 67.3 cc). A study by Peng et al., concluded that, despite daily correction, the repositioning scheme was not always sufficient to correct for the residual variations in shape and volume of the rectum which occur during treatment. The authors calculated the daily rectal DVH for 20 patients using CT on rails and reported rectal volume changes during treatment could be as large as 650 % (maximum volume relative to minimum volume). For 5.6 % of all fractions analysed, the rectal V45 was 15 % and was greater than planned, and between 5 and 10 % greater than planned in 9.3 % of fractions. In 11.6 % of all fractions analysed, the V70 was 5 % greater than planned and 10 % greater than planned in 4.7% of all fractions.

Huang et al., concluded from their study that the variation in rectal volume was greater than variation in mean rectal dose during treatment. For all 28 patients with 112 data sets studied, the mean percentage differences (standard deviation) in the volume and mean dose for the rectum calculated as the difference divided from planning were 36 % (\pm 29) and 22 % (\pm 15) for the rectum, respectively [132]. Hatton et al., undertook a study comparing delivered daily doses calculated using twice weekly CBCTs and concluded that in 65 % of all treatments, the planning DVH underestimated the delivered dose [32]. Chen et al., analysed once weekly CT on rails from 20 patients and noted that the interfraction volume variation varied from patient to patient [31]. The authors initially hypothesised that a small rectal volume at CT would result in a decrease in rectal dose during treatment when the rectal volume increased. It was discovered, however, that in some cases, the rectal wall expanded laterally which increased the rectal dose. In 28 % of daily fractions analysed, the rectal dose did not meet V40<35%, and 26.6 % of fractions did not meet V65<17%. However, overall 70 % of patient treatments showed lower rectal doses than planned.

Though these studies show a difference between planned and delivered dose, they are limited as the delivered dose was analysed either by fraction [32], or the daily dose was summed or averaged [31], which can lead to overestimation of delivered dose [129, 217]. There have been a small number of studies which have accumulated rectal dose using either deformable image registration (DIR) or dose surface maps (DSM) based analysis. These studies also show a difference between planned and delivered rectal dose. Wen et al., who investigated the accumulated dose using daily CBCT and DIR, reported that the rectal generalized equivalent uniform dose (gEUD) was greater than planned. However, they only analysed the data for 5 patients [140]. Akino et al., used DIR and MV-CBCT from 8 patients and reported a greater accumulated than planned gEUD in 1 patient, lower in 6 patients, and one patient was within 1 Gy [217]. This finding was similar to that of Scaife and colleagues (the VoxTox group) who used DSM and MVCT to accumulate dose for 10 patients [92]. They reported a greater accumulated gEUD than planned in 1 patient, a lower accumulated gEUD in 6 patients, and in 3 patients the gEUDs were within 1 Gy of each other. Bostel et al., accumulated the dose for 10 patients using in-room CT and DIR [138]. The study reported large individual variations in patient rectal filling during treatment. Despite the variation in rectal filling, accumulated dose was not significantly greater than planned, with an average accumulated mean dose of 0.14 ± 3.73 Gy higher than planned.

Although the studies discussed have been instrumental in providing early data demonstrating the difference between planned and delivered dose, they are limited by one or more factors. As highlighted most studies did not accumulate the dose, but rather used the fractional or averaged dose from a small number of fractions to evaluate the delivered dose [31, 32, 132]. Studies that did examine accumulated dose using DIR or DSM analysis were limited by small patient study numbers (≤ 20) [92, 138, 140, 217]. Furthermore, direct comparison across these small studies is difficult due to institutional differences including bowel preparation. Some studies used enemas [115] or specific diets [32, 130], while others reported no bowel preparation [31]. Some studies were also limited by inadequacy in the volume of the rectum imaged [115]. For example, in the study by the VoxTox group, which used MVCT images with limited field of view, the inferior and superior rectal length had to be simulated using the CT scan as a model [92]. Although this PhD found that simulating the inferior part of the rectum did not significantly affect the DIR (section 4.5.2),

variation in the inferior rectal length has been shown to be smaller than the variation in the superior rectum [150, 201] and therefore the results of section 4.5.2 cannot be extrapolated to the superior part of the rectum. By simulating the superior rectum, the effect of its deformation was neglected in the VoxTox analysis.

Furthermore, in order to understand the clinical impact of the dose difference between planned and accumulated dose-volume, it is necessary to associate the toxicity with the actual delivered dose. The author is aware of only one study which has associated accumulated dose with toxicity. In a subsequent study, the VoxTox group summed daily DSM to calculate the delivered dose in 109 patients [93]. In addition to using spatial metrics derived from DSM to describe the delivered dose, they also calculated the gEUD. The study found that for all patients, accumulated gEUD was lower than planned (mean difference -2.2 Gy, standard error 0.3 Gy (range: -0.3 to -7.1Gy). When associated with toxicity, there was a stronger association between accumulated gEUD with rectal bleeding and proctitis than the planned gEUD. In addition to the need to simulate missing rectal length on the treatment images (discussed above), the study was limited by the need to collect toxicity data retrospectively, which meant there was no baseline toxicity data available.

The following two chapters describe an investigation into the variation of the size of the rectal volume during treatment, a comparison of the planned and delivered doses to the rectum, and an assessment of the dose in terms of dose-volume constraints (DVC) met and violated, and the relationship between DVC and toxicity. In this chapter (Chapter 5) the analysis of the variation in rectal volume, and differences between dose-volume between planned and delivered dose in terms of number of DVC met and breached are presented. The second chapter (Chapter 6) presents the analysis of the association of toxicity with planned and accumulated dose.

5.2 Aims

The aim of this study was to evaluate the difference between planned and delivered dose using the planning CT and daily treatment CBCTs using DIR. This study used a larger cohort of patients than has been previously studied. Furthermore, in this current cohort of patients, there was opportunity to investigate difference in rectal volume of patients receiving two different enema regimes. Patients in the standard cohort were prescribed enemas for the first 5 fractions of treatment, and hypofractionated patients

for the first 10 fractions of treatment. In this chapter, delivered dose refers to either the daily calculated dose for each fraction, or the total dose accumulated using DIR. To explore the effect of rectal volume variation on delivered dose, the following aspects have been analysed.

1. Daily CBCTs were used to measure the magnitude of the interfraction variation in rectal volume over the course of treatment. The effectiveness of the different enema regimes used in the DELINEATE study were evaluated by comparing the consistency of rectal volume during treatment in the two different regimes.
2. For each cohort, the number of daily and accumulated DVHs which exceed planning DVCs were measured to determine whether, notwithstanding interfraction rectal variation, planning DVC were routinely met during treatment.
3. To provide a larger study population for future analysis of dose-volume and toxicity (Chapter 6), the cohorts were combined by converting the DVH in the hypofractionated cohort to 2 Gy/# using the Wither's equation [156]. The difference between planned and accumulated dose-volume will be calculated using the combined cohorts.

5.3 Methods and materials

For each patient, the daily and accumulated rectal dose was calculated using daily CBCTs and DIR. A more detailed methodology is given in section 2.3.3. Briefly, the rectum was manually delineated by the author on each CBCT. The CBCTs were assigned mass density values using RayStation's bulk density correction method [27, 125]. This enabled the direct calculation of daily dose using the CBCT and the RayStation treatment planning software. The CBCT was deformably registered to the planning CT using RayStation's ANACONDA hybrid algorithm [146]. Deformation vector fields (DVF) generated during the DIR were used to deform the daily dose calculated on the CBCT, and map it to the CT. Images with poor DIR were excluded from dose accumulation (See section 2.3.2.1). Daily doses were weighted to compensate for missing or excluded CBCTs, and summed to generate the accumulated dose.

5.3.1 Rectal volume variation and comparison of different enema regimes in the DELINEATE trial

For each patient, the rectal volume at planning was calculated using the contour delineated on CT planning scan for the purpose of planning in the clinic. The rectal volume on the CBCT was delineated by the author using the DELINEATE delineation criteria (discussed in Section 2.2.2). The rectal volumes for each patient were manually extracted from RayStation and entered into a Microsoft Excel spreadsheet.

5.3.2 Dose-volume constraints

For each patient the planned, daily, and accumulated DVHs were exported from RayStation to Microsoft Excel. The number of times the daily and accumulated dose-volumes represented by the DVH exceeded the optimal and clinically required constraints was calculated for each cohort. The median and IQR of the rectal dose-volume at both the DELINEATE required and optimal dose constraints were also calculated (section 2.2.2). In addition, it was useful to measure how similar the planned and accumulated dose-volumes were to DVC. For example, it is hypothesised that the likelihood of accumulated DVH exceeding DVC is small if the planning rectal dose volumes are initially much smaller than the DVC. Conversely, if the planned dose-volumes are similar to the DVC, it may be more likely that the DVC will be exceeded during treatment. Therefore, the difference between the accumulated DVH and the planning DVC was calculated by subtracting the accumulated rectal dose-volume from the DVC. This was repeated using the planning DVH to provide the comparison. The daily DVHs were also averaged and the average rectal dose-volume was subtracted from the DVC. This provided a comparison between accumulating dose and averaging the daily dose, which is used in a number of studies in the literature [31, 32].

5.3.3 Difference between planned and accumulated dosimetry for combined cohorts

In this thesis, future association of dose-volume and toxicity in Chapter 6 is performed by combining the cohorts to provide a larger patient population for analysis. For consistency, the cohorts in this study were combined when measuring the difference in rectal-volume between planned and delivered dose. To combine the cohorts, the planned and accumulated DVH for the hypofractionated dose cohort were converted

to equivalent 2 Gy per fraction using the Wither's equation given in section 2.2.5 [156].

The difference in the rectal dose-volume between planned and accumulated treatments was analysed using DVHs. To determine the doses at which the greatest difference between planned and accumulated dose is observed, the accumulated rectal volume was subtracted from the planned rectal volume at 10 Gy intervals (10 to 70 Gy) and analysed. The correlation between planned and accumulated rectal volume at 10 Gy intervals (10 to 70 Gy) was also evaluated.

5.4 Statistical analysis

5.4.1 Rectal volume variation and comparison of different enema regimes in the DELINEATE trial

The median and IQR of the rectal volume during treatment was calculated for each patient. A Mann Whitney U (MWU) test was used to test for a statistically significant difference between the median and IQR (across patients) of the rectal volume between the two cohorts (74 Gy vs 60 Gy). Mann Whitney U was also used to compare the effect of different enema regimes on rectal volume consistency by testing whether there was statistically significant difference in rectal volume between cohorts on a week by week basis. In addition, the median and inter-quartile range (IQR) of the planning volume and weekly median (IQR) for each cohort was calculated. A Pearson correlation coefficient was used to determine if there was a linear relationship between the rectal volume at planning and the median (IQR) rectal volume during treatment.

5.4.2 Dose-volume constraints

The proportion of accumulated and daily DVHs that exceeded constraints was recorded. A Freidman test with post hoc analysis was used to determine if there were statistically significant differences between the difference in the DVC and the planned, accumulated and averaged dose-volumes. These differences were analysed for both optimal and required DVC (Section 2.2.2).

5.4.3 Difference between planned and accumulated dose for combined cohorts

A Wilcoxon signed rank (WSR) test was used to test for a statistically significant difference between the planned and accumulated rectal dose-volumes. A Pearson correlation coefficient was used to examine the correlation between planned and accumulated rectal dose-volumes at different doses.

5.5 Results

Of the 106 patients initially identified for inclusion in the study, 86 were suitable for dose accumulation using DIR (standard cohort = 46 and hypofractionated dose cohort = 40). One patient withdrew, five patients were excluded because their imaging data could not be found in the archive. Two patients were excluded as their 2 year toxicity data would not be available within the time frame of the study. Three patients were excluded as the CBCT field of view did not include the superior rectum, and eight patients were excluded as the CBCT image quality was too poor to confidently delineate the rectal volume. Finally, one patient was excluded as the majority of his CBCT to CT registrations were poor (see section 2.3.2.1).

The clinical characteristics and demographics of the 86 patients analysed are shown in Table 5.1. Characteristics by cohort can be found in the appendices (Appendix 2)

	<u>Combined cohort (%)</u>
Age (years)	72 (57-80)
T Stage	
T1a/b/c/x	39 (45)
T2a/b/c/x	35 (41)
T3/x	12 (14)
Combined Gleason Score	
6	21 (24)
7	60 (70)
8	5 (6)
NCCN Risk Group	
Intermediate	60 (70)
High	26 (30)
Diabetes	
Yes	13 (15)
No	73 (85)
Hypertension	
Yes	41 (48)
No	45 (52)
Inflammatory bowel or diverticular disease	
Yes	7 (8)
No	79 (92)
Pelvic surgery	
Yes	14 (16)
No	72 (84)
Symptomatic Haemorrhoids in the last 12 months	
Yes	13 (15)
No	72 (84)
Unknown	1 (1)
Previous transurethral resection of the prostate	
Yes	7 (8)
No	79 (92)
Statins	
Yes	35 (41)
No	51 (59)
Current Smoker	
Yes	7 (8)
No	78 (91)
Unknown	1 (1)

Table 5.1 Clinical characteristics and demographics of the patients analysed (n=86). There were 46 patients in the standard prescription cohort, and 40 patients in the hypofractionated cohort.

5.5.1 Rectal volume variation and comparison of different enema regimes in the DELINEATE trial

The median and IQR of the rectal volume over the course of treatment was calculated for all patients and is shown in Figure 5.1. A MWU test showed there was no statistical difference in the median ($p = 0.84$), or IQR ($p = 0.13$) of the rectal volume of patients between the two cohorts

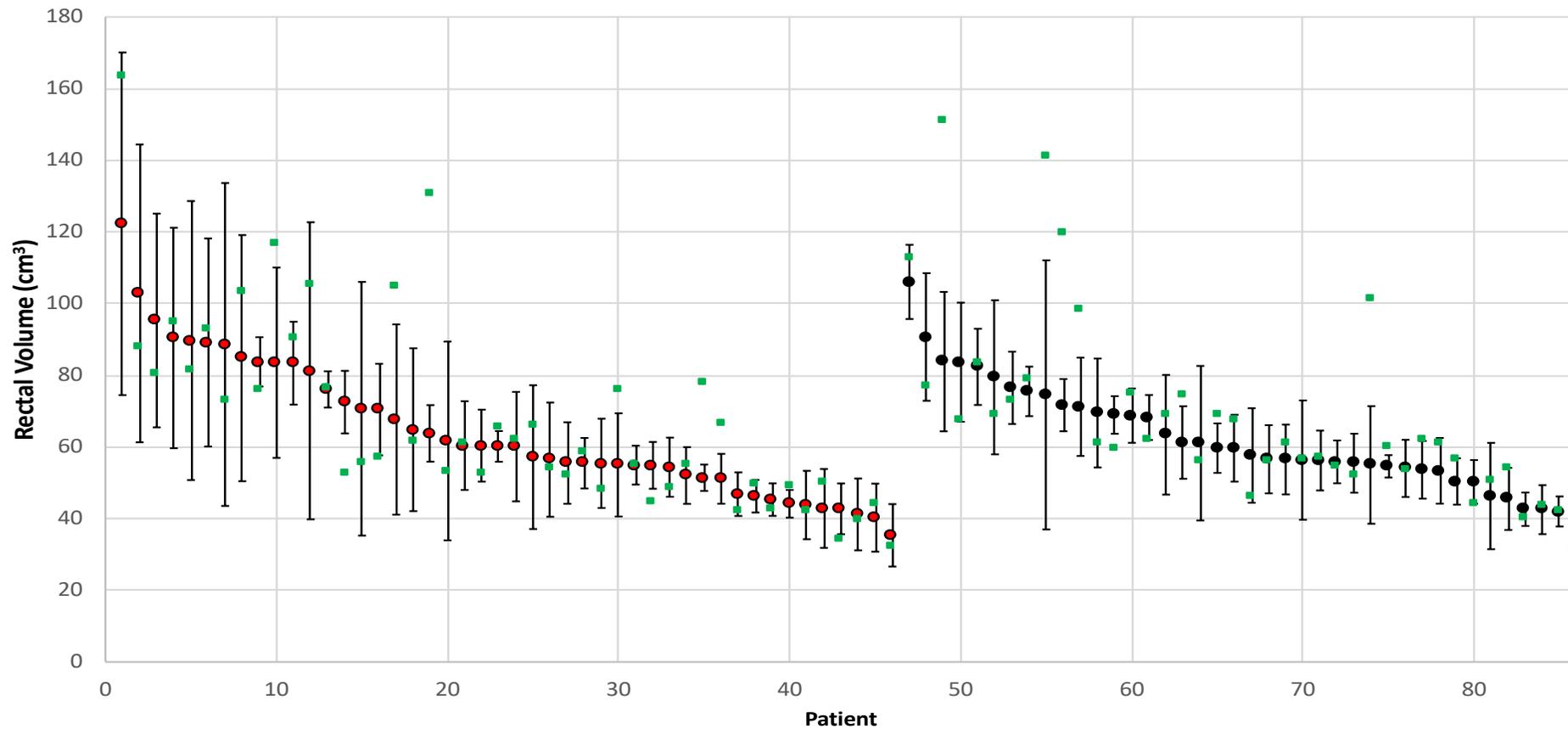


Figure 5.1 Median (markers) and IQR (shown by the error bars) of the rectal volume during the course of treatment. The red markers are the median rectal volume for patients in the standard cohort (n=46), and black for patients in the hypofractionated cohort (n=40), the planning volume for each patient is indicated by the green squares. The median rectal volume for each cohort is ordered from greatest to smallest.

A MWU test showed no statistically significant difference in rectal volume between cohorts when compared on a week by week basis, including comparison of the final week. For the standard fraction cohort, the final week was defined as the last 5 # of treatment. The median (IQR) planning volume and the median (IQR) for each cohort by week (weeks 1 to 3 and final week) is given below in Table 5.2. In both cohorts the median (IQR) treatment rectal volume is consistent across treatment, and the median (IQR) treatment volumes are similar in magnitude to the median (IQR) planning volume.

Cohort	Week				
	Planning	1	2	3	Final week
Standard (74Gy/37#)	60.1 (27.0)	60.5 (30.9)	62.7 (31.7)	61.0 (37.8)	58.1 (26.9)
Hypofractionated (60Gy/20#)	61.1 (20.3)	61.1 (21.0)	59.9 (19.4)	62.5 (22.0)	62.7 (27.7)

Table 5.2 Median (IQR) planning rectal volume and the median (IQR) rectal volume on a week by week basis for each cohort. Treatment weeks 1 to 3 are presented alongside the final week which is defined as the last 5# of treatment.

The difference in rectal volume between planning and treatment was calculated for each fraction. The rectum was larger at treatment than planning in 1,097 (48.5 %) of the 2,261 fractions analysed. On average the difference between the planning rectal volume and the daily rectal volume was small (median (IQR) 0.75 (12.73) cm³). There was a statistically significant, strong positive correlation between the rectal volume at planning and the median treatment volume, $r = 0.74, p < 0.005$ (see Figure 5.2) [218, 219]. There was a statistically significant, but moderate positive correlation between the planned rectal volume and the IQR of the treatment volume $r = 0.54, p < 0.005$ (see Figure 5.3) [218, 219].

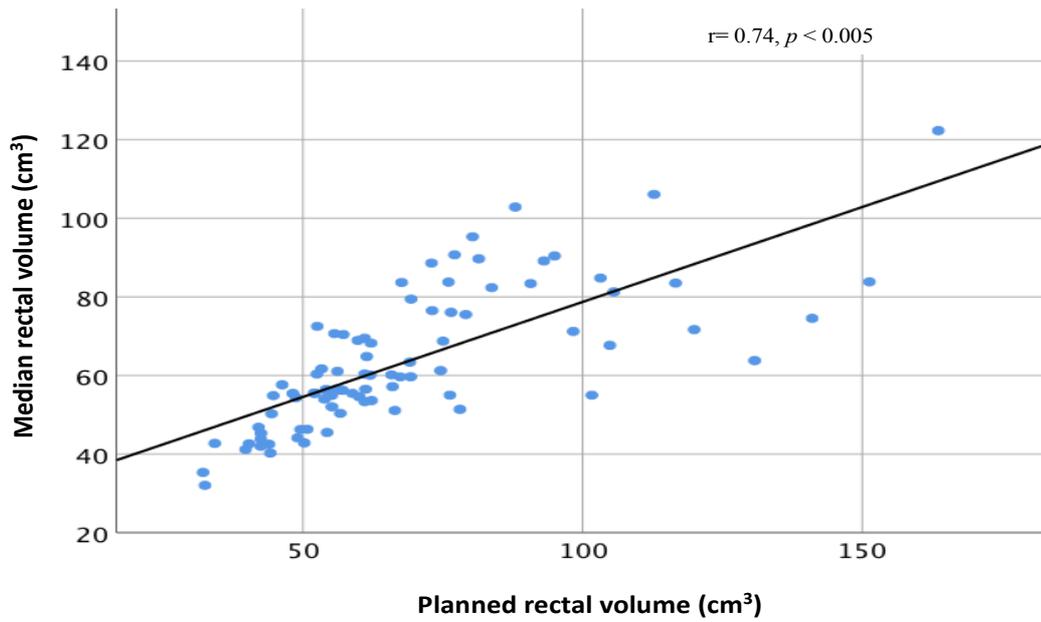


Figure 5.2 Scatter plot illustrating the correlation between the planned and median treatment rectal volume. The line of best fit, Pearson correlation coefficient, r , and p value are given.

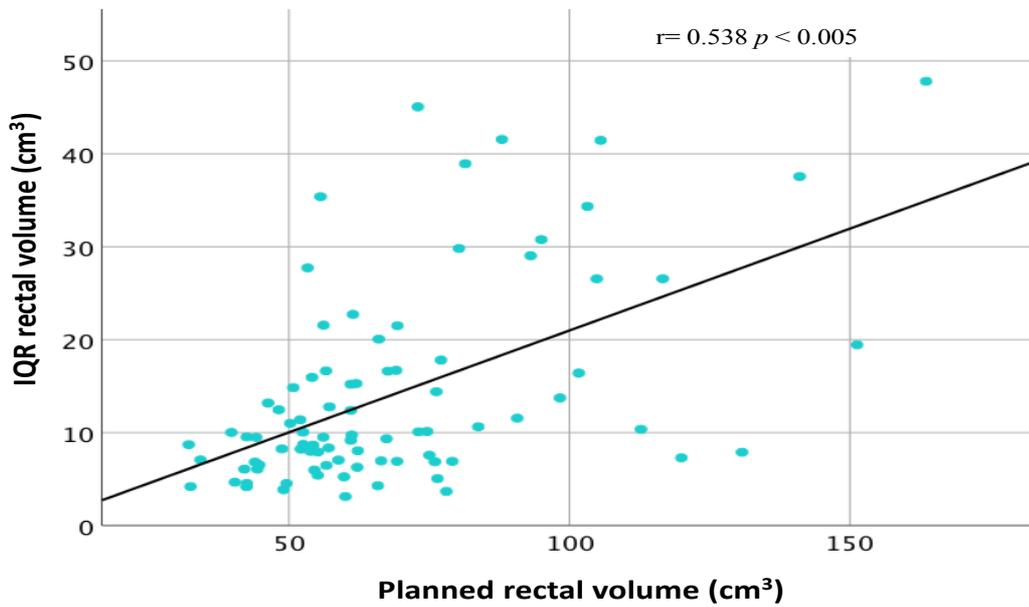


Figure 5.3 Scatter plot illustrating the correlation between the planned rectal volume and IQR of the treatment rectal volume. The line of best fit, Pearson correlation coefficient, r , and p value are given.

5.5.2 Dose constraint analysis

Planned and accumulated DVHs for each cohort are shown in Figures 5.4 and 5.5. The hypofractionated cohort is shown without EQD2 correction. The number of fractions where the daily and accumulated rectal volumes exceeded optimal and required dose constraints are given in Tables 5.3 and 5.4, respectively.

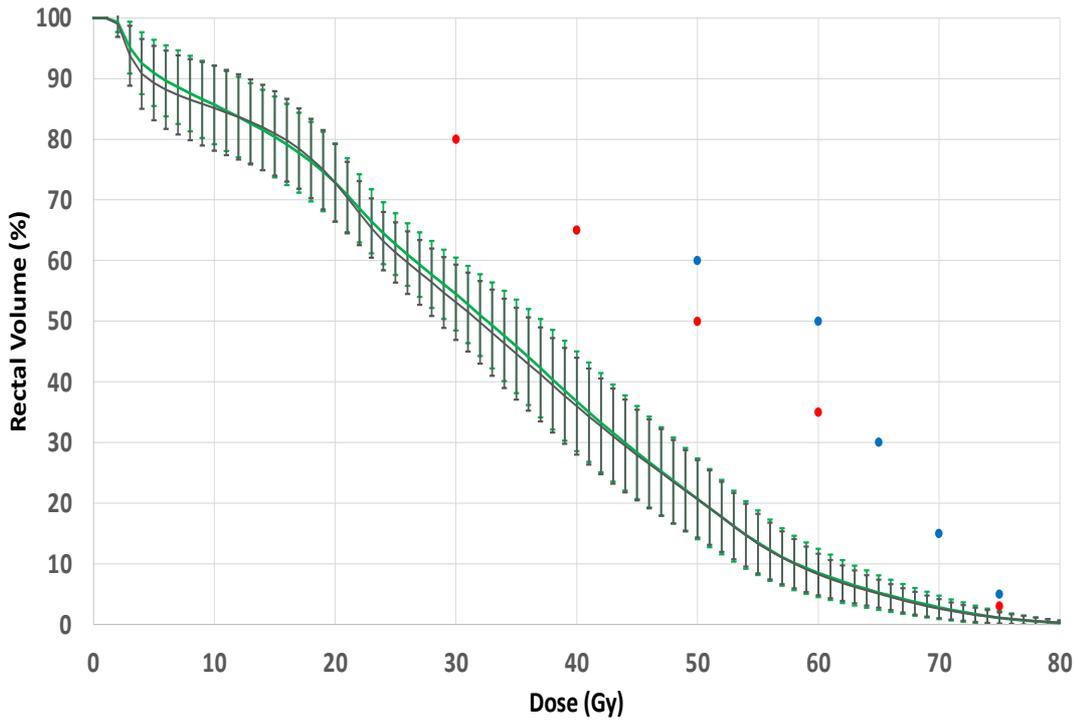


Figure 5.4 The average planned DVH (grey) and averaged accumulated DVH (green) for the standard cohort. The standard deviations is given as an error bar. The required constraints (blue), and optimal dose-volume constraints (red) are also plotted.

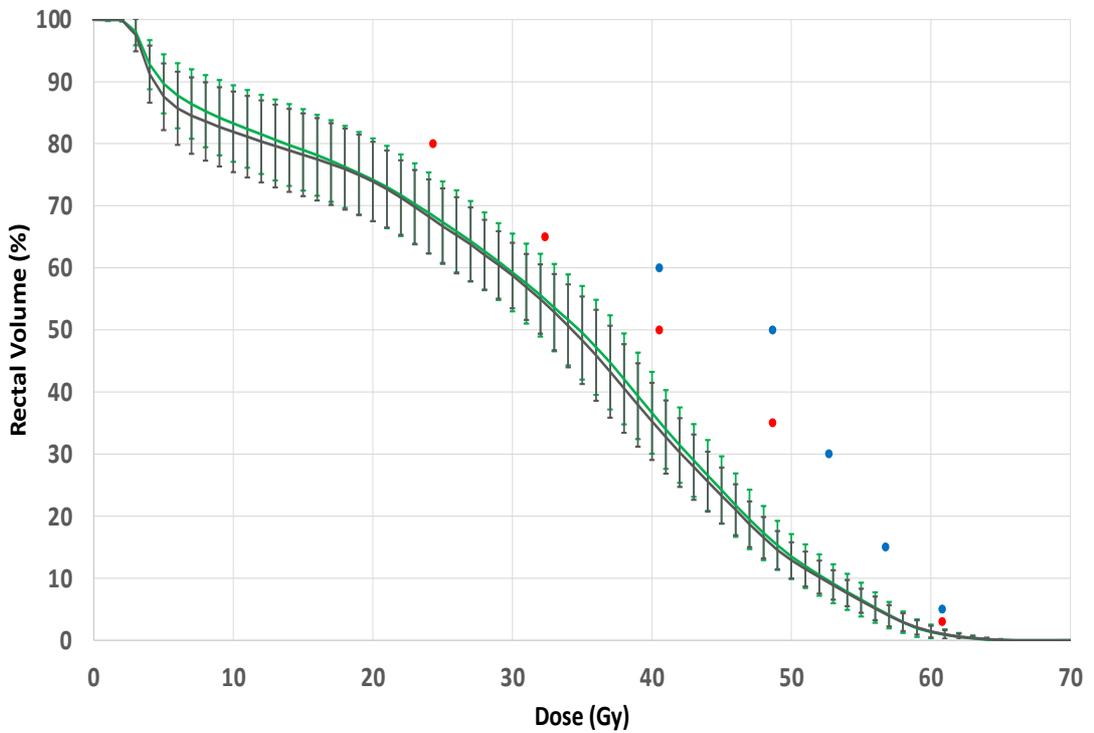


Figure 5.5 The average planned (grey) and average accumulated DVHs (green) for patients in the hypofractionated cohort (no EQD2 correction). Required (blue) and optimal (red) dose-volume constraints are also plotted.

Standard arm (accumulated)						
Vd	Optimal Volume	Number exceeding DVC (%)	Median Rectal Volume (IQR)	Required Volume	Number exceeding DVC (%)	Median Rectal Volume (IQR)
Volume to 30 Gy	< 80%	0	68.2 (7.2)			
Volume to 40 Gy	< 65%	0	49.1 (8.6)			
Volume to 50 Gy	< 50%	0	32.2 (11.1)	< 60 %	0	32.2 (11.1)
Volume to 60 Gy	< 35%	0	13.1 (7.6)	< 50%	0	13.1 (7.6)
Volume to 65 Gy				< 30%	0	6.9 (5.2)
Volume to 70 Gy				<15%	0	3.2 (3.2)
Volume to 75 Gy	< 3%	3 (6.5)	0.7 (1.2)	< 5%	0	0.7 (1.2)

Hypofractionated arm (accumulated)						
Vd	Optimal Volume	Number exceeding DVC (%)	Median Rectal Volume (IQR)	Required Volume	Number exceeding DVC (%)	Median Rectal Volume (IQR)
Volume to 24.32 Gy	< 80%	2 (5)	67.2 (5.0)			
Volume to 32.43 Gy	< 65%	3 (7.5)	54.4 (7.5)			
Volume to 40.54 Gy	< 50%	0	34.9 (7.7)	< 60%	0	34.9 (7.7)
Volume to 48.65 Gy	< 35%	0	15.4 (6.4)	< 50%	0	15.4 (6.4)
Volume to 52.70 Gy				< 30%	0	9.2 (4.8)
Volume to 56.76 Gy				< 15%	0	4.0 (3.1)
Volume to 60.81 Gy	< 3 %	1 (2.5)	0.7 (1.3)	< 5%	0	0.7 (1.3)

Table 5.3 The number of fractions where the accumulated rectal dose-volume exceeded the required DVC and optimal DVC. The median and IQR (%) for the accumulated rectal volume is given at each dose level. The hypofractionated cohort was not EQD2 corrected for these analyses. Standard cohort (n=46) and hypofractionated cohort (n=40).

Standard arm (daily)						
Vd	Optimal Volume	Number of fractions exceeding DVC (%)	Median Rectal Volume (IQR)	Required Volume	Number of fractions exceeding DVC (%)	Median Rectal Volume (IQR)
Volume to 30 Gy	< 80%	48 (3.2)	69.7 (7.7)			
Volume to 40 Gy	< 65%	114 (7.6)	53.3 (11.7)			
Volume to 50 Gy	< 50%	63 (4.2)	34.2 (14.2)	< 60 %	2 (0.1)	34.2 (14.2)
Volume to 60 Gy	< 35%	4 (0.3)	12.9 (9.2)	< 50%	0 (0)	12.9 (9.2)
Volume to 65 Gy				< 30%	1 (0.1)	6.2 (5.7)
Volume to 70 Gy				<15%	9 (0.6)	2.8 (3.2)
Volume to 75 Gy	< 3%	100 (6.8)	0.7 (1.4)	< 5%	5 (0.3)	0.7 (1.4)

Hypofractionated arm (daily)						
Vd	Optimal Volume	Number of fractions exceeding DVC (%)	Median Rectal Volume (IQR)	Required Volume	Number of fractions exceeding DVC (%)	Median Rectal Volume (IQR)
Volume to 24.32 Gy	< 80%	89 (11.7)	71.2 (9.6)			
Volume to 32.43 Gy	< 65%	160 (21.1)	59.6 (11.8)			
Volume to 40.54 Gy	< 50%	71 (9.4)	37.7 (12.4)	< 60%	1 (0.1)	37.7 (12.4)
Volume to 48.65 Gy	< 35%	0	16.2 (7.2)	< 50%	0	16.2 (7.2)
Volume to 52.70 Gy				< 30%	0	9.2 (5.0)
Volume to 56.76 Gy				< 15%	0	4.0 (3.2)
Volume to 60.81 Gy	< 3 %	38 (5.0)	0.7% (1.4%)	< 5%	3 (0.4)	0.7 (1.4)

Table 5.4 Table showing the number of fractions where the daily rectal dose-volume exceeded the required DVC and the optimal DVC. The median and IQR (%) for the daily rectal volume is given at each dose level. The hypofractionated cohort was not EQD2 corrected for this analysis. Standard cohort (n=1482 fractions) and hypofractionated cohort (n=758 fractions).

The accumulated DVHs did not breach any of the required dose-volume constraints for either cohort. For the standard fractionation cohort, the only optimal constraint exceeded by the accumulated DVH was the $V_{75} < 3\%$ which was exceeded for 3 patients. One patient in the hypofractionated cohort exceeded the optimal constraint of $V_{60.81} < 3\%$. A small proportion of daily DVHs exceeded the required DVC in both cohorts (less than 1%). A greater proportion of daily DVHs exceeded the optimal DVC, especially in the mid dose region.

Differences in rectal dose-volume between the optimal DVC and the planning, accumulated, and average daily DVHs are given below in Figures 5.6 for the standard prescription cohort, and 5.7 for the hypofractionated cohort.

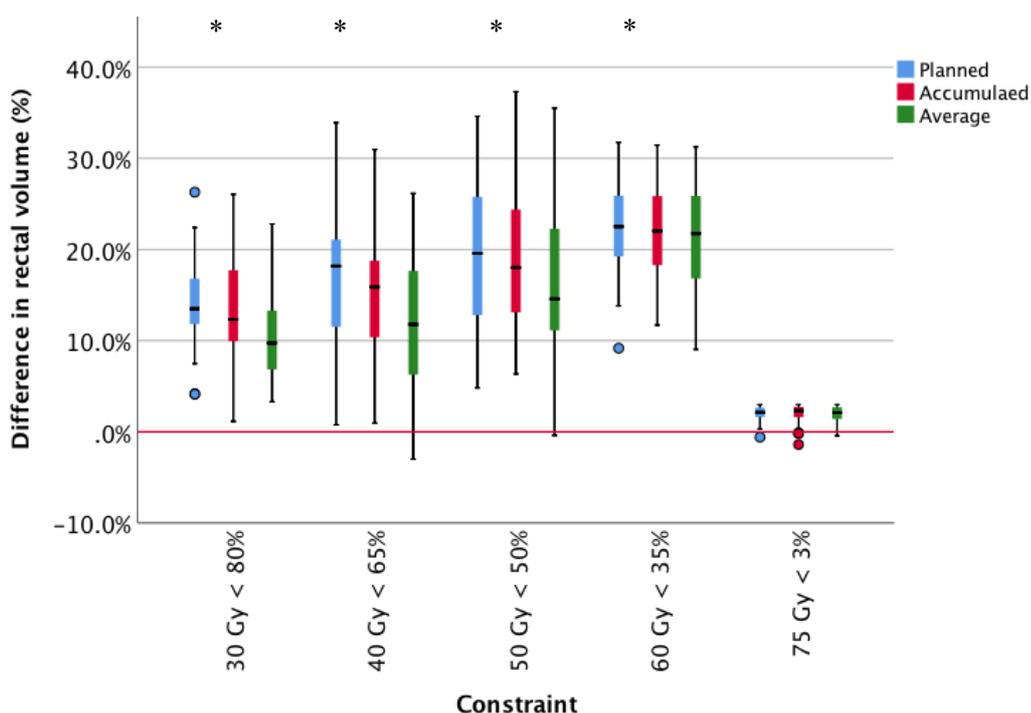
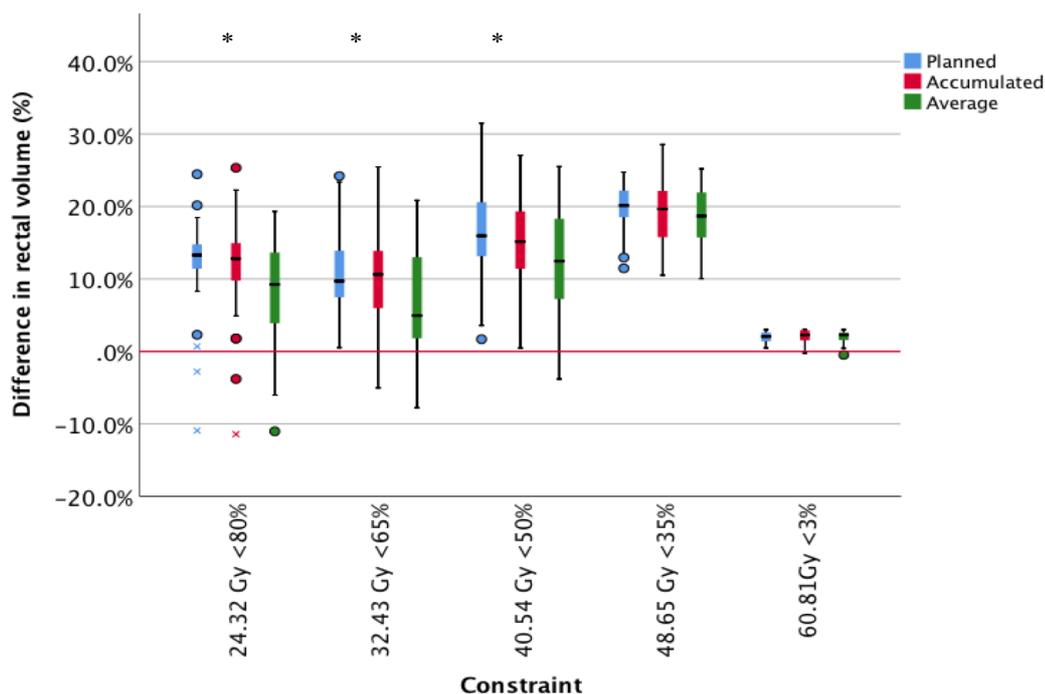


Figure 5.6 A boxplot of the difference in rectal volume between the optimal planning constraint and the planned accumulated and averaged DVH for standard cohort is shown. The volumes from planning, accumulated and averaged DVH were subtracted from the DVC, with negative results being greater than the DVC. DVC where a statistically significant difference between volumes are highlighted with an *

For the standard fractionation cohort, the difference in rectal volume between the constraint volume and planned, accumulated and averaged dose-volume were shown to be statistically significantly different at 30 Gy, 40 Gy, 50 Gy and 60 Gy ($p < 0.005$).

Post hoc analysis showed no significant difference between planned and accumulated differences from DVC in volumes ($p > 0.05$). There was a statistically significant difference between the averaged daily volume, and both the planning and accumulated volumes differences from DVC ($p < 0.05$), where averaged daily dose-volumes were greater than planned and accumulated, and therefore had smaller differences from DVC. There was no statistically significant difference between any volume at 75 Gy.



*Figure 5.7 Difference in rectal dose between the optimal constraint volume and the planned, accumulated and averaged DVH for the hypofractionated cohort. The volumes from planning, accumulated and averaged DVH were subtracted from the DVC, with negative results being greater than DVC. DVC where a statistically significant difference between volume are highlighted with an *.*

A similar trend was shown for the hypofractionated cohort with the difference in rectal volume between the DVC and planned, accumulated and averaged dose-volumes were shown to be statistically significantly different at 24.32 Gy, 32.43 Gy, 40.54 Gy ($p < 0.005$). Post hoc analysis showed no significant difference between planned and accumulated volumes ($p > 0.05$). There was, however, a statistically significant difference between the averaged daily dose-volumes and both the planning and accumulated dose-volumes ($p < 0.05$). Again, averaged daily rectal volumes were larger than both planned and accumulated volumes, and therefore the difference from the DVC was not as great. There was no statistically significant difference between any volume at 48.65 Gy and 60.81 Gy.

5.5.3 Difference between planned and accumulated dose for combined cohorts

Small but statistically significant differences in planned and accumulated rectal dose-volume were detected at 30 Gy ($p < 0.003$, median difference -1.3 %), 40 Gy ($p < 0.002$, median difference -1.84 %) and 50 Gy ($p < 0.04$, median difference -1.16 %). No statistically significant difference in rectal dose-volume was observed for 10 Gy ($p > 0.111$), 20 Gy ($p > 0.531$), 60 Gy ($p > 0.200$) and at 70 Gy ($p > 0.841$). After Holm-Bonferroni correction for multiple testing was applied, only differences at 30 Gy and 40 Gy remained statistically significant between planned and accumulated rectal volume.

The accumulated DVHs were subtracted from the planned DVH, and the difference at 10 Gy dose intervals was calculated (10 Gy to 70 Gy). The differences in the percentage of rectal dose-volume at the dose investigated are presented in Figure 5.8

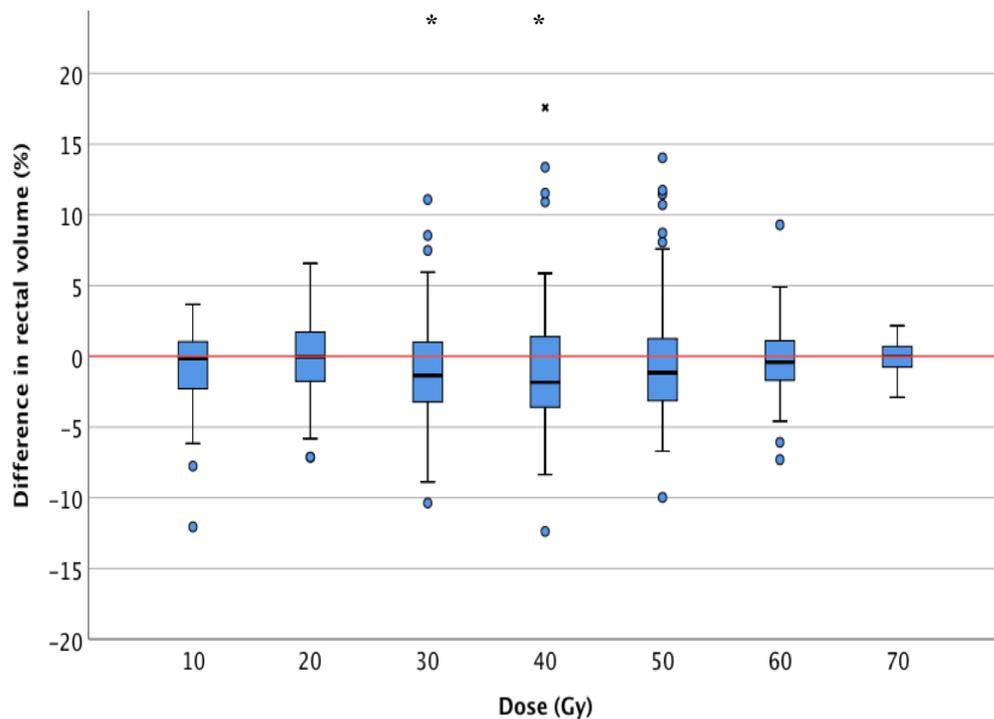


Figure 5.8 Boxplot of the difference in rectal volume (%) between planned and accumulated dose for 10 Gy dose intervals. The negative difference indicates that the accumulated rectal volume is greater than calculated at planning. An * denotes dose levels where there is a statistically significant difference between planned and accumulated rectal volume (after correction using Holm Bonferroni).

At 10 Gy, 52 % (45 patients) of the DVHs analysed had an accumulated rectal dose-volume which was greater than planned. At 20 Gy, the accumulated rectal dose-volume was greater than planned in 50 % (43) patients, at 30 Gy it was 62 % (53), at 40 Gy it was 66 % (57), at 50 Gy it was 63 % (54), and 70 Gy it was 48 % (41) greater than planned. The median (IQR) of the difference in rectal volume (planned minus accumulated volumes) at 10 Gy intervals were as follows: at 10 Gy the median rectal volume difference is -0.1 % (3.34 %), at 20 Gy 0.03 % (3.44 %), at 30 Gy the median is -1.34 % (4.24 %), at 40 Gy the median is -1.77 (4.95 %), at 60 Gy the median is -0.41 % (2.8 %) and at 70 Gy the median is 0.03 (1.5 %).

Analysis of the correlation between the planned and accumulated rectal volumes at 10 Gy dose intervals showed a statistically significant and very strong positive correlation ($r \geq 0.82$) at the low dose levels (10 Gy, 20 Gy and 30 Gy), as well as a statistically significant, strong positive correlation from the mid to high dose levels (40 Gy to 70 Gy) ($r \geq 0.71$) [218, 219]. See Figure 5.9.

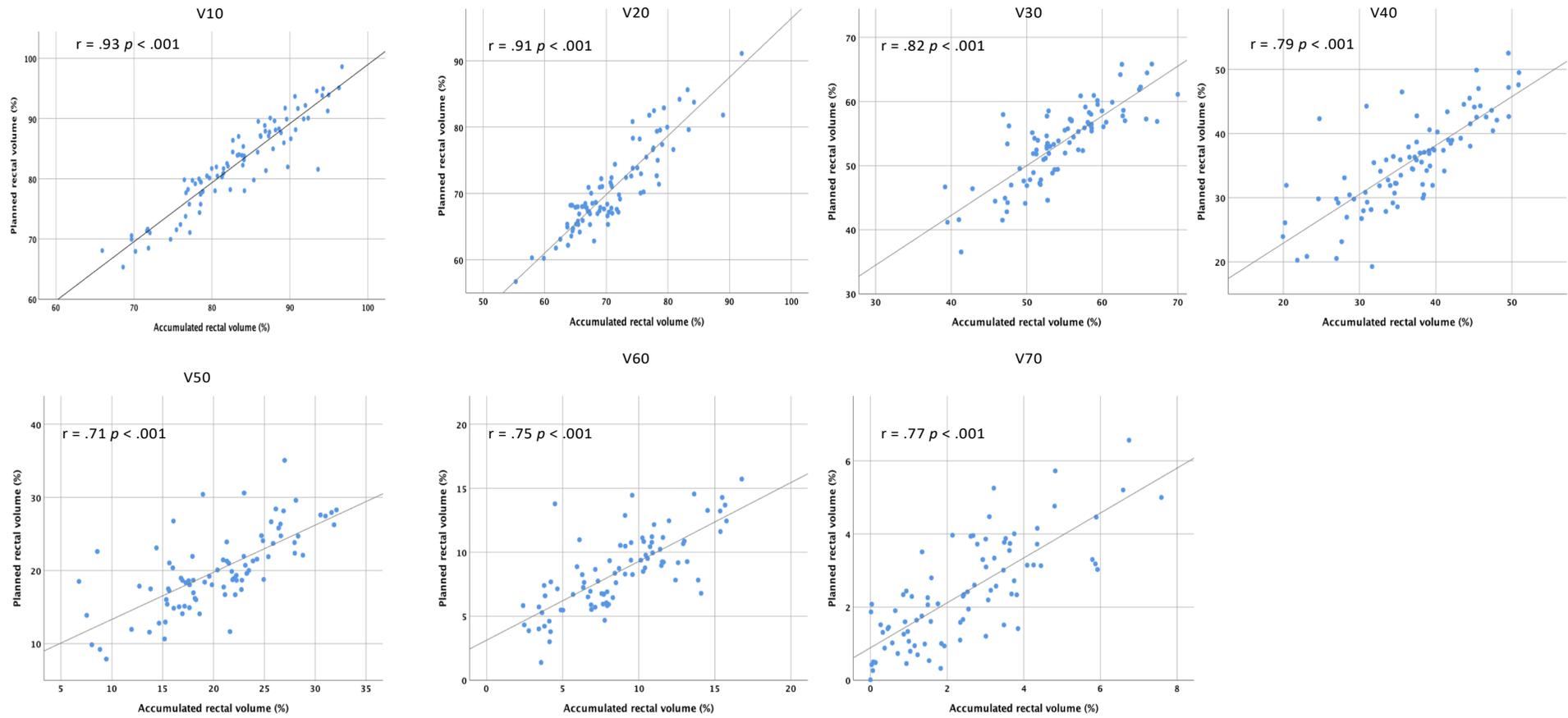


Figure 5.9 Scatter plots illustrating the correlation between planned and accumulated rectal volume at 10 Gy intervals. A line of best fit, Person correlation coefficient (r) and p value is given for each plot.

5.6 Discussion

This study shows that despite interfraction rectal volume variation, differences between the accumulated dose to the rectum and the initial treatment plan remain small (for 50 % of DVHs analysed the difference was < 5% at specific dose-volumes). This indicates that in the context of IMRT to the prostate using on-line image guidance (CBCT and fiducial markers), the planned dose provides a good estimation of delivered dose and the number of dose-volume constraints that may be breached during delivery.

5.6.1 Rectal volume variation and comparison of different enema regimes in the DELINEATE trial

In agreement with the literature, this study found some patients had stable and consistent rectal volumes during treatment, while others had larger variations [31, 216] (see Figure 5.1). There has been limited success in previous studies in identifying, prior to treatment, which patients will have a large interfraction variation during treatment, with the conclusion that rectal volume variation is random and individual to the patient [105, 201, 216]. In this study a statistically significant, strong positive correlation between the planning volume and the median treatment volume ($r = 0.74$, $p < 0.005$) was observed, however there was also a weaker correlation between the planning volume and the variation during treatment suggested by the IQR ($r = 0.54$, $p < 0.005$). This suggests that the planning volume may be indicative of the overall median treatment volume, but it is not so strongly predictive of the magnitude or frequency of rectal volume variation over treatment. Inspection of Figure 5.1. also suggests that the IQR tends to be greater for patients with a larger median rectal volume, particularly for the standard cohort patients. A study by Sripadam et al., reported that daily variations occurred in all patients irrespective of the planning volume size [105]. However, it was also noted that changes in volume tended to be smaller in patients with a small planning volume. As with the current study, this was not observed for all patients.

Patients included in the DELINEATE trial were asked to use enemas prior to treatment delivery to improve consistency in the rectal volume over the course of treatment. Standard cohort patients were prescribed enemas for the first 5 # of treatment, and the hypofractionated cohort patients were prescribed enemas for the first 10 # of

treatment. Patients were given enemas for the first part of treatment as it is assumed that as acute toxicity (rectal inflammation and bowel frequency) develops, this will assist in keeping the rectum empty. The incidence of acute toxicity was not analysed in this thesis. However the CHHiP trial showed acute toxicity for the hypofractionated cohort was greater than the standard fraction cohort and presented at a different time point [52]. Acute toxicity in the hypofractionated cohort peaked 3 to 4 weeks after commencement of treatment, and for the standard fraction cohort it peaked 7 to 8 weeks after the commencement of treatment. Comparison of the median rectal volume by week (week 1 to 3 and final week) between cohorts showed no statistically significant difference in the median rectal volume. The results indicate that giving 1 week of enemas in the standard fraction group is sufficient to maintain a consistent rectal volume. This is also mirrored in the hypofractionated cohort. However, as acute toxicity has been shown to occur earlier in the hypofractionated cohort, it is reasonable to suggest that 1 week (5#) rather than 2 weeks (10#) of enemas is sufficient to maintain a consistent rectal volume over treatment. As insertion of enemas can lead to anxiety for some patients, it follows that reducing the number of times enemas need to be inserted may improve patient experience. As fewer enemas are required it also has the potential to reduce costs.

In addition to evaluating the optimal number of enemas, it may be useful to determine whether, with the advent of online imaging protocols, the use of bowel preparation strategies remain beneficial. Bowel preparation strategies are used in an attempt to maintain a consistent rectal volume, so as to reduce prostate movement and avoid sub-optimal prostate target dosimetry and “geographic miss”. Such strategies may also reduce the variation in rectal dose during treatment, but are a result of studies which showed that a large rectal volume at CT resulted in poorer outcome for patients receiving radical radiotherapy [32, 87, 220]. It is likely this was a consequence of a smaller rectum at treatment resulting in partial coverage of the prostate as it moved posteriorly out of the PTV. These studies were conducted using offline imaging protocols, and at a time when image registration of the prostate was often limited to bony matches, which is a poor surrogate for prostate position [87, 122, 138, 220]. Subsequent studies have shown that using volumetric imaging, and a daily online imaging protocol to correct for prostate position, results in the prostate receiving the intended dose and reduces dose to the rectum [115, 221]. Furthermore, a study by Huang et al., found that the variation in mean rectal volume (36 %, SD \pm 29) was

greater than the subsequent variation in mean rectal dose (22 %, SD ± 15) [132]. However, while the geometric uncertainty of the prostate can be corrected by using IGRT and fiducial markers, the seminal vesicles can also be displaced by rectal gas and this would also need to be considered [99, 222].

5.6.2 Dose-volume constraints

A number of studies have reported differences between the planned and delivered dose to the rectum, with delivered dose often greater than planned [31, 32, 130, 216]. Calculating whether delivered dose exceeds DVC has been used to evaluate the dosimetric impact of rectal volume variation [31, 32, 130]. In this study, the analysis of accumulated dose showed no patient exceeded the *required* DVCs in either cohort. However, both cohorts had a small number of patients whose accumulated DVHs exceeded the *optimal* DVC at high doses, ($V_{75} < 3\%$ in the standard cohort (3 patients), and the arithmetic equivalent $V_{60.81} < 3\%$ in the hypofractionated cohort (1 patient)). It is possible that differences between planned and accumulated dose-volumes may be attributed to the uncertainty of deformable registration in high dose region (> 60 Gy), as well as the small volume of the rectum receiving high dose (0-3 %). As previously discussed in the Section 1.5.2 and Section 3.6, a study by Nassef et al., demonstrated that small differences in registration when deforming the image can result in large differences in dose when there is a steep dose gradient, such as the prostate rectal interface [148]. One of the few studies to accumulate rectal dose using deformable image registration reported no statistically significant deviations from the planning DVH [138]. Furthermore, the authors reported that the greatest (non-statistically significant) variation between planned and accumulated dose was in the high dose region. However, they did not report whether verification of the deformable image registration was performed which is recommended in the literature [135, 142] and was performed for the current study (see section 2.3.2.1).

In addition, the hypofractionated cohort only had a small number of patients where the accumulated DVHs exceeded optimal dose constraints in the low dose region, with 2 patients exceeding $V_{24.32} > 80\%$ and 3 patients exceeding $V_{32.43} > 65\%$. The patients whose accumulated DVH exceeded the DVC $V_{24.32}$ were also found to have planning DVHs which exceeded this constraint. However, for $V_{32.43}$, the planning DVH met the DVC, but only just. This suggests that despite interfraction rectal volume

variation, the planned dose-volume is sufficiently predictive of the accumulated dose-volume.

Approximately 1 % of daily fractions exceeded the *required* DVC for both dose cohorts, with a larger number exceeding the *optimal* DVC. In the standard cohort the *optimal* constraint which had the greatest percentage of daily DVHs exceeding it was at $V_{40}<65\%$ (114 fractions or 7.6 % of fractions). This is the arithmetic equivalent of the hypofractionated cohort constraint $V_{32.4}<65\%$, which was also the constraint with the greatest number of fractions exceeding the DVC (160 fractions or 21.1 %). Other authors have reported similar results. Chen reported that 28 % (39/139) of weekly DVHs analysed in their study exceed the more stringent $V_{40}<35\%$ dose constraint. Hatton et al. reported that differences between planned and daily DVHs were found at 40 Gy for all 12 patients when the twice weekly CBCTs were analysed [32]. However, they did not specify the magnitude of the dose-volume differences.

On average, the differences in dose-volume between planned and accumulated dose were small ($< 3\%$). However, there was variation between patients and in some cases the accumulated dose-volumes were larger than planned with the IQR $> 5\%$ (see Figure 5.8). Despite this variation, DVC were not exceeded because planned dose-volumes were much lower (on average between 10 and 20 % lower) than the optimal DVC (see Figure 5.6 and 5.7). This was also observed by Collery et al., who calculated the mean rectal volume and compared it to the planning volume [130]. For all 10 patients there was a significant difference between the mean plan and mean treatment volume for all constraints, with the mean on treatment volume greater than planned. However, at $V_{50}<50\%$, $V_{60}<35\%$ and $V_{65}<25\%$ the mean treatment volume was within DVC. At $V_{70}<20\%$ and $V_{75}<15\%$, the mean on-treatment volume was greater than the DVC. This led Collery et al., to postulate that by ensuring the planning DVH is kept well below DVC, the subsequent interfraction rectal variation is unlikely to cause the total delivered dose to exceed prescribed constraints. In this study (PhD) there were no statistically significant differences between the between the planned and accumulated rectal dose-volumes when subtracted from the DVC (see Figure 5.6 and 5.7). The lack of a statistically significant difference between the planned and accumulated rectal volume supports Collery's suggestion that keeping the planned DVH well below the prescribed DVC will ensure that the accumulated dose will remain predominantly within prescribed limits despite rectal volume variation.

In this study, the difference between planned and delivered dose-volume was not as great as the literature suggested. This may be because the majority of previous studies that measured differences between planned and delivered dose frequently used daily or averaged DVHs [31, 32, 92, 130, 216]. This can result in overestimation of delivered dose as illustrated in a study by Rigaud et al., and discussed in Section 1.4. [129]. Summing or averaging DVHs does not establish the spatial correspondence between images, which is important when calculating the delivered dose in a heterogenous dose distribution such as the rectum [139]. Deforming the treatment image to the CT establishes spatial correspondence enabling the dose to be mapped on a voxel by voxel level. This ensures that hot spots from one fraction are mapped to the corresponding part of the rectum for earlier and subsequent fractions which may have received lower doses. This mitigates the impact of the hotspots, which may have accounted for an exceeded DVC, over the course of treatment

5.6.3 Difference between planned and accumulated dose for combined cohorts

There was a strong positive correlation between planned and accumulated rectal volumes at low doses (10 to 40 Gy). Treatment with an IMRT technique would result in large portions of the rectum to be encompassed in a low dose bath where there is a shallow dose gradient. As a consequence differences in rectal shape and volume in this homogenous dose region are not as pronounced. As expected the correlation weakened as the dose increased. A weakened relationship at higher doses may be attributed to uncertainty when delineating the rectum (higher dose-volumes tend to be smaller and at the surface of the rectum), or uncertainty associated with the DIR [148]. Greater caution must be exercised when accumulating dose in areas where there is a steep dose gradient, such as at the prostate rectal interface compared to areas in which the dose gradient is shallow, i.e. small differences in registration can lead to large differences in dose [128, 148]. Nevertheless, studies have shown the anterior rectal wall, which is closest to the high dose region, moves more than the posterior wall with Hoogeman et al., reporting variations as great as 8mm for the anterior wall compared to 2mm for the posterior wall [150]. Scaife et al., also noted greater movement in the anterior rectal wall with the variation in the posterior rectal constrained by the sacrum [150, 201]. Chen et al. suggested that it was the effect on the shape of the rectum which had a greater impact on rectal dose [31].

Based on the literature and the variation in the rectal volume observed in the current patient cohort, it was expected that when the dose was accumulated it would be greater than planned [31, 32, 92]. This was true for V30 and V40 only, i.e., low to mid-dose ranges (see Figure 5.8). Despite a greater proportion of rectum than planned receiving dose at these levels, the DVH did not exceed the required or optimal dose constraints except for 4 of the hypofractionated patients at V32.43. The next step is to determine if these variations in volume and dose are clinically relevant by associating toxicity with both planned and accumulated dose-volume data and which is presented in Chapter 6.

5.7 Conclusion

This study shows that despite interfraction rectal variation, the planned dose-volume is a good estimate of accumulated dose-volume during IG-IMRT prostate radiotherapy. Furthermore, by keeping the planned dose to the rectum well below DVC (10 to 20 % lower) it is unlikely that the accumulated rectal volume will exceed the DVC despite rectal deformation during treatment. As a result of the ability to deliver treatment within dose constraints, rectal volume variation may no longer be as great a clinical concern, and more stringent, motion-compensated dose-volume constraints could be developed which may reduce toxicity further. In addition, the study demonstrates that a longer course of 10 # enemas is not more effective than a shorter course of 5 #, which could improve patient experience.

Chapter 6 Association between planned and accumulated dose-volume with toxicity

6.1 Introduction

The introduction to this chapter is presented in Section 5.1

6.2 Aims

The aim of this chapter is to determine whether accumulated dose is a better predictor of toxicity than planned dose.

6.3 Methods and materials

The planned and accumulated dosimetric data of the eighty-six patients described in Chapter 5 are used in this study. The cumulative peak toxicity was determined for nineteen endpoints by calculating the maximum score for each patient (per endpoint) over the first two-year follow-up period. The nineteen endpoints used and the toxicity scoring is described in more detail in Section 2.2.4. If prior to treatment a patient exhibited symptoms similar to the endpoint being analysed (i.e., they were given a score for that endpoint at baseline), they were excluded from analysis for that endpoint only. This approach was also described in Section 2.2.4 [70, 223]. For the work described in the current study, the association between dose-volume and outcome was analysed by combining the cohorts to provide a larger dataset. To combine the doses from the two different dose prescriptions and fractionation schedules, the planned, daily and cumulative hypofractionated DVH were converted to equivalent dose of 2 Gy fractions using an EQD2 correction with an α/β of 3 [9]. This is described in full in Section 2.2.5.

6.4 Statistical analysis

6.4.1 Patient and clinical data

Descriptive and quantitative methods were used to analyse the reported toxicity. A Fishers exact test was used to determine if there was a statistically significant difference in the incidence of toxicity at 2 years between the two cohorts before they were combined.

6.4.2 Association of dose and toxicity

Univariate analysis was performed using Mann Whitney U test to test for a difference in both the planned and accumulated rectal volume at specific doses (10 to 70 Gy) of patients that did or did not experience toxicity.

The receiver operating characteristic (ROC) curve can be used to identify thresholds in the percentage of rectal volume at specific doses which may be used to accurately separate patients that experienced toxicity and those who did not [70]. For each endpoint, a ROC was generated for both planned and accumulated DVHs. The ROC curve was generated by plotting sensitivity against 1- specificity. The test variable was the dose-volume, analysed at 10 Gy intervals (10 Gy to 70 Gy). The outcome was the binarised toxicity endpoint (e.g. 1 = Grade 1⁺ Diarrhoea, 0 = no Diarrhoea experienced). If the area under the curve (AUC) was ≥ 0.6 , with a 95% confidence interval lower limit > 0.5 , it was considered statistically significant [56, 70, 93]. For ROC curves which were statistically significant, the Youden index (J) was used to determine a cut point of the rectal volume that may be used to predict for rectal toxicity [70, 173]. The Youden index was calculated using the following formula [224].

$$J = \text{sensitivity} + \text{specificity} - 1$$

An Atlas of complication incidence (ACI) was used to qualitatively assess the relationship between dose and toxicity by seeking to present visually how dose and the irradiated percentage volume may relate to toxicity [70, 174]. To generate the ACI, the planned and cumulative DVHs (EQD2 corrected), and toxicity were exported to MATLAB (software written by Dr Gulliford). Each atlas was based on data for a specific endpoint. The axes of the atlas were the same as for a DVH, with percentage of the rectum plotted on the vertical axis, and dose on horizontal axis. The plot was divided into dose grids of 10 Gy by 5 %. DVHs were plotted through the grid and a numeric fraction was given for each grid box. The numerator was the number of patients who were experiencing that endpoint, and the denominator was the total number of patients whose DVH was passing through that dose box. For each endpoint an ACI was generated for the planned and the accumulated DVH for both Grade 1⁺ and Grade 2⁺ toxicity.

6.5 Results

6.5.1 Patient and clinical data

The clinical characteristics and demographics of the 86 patients analysed were presented in Section 5.5. A chart showing the incidence of patients excluded at baseline for each endpoint is given in Figure 6.1, and the combined cumulative peak toxicity for all patients is given in Figure 6.2. Peak toxicity by cohort can be found in the appendices (Appendix 3)

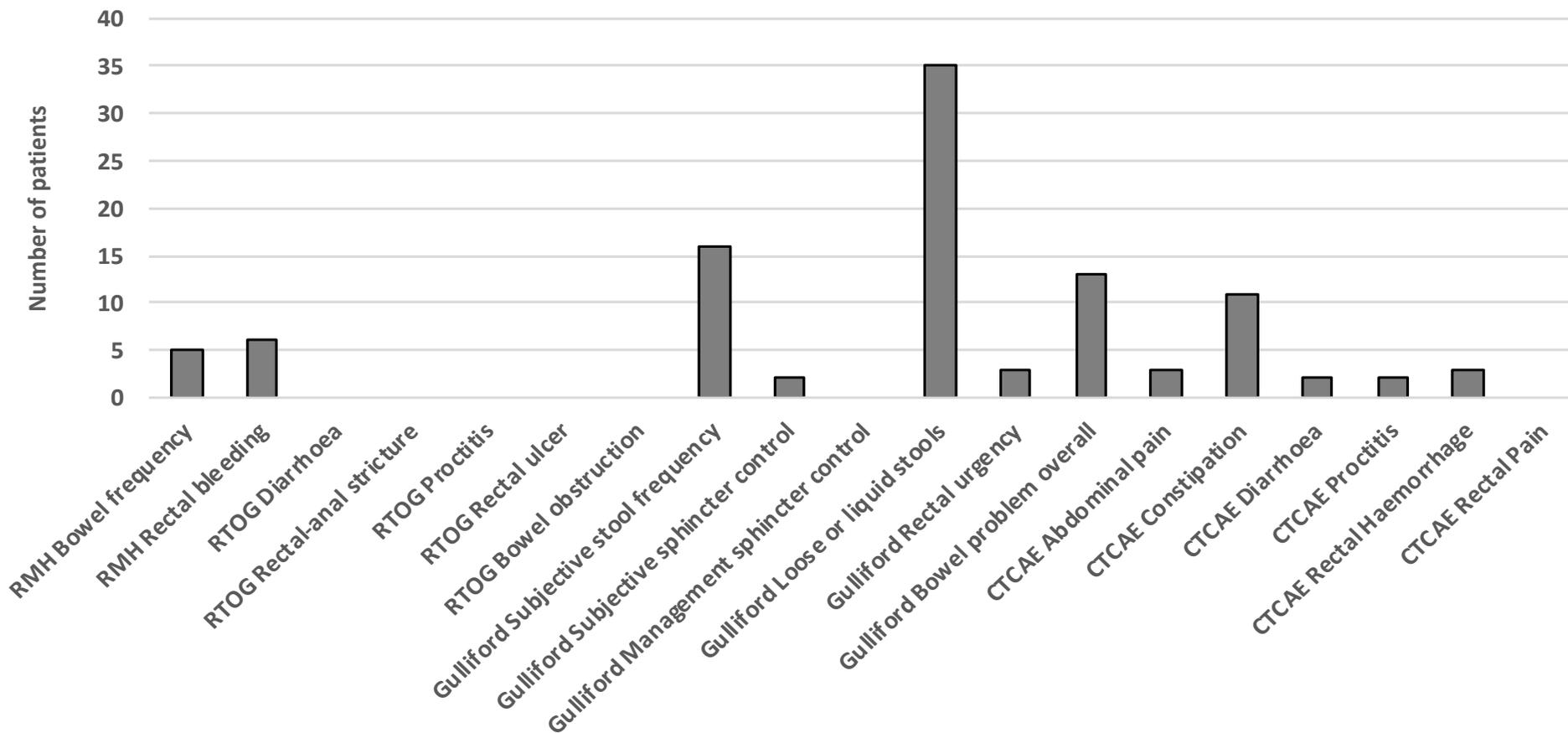


Figure 6.1 Histogram of patients excluded at baseline for each endpoint. Patients were excluded at baseline for that endpoint only if they presented with symptoms similar to the endpoint being investigated.

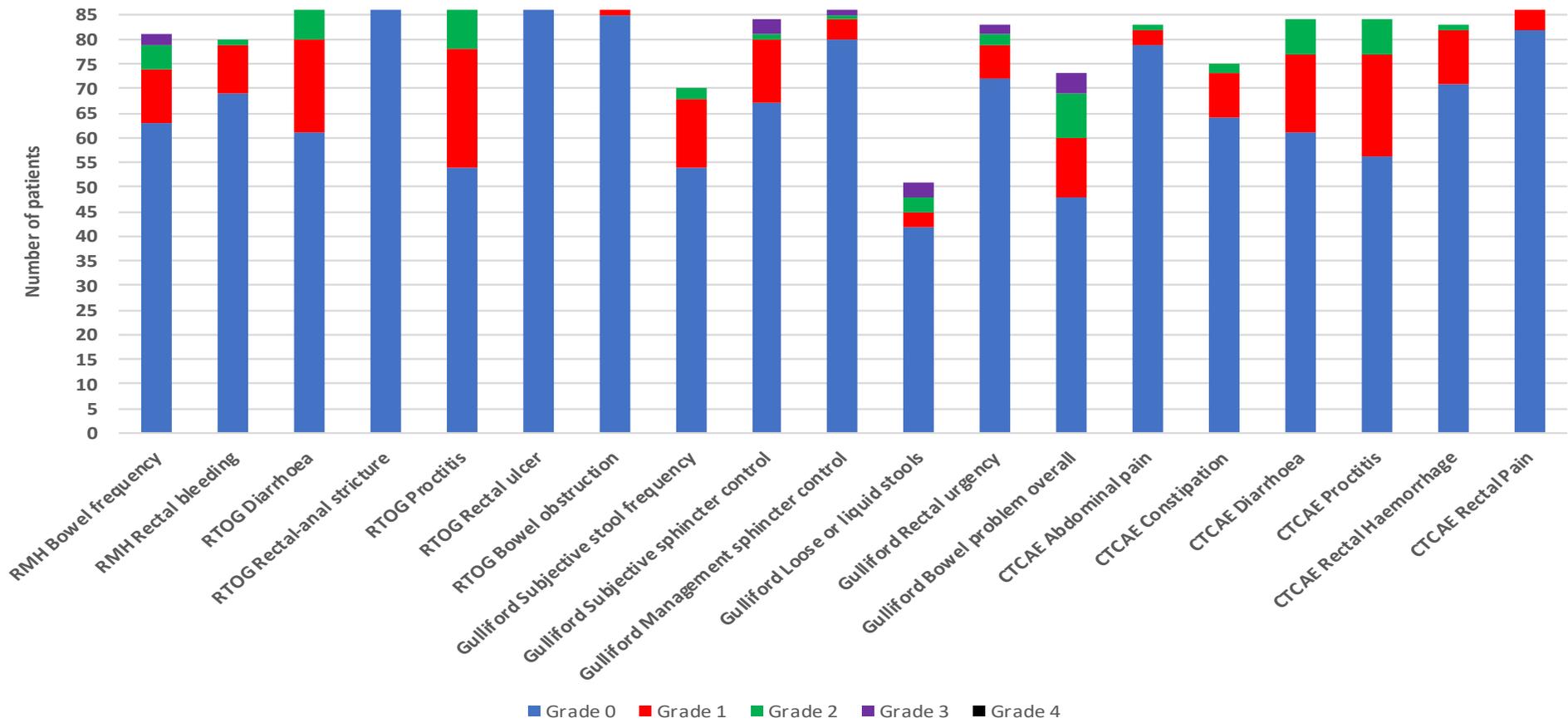


Figure 6.2 The incidence of peak toxicity for each endpoint when the cohorts are combined (n=86). Patients presenting symptoms at baseline similar to the endpoint being analysed were excluded for that endpoint only.

The endpoint with the most patients excluded at baseline was the Gulliford loose or liquid stools, with 41% (35) patients excluded at baseline (baseline was defined as 4 weeks prior to commencement of RT). Gulliford subjective stool frequency had the second greatest number with 19% (16) patients excluded, and Gulliford bowel bother overall had 15% (13) of patients excluded. No patients reported symptoms for RTOG rectal anal stricture, or RTOG rectal ulcer. One patient reported Grade 1 toxicity for RTOG bowel obstruction at 18 months but this had resolved by 24 months. The percentage of patients expressing Grade 1⁺ and Grade 2⁺ toxicity is shown below in Table 6.1. The incidence of Grade 2⁺ toxicity is small, with all endpoints except Gulliford bowel bother reporting an incidence of less than 10 %.

Toxicity endpoint - all patients	Scoring system	Grade 1+	Grade 2+
Proctitis *	RTOG	32 (37%)	8 (9%)
Proctitis	CTCAE	28 (33%)	7 (8%)
Diarrhoea	RTOG	25 (29%)	6 (7%)
Bowel problem overall	Gulliford	25 (29%)	13 (15%)
Diarrhoea**	CTCAE	23 (27%)	7 (8%)
Bowel frequency **	RMH	18 (21%)	7 (8%)
Subjective sphincter control	Gulliford	17 (20%)	4 (5%)
Subjective stool frequency	Gulliford	16 (19%)	2 (2%)
Rectal haemorrhage	CTCAE	12 (14%)	1 (1%)
Rectal bleeding	RMH	11 (13%)	1 (1%)
Rectal urgency	Gulliford	11 (13%)	4 (5%)
Constipation	CTCAE	11 (13%)	2 (2%)
Loose or liquid stools	Gulliford	9 (10%)	6 (7%)
Management sphincter control	Gulliford	6 (7%)	2 (2%)
Abdominal pain	CTCAE	4 (5%)	1 (1%)
Rectal pain	CTCAE	4 (5%)	0 (0%)
Bowel obstruction	RTOG	1 (1%)	0 (0%)
Rectal-anal stricture	RTOG	0 (0%)	0 (0%)
Rectal ulcer	RTOG	0 (0%)	0 (0%)

*Table 6.1 The frequency of recorded peak toxicity (%) (n = 86) is given for all endpoints for both Grade 1⁺ and Grade 2⁺ toxicity. The toxicity is presented in descending order from the endpoint with the most frequent Grade 1⁺ toxicity. All endpoints were associated with toxicity but the endpoint with * denotes Grade 1⁺, and ** denotes the Grade 2⁺ endpoints which are used in the results for illustrative purposes.*

All endpoints for which adverse events were reported were analysed. Only those endpoints which had a statistically significant difference in rectal dose-volume between patients with and without the toxicity prior to correction for multiple testing are presented for illustrative purposes in Section 6.5.2

6.5.2 Association of dose and toxicity

6.5.2.1 Differences in rectal dose-volumes between patients experiencing toxicity and those that did not

Mann Whitney U analysis of Grade 1⁺ toxicity found no statistically significant difference in the rectal volume for any endpoint for either planned or accumulated dose, except for accumulated rectal dose at V70 ($p = 0.047$) for patients expressing Grade 1⁺ RTOG proctitis. This difference in V70 was not observed for the planned dose ($p = 0.406$) expressing Grade 1⁺ RTOG proctitis. The accumulated median (IQR) rectal volume for patients expressing proctitis was 3 % (2 %), compared to 1.8 % (2.6 %) for patients who did report proctitis.

For Grade 2⁺ toxicity the only statistically significant difference was found in the planned V30 of patients experiencing RMH Bowel Frequency (planned $p = 0.044$ vs accumulated $p = 0.100$). Patients expressing bowel frequency had a larger median treated rectal volume of 58.5 % (6.6 %) compared to a median of 53.9 % (8 %) for those who did not express symptoms. There was also a statistically significant difference in the planned V20 of patients experiencing CTCAE Diarrhoea (planned $p = 0.025$ vs accumulated $p = 0.701$). Patients experiencing diarrhoea had a larger median treated rectal volume of 73.8 % (10.5 %) compared to 68.2 % (6.4 %) for those who did not experience diarrhoea. However, post Holm Bonferroni correction for multiple testing, there was no statistically significant difference in rectal volume between patients experiencing toxicity at any grade, dose-volume analysed, or between planned and accumulated dosimetry. In order to highlight the difference between planned and accumulated dose-volume in patients and without toxicity, an average planned and accumulated DVH was calculated for each endpoint which showed statistical significance when dose was associated with toxicity, and is presented below in Fig 6.3 to Fig. 6.8.

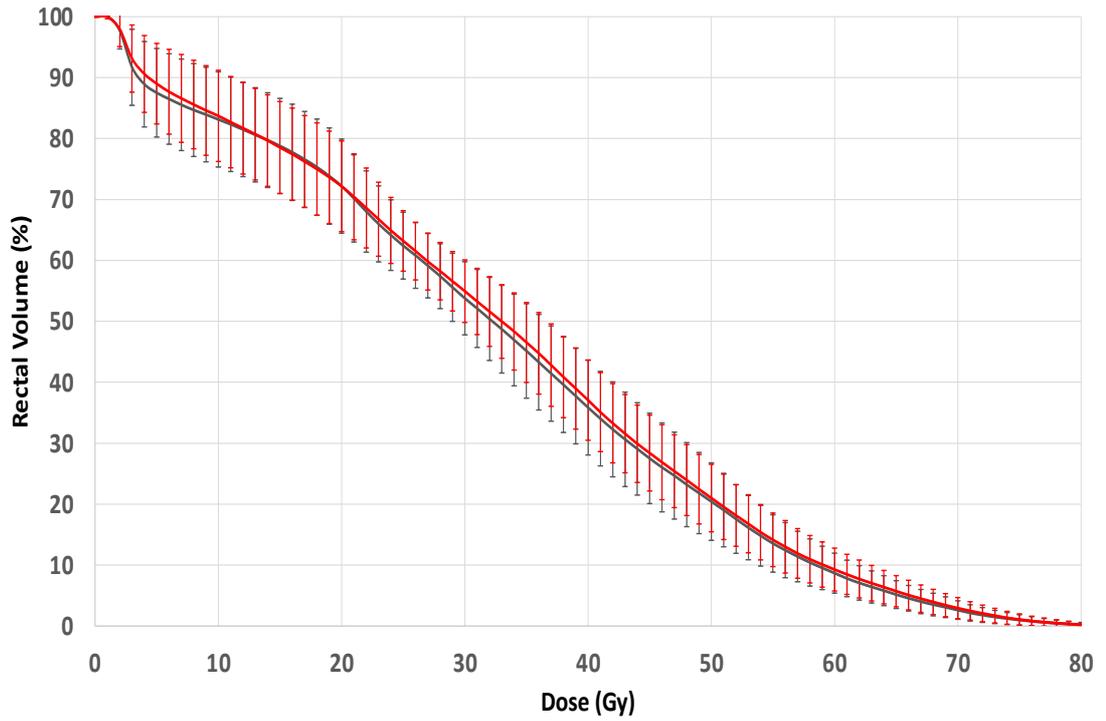


Figure 6.3 The average planned (grey) and average accumulated (red) DVH for patients expressing Grade 1⁺ proctitis (RTOG) is plotted. The standard deviation is given as an error bar.

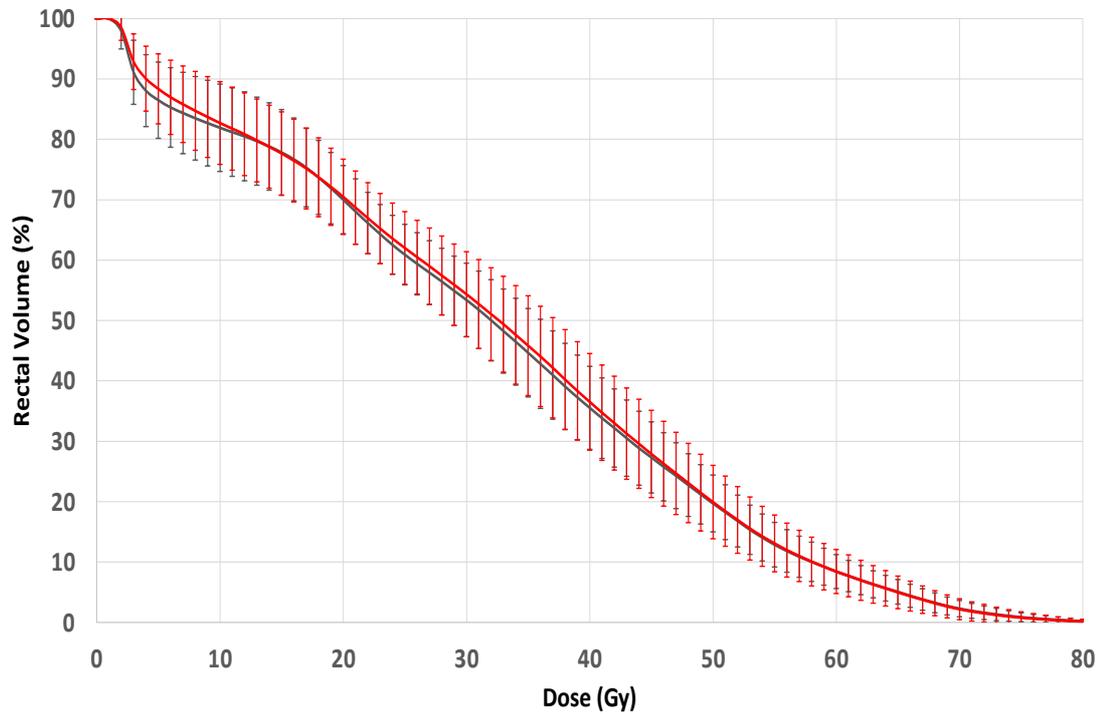


Figure 6.4 The average planned (grey) and accumulated (red) DVH for patients who did not express Grade 1⁺ proctitis (RTOG).

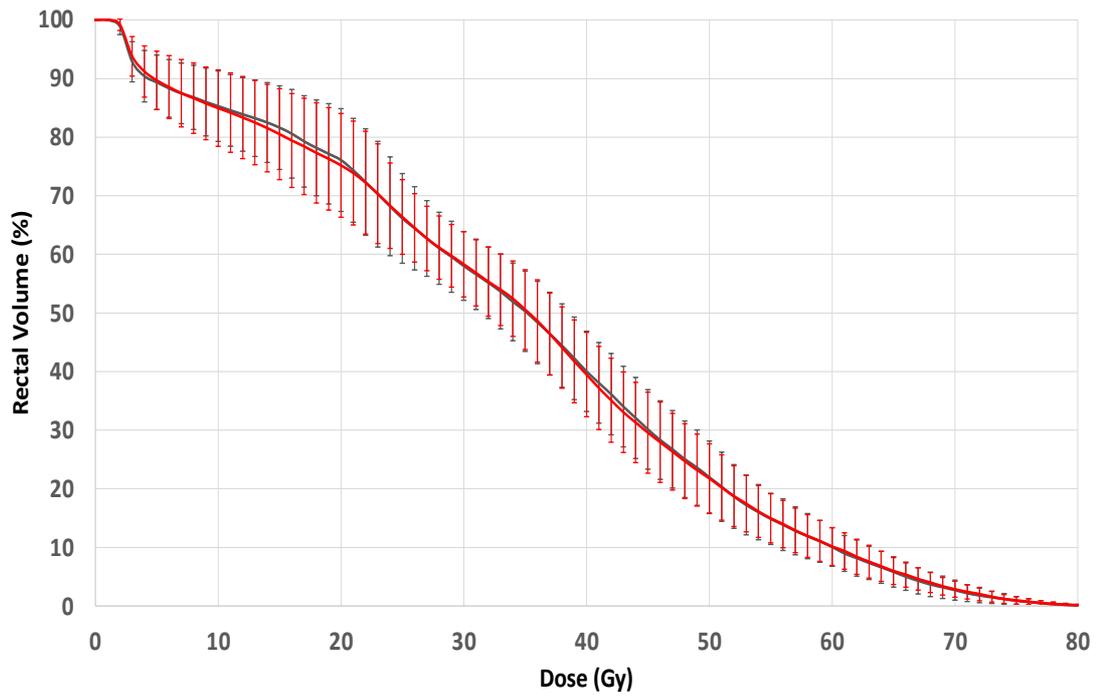


Figure 6.5 The average planned (grey) and accumulated (red) DVH for patients who expressed Grade 2⁺ bowel frequency (RTOG).

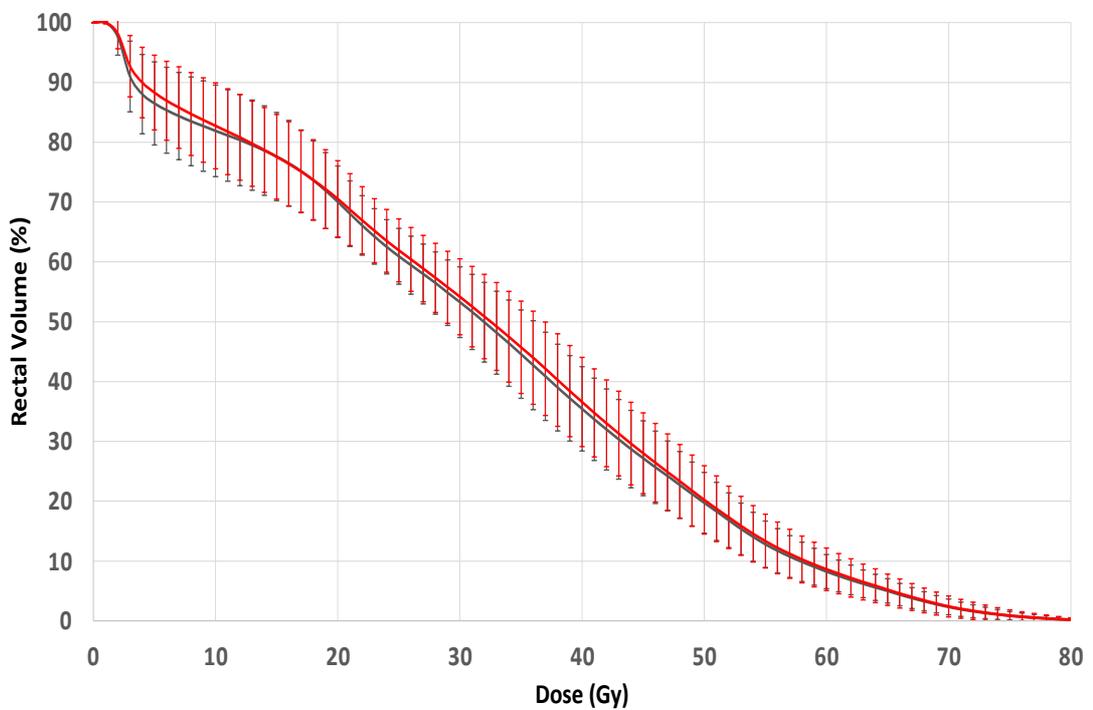


Figure 6.6 The averaged planned (grey) and accumulated (red) DVH for patients who did not express Grade 2⁺ bowel frequency (RTOG)

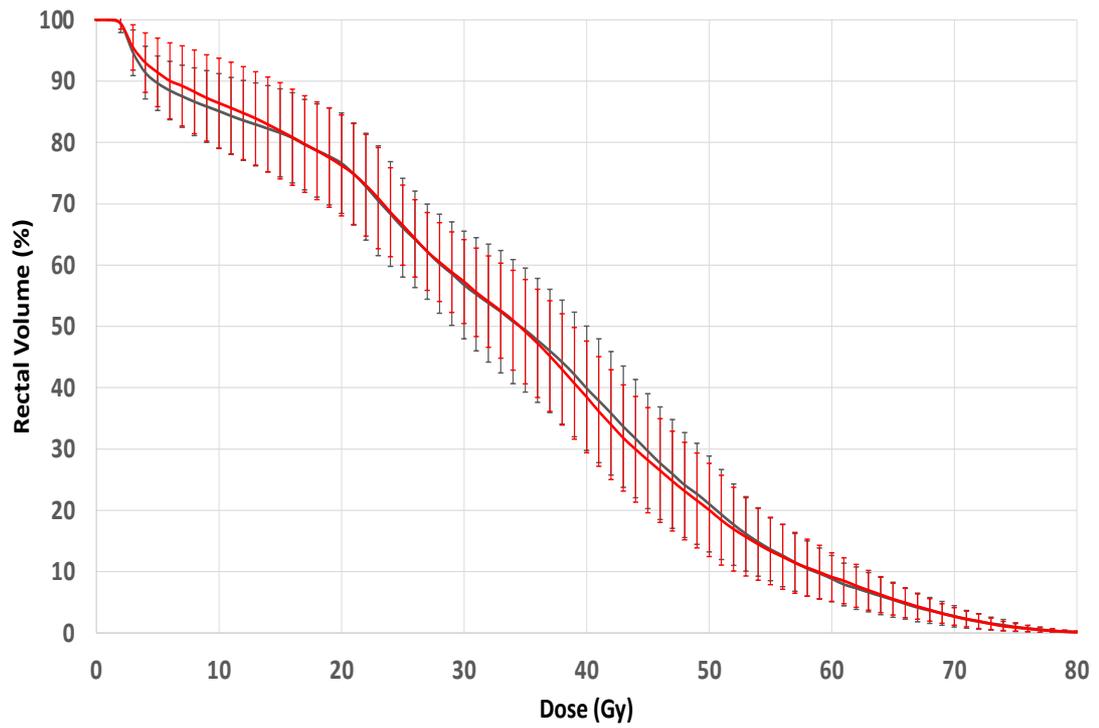


Figure 6.7 The averaged planned (grey) and accumulated (red) DVH for patients who expressed Grade 2⁺ diarrhoea (CTCAE).

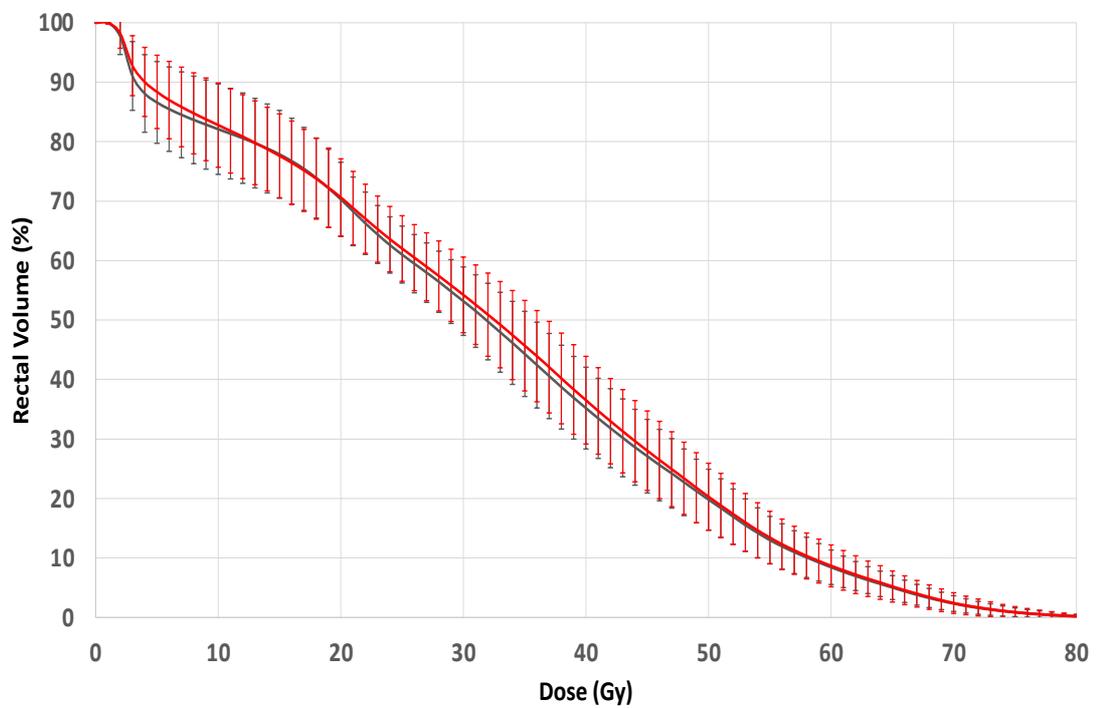
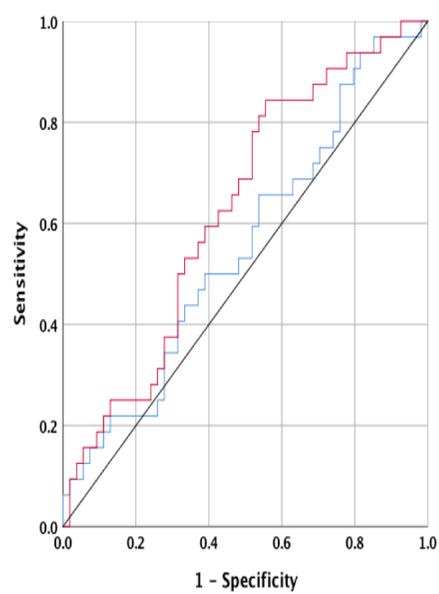


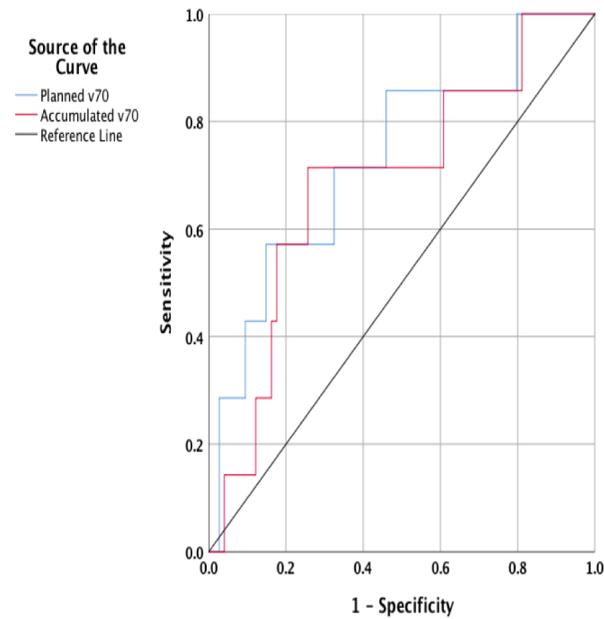
Figure 6.8 The averaged planned (grey) and accumulated (red) DVH given for patients that did not express Grade 2⁺ diarrhoea.

6.5.2.2 Receiver operator analysis for planned and accumulated dose

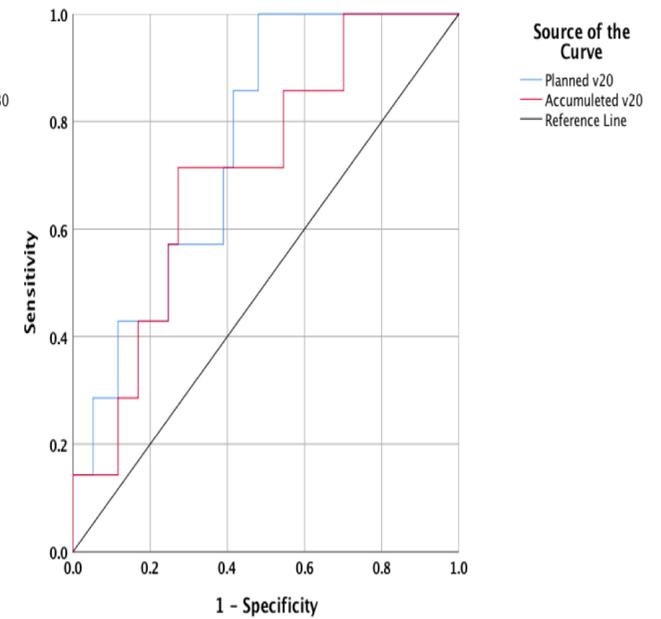
For both Grade 1⁺ and Grade 2⁺ ROC curve analysis, those endpoints which had statistically significant AUC were the same endpoints and dose-volumes which were significant for Mann Whitney U analysis. For Grade 1⁺ Proctitis (RTOG), the accumulated V70 had an AUC of 0.63 ($p < 0.047$). For Grade 2⁺ bowel frequency (RTOG) the planned V30 had an AUC of 0.732 ($p < 0.04$), and Grade 2⁺ CTCAE diarrhoea had a planned V20 with a significant AUC of 0.76 ($p < 0.03$). The ROC curves are presented in Figure 6.3 a.-c. However, again after correction for multiple testing using Holm-Bonferroni no endpoints were statistically significant.



a.



b.



c.

Figure 6.9 Comparison of planned and accumulated curves for ROC curves which had $AUC \geq .06$ and significant p values (uncorrected for multiple testing). Fig a. is for Grade 1+ proctitis (RTOG). Fig b. shows the ROC curves for planned and accumulated Grade 2+ bowel frequency (RTOG). Fig c. shows the ROC curves for the planned and accumulated volumes at V20 for Grade 2+ CTCAE diarrhoea.

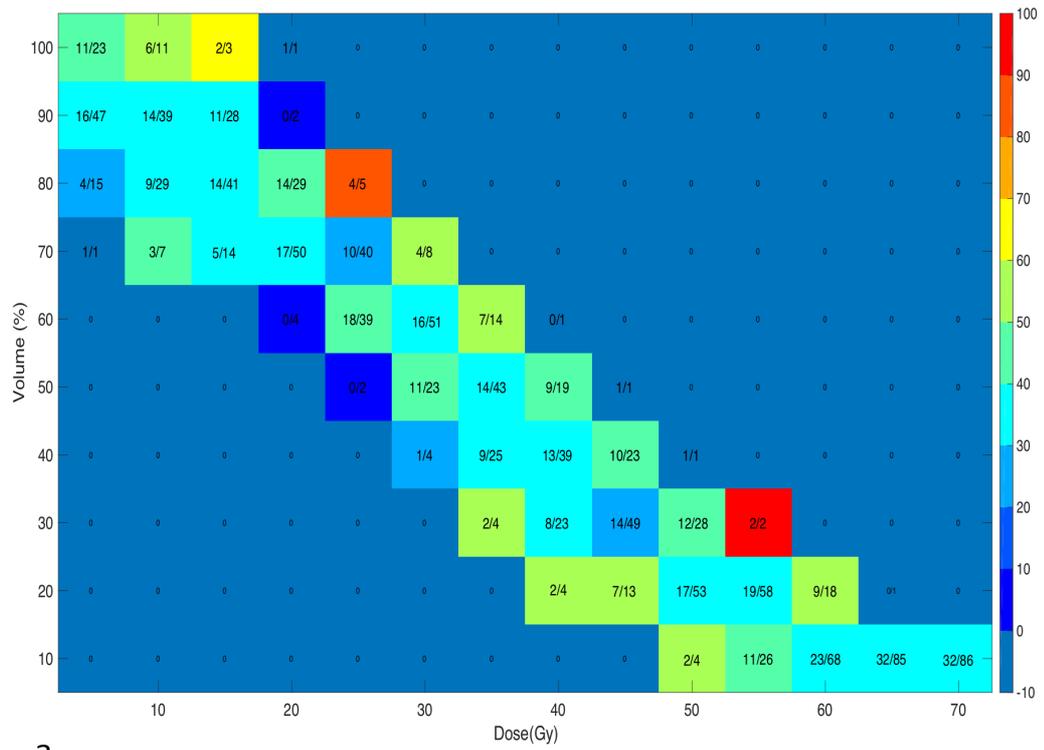
A table of the statistically significant results from ROC analysis (before correction) are presented below alongside the rectal volume cutpoint derived using the Youden Index (Table 6.2) .

<u>Endpoint</u>	<u>Planned</u>					<u>Accumulated</u>						
	Dose Level	Area	P value	Confidence Interval Lower Bound	Confidence Interval Upper Bound	Rectal volume cutpoint	Dose Level	Area	P value	Confidence Interval Lower Bound	Confidence Interval Upper Bound	Rectal volume cutpoint
Grade 1+ RTOG proctitis	V70	0.55	0.41	0.43	0.68		V70*	0.63	0.05	0.51	0.75	1.4%
Grade 2+ RMH Bowel frequency	V30*	0.73	0.04	0.53	0.93	58.5%	V30	0.69	0.10	0.48	0.90	
CTCAE diarrhoea	V20*	0.76	0.02	0.61	0.90	68.4%	V20	0.71	0.07	0.52	0.89	

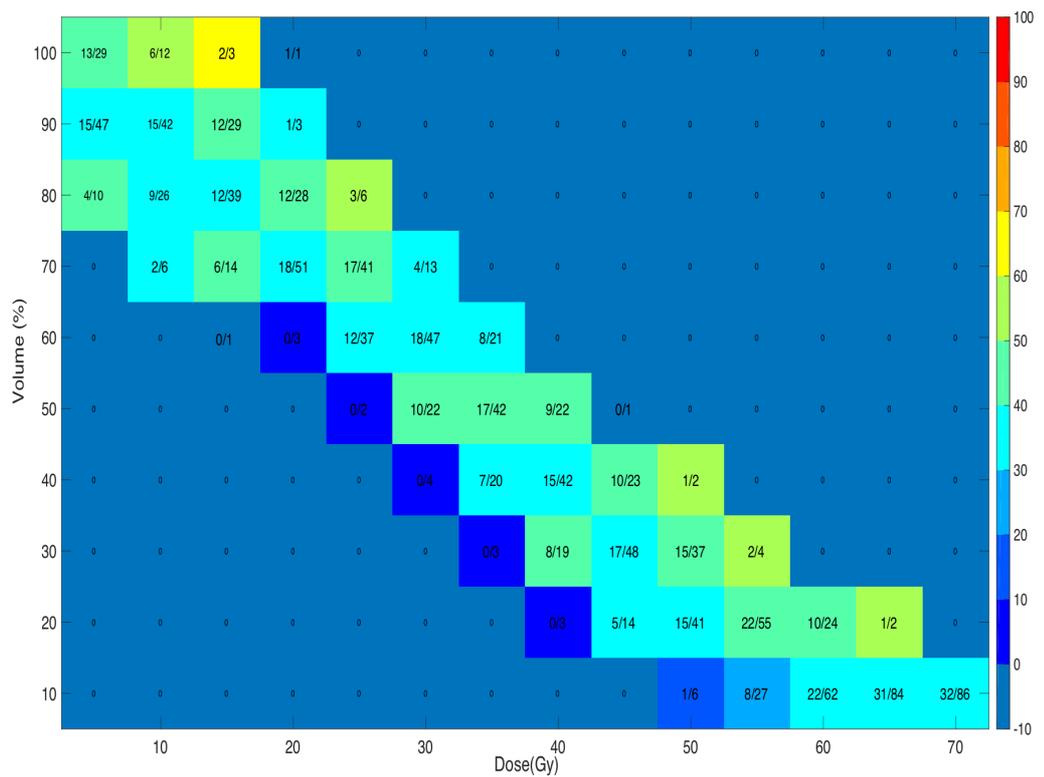
Table 6.2 Table presenting those endpoints and dose levels which were statistically significant (prior to correction for multiple testing) using ROC analysis for both planned and accumulated dose-volume. The rectal volume cutpoint was calculated using the Youden index and is shown for significant ROC curves only.

6.5.2.3 Atlas of complication incidence

An atlas of complication incidence was generated for each endpoint but only those endpoints which were shown to be statistically significant using MWU and ROC analysis (before correction for multiple testing) are presented below (Figure 6.4. to Figure 6.6.)

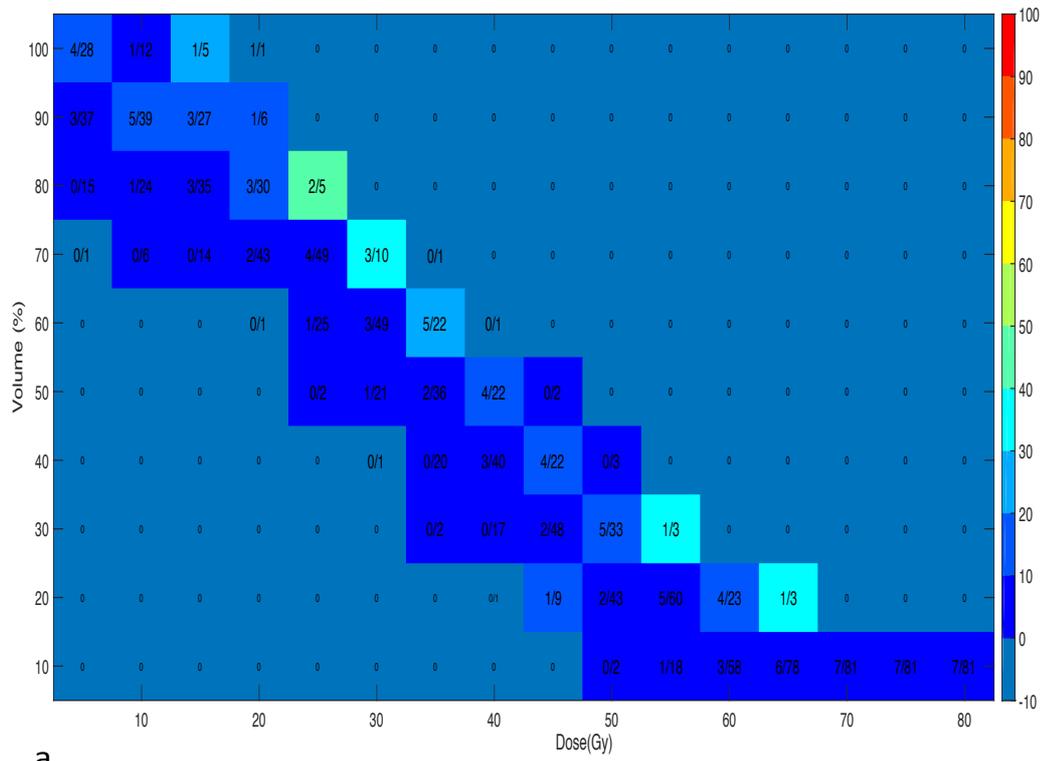


a.

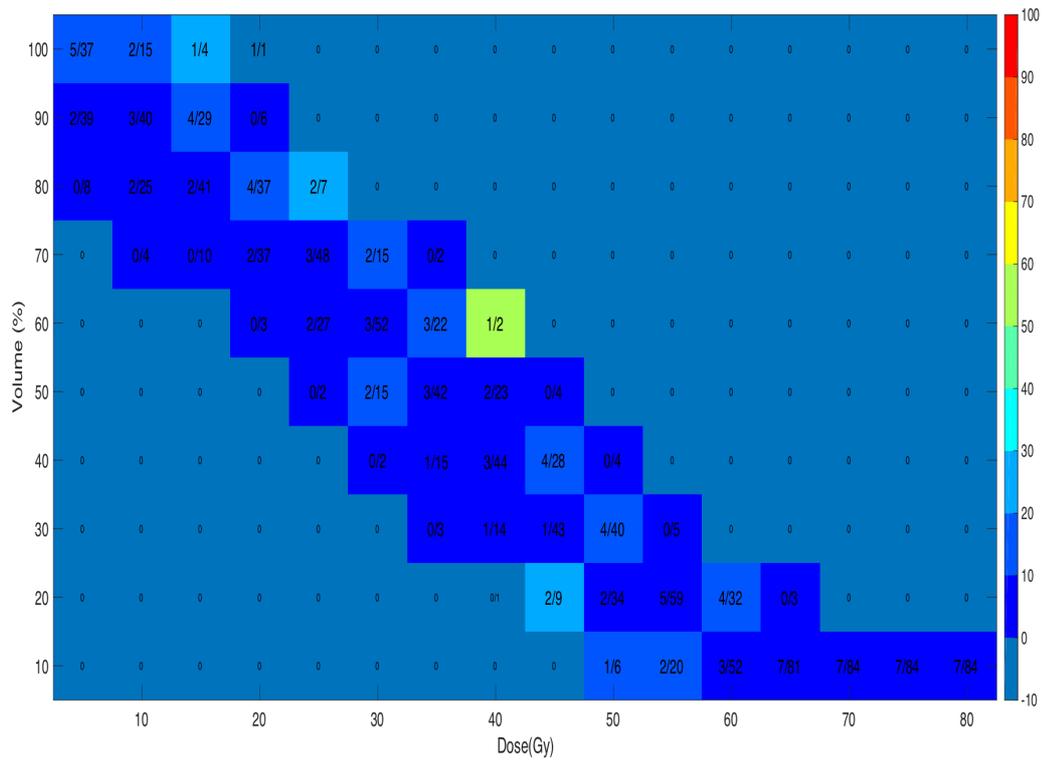


b.

Figure 6.10 ACI for RTOG Grade 1+ Proctitis for planned (a.) and accumulated (b.).



a.



b.

Figure 6.11 ACI for RMH Grade 2+ Bowel frequency for planned (a.) and accumulated (b.).

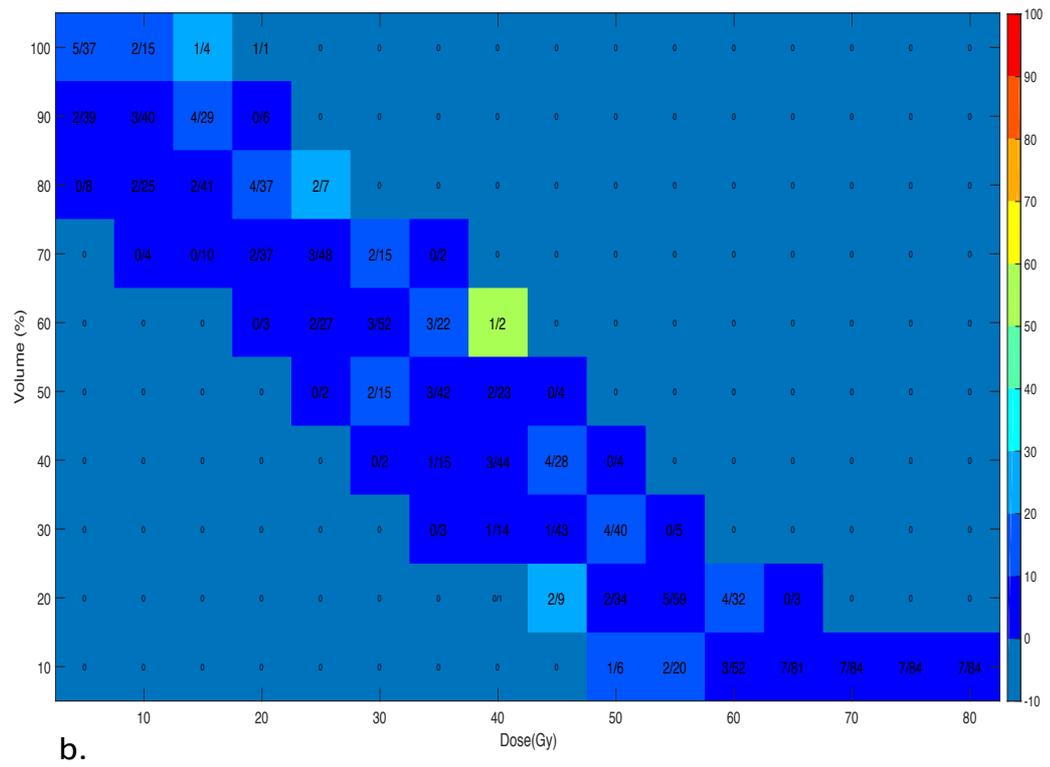
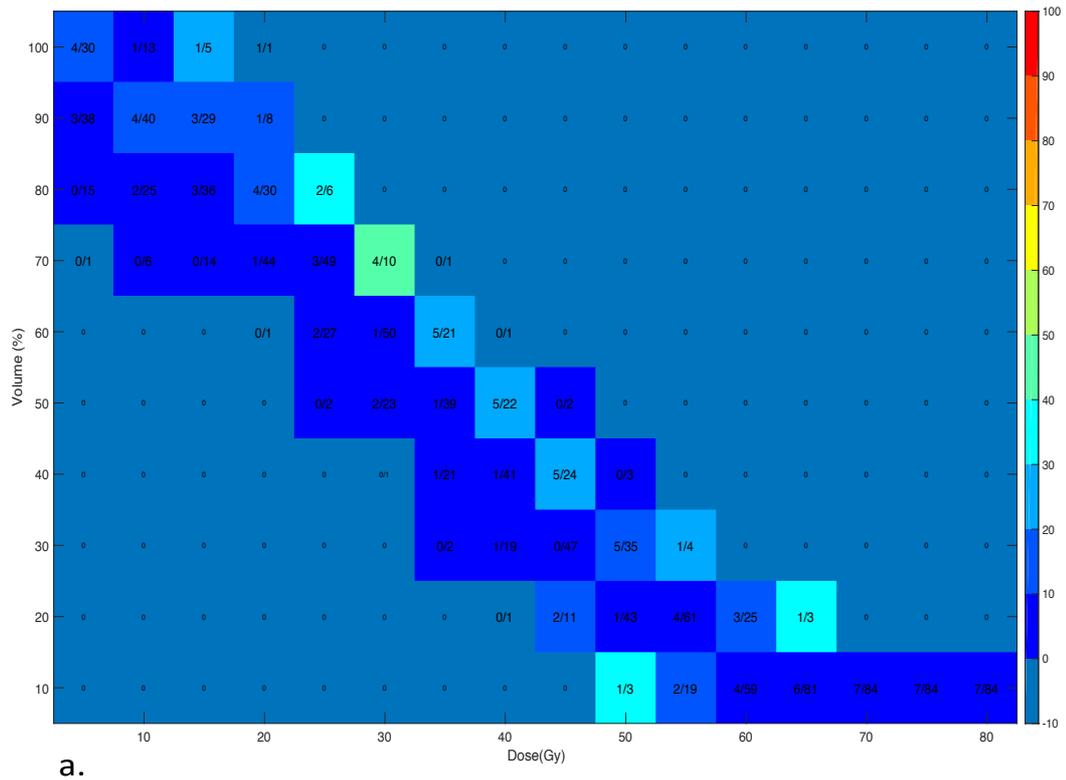


Figure 6.12 ACI for CTCAE Grade 2+ Diarrhoea for planned (a.) and accumulated (b.).

6.6 Discussion

The primary aim of this study was to determine whether accumulated dose, which considers the effect of interfraction variation of the rectum on delivered dose, could better predict toxicity. To the author's knowledge, this is the only study which has accumulated the rectal dose for a large number of patients (>20) using DIR to account for daily variation in rectal size and shape and maintain spatial correspondence between tissues. It is also the only study to associate accumulated dose with prospectively collected toxicity data which is key to determining whether current DVC are clinically relevant.

Statistical analysis using Mann Whitney U (MWU) and receiver operating characteristic curves (ROC) reported statistical significance for the same endpoints, doses and group. For Grade 1⁺ toxicity, there was only one statistically significant result which was for the accumulated rectal volume at V70 for RTOG proctitis ($p = 0.047$, calculated using MWU). No planned dose-volume was significantly associated with any toxicity endpoint for Grade 1⁺. Conversely, no accumulated dose-volume was significantly associated with any Grade 2⁺ toxicity. For planned dose however, there were statistically significant results for Grade 2⁺ RMH bowel frequency at V30 ($p = 0.04$, calculated using MWU) and CTCAE diarrhoea at V20 ($p = 0.02$, calculated using MWU). This would suggest that for those endpoints, planned dose-volume has greater predictive value than accumulated dose-volume for Grade 2⁺ toxicity. It was hypothesized that the accumulated dose-volume was more predictive of the actual delivered dose. The fact that the planned dose-volume was a better predictor of toxicity suggests that there is variation from possible uncertainties in the dose accumulation methodology which prevent it from providing an accurate estimate of delivered dose. However, when the results were corrected for multiple testing, there was no longer any statistical significance, suggesting that neither planned nor accumulated dose were suitable predictors in this patient cohort. Overall the results from the work presented in this chapter lead the author to conclude that there is no evidence (from this study) to support the hypothesis that accumulated dose-volume calculated using DIR is a better predictor of toxicity than planned dose.

This result may be a consequence of the low number of side effect events and relatively small difference between accumulated and planned dose-volumes for one to be clearly a better predictor of toxicity than the other. As demonstrated and discussed in Section

5.6.2, there is good correlation between the accumulated and planned dose-volumes (50 % of accumulated DVHS were within 5 % of planned). When the accumulated DVH was examined, all patients met the required planning DVC, with only 3 patients exceeding some of the optimal DVC. This strict adherence to the dose-volume constraints is likely related to the favourable side-effect profile. Much of the early literature analysing delivered dose suggested there would be a greater difference in delivered rectal dose than was observed in the current study. Previous studies however, were limited by small patient numbers (10 to 12 patients) and often only a small subset of images. Most notably, the delivered dose was not analysed by accumulating the dose, but rather by fraction or by summing the dose which can lead to overestimation of delivered dose [129]. The VoxTox group is the only group the author is aware of that has associated delivered dose with toxicity [93]. The delivered dose was calculated by summing DSM and associated dose with retrospectively collected toxicity data. A stronger correlation between accumulated spatial metrics and toxicity for rectal bleeding and proctitis than for planned was reported [5]. However, the incidence of toxicity reported by VoxTox was much greater than found in this study (Rectal bleeding \geq Grade 1, 25.7 % vs 13 %). This may partially be due to the retrospective collection of toxicity data in the VoxTox study which may have prevented the exclusion of patients at baseline (i.e. those patients who exhibited symptoms similar to the endpoint being analysed). In the current study, 6 patients (7 %) were excluded at baseline for presenting with occasional rectal bleeding, which helped prevent the overestimation of toxicity when calculating the peak toxicity [28]. Additionally, the VoxTox study did not use fiducial based IGRT with reduced margins. Bostel et al., deformed dose using DIR for 10 patients, but used NTCP models instead to calculate the complication risk between planned and accumulated dosimetry. No significant difference was found in the NTCP values for the rectum between planned and delivered dose ($p= 0.475$) [138].

The current study may also be underpowered as the number of patients reporting toxicity for a large number of endpoints is low, although it is comparable to contemporary studies [43, 151]. FLAME is the largest published trial investigating the benefit of a boost to the dominant intraprostatic lesion (DIL) to 95 Gy. No statistically significant difference in the cumulative Grade 2⁺ GI toxicity at two years between the boost (11.2n%) and non-boost technique (10.2 %) was shown. Comparison of the toxicity reported in the current study with the 2-year CHHiP results also show

comparable Grade 2⁺ GI toxicity [151]. For standard dose prescription (74 Gy) the CHHiP trial reported 4.3 % (6/138 patients) versus 15 % (7/46 patients) in this study. The CHHiP hypofractionated arm reported 3.6 % (5/137 patients) vs 3% (1/40) patients in this study. Recently, preliminary toxicity data from the DELINEATE trial was published [42]. Cumulative 1-year RTOG Grade 2⁺ rectal toxicity of 3.6 % (74 Gy/37 #) and 8 % (60 Gy/20 #) respectively was reported. Although a similar subset of patients from the DELINEATE trial was used, there are small differences in reported toxicity between the reports. Firstly, the DELINEATE paper reports the GI toxicity at 1 year for 105 patients, whereas the present study analysed toxicity at 2 years. Most notable though is that the current study excluded a number of patients, mainly due to the dose accumulation methodology (discussed in Section 2.3.2.1. and Section 5.3.). As a result, only 86 of the possible 105 patients included in the preliminary DELINEATE analysis were included in the current study.

As illustrated, significant toxicity (\geq Grade 2) after radiotherapy is now rare, and is a result of improved planning and image guided treatment delivery such as IMRT and IGRT. It is difficult to separate the benefit of IGRT from IMRT as they have evolved contemporaneously. A retrospective study by Zelefsky et al., compared patients treated with IMRT or IGRT guided IMRT [10]. They found no difference in the incidence of late GI Grade 2⁺ toxicity, with this being low in both groups (1.0 % and 1.6 % respectively). However, de Crevoisier reported that GI toxicity was significantly lower in 236 patients who had been imaged daily compared to the 234 patients imaged weekly (70-80 Gy). Late Grade 1⁺ rectal toxicity was significantly lower in the daily group (HR=0.71 [95 % CI, 0.53-0.96]; $p = .027$) [225]. The study also reported an improvement in biochemical progression free survival in the group imaged daily (HR=0.45 [95 % CI, 0.25 - 0.80]; $p = .007$). Only one randomised trial, a sub-study of the CHHiP trial, has recently reported results evaluating the role of IGRT on reducing dose to the rectum. The study by Murray et al., demonstrated that the introduction of IGRT was feasible in a national randomised trial [120]. The IGRT sub study compared no daily IGRT, and daily IGRT with standard and reduced margins. This is of interest as the reduced posterior margins used in the IGRT arm are the same margins used in the DELINEATE trial. Patients were randomised to one of three groups. The first group used standard margins and offline IGRT protocol and bony match (no-IGRT). The second group used daily online IGRT with correction to fiducials and standard posterior margins (IGRT-S) of 10mm/5mm/0mm. The final group had daily online

IGRT with fiducials and reduced margins of 6mm/3mm/0mm (IGRT-R). The rectal dose-volume was significantly statistically lower in the IGRT-R arm compared to the IGRT-S (5-65 Gy $p < 0.0001$). This, however, did not translate into a statistically significant benefit in acute or late toxicity, although late GI toxicity was reported less often. The similarity in GI toxicity across all arms of the IGRT study was suggested to be as a result of the strict normal tissue dose-volume constraints used in the main CHHiP trial [52]. The use of daily IGRT has enabled the reduction of the CTV to PTV margin, resulting in smaller volume of normal tissue in the high dose region. It has been suggested that daily online imaging and correction may improve biochemical control as the daily correction for prostate motion ensures the target is adequately covered [10, 115].

ACIs are recommended by QUANTEC as a method of qualitatively assessing the relationship between dose and toxicity by presenting visually how dose and the irradiated percentage volume may relate to toxicity [29]. QUANTEC recommend the use of reporting tools such as ACIs to enable comparison of results from studies where it would otherwise be difficult to directly compare results because of different dose-volume constraints or model parameters. In a study by Murray et al., (previously described in section 2.4.3) ACIs were used to explore the relationship between dose-volume and erectile potency [171]. The ACI showed an increase in the incidence of men with erectile dysfunction as the volume irradiated and dose to the penile bulb increased. However, the use of ACIs in this study is limited as there was not a strong statistically significant relationship between dose-volume and toxicity, as well as only small differences in in dose-volume between patients.

The reduction in toxicity as a result of reduced dose to the rectum through the use of DVC is further illustrated by comparing the planning DVH from the Medical Research Council RT01 trial (74 Gy/37 # n= 422) to the CHHiP trial (74 Gy/37 # n=1065) and the DELINEATE patients (EQD2 corrected n=86) included in this chapter (Figure 6.7). For the MRC RTO1 trial the only rectal dose constraint used was that the dose to the rectum outside of the PTV should not exceed that of the prescription [28]. The rectal DVC used in the CHHiP trial were also implemented in the DELINEATE trial (see section 2.2.2.) [13]. As can be appreciated by the plot, the rectal dose-volume for the DELINEATE and CHHiP trial are substantially lower than the dose-volume for the MRC RT01 trial which used the then standard CFRT rather than IMRT. The

smaller difference between DELINEATE and CHHiP can also be seen due to the reduced planning margins despite extra dose from the intra-prostatic boost. CHHiP was associated with at least a halving of treatment related side effects compared with RT01, 2 year patient reported outcomes using the Expanded Prostate Cancer Index Composite-26 (EPIC) score for moderate or more bowel distress being 12 % and 4 % in RT01 and CHHiP respectively [226].

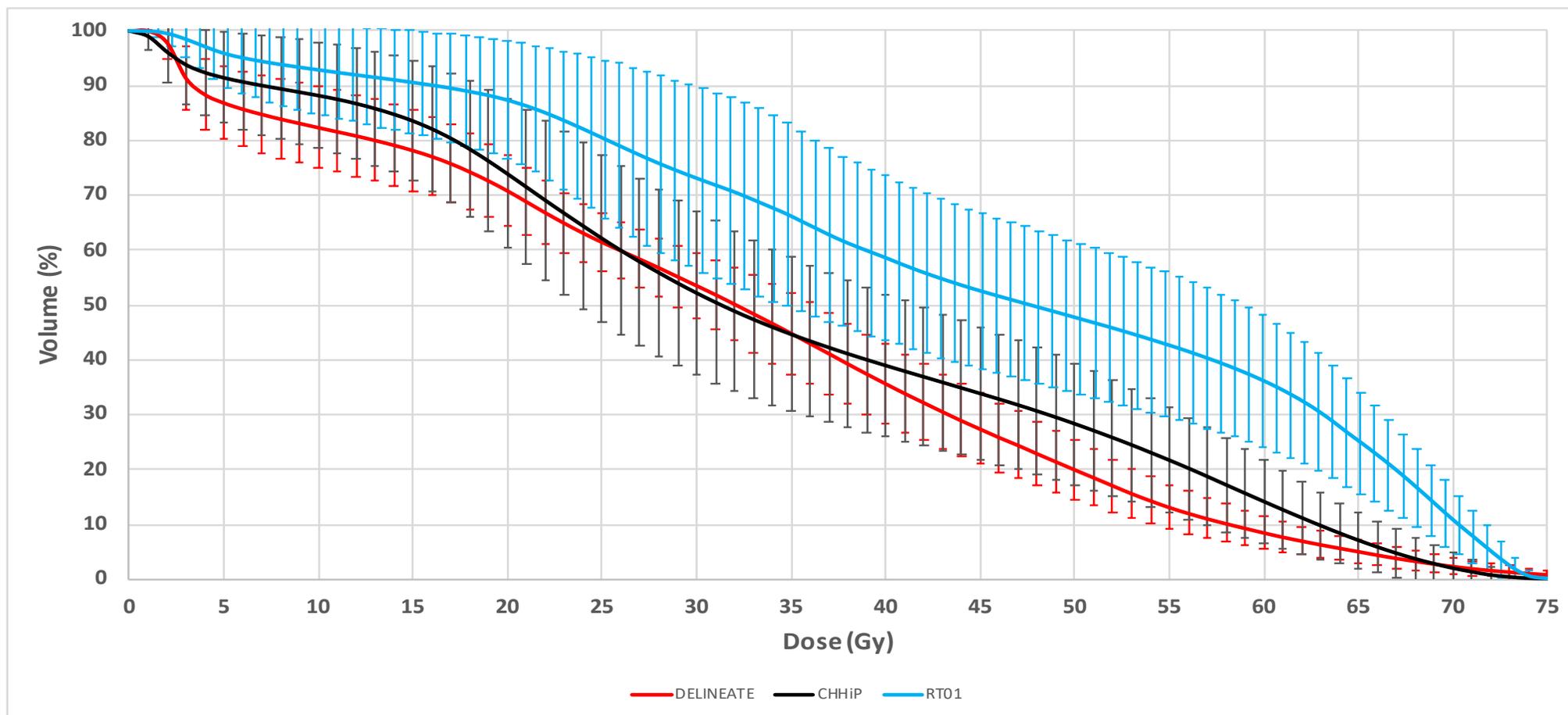


Figure 6.13 Average planning DVH for all patients included in RT01 (74 Gy/3 7#), CHHiP (74 Gy/37 # arm), and DELINEATE (EQD2 corrected) patients analysed in this study (hypofractionated arm EQD2 corrected). The standard deviation is given by the error bars.

Finally, a limitation of this study was that the relationship between the intra-prostatic nodules being boosted, and the subsequent effect on the relationship between rectal dosimetry and toxicity was not evaluated. The number of intra-prostatic nodules, the volume of the nodules and, most importantly, the position of the nodules within the prostate can affect the rectal dosimetry. For example, boosting a nodule in the posterior aspect of the prostate rather than anterior, will increase the maximum rectal dose, potentially increasing the risk of toxicity. However, studies evaluating the impact of simultaneous integrated boost on GI toxicity did not show a statistically significant increase in GI toxicity compared to those who received a standard dose to the prostate only [42, 43, 227].

6.7 Conclusion

This study shows that the accumulated dose is not more predictive of rectal toxicity than planned dose in the patient cohort studied. This is likely as a result of reduced CTV-PTV margins, stricter DVC and correction for geometric uncertainty in prostate position using daily online imaging with fiducial markers. However, this study used DVH to summarise the dose-volume to the rectum which has limitations as the rectal contents are also included. The literature has shown better association between dose and toxicity in some studies which have calculated the dose to the surface of the rectum, and described the distribution of dose using spatial metrics. In the next chapter DSM will be used to measure the difference between planned and accumulated dose, and spatial metrics will be associated with toxicity.

Chapter 7 Association between planned and accumulated spatial characteristics with toxicity

7.1 Introduction

The risk of normal tissue complications is usually evaluated by measuring the amount of tissue receiving a given dose [15]. In the clinical setting, this is most commonly described by using dose volume histograms (DVH). Studies have associated rectal DVH with GI toxicity to derive rectal dose volume constraints which are used to optimise the dose distribution during planning. The use of dose volume constraints has been instrumental in decreasing the dose to the rectum during prostate radiotherapy, and subsequently reducing GI toxicity [9, 52, 65, 226]. A limitation of DVHs however, is that it treats the rectum as a solid rather than a hollow organ. A DVH generated for the rectum includes the rectal contents. However, it is the dose to the rectal wall which is important in understanding the dose response of the rectum [84]. In addition, the DVH reduces the 3D dose distribution within an organ to a one dimensional histogram, and consequently spatial information of how the dose is distributed across the rectal wall is lost [150]. Other methods of quantifying dose which exclude rectal contents are Dose Surface Histograms (DSH) and Dose Wall Histograms (DWH) [85, 86]. However, neither method is able to describe the distribution of dose to the rectal wall. An alternative is to use dose surface maps (DSM) where a hollow organ, such as the rectum, can be virtually unfolded, and the dose distribution to the rectal wall represented in 2 dimensions.

DSM can be used to help understand the relationship between toxicity and organ motion (a series of maps required), with studies showing that shape of the dose distribution may be a better predictor of toxicity than DVH alone [84, 150, 228]. A study by Casares-Magaz et al., explored the association of both spatial and volume dose metrics with toxicity in 200 patients. The use of spatial metrics, such as lateral extent (described in section 2.3.4), improved the prediction of patient reported GI toxicity with a significant association found for six toxicity endpoints versus four endpoints for the DVH based model [228]. Buettner et al., associated clinician reported toxicity with planning DSM for 388 patients from the MRC RTO1 trial [149]. The study showed that there was a stronger correlation between the measures of the shape of the dose distribution and some toxicity outcomes than there was for metrics

derived from DSH. The lateral extent of the dose distribution between doses of 39 Gy and 61 Gy correlated strongly with rectal bleeding, and the longitudinal extent of the dose distribution between 21 Gy and 33 Gy correlated strongly with loose stools. Wilkins et al., associated the spatial metrics of 3,216 men in the CHHiP trial using DSM created from the planning CT scan with both clinician and patient reported outcomes [70]. The study reported DSM-based predictors for rectal bleeding for both DSH and lateral extent at intermediate to high doses, providing further evidence to support the work by Buettner et al.

Some studies have also attempted to discriminate between the spatial metrics of patients with and without toxicity by creating a mean DSM of all patients reporting a particular adverse event, and then comparing it to a mean dose map of patients without [95, 229, 230] (see Figure 7.1). The difference can be described by measuring the spatial metrics in each of the mean DSM and subtracting one value from another, or by creating a map of the differences between means known as a dose difference map. Heemsbergen et al., reported differences in the superior extent of the rectal dose from average DSM created for patients complaining of bleeding vs no bleeding, and in the inferior area for soiling vs no soiling (n=197) [167]. In addition, some studies have chosen to further divide the dose surface map into geometrical regions, or create DSM of specific anatomical sub-regions. Dividing the DSM into geometrical regions enables analysis of organs with stable sub-regions which may increase the chance of finding statistically significant differences between patients with and without toxicity [89]. The bladder is illustrative of this, with the cranial portion showing large variation as a result of filling, while the base remains more stable [231]. A study by Moulton et al., defined sub-regions of the anorectum by increasing the percentage of the area of map (10 % increments) starting with the most inferior 10 %, and associating each increment with toxicity [232]. The study established associations between different endpoints and sub-regions. For instance, for bleeding and mucus loss the strongest correlation was found for dose received in the upper 70 to 80% of the anorectal map ($p < 0.01$). In the lower 40 to 50 % of the anorectal region, dose was most strongly correlated with soiling and faecal incontinence ($p < 0.05$). Another approach is to create DSM of specific anatomical sub-regions such as the bladder trigone [233], or the anorectum, rectum and anal canal [70]. Furthermore, studies have begun to assess the relationship between toxicity and dose at a voxel level which may improve the ability to identify radiosensitive symptom related sub-regions [234, 235].

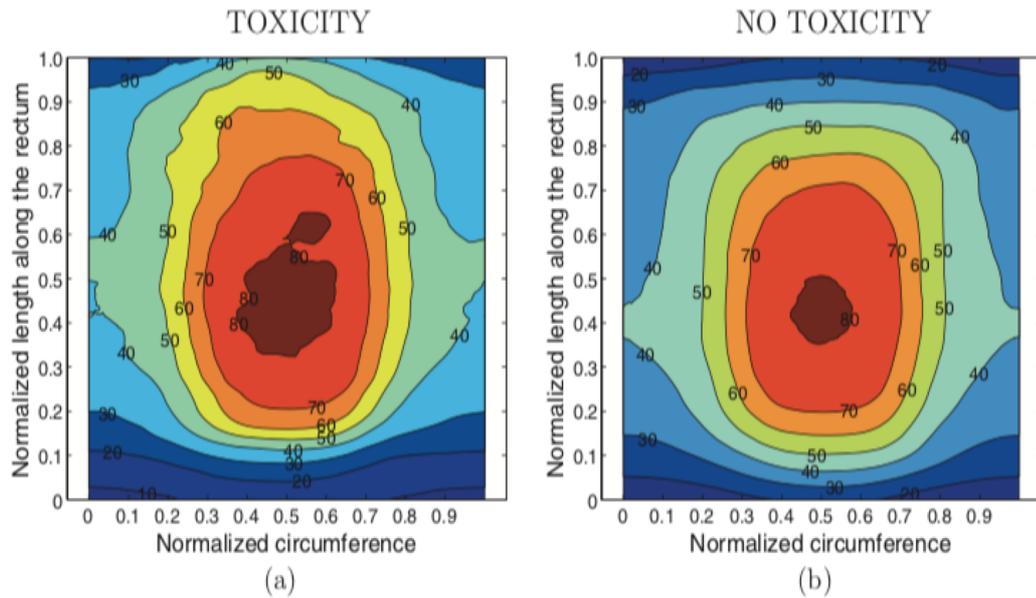


Figure 7.1 Figure showing the average DSM for patients with RTOG Grade 2+ toxicity (a.) and no toxicity (b.). The maps were generated from 49 patients treated with IMRT to a dose of 81 Gy. The average DSM of patients with toxicity is shifted upward compared to the average DSM for patients without toxicity. Figure courtesy of Mundboh et al., which was one of the first studies to evaluate the relationship between the spatial metrics of dose to the rectal surface and toxicity [230].

DSM can also be used to accumulate dose independent of DIR by averaging or summing daily DSMs. This approach was used by Murray et al., who proposed a method to accumulate dose by creating daily DSM using daily CBCTs [91]. Daily DSM were generated for 3 patients using the rectum manually delineated on the CBCT and dose calculated after CBCT bulk density correction in RayStation. The DSM were then normalised in the lateral and longitudinal direction to create maps 21 x 21 pixels which could then be stacked together and summed. The study reported a mean accumulated rectal surface dose which was on average $7.5(\pm 3.5)$ % lower than the planned dose. This method was used by the VoxTox group to accumulate the dose for 109 patients using daily MVCTs, although only the lateral direction was normalised. The longitudinal direction was restricted by the height of the field of view (FOV) of the MVCT which meant that the inferior and superior extent of the rectum were not included in the analysis. Instead the missing rectal length was simulated using the CT scan as a proxy. The lateral extent and equivalent uniform dose (EUD) from both the planned and accumulated DSMs were found to be associated with toxicity data. The group concluded that there was a stronger correlation between rectal bleeding and proctitis with the lateral extent extracted from accumulated DSMs than planned DSM [93]. However, the longitudinal extent of the dose distribution was not analysed due

to not being fully described by the shorter DSM. Another limitation of that study was that the toxicity data was retrospectively collected and did not include baseline data.

A summary of the studies which specify spatial metrics at defined doses and adverse events is given in Table 7.1

<u>Author</u>	<u>Patient population</u>	<u>Dose and treatment technique</u>	<u>Planned or accumulated</u>	<u>Structure</u>	<u>Clinician reported outcome (CRO) or patient reported outcome (PRO)</u>	<u>Symptom</u>	<u>Metric</u>
Buettner et al. 2009	843 (3D-CRT)	421 (64Gy/32#) and 422 (74Gy/37#)	Planned	Rectum	CRO and PRO	Rectal bleeding Loose stools	Lateral extent (39 - 61 Gy) Longitudinal extent (21 - 29 Gy) & DSH (21-27 Gy)
Shelley et al. 2017	109 (3D-CRT)	74 Gy/37# (3DCRT)	Planned and Accumulated	Rectum	CRO	Rectal bleeding	Accumulated lateral extent (strongest correlation at 65Gy)
Onjukka et al. 2019	213	74-80 Gy or 65-75.2 Gy	Planned	Ano-rectum, rectum and anal canal	PRO	Rectal bleeding Faecal incontinence	Lateral extent(59, 61 and 71 Gy) Lateral extent (5-49Gy) & Longitudinal extent (60Gy)
Wilkins et al 2020	3216 CHHiP trial	(74Gy/37#, 60Gy/20# and 57 Gy/19#)	Planned	Ano-rectum, rectum and anal canal	CRO and PRO	Rectal bleeding	Lateral extent (30-70GY) & DSH (20-70GY)

Table 7.1 Table of studies which have reported strong correlations at given doses between spatial metrics with toxicity.

7.2 Aim

This chapter tests the hypothesis that the spatial metrics of delivered dose are better predictors of toxicity than spatial metrics of planned dose. The aim of this chapter is to compare the spatial characteristics between planned and delivered dose using DSM, and associate each metric with reported and prospectively collected toxicity data. The following objectives were addressed:

1. Measurement of the difference in the spatial characteristics between planned and accumulated DSM using pixel count, longitudinal and lateral extent.
2. Analysis of the strength of the association between measures which describe the shape of dose distribution and different toxicity endpoints. To characterise the dose distribution, the pixel count, longitudinal and lateral extent were calculated at different doses for both planned and accumulated dose, and associated with reported toxicity.

7.3 Methods and materials

The cohort of patients was similar to those reported in Chapters 5 and 6 but patients were excluded when a rectal DSM could not be generated from the CT, or when there were a large number of fractions where a DSM could not be generated from the associated CBCT. This occurred when the rectum was contoured as two separate regions on the same slice and usually arose in the superior portion of the rectum when a large pocket of gas distorted the image of the rectum, or when the superior extent of the rectum curved inferiorly forming a loop. This has also been reported by Hoogeman et al., [150]. For each patient the undeformed rectal contour was used to generate daily DSM which were then summed to generate an accumulated DSM. In addition, daily DSM were summed only if they had previously been included in the DIR accumulation methodology (Chapter 6), i.e. if the image had shown good deformable registration ($DSC \geq 0.7$ and $MDA \leq 3$ mm), and had subsequently been accumulated.

The DSM were created using the methodology developed by Dr Buettner and Dr Murray, which is discussed in more detail in Section 2.3.4 [91, 149]. Briefly, a binary mask was created in Raystation for the contoured rectum. The dose was sampled at 2° along the circumference of each axial slice. The longitudinal extent was sampled at

0.3 cm intervals. The cylindrical array was exported to MATLAB (the script used in Buettner et al., was modified by Dr Gulliford), and virtually unfolded by cutting each slice at its most posterior extent [94]. This produced a 2D array of 180 x n slices. The length of the rectum varied from image to image within individual patient's dataset, and between patients. To enable comparison between fractions and patients, the rectal the lateral and longitudinal dimensions were normalised to create maps, creating DSM of 21 x 21 pixels. This resolution was chosen as it has previously been successfully used for DSM analysis [91, 149]. In addition, the hypofractionated doses were also converted to 2 Gy fractions using EQD2 correction and an α/β ratio of 3 Gy [9] (please see Section 2.2.5 for EQD2 equation).

Feature extraction was performed using the method described by Buettner et al., and is described in more detail in Section 2.3.4 [149]. Binary images were generated by thresholding the DSM at 10 Gy intervals from 10 to 70 Gy. Pixels receiving dose equal to or above a particular dose were allocated a value of 1, and pixels below this dose were allocated a value of 0. Doses greater than 70 Gy were not tested as only small regions of the rectal wall received doses greater than this [93]. All pixels with a value of 1 in the binary image were grouped into clusters. Using this method revealed the main cluster which corresponded to the high dose region where the anterior rectal wall abutted the prostate [149]. Consequently, for each binary image the features were extracted as follows:

Pixel count - refers to the number of pixels with a value of 1 in each binary image. To analyse the percentage of the surface area receiving a specified dose the number of pixels with a value of 1 was divided by the total number of pixels in the DSM (441 pixels).

Longitudinal extent - was the maximum number of pixels with a value of 1 across the longitudinal extent of the largest cluster.

Lateral extent - was the maximum number of 1-valued pixels across the lateral extent of the largest cluster. However, analyses of the lateral extent were only performed for doses greater than 30 Gy as the number of 1-valued pixels across the lateral extent were consistently 1 for doses less than 30 Gy. This indicated that at lower dose levels the entire rectal circumference was receiving the selected isodose level. This has previously been reported by Shelley et al., [93].

A schematic illustrating extraction of the lateral and longitudinal features across the greatest extent of pixels can be found in Figure 2.9 in Section 2.3.4.

The cumulative peak toxicity was determined for nineteen endpoints by calculating the maximum score for each patient (per endpoint) over the first two-year follow-up period. The nineteen endpoints used and the toxicity scoring is described in more detail in Section 2.2.4.

7.4 Statistical Analysis

Results were analysed by combining the dose prescription cohorts using EQD2 corrections for the hypofractionated cohort. The Holm-Bonferroni method was used to adjust for multiple testing [178].

7.4.1 Clinical characteristics and incidence of toxicity

Descriptive and quantitative methods were used to analyse the reported toxicity. A Fishers exact test was used to test for significant differences in incidence of toxicity between cohorts for both Grade 1⁺ and Grade 2⁺ toxicity.

7.4.2 Difference between planned and accumulated spatial characteristics

For each patient the difference in spatial characteristics between planned and the accumulated DSM was calculated at 10 Gy intervals (10 to 70 Gy). For each of the spatial characteristics, the accumulated value was subtracted from the planned value. Finally, a Wilcoxon Signed Rank test was used to test for significance between the pixel count, longitudinal and lateral extent of the planned and accumulated DSM.

7.4.3 Association of spatial characteristics and toxicity

The planned and accumulated spatial metrics at 10 Gy intervals (10 to 70 Gy) for patients with and without toxicity was tested using Mann Whitney U for each toxicity endpoint. Receiver Operator Curves (ROCs) were generated for each endpoint for both planned and accumulated DSMs. ROC curves were generated by plotting sensitivity against 1-specificity. The test variable was the spatial metric (pixel count, longitudinal or lateral extent) at a specific dose, with dose analysed at 10 Gy intervals. The outcome was the toxicity endpoint being analysed (e.g. 1 = Grade 1⁺ Diarrhoea, 0 = no diarrhoea

experienced). ROC curves were used to identify thresholds in the test variable at specific doses between patients who expressed toxicity compared to those who did not. If the area under the curve (AUC) was ≥ 0.6 , with a lower bound confidence interval of > 0.5 [93, 175], with $p > 0.05$, it was considered to be statistically significant. For ROC curves, which were statistically significant, the Youden index (J) was used to determine a cut point of the spatial characteristic that may be used to predict for rectal toxicity [70, 173]. The formula used to calculate the Youden index can be found in Section 6.4.2.

7.5 Results

7.5.1 Clinical characteristics and incidence of toxicity

In addition to the 19 patients that were excluded from the DVH analysis presented in Chapters 5 and 6, a further 6 patients were excluded when the accumulated dose was calculated using DSM because of the difficulties in creating the DSM as described above in section 7.1.3. This resulted in the analysis of 80 patients (44 patients from the standard cohort and 36 from the hypofractionated cohort). The demographic and clinical characteristics of the 80 patients included are shown in Table 7.2. Characteristics by cohort can be found in Appendix 4.

	Combined cohorts(%)
<u>Age (years)</u>	70.5 (57-80)
<u>T Stage</u>	
T1a/b/c/x	36 (45)
T2a/b/c/x	32 (40)
T3/x	12 (15)
<u>Gleason Score</u>	
6	19 (24)
7	56 (70)
8	5 (6)
<u>NCCN Risk Group</u>	
Intermediate	26 (33)
High	54 (67)
<u>Diabetes</u>	
Yes	11 (14)
No	69 (86)
<u>Hypertension</u>	
Yes	39 (49)
No	41 (51)
<u>Inflammatory bowel or diverticular disease</u>	
Yes	6 (8)
No	74 (93)
<u>Pelvic surgery</u>	
Yes	12 (15)
No	68 (85)
<u>Symptomatic Haemorrhoids in the last 12 months</u>	
Yes	13 (16)
No	66 (83)
Unknown	1 (1)
<u>Previous transurethral resection of the prostate</u>	
Yes	7 (9)
No	73 (91)
<u>Statins</u>	
Yes	32 (40)
No	48 (60)
<u>Current Smoker</u>	
Yes	6 (8)
No	73 (91)
Unknown	1 (1)

Table 7.2 Combined clinical characteristics for the 80 patients analysed. Standard cohort n = 44 and hypofractionated cohort n=36.

There was no statistically significant difference in the incidence of peak toxicity between cohorts, for either Grade 1⁺ or Grade 2⁺ toxicity. A histogram of patients excluded at baseline for each endpoint can be found in Figure 7.2 A histogram of the peak toxicity for all grades is shown in Figure 7.3. The toxicity for separate cohorts can be found in the appendices (Appendix 5).

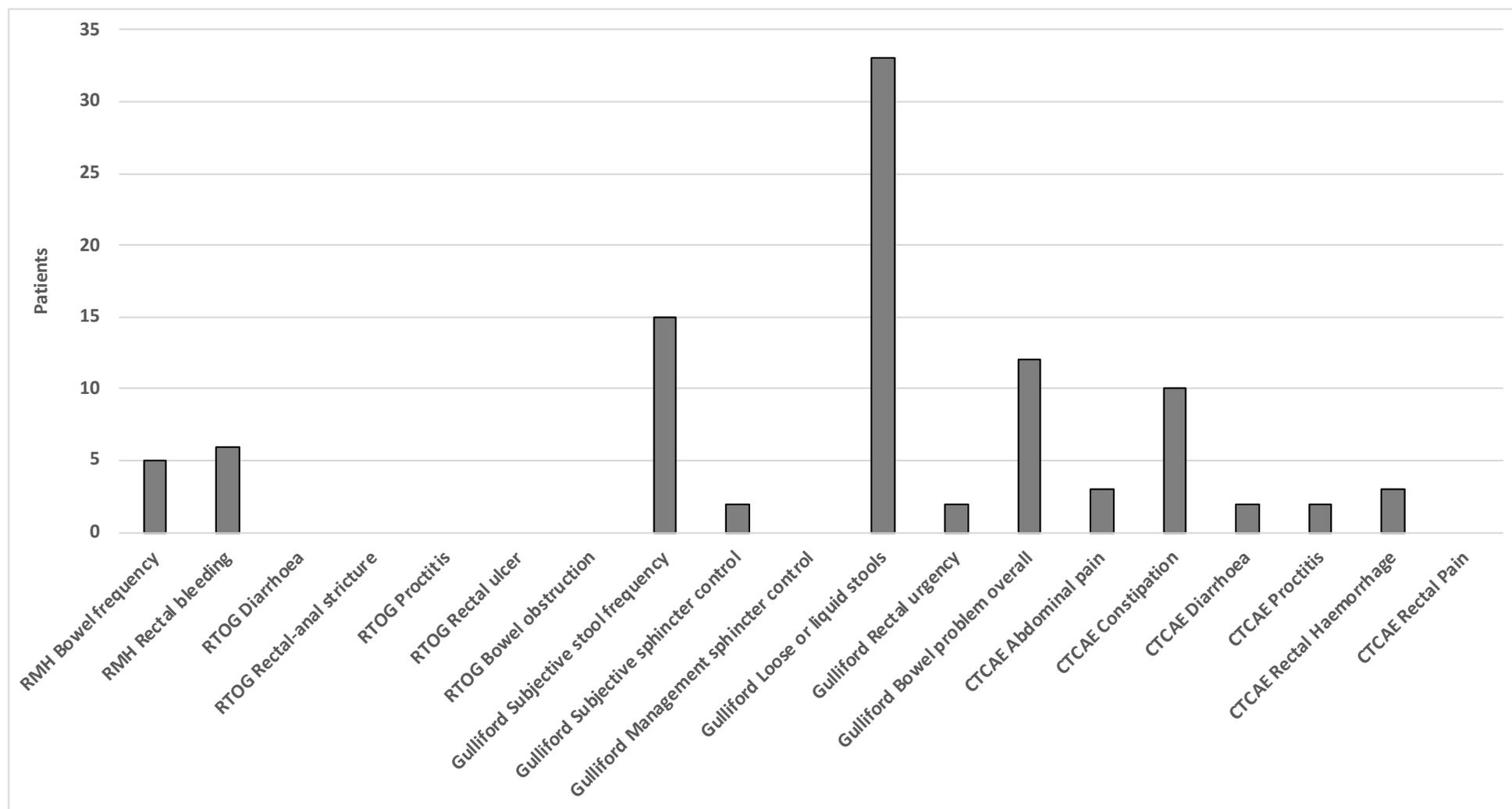


Figure 7.2 Histogram of patients excluded at baseline for each endpoint. Patients were excluded at baseline for that endpoint only if they presented with symptoms similar to the endpoint being investigated.

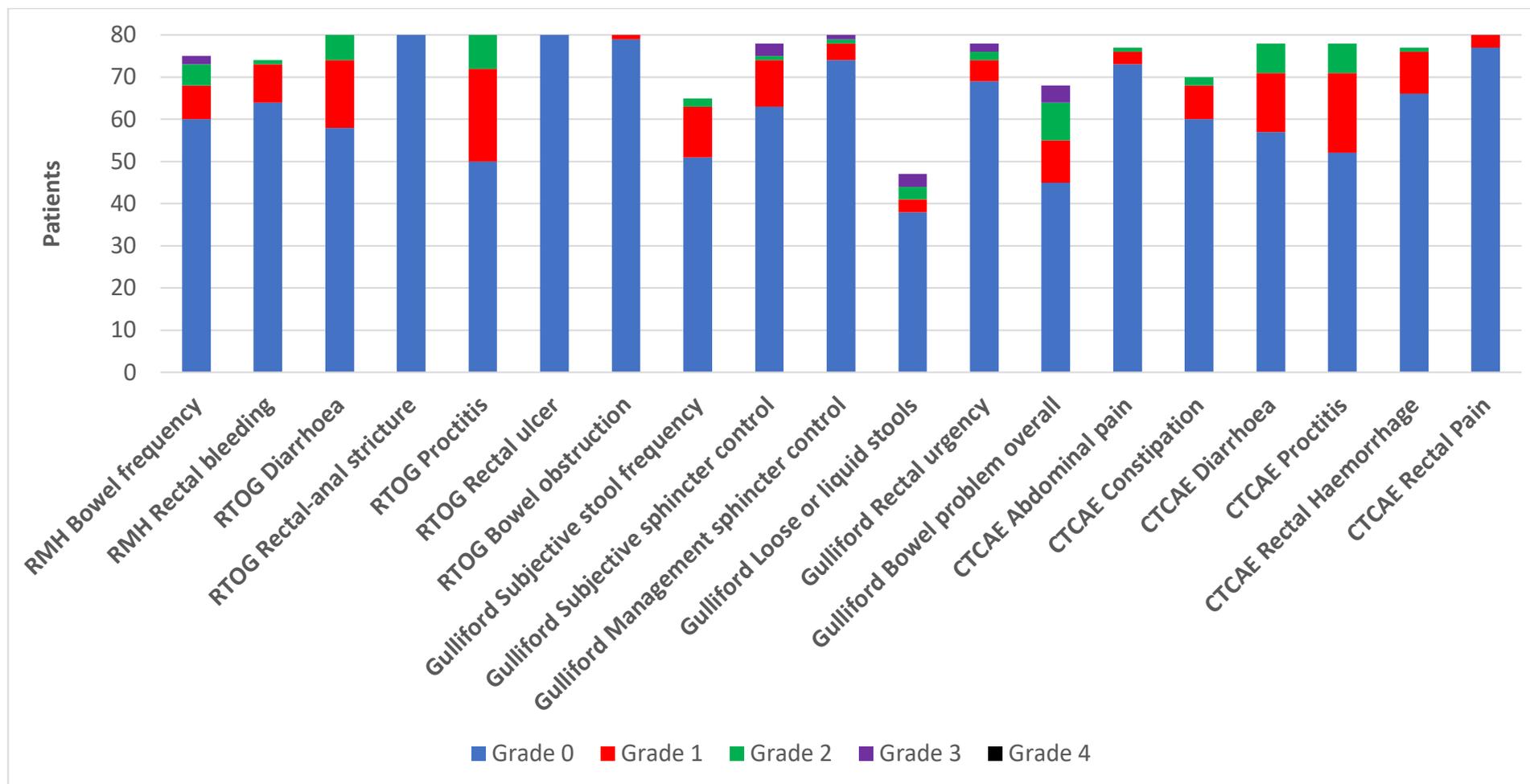


Figure 7.3 Histogram showing the incidence of peak toxicity at two years for the patients included in this DSM study with cohorts combined. Patients presenting symptoms at baseline similar to the endpoint being analysed were excluded for that endpoint only.

The endpoints with the greatest number of patients excluded at baseline were Gulliford loose and liquid stools (33 patients), Gulliford subjective stool frequency (15 patients) and Gulliford bowel problem overall (12 patients), which is consistent with the exclusion of patients at baseline in Chapter 6. However, removal of an additional 6 patients in the current study as a result of the DSM methodology meant the exclusion of patients reporting bowel frequency, proctitis and diarrhoea. As a result, the incidence of toxicity for those endpoints in this study is lower than that reported in Chapter 6. The percentage of patients expressing Grade 1⁺ and Grade 2⁺ is presented below (Table 7.3). It is calculated as a percentage of the total number of patients (n=80).

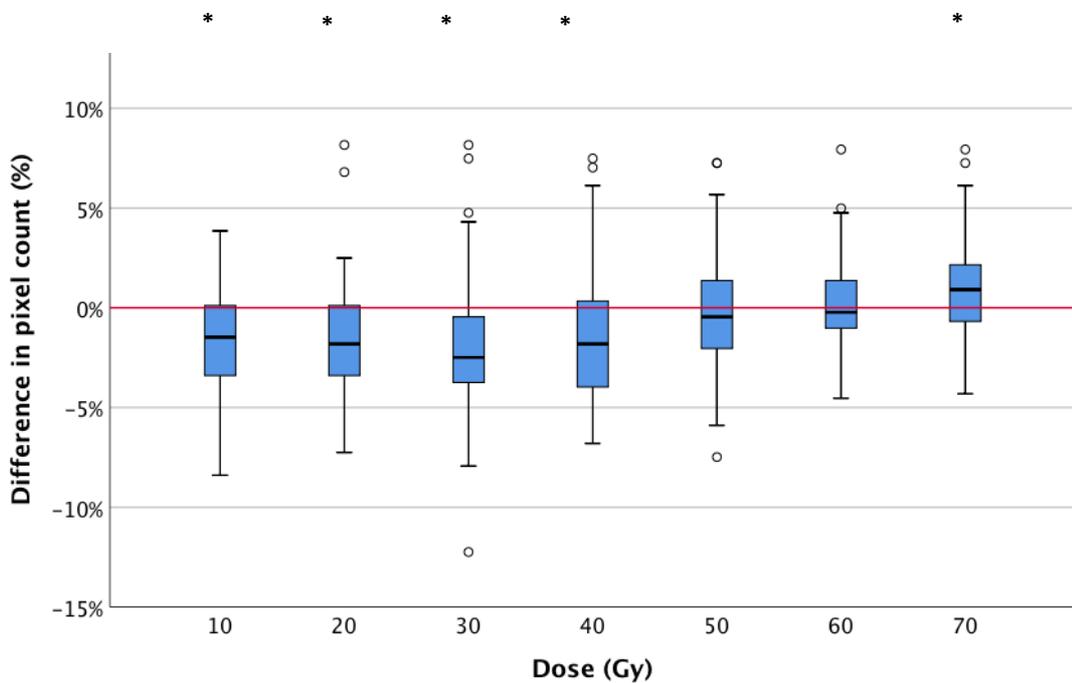
<u>Toxicity endpoint</u>	<u>Scoring system</u>	<u>Grade 1+</u>	<u>Grade 2+</u>
Proctitis *	RTOG	30 (38%)	8 (10%)
Proctitis *	CTCAE	25 (33%)	7 (9%)
Bowel problem overall *	Gulliford	23 (29%)	13 (16%)
Diarrhoea	RTOG	22 (28%)	6 (8%)
Diarrhoea *	CTCAE	21 (26%)	7 (9%)
Bowel frequency*	RMH	15 (19%)	7 (9%)
Subjective sphincter control *	Gulliford	15 (19%)	4 (5%)
Subjective stool frequency	Gulliford	14 (18%)	2 (3%)
Rectal haemorrhage	CTCAE	11 (14%)	1 (1%)
Rectal bleeding	RMH	10 (13%)	1 (1%)
Constipation	CTCAE	10 (13%)	2 (3%)
Loose or liquid stools*	Gulliford	9 (11%)	6 (8%)
Rectal urgency*	Gulliford	9 (11%)	4 (5%)
Management sphincter control	Gulliford	6 (8%)	2 (3%)
Abdominal pain	CTCAE	4 (5%)	1 (1%)
Rectal pain	CTCAE	3 (4%)	0 (0%)
Bowel obstruction	RTOG	1 (1%)	0 (0%)
Rectal-anal stricture	RTOG	0 (0%)	0 (0%)
Rectal ulcer	RTOG	0 (0%)	0 (0%)

*Table 7.3 Table of toxicity (%) (n = 80) is given for all endpoints for both Grade 1⁺ and Grade 2⁺ toxicity. The toxicity is presented in descending order from the endpoint with greatest Grade 1⁺ toxicity. Endpoints above the redline are those endpoints with Grade 1⁺ toxicity which are presented in this study, and additionally * denotes the Grade 2⁺ endpoints included.*

All endpoints for which adverse events were reported were analysed. Grade 1⁺ endpoints with a toxicity incidence of $\geq 10\%$, and endpoints with a Grade 2⁺ toxicity incidence $\geq 5\%$ are presented in the results for illustration. There were no statistically significant results for less common toxicities.

7.5.2 Difference between planned and accumulated spatial characteristics

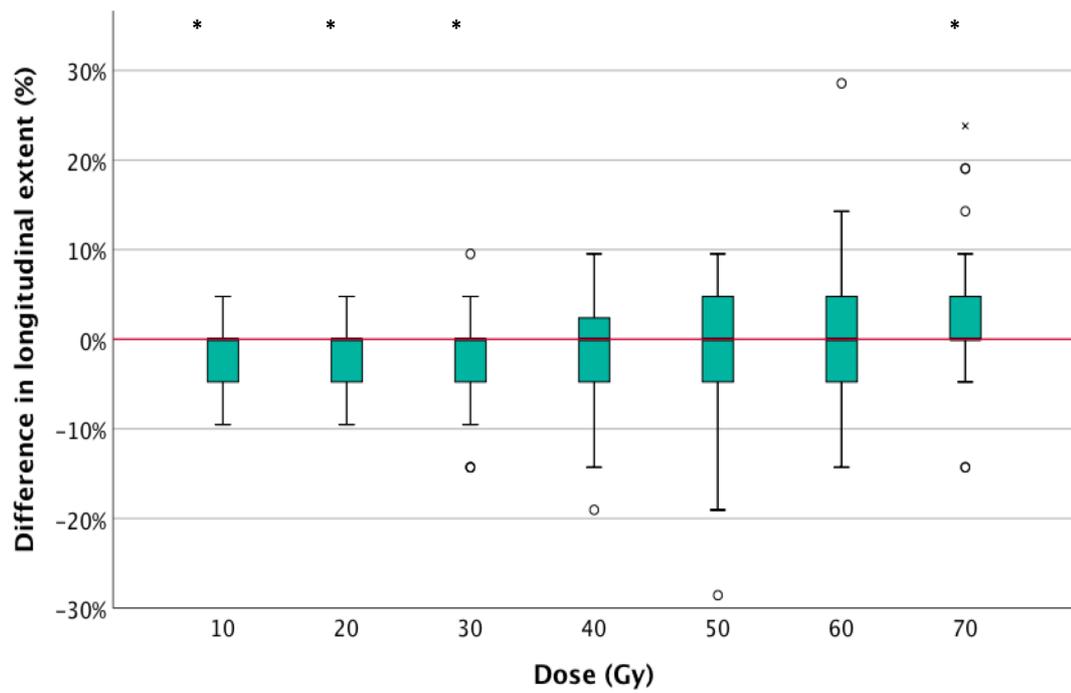
There was a statistically significant difference (Wilcoxon Signed Rank test) between the pixel count of the planned and accumulated DSM between 10 Gy to 40 Gy, higher for accumulated dose. At 70 Gy, the accumulated pixel count was statistically significantly lower than the planned ($p < 0.005$). There were no statistically significant differences at 50 Gy ($p = 0.190$) and 60 Gy ($p = 0.914$). Statistical significance remained after correction for multiple testing.



*Figure 7.4 Boxplot of the difference (%) between the planned pixel count and the accumulated pixel count at 10 Gy intervals (10 Gy to 70 Gy) for the combined cohorts ($n=80$). The box represents the interquartile range (IQR), the median is denoted by the black line. The whiskers extending from the box indicate the lowest and highest values which are no greater than 1.5 times of the IQR. Outliers which are greater than 1.5 to 3 times the IQR are denoted by circles. Extreme values which are greater than 3 times the IQR are denoted by x. An * denotes dose levels where there is a statistically significant difference between planned and accumulated pixel count after correction using Holm Bonferroni. A negative difference indicates that the accumulated pixel count is greater than the planned*

A similar pattern was observed for the longitudinal extent of planned and accumulated DSM with statistically significant differences observed for 10 Gy ($p < 0.005$), 20 Gy to 30 Gy ($p = 0.002$), 40 Gy ($p = 0.049$), with higher doses in accumulated DSM but a statistically significant lower accumulated longitudinal extent was apparent at 70 Gy ($p = 0.001$). There was no statistically significant difference between planned and

accumulated DSM at 50 Gy ($p = 0.309$) and 60 Gy ($p = 0.078$). After Holm Bonferroni correction the difference in longitudinal extent at 40 Gy was no longer significant.



*Figure 7.5 Boxplot of the difference in the longitudinal extent of the planned DSM and the accumulated DSM at 10 Gy intervals (10-70 Gy) for combined cohorts (n=80). An * denotes dose levels where there is a statistically significant difference between planned and accumulated pixel count after correction using Holm Bonferroni. A negative difference indicates that the accumulated pixel count is greater than the planned*

There were statistically significant differences in the lateral extent with higher planned than accumulated DSM at 40 Gy ($p = 0.04$), 50 Gy ($p = 0.001$), and both 60 and 70 Gy ($p < 0.005$). No statistical difference was noted at 30 Gy ($p = 0.036$). After Holm Bonferroni correction, the difference in the lateral extent at 40 Gy was no longer significant but the differences at 50 to 70 Gy remained statistically significant.

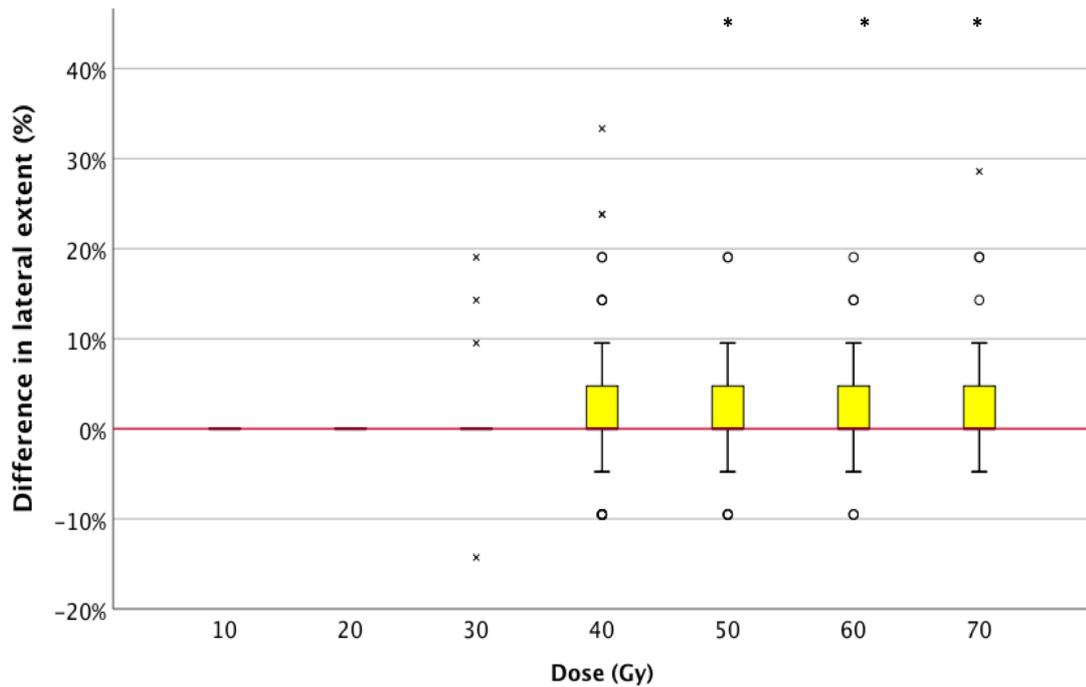


Figure 7.6 Boxplot of the difference in the lateral extent of the planned DSM and accumulated DSM at 10 Gy intervals (10 to 70 Gy) for combined cohorts (n=80). An* denotes dose levels where there is a statistically significant difference between planned and accumulated pixel count after correction using Holm Bonferroni. A negative difference indicates that the accumulated pixel count is greater than planned

7.5.3 Association of spatial metrics and toxicity

Association of Grade 1⁺ toxicity and spatial metrics using Mann Whitney U and ROC curves showed statistical significance for bowel frequency and subjective stool frequency only. Both planned and accumulated pixel count and longitudinal extent showed statistically significant association across a range of doses in the low to mid dose range for these endpoints. However, there was no significant result when the lateral extent was associated with Grade 1⁺ toxicity for any endpoint.

Grade 2⁺ bowel frequency, diarrhoea and subjective stool frequency were also statistically significantly associated with both planned and accumulated pixel count and longitudinal extent across a range of low to mid dose regions. No toxicity endpoint was statistically significantly associated with the planned lateral extent. However, there was a statistically significant association between the accumulated lateral extent and subjective sphincter control in the mid to high dose range.

A Holm Bonferroni correction for multiple testing was applied for both Mann Whitney U test and ROC curves. After correction for multiple testing, none of the previous

statistically significant association remained significant. The uncorrected results are presented below.

7.5.3.1 Mann Whitney U

A table of p values from association for Grade 1⁺ and Grade 2⁺ toxicity with spatial metrics are given below in Table 7.4 and Table 7.5in .

Toxicity Endpoint	Scoring System	Planned Pixel Count						Accumulated Pixel count							
		10	20	30	40	50	60	70	10	20	30	40	50	60	70
Proctitis	RTOG	0.69	0.35	0.37	0.87	0.74	0.75	0.61	0.48	0.28	0.30	0.87	0.72	0.61	0.16
Proctitis	CTCAE	0.82	0.81	0.90	0.90	0.59	0.60	0.64	0.89	0.63	0.70	0.90	0.60	0.46	0.14
Rectal bleeding	RMH	0.45	0.92	0.28	0.34	0.57	0.84	0.76	0.95	0.54	0.18	0.13	0.45	0.29	0.16
Rectal haemorrhage	CTCAE	0.49	0.87	0.40	0.42	0.85	0.82	0.92	0.88	0.78	0.26	0.19	0.59	0.22	0.41
Rectal urgency	Gulliford	0.68	0.40	0.34	0.34	0.52	0.78	0.65	0.47	0.25	0.37	0.62	0.85	0.90	0.47
Diarrhoea	CTCAE	0.58	0.39	0.53	0.32	0.43	0.32	0.18	0.59	0.26	0.59	0.69	0.45	0.52	0.40
Diarrhoea	RTOG	0.38	0.20	0.33	0.35	0.24	0.29	0.36	0.44	0.13	0.40	0.76	0.64	0.68	0.62
Bowel frequency	RMH	0.05	0.09	0.27	0.19	0.05	0.10	0.29	0.04	0.02	0.16	0.36	0.22	0.21	0.28
Subjective stool frequency	Gulliford	0.01	0.07	0.45	0.16	0.04	0.71	0.21	0.03	0.04	0.81	0.37	0.07	0.38	0.62
Loose or liqued stools	Gulliford	0.83	0.31	0.39	0.33	0.14	0.98	0.46	0.38	0.13	0.44	0.43	0.52	0.79	0.24
Subjective sphincter control	Gulliford	0.72	0.28	0.45	0.80	0.52	0.90	0.61	0.61	0.28	0.62	0.88	0.47	0.43	0.31
Constipation	CTCAE	0.85	0.99	0.49	0.62	0.17	0.90	0.66	0.91	0.57	0.96	0.71	0.75	0.21	0.16
Bowel problem overall	Gulliford	0.86	0.14	0.71	0.67	0.95	0.74	0.37	0.54	0.12	0.66	0.93	0.80	0.97	0.42

Toxicity Endpoint	Scoring System	Longitudinal Extent						Longitudinal Extent							
		10	20	30	40	50	60	70	10	20	30	40	50	60	70
Proctitis	RTOG	0.93	0.71	0.82	0.69	0.59	0.96	0.68	0.54	0.79	0.51	0.96	0.67	0.20	0.07
Proctitis	CTCAE	0.68	0.92	0.26	0.87	0.45	0.97	0.80	0.94	0.94	0.96	0.97	0.86	0.22	0.09
Rectal bleeding	RMH	0.48	0.93	0.95	0.99	0.59	0.15	0.94	0.94	0.99	0.85	0.87	0.53	0.28	0.33
Rectal haemorrhage	CTCAE	0.41	0.62	0.95	0.66	0.87	0.43	0.94	0.79	0.76	0.84	0.99	0.70	0.14	0.55
Rectal urgency	Gulliford	0.86	0.60	0.69	0.16	0.61	0.94	0.86	0.43	0.76	0.39	0.53	0.65	0.49	0.26
Diarrhoea	CTCAE	0.72	0.33	0.83	0.31	0.20	0.64	0.23	0.53	0.43	0.26	0.52	0.72	0.26	0.17
Diarrhoea	RTOG	0.30	0.12	0.29	0.07	0.06	0.54	0.14	0.30	0.27	0.17	0.22	0.47	0.17	0.28
Bowel frequency	RMH	0.04	0.03	0.15	0.04	0.08	0.88	0.43	0.13	0.01	0.11	0.04	0.42	0.38	0.33
Subjective stool frequency	Gulliford	0.02	0.01	0.06	0.04	0.02	0.21	0.27	0.02	0.01	0.06	0.06	0.01	0.19	0.54
Loose or liqued stools	Gulliford	0.96	0.48	0.93	0.90	0.19	0.97	0.74	0.41	0.59	0.79	0.67	0.62	0.73	0.21
Subjective sphincter control	Gulliford	0.97	0.60	0.62	0.88	0.86	891.00	0.81	0.65	0.76	0.76	0.97	0.35	0.79	0.47
Constipation	CTCAE	0.99	0.75	0.68	0.70	0.37	0.94	0.92	0.85	0.82	0.50	0.54	0.47	0.39	0.24
Bowel problem overall	Gulliford	0.94	0.40	0.73	0.48	0.93	0.96	0.13	0.54	0.60	0.65	0.96	0.84	0.47	0.17

Toxicity Endpoint	Scoring System	Lateral Extent					Lateral Extent				
		30	40	50	60	70	30	40	50	60	70
Proctitis	RTOG	0.55	0.37	0.76	0.80	0.52	0.39	0.52	0.41	0.37	0.99
Proctitis	CTCAE	0.49	0.71	0.95	0.70	0.59	0.72	0.87	0.80	0.40	0.74
Rectal bleeding	RMH	0.76	0.47	0.83	0.76	0.93	0.66	0.38	0.80	0.69	0.49
Rectal haemorrhage	CTCAE	0.63	0.31	0.99	0.91	0.63	0.54	0.32	0.92	0.87	0.90
Rectal urgency	Gulliford	0.89	0.36	0.59	0.92	0.68	0.79	0.86	0.34	0.47	0.97
Diarrhoea	RTOG	0.81	0.84	1.00	0.43	0.68	0.99	0.54	0.63	0.49	0.67
Diarrhoea	CTCAE	0.63	0.29	0.54	0.24	0.26	0.84	0.88	0.87	0.56	0.76
Bowel frequency	RMH	0.49	0.63	0.65	0.26	0.90	0.98	0.86	0.70	0.55	0.52
Subjective stool frequency	Gulliford	0.78	0.52	0.48	0.57	0.69	0.07	0.68	0.90	0.78	0.53
Loose or liqued stools	Gulliford	0.92	0.33	0.25	0.28	0.24	0.94	0.34	0.35	0.88	0.29
Subjective sphincter control	Gulliford	0.45	0.82	0.29	0.25	0.31	0.98	0.96	0.33	0.30	0.45
Constipation	CTCAE	0.92	0.79	0.37	0.53	0.40	0.93	0.89	0.61	0.31	0.06
Bowel problem overall	Gulliford	0.30	0.65	0.50	0.77	0.67	0.93	0.87	0.62	0.13	0.74

 p value statistically significant prior to correction for multiple testing using Holm Bonferroni
 p value not statistically significant

Table 7.4 Table of the p values calculated using MWU to test for significance between each of the spatial characteristics and Grade 1+ toxicity for both planned and accumulated DSM (n=80). Boxes highlighted in orange are those which are statistically significant (pre-Holm Bonferroni uncorrected alpha of 0 .05). The p values of 0.05 which have been highlighted by an * are statistically significant prior to rounding up to two decimal places. After correction for multiple testing using Holm Bonferroni there was no statistical significance for any endpoint.

Toxicity Endpoint	Scoring System	Planned Pixel Count						Accumulated Pixel count							
		10	20	30	40	50	60	70	10	20	30	40	50	60	70
Proctitis	RTOG	0.68	0.69	0.44	0.52	0.61	0.90	0.36	0.31	0.41	0.96	0.89	0.79	0.65	0.25
Proctitis	CTCAE	0.89	0.91	0.20	0.40	0.41	0.97	0.54	0.45	0.57	0.70	0.62	0.98	0.89	0.44
Rectal urgency	Gulliford	0.90	0.99	0.70	0.87	0.92	0.92	0.41	0.91	0.77	0.87	0.59	0.43	0.89	0.68
Diarrhoea	RTOG	0.12	0.03	0.11	0.40	0.43	0.64	0.30	0.10	0.02	0.08	0.69	0.86	0.79	0.51
Diarrhoea	CTCAE	0.13	0.06	0.14	0.20	0.22	0.40	0.18	0.05	0.02	0.09	0.40	0.51	0.47	0.16
Bowel frequency	RMH	0.03	0.03	0.03	0.08	0.06	0.04	0.09	0.04	0.02	0.03	0.46	0.69	0.26	0.35
Subjective sphincter control	Gulliford	0.62	0.13	0.05	0.05	0.31	0.25	0.40	0.62	0.22	0.06	0.03	0.04	0.03	0.01
Bowel problem overall	Gulliford	0.83	0.55	0.90	0.73	0.73	0.26	0.54	0.85	0.53	0.80	0.78	0.53	0.21	0.73

Toxicity Endpoint	Scoring System	Longitudinal Extent						Longitudinal Extent							
		10	20	30	40	50	60	70	10	20	30	40	50	60	70
Proctitis	RTOG	0.95	0.75	0.90	0.64	0.23	0.97	0.39	0.19	0.45	0.49	0.15	0.54	0.45	0.11
Proctitis	CTCAE	0.66	0.94	0.48	0.99	0.06	0.75	0.38	0.28	0.63	0.86	0.38	0.94	0.56	0.12
Rectal urgency	Gulliford	0.82	0.95	0.94	0.50	0.70	0.40	0.85	0.64	0.87	0.94	0.84	0.16	0.75	0.54
Diarrhoea	RTOG	0.09	0.04	0.19	0.16	0.46	0.26	0.51	0.06	0.09	0.11	0.05	0.36	0.91	0.75
Diarrhoea	CTCAE	0.19	0.06	0.51	0.22	0.47	0.74	0.71	0.09	0.03	0.19	0.07	0.41	0.71	0.23
Bowel frequency	RMH	0.04	0.02	0.07	0.05	0.06	0.60	0.10	0.06	0.01	0.03	0.01	0.79	0.19	0.36
Subjective sphincter control	Gulliford	0.79	0.60	0.42	0.27	0.89	0.69	0.40	0.34	0.79	0.84	0.56	0.86	0.38	0.10
Bowel problem overall	Gulliford	0.76	0.96	0.55	0.89	0.91	0.31	0.90	0.67	0.82	1.00	0.74	0.34	0.48	0.61

Toxicity Endpoint	Scoring System	Lateral Extent					Lateral Extent				
		30	40	50	60	70	30	40	50	60	70
Proctitis	RTOG	0.93	0.60	0.79	0.81	0.61	0.84	0.29	0.69	0.59	0.31
Proctitis	CTCAE	0.88	0.76	0.89	0.77	0.68	0.98	0.38	0.98	0.52	0.48
Rectal urgency	Gulliford	0.44	0.66	0.76	0.85	0.67	0.41	0.81	0.67	0.90	0.71
Diarrhoea	RTOG	0.34	0.93	0.79	0.57	0.89	0.31	0.46	0.40	0.58	0.73
Diarrhoea	CTCAE	0.32	0.62	0.97	0.59	0.52	0.29	0.85	0.62	0.84	0.52
Bowel frequency	RMH	0.31	0.82	0.96	0.52	0.89	0.28	0.58	0.82	0.79	0.57
Subjective sphincter control	Gulliford	0.44	0.17	0.18	0.91	0.71	0.41	0.04	0.04	0.02	0.03
Bowel problem overall	Gulliford	0.62	0.75	0.60	0.58	0.37	0.52	0.83	0.98	0.08	0.29

	p value statistically significant prior to correction for multiple testing using Holm Bonferroni
	p value not statistically significant

Table 7.5 Table of the p values calculated using MWU to test for significance between each of the spatial characteristics and Grade 2+ toxicity for both planned and accumulated DSM (n=80). Boxes highlighted orange are those which are statistically significant (pre-Holm Bonferroni uncorrected alpha of 0.05). The p values of 0.05 which have been highlighted by an * are statistically significant prior to rounding up to two decimal places. After correction for multiple testing using Holm Bonferroni there was no statistical significance for any endpoint.

7.5.3.2 Receiver operating characteristic curves

A table is presented for each spatial metric for both Grade 1⁺ and Grade 2⁺ toxicity endpoints which had a statistically significant ROC curve for at least one dose level (Table 7.6 to Table 7.10).

Grade 1⁺ Pixel Count

	Planned						Accumulated					
	Dose Level	AUC	p Value	Confidence Lower Bound	Confidence Upper Bound	Youden (%)	Dose Level	AUC	p Value	Confidence Lower Bound	Confidence Upper Bound	Youden (%)
<u>RMH Bowel frequency</u>	10	0.66	0.05	0.51	0.81	-	10	0.66*	0.04	0.52	0.82	82.8%
	20	0.64	0.09	0.51	0.78	-	20	0.70*	0.02	0.57	0.83	65.9%
	30	0.59	0.27	0.42	0.76	-	30	0.62	0.16	0.46	0.77	-
	40	0.61	0.19	0.44	0.78	-	40	0.58	0.36	0.42	0.73	-
	50	0.67*	0.04	0.50	0.83	25.6%	50	0.60	0.22	0.44	0.76	-
	60	0.64	0.10	0.48	0.80	-	60	0.61	0.21	0.45	0.76	-
	70	0.59	0.29	0.44	0.74	-	70	0.59	0.28	0.43	0.75	-
<u>Gulliford Subjective stool frequency</u>	10	0.72*	0.01	0.58	0.86	78.3%	10	0.69*	0.03	0.54	0.84	78.8%
	20	0.66	0.07	0.53	0.80	-	20	0.68*	0.04	0.54	0.82	65.5%
	30	0.57	0.45	0.38	0.75	-	30	0.52	0.80	0.33	0.71	-
	40	0.62	0.16	0.45	0.79	-	40	0.58	0.37	0.41	0.75	-
	50	0.68*	0.04	0.51	0.85	27.6%	50	0.66	0.07	0.51	0.81	-
	60	0.66	0.07	0.50	0.82	-	60	0.58	0.38	0.40	0.75	-
	70	0.61	0.21	0.46	0.76	-	70	0.54	0.62	0.35	0.74	-

*Table 7.6 The AUC, p value and 95 % confidence intervals are presented for Grade 1⁺ toxicity endpoints for pixel count both planned and accumulated dose. Those ROC curves which are statistically significant prior to Holm- Bonferroni correction are highlighted by * (AUC of ≥ 0.6 and lower confidence interval of >0.5). The p values of 0.05 which have been highlighted by an * are statistically significant prior to rounding up to two decimal places. The Youden index has also been calculated for doses which are statistically significant prior to correction with Holm-Bonferroni.*

Grade 1⁺ Longitudinal Extent

	Planned						Accumulated					
	Dose Level	AUC	p Value	Confidence Lower Bound	Confidence Upper Bound	Youden (%)	Dose Level	AUC	p Value	Confidence Lower Bound	Confidence Upper Bound	Youden (%)
<u>RMH Bowel frequency</u>	10	0.67*	0.04	0.53	0.82	78.6%	10	0.62	0.14	0.47	0.78	-
	20	0.68*	0.03	0.54	0.83	78.6%	20	0.71*	0.01	0.56	0.86	78.6%
	30	0.62	0.15	0.45	0.79	-	30	0.63	0.11	0.48	0.79	-
	40	0.67*	0.05	0.51	0.83	64.3%	40	0.67*	0.04	0.52	0.82	69.0%
	50	0.65	0.08	0.48	0.81	-	50	0.57	0.44	0.41	0.72	-
	60	0.51	0.88	0.33	0.69	-	60	0.57	0.40	0.41	0.73	-
	70	0.56	0.44	0.41	0.72	-	70	0.58	0.34	0.41	0.75	-
<u>Gulliford Subjective stool frequency</u>	10	0.68*	0.03	0.55	0.84	78.6%	10	0.69*	0.03	0.54	0.84	83.3%
	20	0.72*	0.01	0.58	0.86	78.6%	20	0.73*	0.01	0.59	0.88	78.6%
	30	0.66	0.06	0.50	0.82	-	30	0.66	0.07	0.52	0.80	-
	40	0.68*	0.04	0.52	0.85	64.3%	40	0.66	0.06	0.52	0.81	-
	50	0.69*	0.03	0.53	0.86	69.0%	50	0.71*	0.01	0.56	0.87	64.3%
	60	0.61	0.23	0.43	0.78	-	60	0.61	0.21	0.45	0.77	-
	70	0.59	0.28	0.43	0.75	-	70	0.55	0.55	0.37	0.74	-

Table 7.7 The AUC, p value and 95 % confidence intervals are presented for toxicity endpoints for longitudinal extent for both planned and accumulated dose.

Grade 1⁺ Lateral Extent

No table is presented for the lateral extent as there was no statistical significance for any endpoint when Grade 1⁺ toxicity was associated with the planned or accumulated lateral extent using ROC curves.

Grade 2+ Pixel Count

	Planned						Accumulated					
	Dose Level	AUC	p Value	Confidence Interval Lower Bound	Upper Bound	Youden (%)	Dose Level	AUC	p Value	Confidence Interval Lower Bound	Upper Bound	Youden (%)
<u>RTOG Diarrhoea</u>	10	0.69	0.12	0.51	0.88	-	10	0.70	0.10	0.51	0.89	-
	20	0.77*	0.03	0.61	0.94	75.1%	20	0.78*	0.02	0.63	0.93	66.1%
	30	0.70	0.11	0.41	0.99	-	30	0.72	0.08	0.50	0.93	-
	40	0.60	0.40	0.30	0.91	-	40	0.55	0.69	0.30	0.80	-
	50	0.60	0.43	0.27	0.92	-	50	0.52	0.86	0.24	0.81	-
	60	0.56	0.64	0.31	0.80	-	60	0.53	0.79	0.27	0.80	-
	70	0.63	0.30	0.43	0.82	-	70	0.58	0.51	0.33	0.84	-
<u>CTCAE Diarrhoea</u>	10	0.67	0.13	0.51	0.84	-	10	0.72	0.05	0.55	0.90	-
	20	0.72	0.06	0.56	0.88	-	20	0.77*	0.02	0.61	0.92	66.1%
	30	0.67	0.13	0.42	0.92	-	30	0.70	0.09	0.50	0.89	-
	40	0.65	0.20	0.38	0.92	-	40	0.60	0.40	0.35	0.84	-
	50	0.64	0.22	0.35	0.93	-	50	0.58	0.51	0.34	0.81	-
	60	0.60	0.40	0.37	0.82	-	60	0.58	0.47	0.36	0.80	-
	70	0.65	0.18	0.45	0.85	-	70	0.66	0.16	0.48	0.85	-
<u>RMH Bowel frequency</u>	10	0.74*	0.03	0.60	0.89	78.2%	10	0.74*	0.04	0.57	0.90	78.6%
	20	0.75*	0.03	0.59	0.91	65.1%	20	0.76*	0.02	0.61	0.91	66.1%
	30	0.75*	0.03	0.55	0.96	58.7%	30	0.75*	0.03	0.60	0.90	51.5%
	40	0.70	0.08	0.48	0.93	-	40	0.59	0.46	0.39	0.79	-
	50	0.71	0.06	0.49	0.94	-	50	0.55	0.69	0.31	0.78	-
	60	0.74*	0.04	0.56	0.92	26.3%	60	0.63	0.26	0.44	0.82	-
	70	0.69	0.09	0.52	0.87	-	70	0.61	0.35	0.39	0.82	-
<u>Gulliford Subjective sphincter control</u>	10	0.57	0.62	0.21	0.94	-	10	0.57	0.62	0.17	0.97	-
	20	0.73	0.13	0.46	0.99	-	20	0.68	0.22	0.35	1.00	-
	30	0.79	0.05	0.58	0.99	-	30	0.78	0.06	0.59	0.97	-
	40	0.79	0.05	0.61	0.97	-	40	0.83*	0.03	0.66	0.99	39.8%
	50	0.65	0.31	0.39	0.91	-	50	0.81*	0.04	0.66	0.96	25.1%
	60	0.67	0.25	0.40	0.94	-	60	0.83*	0.02	0.61	1.00	18.5%
	70	0.63	0.40	0.43	0.82	-	70	0.87*	0.01	0.70	1.00	12.1%

Table 7.8 The AUC, p value and 95 % confidence intervals are presented for Grade 2+ toxicity endpoints for pixel count both planned and accumulated dose.

Grade 2+ Longitudinal Extent

	Planned						Accumulated					
	Dose Level	AUC	p Value	Confidence Interval		Youden (%)	Dose Level	AUC	p Value	Confidence Interval		Youden (%)
				Lower Bound	Upper Bound				Lower Bound	Upper Bound		
<u>RTOG Diarrhoea</u>	10	0.71	0.09	0.55	0.86	-	10	0.73	0.06	0.52	0.94	-
	20	0.74*	0.05	0.57	0.92	73.8%	20	0.71	0.09	0.53	0.89	-
	30	0.66	0.19	0.40	0.92	-	30	0.69	0.12	0.48	0.90	-
	40	0.67	0.17	0.40	0.94	-	40	0.74	0.06	0.57	0.90	-
	50	0.59	0.46	0.28	0.90	-	50	0.61	0.37	0.40	0.82	-
	60	0.36	0.27	0.16	0.57	-	60	0.51	0.91	0.31	0.72	-
	70	0.58	0.52	0.36	0.80	-	70	0.54	0.76	0.29	0.78	-
<u>CTCAE Diarrhoea</u>	10	0.65	0.20	0.50	0.80	-	10	0.69	0.10	0.49	0.89	-
	20	0.71	0.07	0.54	0.88	-	20	0.74*	0.04	0.58	0.90	73.8%
	30	0.57	0.52	0.35	0.80	-	30	0.65	0.20	0.47	0.83	-
	40	0.64	0.23	0.42	0.86	-	40	0.70	0.08	0.56	0.85	-
	50	0.58	0.48	0.31	0.85	-	50	0.59	0.43	0.38	0.80	-
	60	0.46	0.75	0.23	0.70	-	60	0.54	0.72	0.36	0.73	-
	70	0.54	0.71	0.34	0.75	-	70	0.63	0.25	0.45	0.82	-
<u>RMH Bowel frequency</u>	10	0.73*	0.05	0.59	0.87	78.6%	10	0.71	0.07	0.51	0.91	-
	20	0.76*	0.02	0.61	0.92	73.8%	20	0.80*	0.01	0.66	0.94	78.6%
	30	0.70	0.08	0.47	0.94	-	30	0.74*	0.04	0.55	0.93	73.8%
	40	0.72	0.06	0.48	0.96	-	40	0.78*	0.02	0.64	0.92	69.0%
	50	0.71	0.07	0.50	0.93	-	50	0.53	0.79	0.33	0.73	-
	60	0.56	0.61	0.35	0.76	-	60	0.65	0.21	0.49	0.80	-
	70	0.68	0.11	0.50	0.87	-	70	0.60	0.37	0.40	0.81	-

Table 7.9 The AUC, p value and 95 % confidence intervals are presented for Grade 2+ endpoints for longitudinal extent for both planned and accumulated dose.

Grade 2+ Lateral Extent

	Planned					Accumulated						
	Dose Level	AUC	p Value	Confidence Interval		Youden (%)	Dose Level	AUC	p Value	Confidence Interval		Youden (%)
				Lower	Upper				Lower	Upper		
<u>Gulliford Subjective sphincter control</u>	30	0.57	0.65	0.31	0.82	-	30	0.57	0.62	0.32	0.83	-
	40	0.71	0.17	0.54	0.87	-	40	0.81*	0.04	0.67	0.95	73.8%
	50	0.70	0.19	0.57	0.83	-	50	0.81*	0.04	0.69	0.93	54.8%
	60	0.52	0.91	0.29	0.74	-	60	0.83*	0.03	0.59	1.00	45.2%
	70	0.55	0.72	0.31	0.79	-	70	0.83*	0.03	0.58	1.00	35.7%

Table 7.10 The AUC, p value and 95 % confidence intervals are presented for Grade 2+ endpoints for lateral extent for both planned and accumulated dose.

7.6 Discussion

To the author's knowledge, this is the first study to associate both planned and accumulated spatial characteristics with prospectively collected toxicity data (including baseline toxicity) for a substantial number of patients. The use of prospective data enabled the exclusion of patients with baseline symptoms, reducing the risk of including non-radiotherapy toxicity in the analysis [28, 95].

7.6.1 Difference between planned and accumulated spatial characteristics

Analyses of the difference between planned and accumulated spatial metrics showed the accumulated pixel count and longitudinal extent were greater than planned in the low to mid dose regions (10 to 40 Gy). This was in partial agreement with DVH data (Chapter 5) which showed that the accumulated rectal volume was statistically significantly greater than planned at 30 Gy ($p < 0.003$) and 40 Gy ($p < 0.002$). At 70 Gy, however, the accumulated pixel count ($p < 0.005$) and longitudinal extent ($p = 0.001$) was statistically significantly smaller than planned. For the lateral extent the accumulated dose was statistically significantly smaller than the planned lateral extent between 50 to 70 Gy. At the higher doses, DVH analysis did not show any statistically significant difference between planned and accumulated dose, suggesting that the DSM analysis may be more sensitive to changes in the dose distribution at higher doses during treatment. Overall, the difference between planned and accumulated spatial metrics across all doses was small. Approximately 50 % of patients had a difference of less than 5 % for all metrics and across all doses when the accumulated DSM was compared to the planned DSM. This is of a similar magnitude to the difference in dose-volume parameters between planned and accumulated dose reported in Section 5.5.3. As with the DVH analysis, this would indicate that planned DSM is a good estimate of delivered dose in the cohort of patients studied in this thesis. This is likely due to the daily online correction of the prostate position in the DELINEATE trial using CBCT which would reduce the variability of the volume of the rectum in the high dose region.

Due to the similarity in spatial metrics found between planned and accumulated dose, a more detailed DSM may improve the ability to describe the small differences reported. In the current study, the DSM were normalised to an array of 21 x 21 pixels using previously published methodology [94]. Differences in rectal length between

fractions and between patients required normalisation of the rectal length in order for comparison. Hoogeman et al., also normalised the length of their maps [150]. They reported the average (SD) rectal length was 13.3cm \pm 0.6cm. As with Hoogeman there is an assumption in this study that the biomechanical and dose-effect relationships are functions of relative length and not absolute length of the rectum. Hoogeman created inter-patient maps with combined absolute length and compared to maps created by combining relative length with only small differences noted. Furthermore, the DSM used in this thesis were generated by slicing the rectum longitudinally at 0.3 cm intervals. This interval is smaller than that used by Shelley et al., who were limited to 0.6 cm intervals due to the slice thickness of the MVCT [93]. Despite the slice thickness, their study of 109 patients reported association between the lateral extent and rectal bleeding when daily dose maps were accumulated. Nevertheless, the previously reported methodology used to generate the DSM in this thesis could have been amended to mimic the CBCT thickness of 0.15 cm which would support the development of a finer array. However, the small differences observed between planned and accumulated spatial metrics in this thesis may mean this would not to be clinically meaningful.

7.6.2 Association of spatial metrics and toxicity

The association of planned and accumulated pixel count and longitudinal extent showed statistical significance for both Grade 1⁺ and Grade 2⁺ toxicity endpoints in the low to mid dose region, namely those that reflect diarrhoea and stool frequency. Buettner et al., has also reported a strong association between loose stools and the longitudinal extent between 21 to 29 Gy [149]. Previously, rectal bleeding was the most common endpoint reported to be associated with the shape of dose distribution described by DSM, and was associated with the lateral extent at mid to high dose range [70, 93, 95, 149]. The most recent and largest study to date, by Wilkins et al., found statistically significant association between the lateral extent and rectal bleeding at 40 to 65 Gy ($p < 0.005$) in patients in the CHHiP trial (3,216 patients) [70]. However, in the current study there was no statistically significant difference in the lateral extent measured in patients with and without rectal bleeding. The only statistically significant association found between lateral extent and toxicity, was for the accumulated lateral extent and Grade 2⁺ subjective sphincter control at doses of 40 to 70 Gy ($p \leq 0.04$). Previously, using a sub region analysis of the anal sphincter, association between the

lateral extent and subjective sphincter control at 53 Gy ($p = 0.01$) was reported by Buettner et al., in the MRC RT01 trial [236].

A Holm-Bonferroni correction was applied to the p values to avoid detection of random associations occurring by chance. Once a correction for multiple testing had been applied, there was no longer a significant association observed between any of the planned and accumulated spatial metrics, and the toxicity endpoints analysed using either MWU or ROC curves. However, the clustering of statistically significant p -values in clinically plausible dose regions (i.e. bowel frequency in the low to mid dose region), with similar findings for both planned and accumulated doses, strongly suggests the association of these spatial metrics with endpoints reflecting diarrhoea and bowel frequency [70].

As acknowledged in the section 6.6, it should be noted that the incidence of toxicity within this thesis population is low, and subsequently may be underpowered for some of the endpoints being analysed. For example, as previously discussed, studies have consistently reported the lateral extent of dose correlating strongly with rectal bleeding at high doses (See Table 7.1) [70, 93, 95, 149]. In this study, however, there was no correlation between either the planned and accumulated lateral extent and rectal bleeding between 30 to 70 Gy. This may be because the incidence of rectal bleeding in this study was low (Grade 1+ <13 % and Grade 2+ 1 %). For comparison, Shelley et al., reported an incidence of rectal bleeding of 25.7 % [10]. However, they did not perform a Mann Whitney U test analysis, restricting their analysis to ROC only. The low incidence of rectal bleeding in the current study, despite dose escalation, is probably due to adherence to stringent DVC, though reduced posterior PTV margins and a daily on-line IGRT protocol (discussed in section 6.6). All of which contributes to reducing the volume of the rectum in the high dose region which is associated with an increased incidence of rectal bleeding [9]. Finally, for endpoints that were found to have an association with pixel count and longitudinal extent, greater patient numbers may have also provided more confidence in detected differences (smaller p -values) making the findings more robust to correction for multiple testing (Holm-Bonferroni).

7.6.3 Spatial dose metrics as predictors of toxicity.

The statistically significant AUC values for ROCs reported in the current study were greater than those reported by others. Shelly et al. [10] found the accumulated dose

correlated most strongly with rectal bleeding at 65 Gy with a accumulated AUC of 0.67 versus a planned AUC of 0.64. Similarly, AUCs reported by Wilkins et al., were all < 0.68. In the current study, all statistically significant AUC were greater than 0.67, with median AUC 0.72 and 0.76 for planned and accumulated ROC, respectively ($p = 0.076$).

The association of pixel count and longitudinal extent showed statistical significance for both Grade 1⁺ and Grade 2⁺ toxicity endpoints that reflect diarrhoea and stool frequency. Significant association for these endpoints was found in the low to mid dose region which has been shown to be an important dose region for these particular endpoints [68, 70, 210]. In the current study, statistically significant associations were found for the same endpoints for both planned and accumulated DSM, and across similar doses indicating that neither is more predictive than the other. Association of dose-volume and toxicity in Chapter 6 also found statistically significant associations for planned dose and RMH bowel frequency and CTCAE diarrhoea. However, these associations were limited to planned V20 and V30. Overall, using DSM has resulted in a greater number of statistically significant associations between dose metrics (pixel count and longitudinal extent) for both planned and accumulated dose using DSM (significance at V20 to V50).

A limitation of the work undertaken in this chapter is that DSM were not subdivided into geometrical regions or specific anatomical rectal sub-regions. Dividing the DSM into geometrical regions enables analysis of organs with stable sub-regions which may have increased the chance of finding statistically significant differences between patients with and without toxicity. Some studies have shown that sub-regions of the rectum may be associated with different toxicity endpoints [95, 167, 229]. Heemsbergen et al., reported that when the total region was analysed, the only significant dose area effect was for rectal bleeding (n=266 patients) [167]. However, when sub-regions were defined, it was reported that different effects were associated with specific regions of the rectum. Bleeding and mucus loss were related to the upper part of rectum, and soiling and faecal incontinence were more likely associated with dose to anal canal and lower part of rectum. Conversely a recent paper by Wilkins et al., analysed a considerably larger patient cohort (n= 3,216) and described the separation of the anorectum into the anal canal and rectum using both DVH and DSM [70]. They concluded that there was no benefit from separation of the anorectum after

the study found that the dose to the entire anorectum or rectum was a stronger predictor of faecal incontinence than the anal canal.

Finally, the main hypothesis was that delivered dose metrics were better predictor of than planned. The median AUC for accumulated dose (0.76 range = 0.67 to 0.87) was not statistically significantly greater than for planned dose AUC (0.72 range = 0.67 to 0.77). However, there were a greater number of accumulated dose metrics, which had statistically significant AUC (19 planned versus 26 accumulated, See Tables 7.6 to 7.10). Overall, this result is not that dissimilar to that of Shelly et al., who reported a small difference in AUC between planned and accumulated dose (0.03) when the lateral extent of 109 patients was associated with Grade 2⁺ rectal bleeding (LENT SOMA). However, the statistical significance of this result was not given [93]. The small difference observed between planned and accumulated AUC in the current study is unsurprising, and is a result of the small differences in metrics calculated between planned and accumulated DSM, which again confirms the efficiency of daily on-line correction strategy.

7.7 Conclusion

Results for accumulated and planned DSM were broadly consistent with relationships between dosimetry and toxicity for a cluster of symptoms that are often related, for example bowel frequency, loose stools and diarrhoea. This consistency is a consequence of the small differences between planned and accumulated spatial metrics as a result of daily IGRT, small PTV margins and stringent DVC (discussed in Chapters 5 and 6) resulting in low incidence of toxicity.

Chapter 8 Discussion and final conclusion

The aim of this thesis was to determine whether delivered dose to the rectum is a more accurate predictor of rectal toxicity than planned dose. The delivered dose was calculated using deformably registered dose volume histograms (DVH), and dose surface maps (DSM), to accumulate the dose using daily treatment CBCTs. The work undertaken was divided into three key objectives:

- i. To evaluate the geometric uncertainties in the dose accumulation methodology.
- ii. To determine the difference between planned and delivered dose using dose-volume and spatial metrics.
- iii. To evaluate and compare the association of planned and accumulated dose with toxicity.

8.1 Geometric uncertainties in the dose accumulation methodology

8.1.1 Delineation of the rectum

Both of the dose accumulation methodologies used in this thesis relied on the accurate delineation of the rectum. It is already well established that manual delineation of structures is subject to variation between observers which can lead to differences in dosimetry [119, 142]. In addition, the rectum was delineated on CBCTs which have poorer soft tissue contrast than CT scans, potentially increasing inter- and intraobserver variation [191]. The author who delineated the rectum on all CBCTs, had no previous experience in delineation. Therefore, initial work in this thesis focused on measuring inter- and intraobserver variation in manual delineation of the rectum on CBCTs and its influence on deformable image registration (DIR) and calculation of dose. Comparison of the author's manual contours to the experts showed good concordance, and despite the CBCTs being generally a lower image quality, the interobserver variation was similar to that obtained for delineation of the rectum on CT. This study also highlighted that the observers' contours are most likely to vary at the superior end of the rectum, and therefore clearer guidance on the delineation of the superior rectum is required to help reduce interobserver variation

To the author's knowledge, there have been no published studies which have measured the effect of interobserver variation on deformable image registration or the resulting deformed dose. Work undertaken in this thesis showed that variation in manual delineation of the rectum could lead to poor performance of the DIR which resulted in large differences in the deformed dose-volume, as quantified by similarity and overlap measures. The images which were poorly deformed were those images where there was poor concordance between observers in defining the superior rectum. Visual analysis showed that this resulted in ambiguity in correspondence between tissues contained within the contours which were used as regions of interest (ROIs) to control the deformation. Exclusion of the poorly deformed images ensured that the variation in deformed dose-volume between observers was within acceptable limits [135, 142]. It should be acknowledged that Dice similarity coefficient (DSC) and mean distance to agreement (MDA), which were used to quantify how well the image was deformed are only able to verify the deformation locally i.e. close to the rectal contour, and not globally i.e. the whole image. Another limitation of this method is that the deformation is only evaluated at the boundary of the rectum. However, as previously discussed, both DSC and MDA are recommended as means of verifying how well an image has been deformed [135, 148]. The integration of these measures into the RayStation dose accumulation workflow enables the verification of large numbers of deformed images quickly. No study was performed comparing the effect of interobserver variation on DSM as there was good concordance between observers. However, future work should assess the sensitivity of DSM to contour variation because the contour is used to generate a binary mask from which dose and subsequent spatial metrics are measured.

Auto-segmentation models (ASM) have been developed to automatically segment structures to reduce the time required for manual delineation and reduce observer variation. Three different auto-segmentation models for segmenting the rectum on CBCTs were evaluated in this thesis: Atlas-based segmentation; model-based segmentation; and contour propagation using deformable image registration. None of the auto-segmentation models evaluated were able to generate rectal contours which were comparable to those which were manually delineated. This may be because ASM have historically been primarily developed to ease the burden of delineation on the planning CT. As a result, the algorithms may be affected by the poorer image quality of the CBCT [237]. Consequently, rectums were manually delineated on all CBCTs used in this thesis. If dose accumulation is to be implemented clinically, it is important

that algorithms are developed that can quickly and accurately segment organs on treatment images such as CBCT. One solution is the development of artificial intelligence (AI) contouring tools which were unavailable at the time of this PhD. AI contouring tools may be better at segmenting structures on CBCTs than other established auto segmentation tools, with emerging studies showing good concordance between AI segmented pelvic structures and physician delineated structures on CBCTs [237, 238].

8.1.2 Missing data

A key element to the accumulation of dose is a time dependent description of the patient's anatomy [134]. It was identified early in the thesis that the majority of patients had one or more CBCT images missing, or images which were unsuitable for dose accumulation. To evaluate the effect of missing image data on accumulated dose, a study was undertaken to compare dose calculated using a full (i.e. an image for each fraction), and an artificially reduced imaging dataset. To the authors knowledge, this is only the second study to evaluate the effect of missing images on accumulated dose, with the only other study limited to 3 patients [148]. The current study is also the only study to evaluate the effect of missing images on accumulated dose using spatial metrics. The work undertaken showed that difference in dose between complete and reduced datasets increased as the number of CBCTs being accumulated reduced. However, even when half the CBCTs were removed from the dataset, the difference in rectal volume did not exceed 2.5 % at specified dose levels. Most importantly, the difference between the complete and reduced imaging datasets was smaller than the difference between accumulated and planned dose. This suggests that even if only half of the imaging data is available, using this subset of images will provide a more accurate estimate of the actual delivered dose than using the plan alone.

It should also be noted that the current study used a moderately hypofractionated dose prescription (60 Gy/20 #), and subsequently there are sufficient fractions for the effect of rectal deformation on dose to be "blurred". Further work would need to be undertaken to evaluate the impact of missing images when treating with extreme hypofractionation (2 to 5 #) [239, 240]. This is firstly because large doses in the small number of fractions reduce the opportunity for "blurring" of dose, and secondly because the effect of a missing image is amplified in a smaller fractionation schedule. For example, if one image is missing out of 20 #, only 5 % of the imaging data is

missing. However, if one image is missing from 5 #, 20% of the imaging data is missing.

The work undertaken in this thesis will support future prostate dose accumulation studies by demonstrating that a missing image(s) will not affect the validity of the work and that for moderately hypofractionated schedules, now standard of care in the UK, delivered dose can be estimated using as little as 50 % of the daily imaging data. Different tumour sites, or OAR, will require further evaluation and the number of images required will be influenced by the magnitude of organ deformation and whether this is reproducible over the time course of treatment delivery. The current study also has the potential to reduce the burden of manual delineation by reducing the number of images needed to calculate the accumulated dose. For example, accumulating the dose using every other image essentially reduces the workload by half. Importantly, studies such as this one will be required to support evidence-based decisions in future clinical practice regarding additional exposures (i.e. the need to re-image or not), as we move from image guided to dose guided radiotherapy.

8.2 Difference between planned and delivered dose

This study showed that despite interfraction rectal volume variation, differences between planned and delivered dose-volume and spatial metrics remained small. For 50 % of the population analysed, the difference was < 5 % for specific dose-volumes and spatial metrics. This indicates that in the context of IMRT to the prostate using on-line image guidance (CBCT and fiducial markers), the planned dose provides a good approximation of delivered dose. The small differences in dose-volume observed also meant that the accumulated dose adhered to the required planning dose volume constraints. Previously in the literature, variation in rectum dose-volume was reported as often being greater than planned, with delivered dose frequently reported as exceeding DVC [31, 32, 130]. However, the delivered dose was usually analysed by fraction or summed which may have led to overestimation of the difference between planned and delivered dose [129, 134, 148]. In this thesis, spatial correspondence was established between images enabling dose to be accumulated on a voxel by voxel basis, reducing the overestimation of delivered dose. In addition, this PhD has drawn data from a contemporary trial, which benefits from improvements in image guidance and technique, which have been shown to reduce dose to normal tissues [120]. The

results are therefore most relevant to present “state of the art” treatment delivery rather than previously used technologies.

The adherence of delivered dose to the planning DVC suggests that even tighter DVC could be developed and tested. A recent study by Wilkins et al., has published more stringent DVC for moderate hypofractionation, based on analysis of the CHHiP trial which was a forerunner to the DELINEATE trial [70]. The DVC were derived from association of the planning dose-volume data with toxicity. Comparison of accumulated data from the current study with the newly proposed planning DVC ($V_{20}<85\%$, $V_{30}<57\%$, $V_{40}<38\%$, $V_{50}<22\%$ and $V_{60}<0.01\%$), showed that the accumulated hypofractionated DVHs were already meeting $V_{50}<22\%$. Apart from two patients, the accumulated DVHs met the $V_{20}<85\%$, and approximately half of the DVHs met the $V_{30}<57\%$ and $V_{40}<38\%$. Very few appeared to meet the $V_{60}<0.01\%$. However, as this analysis was influenced by the inclusion of the 57 Gy in 19 fractions group in CHHiP, present ongoing UK National Trials PACE-B (ISRCTN17627211) and PIVOTAL boost (ISRCTN80146950) have modified this high dose “optimal” constraint to 1 % and 3 % respectively. In the current study, the accumulated DVH for 14 patients met the high dose constraint of 1 %. Nevertheless, work undertaken as part of this PhD found that providing the planning DVH are kept well below the DVC, subsequent interfraction variation is unlikely to cause the delivered dose to exceed the prescribed constraints. It would be interesting to determine if this remains true when tighter DVCs are used. The imaging dataset from this thesis could be used to test the robustness of the newly proposed DVC to interfraction motion.

Future work should be undertaken to accumulate dose in patients treated with extreme hypofractionation (2 to 5 #). Although all patients in the current study met the required DVC when the dose was accumulated, approximately 1 % of daily fractions exceeded the required DVC for both dose cohorts with a larger number exceeding the optimal DVC. If extreme hypofractionation was delivered on days when the DVH exceeded the DVC, a greater difference between planned and delivered dose may have been observed. One of the few studies to evaluate the dosimetric impact of organ motion in extreme hypofractionation is by Devlin et al., who analysed the dose to the rectum in 41 patients receiving 35 Gy/5 # [241]. Although they did not report a statistically significant difference in the rectal volume between planning and treatment, they did report that almost 50 % of patients failed the mid-high DVC at treatment. However,

the dose was not accumulated but rather analysed on a fraction by fraction basis. As previously discussed above, not mapping the dose can lead to overestimation of dose in a deformable structure such as the rectum. However, until a dose accumulation study is performed, we will not know the true dosimetric effect of rectal motion on delivered dose in extreme hypofractionation.

8.3 Association of dose and toxicity

This is the first study to associate accumulated dose to the rectum using DVH and spatial metrics with prospectively collected GI toxicity data. The only other study to have previously performed this was the VoxTox group [92, 93]. There are some key differences in the methodology used between the VoxTox study and this study. Firstly, VoxTox used toxicity data retrospectively collected from patient notes, which did not include toxicity at baseline. Inclusion of toxicity at baseline reduces the risk of including non-radiotherapy toxicity in the analysis. Secondly, due to the reduced FOV of the MVCT acquired images, the superior, as well as the inferior, rectum was simulated using the CT scan as a proxy. A limited FOV was sometimes a challenge in the current study, although work undertaken as part of this PhD validated a different method to that used by the VoxTox group of simulating the inferior rectum. However, it was decided not to simulate the superior rectum as it could mask the large variation in volume that can occur in the superior third of the rectum [201], these patients were therefore excluded from analysis in the current study.

The association of planned and delivered dose, using both dose-volume and spatial metrics, with GI toxicity has shown that accumulated dose is not a better predictive of rectal toxicity. This is because planned dose is a close approximation of accumulated dose which is likely a result of reduced CTV-PTV margins, stricter DVC and correction for geometric uncertainty in the prostate position. It should be acknowledged that for some toxicity endpoints, the current study may be underpowered due to the low incidence of side-effects. However, the incidence rates are comparable to contemporary studies [43, 52, 151]. Increasing the number of patients analysed, by including additional DELINEATE patients would also allow for multivariate analysis of additional patient risk factors to be performed. Unfortunately, there were insufficient patient numbers to do so in the current study.

Further work could also be undertaken in evaluating which patients might benefit from adaptive treatment. The MR-Linac has superior soft tissue imaging, and the ability to adapt the dose distribution on a daily basis. However, the workflow can take between 30 to 50 minutes which can result in displacement of the prostate by organ motion and filling [242]. There are also limited numbers of MR-Linacs globally which means access is currently limited. Adaption on a conventional linac is challenging but there have been several techniques proposed including triggered adaption, which would fit well with the work undertaken in this PhD [193, 243]. Triggered adaption refers to the process of adapting the plan when a certain threshold is exceeded e.g. time triggered adaption after 50 % of treatments delivered. However, the rectal dose could be accumulated using the first 5 #, building on the work undertaken in this thesis of accumulating dose using limited numbers of CBCTs. This estimated delivered dose could be used to identify patients who may benefit from adaption.

DSM may be more sensitive than DVH in predicting toxicity as there were more significant DSM based predictors than DVH when both the planned and accumulated metrics were associated with toxicity. This suggests that the shape of the dose distribution also plays an important role in the development of adverse events. At present, the use of DSM is limited to being a research tool as it's not integrated into commercial planning systems. Future work could investigate how to integrate the DSM into the planning workflow. This would also support the long-term development of spatial dose constraints similar to dose volume constraints in current practice.

A limitation of this thesis is that the effect of intrafraction motion of the rectum on delivered dose was not measured. However, this is challenging and has not yet been addressed in the research literature. This has been achieved for the prostate using continuous monitoring by implanted electromagnetic transponders for example but these do not monitor rectal shape or volume [244, 245]. Rectal dose parameters will relate to the position of the anterior rectal wall but also to any changes in the volume and shape of the rectum. Treatment images acquired after treatment would enable comparison in shape and volume over time, however this would not capture the transient, rapid changes in the rectum as gas moves through the rectum. The effect of intrafraction motion could be simulated. However, this would not enable the actual delivered dose to be calculated and associated with toxicity. The MR Linac could be used in the future to acquire continuous images during treatment which could enable

the dosimetric impact of rectal changes to be assessed. However, correlation with treatment related side-effects might remain challenging without prospective collection of substantial patient cohorts.

8.4 Summary of future work and recommendations

This study has shown that accumulated dose is not a better predictor of rectal toxicity than planned dose. However, only conventional (37#) and moderate hypofractionated prescriptions (20#) were analysed. Future work should now focus on accumulating dose using extreme hypofractionation (2-5#) to determine if this holds true when less opportunity for the dose to be ‘blurred’ exists.

As shown in this study the planned dose provides a good estimation of delivered dose. The adherence of delivered dose to the planning DVC suggests even tighter constraints could be developed for the rectum. The current imaging dataset should be used to validate newly proposed DVC such as those derived by Wilkins et al from the CHHiP trial [70]. Future work should also focus on identifying and stratifying the few patients who struggle to meet DVC, and who might benefit from adaptive radiotherapy.

The increased number of statistically significant associations between spatial metrics and toxicity indicate that DSM may be more sensitive than DVH in predicting toxicity. A comparison could be undertaken using data from this study however the incidence of toxicity is low. Nevertheless, methods to integrate DSM into the planning workflow should be explored to enable a larger population to be analysed. Further work also needs to be done to measure uncertainties in the DSM methodology such as sensitivity to contour variation. Data from this study should also be used to develop spatial dose constraints.

Finally, the effect of the position and volume of the intra-prostatic nodules on dose to the rectum, and the correlation with toxicity, was not considered in this work. This should be evaluated in a larger population. Furthermore, dose to the penile bulb, peri-prostatic nerves and bladder trigone should be evaluated, and the use of spatial metrics and voxel-level analysis considered.

8.5 Final conclusion

It has been shown in this PhD that it is possible to accumulate dose using approximately 50 % of delivered fractions. Planned dose closely approximates accumulated dose in patients treated with IGRT using fiducials and daily on-line image correction. Dose surface metrics may have a closer association with late toxicity than dose-volume metrics.

References

1. *About Prostate Cancer.* Prostate Cancer UK: <https://prostatecanceruk.org/prostate-information/about-prostate-cancer>.
2. *Prostate Cancer Statistics - Key Facts.* <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer>.
3. Kamran, S.C. and A.V. D'Amico, *Radiation Therapy for Prostate Cancer.* Hematol Oncol Clin North Am, 2020. **34**(1): p. 45-69.
4. Mayles, W.P. and B. Radiotherapy Development, *Survey of the availability and use of advanced radiotherapy technology in the UK.* Clin Oncol (R Coll Radiol), 2010. **22**(8): p. 636-42.
5. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.
6. Dal Pra, A. and L. Souhami, *Prostate cancer radiation therapy: A physician's perspective.* Phys Med, 2016. **32**(3): p. 438-45.
7. Moon, D.H., J.A. Efstathiou, and R.C. Chen, *What is the best way to radiate the prostate in 2016?* Urol Oncol, 2017. **35**(2): p. 59-68.
8. Warde, P., et al., *Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: a randomised, phase 3 trial.* The Lancet, 2011. **378**(9809): p. 2104-2111.
9. Michalski, J.M., et al., *Radiation dose-volume effects in radiation-induced rectal injury.* Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S123-9.
10. Zelefsky, M.J., et al., *Improved clinical outcomes with high-dose image guided radiotherapy compared with non-IGRT for the treatment of clinically localized prostate cancer.* Int J Radiat Oncol Biol Phys, 2012. **84**(1): p. 125-9.
11. Zelefsky, M.J., 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, 2008. **70**(4): p. 1124-9.
12. Al-Mamgani, A., et al., *Update of Dutch multicenter dose-escalation trial of radiotherapy for localized prostate cancer.* Int J Radiat Oncol Biol Phys, 2008. **72**(4): p. 980-8.
13. Dearnaley, D.P., et al., *Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the MRC RT01 randomised controlled trial.* The Lancet Oncology, 2014. **15**(4): p. 464-473.
14. Beckendorf, V., et al., *70 Gy Versus 80 Gy in Localized Prostate Cancer: 5-Year Results of GETUG 06 Randomized Trial.* International Journal of Radiation Oncology*Biophysics*Physics, 2011. **80**(4): p. 1056-1063.
15. Zelefsky, M.J., et al., *Long-term results of conformal radiotherapy for prostate cancer: impact of dose escalation on biochemical tumor control and distant metastases-free survival outcomes.* Int J Radiat Oncol Biol Phys, 2008. **71**(4): p. 1028-33.
16. Fiorino, C., et al., *Rectal dose-volume constraints in high-dose radiotherapy of localized prostate cancer.* Int J Radiat Oncol Biol Phys, 2003. **57**(4): p. 953-62.
17. Roach, M., 3rd, et al., *Radiation dose-volume effects and the penile bulb.* Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S130-4.
18. Murray, J., et al., *Evaluation of erectile potency and radiation dose to the penile bulb using image guided radiotherapy in the CHHiP trial.* Clin Transl Radiat Oncol, 2020. **21**: p. 77-84.

19. Landoni, V., et al., *Predicting toxicity in radiotherapy for prostate cancer*. *Phys Med*, 2016. **32**(3): p. 521-32.
20. Miyake, M., et al., *Changes in lower urinary tract symptoms and quality of life after salvage radiotherapy for biochemical recurrence of prostate cancer*. *Radiother Oncol*, 2015. **115**(3): p. 321-6.
21. Shadad, A.K., et al., *Gastrointestinal radiation injury: prevention and treatment*. *World J Gastroenterol*, 2013. **19**(2): p. 199-208.
22. James W. Denhama, P.C.O.B., R. Hugh Dunstan, Jorgen Johansen, Andrew Seea Christopher S. Hamilton, Sean Bydder, Suzanne Wright, *Is there more than one late radiation proctitis syndrome?* *Radiat Oncol*, 1999. **51**: p. 43-53.
23. Andreyev, J., *Gastrointestinal symptoms after pelvic radiotherapy: a new understanding to improve management of symptomatic patients*. *The Lancet Oncology*, 2007. **8**(11): p. 1007-1017.
24. Vanneste, B.G., et al., *Chronic radiation proctitis: tricks to prevent and treat*. *Int J Colorectal Dis*, 2015. **30**(10): p. 1293-303.
25. Measurements, I.C.o.R.U.a. *Prescribing, Recording and Reporting Photon Beam Therapy (Report 62) October 2020*]; Available from: <https://www.icru.org/home/reports/prescribing-recording-and-reporting-photon-beam-therapy-report-62>.
26. K. M. Langen, P.D., and D. T. L. Jones, PH.D., *Organ Motion and its management* *International Journal of Radiation Oncology*Biophysics*, 2001. **50**(1): p. 265-275.
27. Chen, S., et al., *Feasibility of CBCT-based dose with a patient-specific stepwise HU-to-density curve to determine time of replanning*. *J Appl Clin Med Phys*, 2017. **18**(5): p. 64-69.
28. Gulliford, S.L., et al., *Dose-volume constraints to reduce rectal side effects from prostate radiotherapy: evidence from MRC RT01 Trial ISRCTN 47772397*. *Int J Radiat Oncol Biol Phys*, 2010. **76**(3): p. 747-54.
29. Jackson, A., et al., *The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome*. *Int J Radiat Oncol Biol Phys*, 2010. **76**(3 Suppl): p. S155-60.
30. Valdagni, R. and T. Rancati, *Reducing rectal injury during external beam radiotherapy for prostate cancer*. *Nat Rev Urol*, 2013. **10**(6): p. 345-57.
31. Chen, L., et al., *Rectal dose variation during the course of image-guided radiation therapy of prostate cancer*. *Radiother Oncol*, 2010. **95**(2): p. 198-202.
32. Hatton, J.A., et al., *Does the planning dose-volume histogram represent treatment doses in image-guided prostate radiation therapy? Assessment with cone-beam computerised tomography scans*. *Radiother Oncol*, 2011. **98**(2): p. 162-8.
33. Kuban, D.A., et al., *Long-term failure patterns and survival in a randomized dose-escalation trial for prostate cancer. Who dies of disease?* *Int J Radiat Oncol Biol Phys*, 2011. **79**(5): p. 1310-7.
34. Zaorsky, N.G., et al., *What is the ideal radiotherapy dose to treat prostate cancer? A meta-analysis of biologically equivalent dose escalation*. *Radiother Oncol*, 2015. **115**(3): p. 295-300.
35. Zaorsky, N.G., et al., *Impact of Radiation Therapy Dose Escalation on Prostate Cancer Outcomes and Toxicities*. *Am J Clin Oncol*, 2018. **41**(4): p. 409-415.
36. Dearnaley, D.P., et al., *Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial*. *The Lancet*, 1999. **353**(9149): p. 267-272.

37. Al-Mamgani, A., et al., *Role of intensity-modulated radiotherapy in reducing toxicity in dose escalation for localized prostate cancer*. Int J Radiat Oncol Biol Phys, 2009. **73**(3): p. 685-91.
38. Tree, A.C., et al., *Biological dose escalation and hypofractionation: what is there to be gained and how will it best be done?* Clin Oncol (R Coll Radiol), 2013. **25**(8): p. 483-98.
39. Arrayeh, E., et al., *Does local recurrence of prostate cancer after radiation therapy occur at the site of primary tumor? Results of a longitudinal MRI and MRSI study*. Int J Radiat Oncol Biol Phys, 2012. **82**(5): p. e787-93.
40. van Lin, E.N., et al., *IMRT boost dose planning on dominant intraprostatic lesions: gold marker-based three-dimensional fusion of CT with dynamic contrast-enhanced and 1H-spectroscopic MRI*. Int J Radiat Oncol Biol Phys, 2006. **65**(1): p. 291-303.
41. Nutting, C.M., et al., *Potential improvements in the therapeutic ratio of prostate cancer irradiation: dose escalation of pathologically identified tumour nodules using intensity modulated radiotherapy*. BJ Radiology 2002. **75**: p. 151-161.
42. Murray, J.R., et al., *Standard and hypofractionated dose escalation to intraprostatic tumour nodules in localised prostate cancer: efficacy and toxicity in the DELINEATE trial*. Int J Radiat Oncol Biol Phys, 2019.
43. Monninkhof, E.M., et al., *Standard whole prostate gland radiotherapy with and without lesion boost in prostate cancer: Toxicity in the FLAME randomized controlled trial*. Radiother Oncol, 2018. **127**(1): p. 74-80.
44. Angela U Pathmanathan, E.J.A., Robert A Huddart & Alison C Tree, *The delineation of intraprostatic boost regions for radiotherapy using multimodality imaging*. 2016. **12**(21): p. 2495-2511.
45. Benjamin, L.C., A.C. Tree, and D.P. Dearnaley, *The Role of Hypofractionated Radiotherapy in Prostate Cancer*. Curr Oncol Rep, 2017. **19**(4): p. 30.
46. Vogelius, I.R. and S.M. Bentzen, *Meta-analysis of the alpha/beta ratio for prostate cancer in the presence of an overall time factor: bad news, good news, or no news?* Int J Radiat Oncol Biol Phys, 2013. **85**(1): p. 89-94.
47. Dasu, A., *Is the alpha/beta value for prostate tumours low enough to be safely used in clinical trials?* Clin Oncol (R Coll Radiol), 2007. **19**(5): p. 289-301.
48. van Leeuwen, C.M., et al., *The alfa and beta of tumours: a review of parameters of the linear-quadratic model, derived from clinical radiotherapy studies*. Radiat Oncol, 2018. **13**(1): p. 96.
49. Catton, C.N., et al., *Randomized Trial of a Hypofractionated Radiation Regimen for the Treatment of Localized Prostate Cancer*. Journal of Clinical Oncology, 2017. **35**(17): p. 1884-1890.
50. Lee, W.R., et al., *Randomized Phase III Noninferiority Study Comparing Two Radiotherapy Fractionation Schedules in Patients With Low-Risk Prostate Cancer*. J Clin Oncol, 2016. **34**(20): p. 2325-32.
51. Incrocci, L., et al., *Hypofractionated versus conventionally fractionated radiotherapy for patients with localised prostate cancer (HYPRO): final efficacy results from a randomised, multicentre, open-label, phase 3 trial*. The Lancet Oncology, 2016. **17**(8): p. 1061-1069.
52. Dearnaley, D., et al., *Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial*. The Lancet Oncology, 2016.

53. Lardas, M., et al., *Quality of Life Outcomes after Primary Treatment for Clinically Localised Prostate Cancer: A Systematic Review*. Eur Urol, 2017. **72**(6): p. 869-885.
54. Donovan, J.L., et al., *Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer*. N Engl J Med, 2016. **375**(15): p. 1425-1437.
55. Hamdy, F.C., et al., *10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer*. N Engl J Med, 2016. **375**(15): p. 1415-1424.
56. Gulliford, S.L., et al., *A comparison of dose-volume constraints derived using peak and longitudinal definitions of late rectal toxicity*. Radiother Oncol, 2010. **94**(2): p. 241-7.
57. U.S. department of health and human services NIOH. National Cancer Institute. *Common Terminology Criteria for Adverse Events (CTCAE) 2010*.
58. *LENT SOMA tables of contents*. Radiotherapy and Oncology, 1995. **35**: p. 17-60.
59. Litwin, M.S., J.M. Henning, and P.R. Carroll, *Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: results of the CAPSURE database*. The Journal of Urology, 1998. **159**: p. 1988-1992.
60. Guyatt, G., et al., *A new measure of health status for clinical trials in inflammatory bowel disease*. Gastroenterology, 1989. **96**(3): p. 804-10.
61. Vaizey, C.J., et al., *Prospective comparison of faecal incontinence grading systems*. Gut, 1999. **44**(1): p. 77-80.
62. Szymanski, K.M., et al., *Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors*. Urology, 2010. **76**(5): p. 1245-50.
63. Skwarchuck, M.W.J., A; Zelefsky, M.J; Venkatraman, E.S; Cowen, D.M.; Levegrun, S., *Late rectal toxicity after conformal radiotherapy of prostate cancer: Multivariate analysis and dose-response*. International Journal of Radiation Oncology*Biology*Physics, 2000. **47**(1): p. 103-113.
64. Hamstra, D.A., et al., *Age and comorbid illness are associated with late rectal toxicity following dose-escalated radiation therapy for prostate cancer*. Int J Radiat Oncol Biol Phys, 2013. **85**(5): p. 1246-53.
65. Fiorino, C., et al., *Dose-volume effects for normal tissues in external radiotherapy: pelvis*. Radiother Oncol, 2009. **93**(2): p. 153-67.
66. Defraene, G., et al., *The benefits of including clinical factors in rectal normal tissue complication probability modeling after radiotherapy for prostate cancer*. Int J Radiat Oncol Biol Phys, 2012. **82**(3): p. 1233-42.
67. Valdagni, R., et al., *Increasing the risk of late rectal bleeding after high-dose radiotherapy for prostate cancer: the case of previous abdominal surgery. Results from a prospective trial*. Radiother Oncol, 2012. **103**(2): p. 252-5.
68. Ebert, M.A., et al., *Gastrointestinal dose-histogram effects in the context of dose-volume-constrained prostate radiation therapy: analysis of data from the RADAR prostate radiation therapy trial*. Int J Radiat Oncol Biol Phys, 2015. **91**(3): p. 595-603.
69. Alsadius, D., et al., *Mean absorbed dose to the anal-sphincter region and fecal leakage among irradiated prostate cancer survivors*. Int J Radiat Oncol Biol Phys, 2012. **84**(2): p. e181-5.

70. Wilkins, A., et al., *Derivation of dose/volume constraints for the anorectum from clinician and patient-reported outcomes in the CHHiP trial of radiotherapy fractionation*. Int J Radiat Oncol Biol Phys, 2020.
71. Coates, J. and I. El Naqa, *Outcome modeling techniques for prostate cancer radiotherapy: Data, models, and validation*. Phys Med, 2016. **32**(3): p. 512-20.
72. Tucker, S.L., et al., *Late rectal toxicity on RTOG 94-06: analysis using a mixture Lyman model*. Int J Radiat Oncol Biol Phys, 2010. **78**(4): p. 1253-60.
73. Drzymala, R.E., et al., *Dose Volume Histograms International Journal of Radiation Oncology*Biology*Physics*, 1991. **21**: p. 71-78.
74. Emami, B., et al., *Tolerance of normal tissue of therapeutic irradiation International Journal of Radiation Oncology*Biology*Physics*, 1991. **21**.
75. Fellin, G., et al., *Clinical and dosimetric predictors of late rectal toxicity after conformal radiation for localized prostate cancer: results of a large multicenter observational study*. Radiother Oncol, 2009. **93**(2): p. 197-202.
76. Matta, R., et al., *Pelvic Complications After Prostate Cancer Radiation Therapy and Their Management: An International Collaborative Narrative Review*. Eur Urol, 2019. **75**(3): p. 464-476.
77. Lam, T.J., D.J. Kuik, and R.J. Felt-Bersma, *Anorectal function evaluation and predictive factors for faecal incontinence in 600 patients*. Colorectal Dis, 2012. **14**(2): p. 214-23.
78. Smeenk, R.J., et al., *Differences in radiation dosimetry and anorectal function testing imply that anorectal symptoms may arise from different anatomic substrates*. Int J Radiat Oncol Biol Phys, 2012. **82**(1): p. 145-52.
79. Smeenk, R.J., et al., *Dose-effect relationships for individual pelvic floor muscles and anorectal complaints after prostate radiotherapy*. Int J Radiat Oncol Biol Phys, 2012. **83**(2): p. 636-44.
80. Dearnaley, D.P., et al., *Escalated-dose versus standard-dose conformal radiotherapy in prostate cancer: first results from the MRC RT01 randomised controlled trial*. The Lancet Oncology, 2007. **8**(6): p. 475-487.
81. Miralbell, R., et al., *Influence of rectal volume changes during radiotherapy for prostate cancer: A predictive model for mild-to-moderate late rectal toxicity*. International Journal of Radiation Oncology*Biology*Physics, 2003. **57**(5): p. 1280-1284.
82. Marks, L.B., et al., *Use of normal tissue complication probability models in the clinic*. Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S10-9.
83. Wheldon, T.E., Deehanb, C., Wheldon, E., Barrett, A., *The linear-quadratic transformation of dose-volume histograms in fractionated radiotherapy*. Radiat Oncol, 1998. **46**: p. 285-295.
84. Tucker, S.L., et al., *Comparison of rectal dose-wall histogram versus dose-volume histogram for modeling the incidence of late rectal bleeding after radiotherapy*. Int J Radiat Oncol Biol Phys, 2004. **60**(5): p. 1589-601.
85. Desrochers, A., et al., *Comparison of dose statistics for bladder wall and rectum wall vs whole organs for VMAT prostate treatment*. Med Dosim, 2019.
86. Meijer, G.J., et al., *Dose-wall histograms for the rectum: a new method to analyse the dose distribution over the rectum in conformal radiotherapy*. Int J Radiat Oncol Biol Phys, 1999. **45**(4): p. 1073-1080.
87. de Crevoisier, R., et al., *Increased risk of biochemical and local failure in patients with distended rectum on the planning CT for prostate cancer radiotherapy*. Int J Radiat Oncol Biol Phys, 2005. **62**(4): p. 965-73.

88. Lu, Y., et al., *A method of analysing rectal surface area irradiated and rectal complications in prostate conformal radiotherapy*. Int J Radiat Oncol Biol Phys, 1995. **33**(5): p. 1121-1125.
89. Casares-Magaz, O., et al., *Towards spatial representations of dose distributions to predict risk of normal tissue morbidity after radiotherapy*. Physics and Imaging in Radiation Oncology, 2020. **15**: p. 105-107.
90. Buettner, F., et al., *Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution*. Phys Med Biol, 2011. **56**(7): p. 2103-18.
91. Murray, J., et al., *A novel Approach to Evaluate the Dosimetric Effect of Rectal Variation During Image Guided Prostate Radiotherapy*. Medical Physics, 2014. **41**(157).
92. Scaife, J.E., et al., *Accumulated dose to the rectum, measured using dose-volume histograms and dose-surface maps, is different from planned dose in all patients treated with radiotherapy for prostate cancer*. Br J Radiol, 2015: p. 20150243.
93. Shelley, L.E.A., et al., *Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy*. Radiother Oncol, 2017. **123**(3): p. 466-471.
94. Buettner, F., et al., *Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach*. Phys Med Biol, 2009. **54**(17): p. 5139-53.
95. Onjukka, E., et al., *Patterns in ano-rectal dose maps and the risk of late toxicity after prostate IMRT*. Acta Oncol, 2019. **58**(12): p. 1757-1764.
96. Dang, A., et al., *Image-guided radiotherapy for prostate cancer*. Transl Androl Urol, 2018. **7**(3): p. 308-320.
97. McNair, H.A., et al., *A systematic review: effectiveness of rectal emptying preparation in prostate cancer patients*. Pract Radiat Oncol, 2014. **4**(6): p. 437-47.
98. Dawson, L.A., et al., *Target variability throughout prostate radiotherapy*. Int J Radiat Oncol Biol Phys, 1998. **42**(5): p. 1155-1161.
99. Roeske, J.C., et al., *Evaluation of changes in the size and location of the prostate, seminal vesicles, bladder, and rectum during a course of external beam radiation therapy*. Int J Radiat Oncol Biol Phys, 1995. **33**(5): p. 1321-9.
100. Langen, K.M., et al., *Observations on real-time prostate gland motion using electromagnetic tracking*. Int J Radiat Oncol Biol Phys, 2008. **71**(4): p. 1084-90.
101. Kupelian, P., et al., *Multi-institutional clinical experience with the Calypso System in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy*. Int J Radiat Oncol Biol Phys, 2007. **67**(4): p. 1088-98.
102. McPartlin, A.J., et al., *MRI-guided prostate adaptive radiotherapy - A systematic review*. Radiother Oncol, 2016. **119**(3): p. 371-80.
103. Hamamoto Y, I.H., Sodeoka N, Nakayama S, Tsuruoka S, Takeda H, Manabe T, Mochizuki T, Umeda M. , *Observation of intrafraction prostate displacement through the course of conventionally fractionated radiotherapy for prostate cancer*. Jpn J Radiol, 2015: p. 187-193.
104. Shelton, J., et al., *Observations on prostate intrafraction motion and the effect of reduced treatment time using volumetric modulated arc therapy*. Pract Radiat Oncol, 2011. **1**(4): p. 243-50.

105. Sripadam, R., et al., *Rectal motion can reduce CTV coverage and increase rectal dose during prostate radiotherapy: A daily cone-beam CT study.* Radiother Oncol, 2009. **90**(3): p. 312-7.
106. Lips, I.M., et al., *A double-blind placebo-controlled randomized clinical trial with magnesium oxide to reduce intrafraction prostate motion for prostate cancer radiotherapy.* Int J Radiat Oncol Biol Phys, 2012. **83**(2): p. 653-60.
107. Oates, R.W., et al., *A randomised study of a diet intervention to maintain consistent rectal volume for patients receiving radical radiotherapy to the prostate.* Acta Oncol, 2014. **53**(4): p. 569-71.
108. Yahya, S., et al., *Which bowel preparation is best? Comparison of a high-fibre diet leaflet, daily microenema and no preparation in prostate cancer patients treated with radical radiotherapy to assess the effect on planned target volume shifts due to rectal distension.* Br J Radiol, 2013. **86**(1031): p. 20130457.
109. McNair, H.A., et al., *Can diet combined with treatment scheduling achieve consistency of rectal filling in patients receiving radiotherapy to the prostate?* Radiother Oncol, 2011. **101**(3): p. 471-8.
110. Vanneste, B.G.L., et al., *Ano-rectal wall dose-surface maps localize the dosimetric benefit of hydrogel rectum spacers in prostate cancer radiotherapy.* Clinical and Translational Radiation Oncology, 2019. **14**: p. 17-24.
111. Smeenk, R.J., et al., *An endorectal balloon reduces intrafraction prostate motion during radiotherapy.* Int J Radiat Oncol Biol Phys, 2012. **83**(2): p. 661-9.
112. Buettner, F., et al., *A Novel Rectal Obturator for Prostate Radiotherapy Improves the Spatial Distribution of Dose and Reduces the Predicted Risk for Rectal Bleeding and Subjective Sphincter Control, in American association of physicist in medicine.* 2012, Medical Physics. p. 3762-3762.
113. Ranjani Padmanabhan, M.P.D.Y.S., *Hydrogel spacers in prostate radiotherapy: a promising approach to decrease rectal toxicity.* Future Oncology, 2017. **13**(29): p. 2697-2708.
114. Hamstra, D.A., et al., *Continued Benefit to Rectal Separation for Prostate Radiation Therapy: Final Results of a Phase III Trial.* International Journal of Radiation Oncology*Biography*Physics, 2017. **97**(5): p. 976-985.
115. Kupelian, P.A., et al., *Daily variations in delivered doses in patients treated with radiotherapy for localized prostate cancer.* Int J Radiat Oncol Biol Phys, 2006. **66**(3): p. 876-82.
116. The Royal College of Radiologists, I.o.P.a.E.i.M., Society and College of Radiographers, *On Target.* 2013.
117. Kuban, D.A., et al., *Long-term results of the M. D. Anderson randomized dose-escalation trial for prostate cancer.* Int J Radiat Oncol Biol Phys, 2008. **70**(1): p. 67-74.
118. Eade, T.N., et al., *What dose of external-beam radiation is high enough for prostate cancer?* Int J Radiat Oncol Biol Phys, 2007. **68**(3): p. 682-9.
119. van Herk, M., *Different styles of image-guided radiotherapy.* Semin Radiat Oncol, 2007. **17**(4): p. 258-67.
120. Murray, J., et al., *A randomised assessment of image guided radiotherapy within a phase 3 trial of conventional or hypofractionated high dose intensity modulated radiotherapy for prostate cancer.* Radiother Oncol, 2020. **142**: p. 62-71.
121. McNair, H.A., et al., *A comparison of the use of bony anatomy and internal markers for offline verification and an evaluation of the potential benefit of online and offline verification protocols for prostate radiotherapy.* Int J Radiat Oncol Biol Phys, 2008. **71**(1): p. 41-50.

122. Pawlowski, J.M., et al., *Reduction of dose delivered to organs at risk in prostate cancer patients via image-guided radiation therapy*. Int J Radiat Oncol Biol Phys, 2010. **76**(3): p. 924-34.
123. Owen, R., et al., *Comparison of CT on rails with electronic portal imaging for positioning of prostate cancer patients with implanted fiducial markers*. Int J Radiat Oncol Biol Phys, 2009. **74**(3): p. 906-12.
124. Richardson, A.K. and P. Jacobs, *Intrafraction monitoring of prostate motion during radiotherapy using the Clarity((R)) Autoscan Transperineal Ultrasound (TPUS) system*. Radiography (Lond), 2017. **23**(4): p. 310-313.
125. Dunlop, A., et al., *Comparison of CT number calibration techniques for CBCT-based dose calculation*. Strahlenther Onkol, 2015. **191**(12): p. 970-978.
126. Richter, A., et al., *Investigation of the usability of conebeam CT data sets for dose calculation*. Radiat Oncol, 2008. **3**: p. 42.
127. Fotina, I., et al., *Feasibility of CBCT-based dose calculation: comparative analysis of HU adjustment techniques*. Radiother Oncol, 2012. **104**(2): p. 249-56.
128. Simon, A., et al., *Roles of Deformable Image Registration in Adaptive RT: from contour Propagation to Dose Monitoring*, in *Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE*. 2015: Milan. p. 5215-5218.
129. Rigaud, B., et al., *Deformable image registration for radiation therapy: principle, methods, applications and evaluation*. Acta Oncol, 2019. **58**(9): p. 1225-1237.
130. Collery, A.F., E. , *Daily Rectal Dose-volume Histogram Variation in Prostate Intensity-modulated Radiation Therapy: Is It Clinically Significant in the Era of Image Guidance?* Journal of Medical Imaging and Radiation Sciences 2017. **48**(4): p. 346-351.
131. Chen, Z., et al., *Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy*. Radiat Oncol, 2016. **11**: p. 103.
132. Huang, T.C., et al., *Fractionated changes in prostate cancer radiotherapy using cone-beam computed tomography*. Med Dosim, 2015. **40**(3): p. 222-5.
133. Abe, T., et al., *Assessing cumulative dose distributions in combined radiotherapy for cervical cancer using deformable image registration with pre-imaging preparations*. Radiat Oncol, 2014. **9**: p. 293.
134. Jaffray, D.A., et al., *Accurate accumulation of dose for improved understanding of radiation effects in normal tissue*. Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S135-9.
135. *Image Processing in Radiation Therapy 2014*, Boca Raton, Florida CRC Press Taylor and Francis Group
136. Ciarmatori, A., et al., *Evaluation of the effectiveness of novel single-intervention adaptive radiotherapy strategies based on daily dose accumulation*. Med Dosim, 2019. **44**(4): p. 379-384.
137. Ma, C., et al., *A study of the anatomic changes and dosimetric consequences in adaptive CRT of non-small-cell lung cancer using deformable CT and CBCT image registration*. Technol Cancer Res Treat, 2014. **13**(2): p. 95-100.
138. Bostel, T., et al., *Dosimetric Impact of Interfractional Variations in Prostate Cancer Radiotherapy-Implications for Imaging Frequency and Treatment Adaptation*. Front Oncol, 2019. **9**: p. 940.
139. Yu, J., et al., *On voxel-by-voxel accumulated dose for prostate radiation therapy using deformable image registration*. Technol Cancer Res Treat, 2015. **14**(1): p. 37-47.

140. Wen, N., et al., *Evaluation of the deformation and corresponding dosimetric implications in prostate cancer treatment*. Phys Med Biol, 2012. **57**(17): p. 5361-79.
141. Crum, W.R., T. Hartkens, and D.L. Hill, *Non-rigid image registration: theory and practice*. Br J Radiol, 2004. **77 Spec No 2**: p. S140-53.
142. Kirtsy Brock, S.M., Todd R. McNutt, Hua LI and Marc L. Kessler *Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Medical Physics 2017*. **44**(7).
143. Chetty, I.J. and M. Rosu-Bubulac, *Deformable Registration for Dose Accumulation*. Semin Radiat Oncol, 2019. **29**(3): p. 198-208.
144. Whitfield, G.A., et al., *Automated delineation of radiotherapy volumes: are we going in the right direction?* Br J Radiol, 2013. **86**(1021): p. 20110718.
145. Thornqvist, S., et al., *Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration*. Acta Oncol, 2010. **49**(7): p. 1023-32.
146. Weistrand, O. and S. Svensson, *The ANACONDA algorithm for deformable image registration in radiotherapy*. Med Phys, 2015. **42**(1): p. 40-53.
147. Takayama, Y., et al., *Evaluation of the performance of deformable image registration between planning CT and CBCT images for the pelvic region: comparison between hybrid and intensity-based DIR*. J Radiat Res, 2017. **58**(4): p. 567-571.
148. Nassef, M., et al., *Quantification of dose uncertainties in cumulated dose estimation compared to planned dose in prostate IMRT*. Radiother Oncol, 2016. **119**(1): p. 129-36.
149. Buettner, F., et al., *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)*. Phys Med Biol, 2009. **54**(21): p. 6535-48.
150. Hoogeman, M.S., et al., *Quantification of local rectal wall displacements by virtual rectum unfolding*. Radiother Oncol, 2004. **70**(1): p. 21-30.
151. Dearnaley, D., et al., *Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: preliminary safety results from the CHHiP randomised controlled trial*. Lancet Oncol, 2012. **13**(1): p. 43-54.
152. Peeters, S.T., et al., *Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling*. Int J Radiat Oncol Biol Phys, 2006. **66**(1): p. 11-9.
153. Litwin MS, H.R., Fink A, Ganz PA, Leake B, Brook RH. , *The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure*. Medical Care, 1998. **36**(7): p. 1002-1012.
154. Sonn, G.A., et al., *Differing Perceptions of Quality of Life in Patients With Prostate Cancer and Their Doctors*. Journal of Urology, 2009. **182**(5): p. 2296-2302.
155. Pilepich, M.V., et al., *Correlation of Radiotherapeutic parameters and treatment related morbidity-analysis of RTOG study 77-06*. Int J Radiat Oncol Biol Phys, 1987. **13**: p. 1007 - 10012.
156. H. Rodney Withers , H.D.T., Jr. and Lester J. Peters *A new isoeffect curve for change in dose per fraction*. Radiat Oncol. **1**: p. 187-191.
157. <RSL-D-RS-6.0-USM-EN-1.0-2016-12-22 RayStation 6 User Manual.pdf>.
158. Laboratories, R., *RayStation User Manual 4.5*. 2014: Sweden.

159. Fiorino, C., et al., *Rectum contouring variability in patients treated for prostate cancer: impact on rectum dose-volume histograms and normal tissue complication probability*. *Radiother Oncol*, 2002. **63**(3): p. 249-55.
160. Onal, C., et al., *Comparison of rectal volume definition techniques and their influence on rectal toxicity in patients with prostate cancer treated with 3D conformal radiotherapy: a dose-volume analysis*. *Radiat Oncol*, 2009. **4**: p. 14.
161. Rasch, C., et al., *Definition of the prostate in CT and MRI: a multi-observer study*. *Int J Radiat Oncol Biol Phys*, 1999. **43**(1): p. 57-66.
162. Jameson, M.G., et al., *A review of methods of analysis in contouring studies for radiation oncology*. *J Med Imaging Radiat Oncol*, 2010. **54**(5): p. 401-10.
163. Mason, S.A., et al., *Towards ultrasound-guided adaptive radiotherapy for cervical cancer: Evaluation of Elekta's semiautomated uterine segmentation method on 3D ultrasound images*. *Med Phys*, 2017. **44**(7): p. 3630-3638.
164. Zhang, L., et al., *The impact of robustness of deformable image registration on contour propagation and dose accumulation for head and neck adaptive radiotherapy*. *J Appl Clin Med Phys*, 2018. **19**(4): p. 185-194.
165. Motegi, K., et al., *Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy*. *J Appl Clin Med Phys*, 2019. **20**(1): p. 229-236.
166. Kadoya, N., et al., *Multi-institutional Validation Study of Commercially Available Deformable Image Registration Software for Thoracic Images*. *Int J Radiat Oncol Biol Phys*, 2016. **96**(2): p. 422-431.
167. Heemsbergen, W.D., et al., *Gastrointestinal toxicity and its relation to dose distributions in the anorectal region of prostate cancer patients treated with radiotherapy*. *Int J Radiat Oncol Biol Phys*, 2005. **61**(4): p. 1011-8.
168. Murray, J., *The evaluation of toxicity modification using image-guided radiotherapy delivery in the treatment of prostate cancer*, in *Department of Radiotherapy and Imaging 2017*, University of London The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.
169. Tucker, S.L., et al., *Cluster model analysis of late rectal bleeding after IMRT of prostate cancer: A case-control study*. *International Journal of Radiation Oncology*Biology*Physics*, 2006. **64**(4): p. 1255-1264.
170. Otter, S., et al., *Evaluation of the Risk of Grade 3 Oral and Pharyngeal Dysphagia Using Atlas-Based Method and Multivariate Analyses of Individual Patient Dose Distributions*. *Int J Radiat Oncol Biol Phys*, 2015. **93**(3): p. 507-15.
171. Murray, J., et al., *Effect of Dose and Image Guided Radiation Therapy (IGRT) on Patient-Reported Sexual Function in Prostate Radiation Therapy*. *International Journal of Radiation Oncology*Biology*Physics*, 2016. **96**(2): p. E232.
172. Vining, D.J. and G.W. Gladish, *Receiver operating characteristic curves: a basic understanding*. *Radiographics*, 1992. **12**(6): p. 1147-54.
173. Youden, W.J., *Index for rating diagnostic tests*. *Cancer*, 1950. **3**(1): p. 32-35.
174. Jackson, A., E.D. Yorke, and K.E. Rosenzweig, *The atlas of complication incidence: a proposal for a new standard for reporting the results of radiotherapy protocols*. *Semin Radiat Oncol*, 2006. **16**(4): p. 260-8.
175. Gulliford, S.L., et al., *Dosimetric explanations of fatigue in head and neck radiotherapy: an analysis from the PARSPORT Phase III trial*. *Radiother Oncol*, 2012. **104**(2): p. 205-12.
176. Ghasemi, A. and S. Zahediasl, *Normality tests for statistical analysis: a guide for non-statisticians*. *Int J Endocrinol Metab*, 2012. **10**(2): p. 486-9.

177. Sedgwick, P., *Multiple hypothesis testing and Bonferroni's correction*. BMJ, 2014. **349**: p. g6284.
178. Holm, S., *A simple sequential rejective multiple test procedure*. Scandinavian Journal of Statistics, 1979. **6**: p. 65-70.
179. Sharp, G., et al., *Vision 20/20: perspectives on automated image segmentation for radiotherapy*. Med Phys, 2014. **41**(5): p. 050902.
180. Vinod, S.K., et al., *Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies*. Radiother Oncol, 2016. **121**(2): p. 169-179.
181. Gao, Z., et al., *A study of prostate delineation referenced against a gold standard created from the visible human data*. Radiother Oncol, 2007. **85**(2): p. 239-46.
182. Livsey, J.E., et al., *Do differences in target volume definition in prostate cancer lead to clinically relevant differences in normal tissue toxicity?* Int J Radiat Oncol Biol Phys, 2004. **60**(4): p. 1076-81.
183. Foppiano, F., et al., *The impact of contouring uncertainty on rectal 3D dose-volume data: Results of a dummy run in a multicenter trial (AIROPROS01-02)*. International Journal of Radiation Oncology*Biography*Physics, 2003. **57**(2): p. 573-579.
184. Batumalai, V., et al., *Interobserver variability in clinical target volume delineation in tangential breast irradiation: a comparison between radiation oncologists and radiation therapists*. Clin Oncol (R Coll Radiol), 2011. **23**(2): p. 108-13.
185. Breunig, J., et al., *A system for continual quality improvement of normal tissue delineation for radiation therapy treatment planning*. Int J Radiat Oncol Biol Phys, 2012. **83**(5): p. e703-8.
186. Gay, H.A., et al., *Pelvic normal tissue contouring guidelines for radiation therapy: a Radiation Therapy Oncology Group consensus panel atlas*. Int J Radiat Oncol Biol Phys, 2012. **83**(3): p. e353-62.
187. Rasch, C., et al., *Comparison of prostate cancer treatment in two institutions: a quality control study*. Int J Radiat Oncol Biol Phys, 1999. **45**(4): p. 1055-62.
188. Weiss, E., et al., *Clinical evaluation of soft tissue organ boundary visualization on cone-beam computed tomographic imaging*. Int J Radiat Oncol Biol Phys, 2010. **78**(3): p. 929-36.
189. McLaughlin, P.W., et al., *Radiographic and anatomic basis for prostate contouring errors and methods to improve prostate contouring accuracy*. Int J Radiat Oncol Biol Phys, 2010. **76**(2): p. 369-78.
190. Seddon, B., Bidmead, M., Wilson, J., Khoo, V. and Dearnaley, D. , *Target volume definition in conformal radiotherapy for prostate cancer: quality assurance in the MRC RT-01 trial*. Radiat Oncol, 2000. **56**(1): p. 73-83.
191. White, E.A., et al., *Inter-observer variability of prostate delineation on cone beam computerised tomography images*. Clin Oncol (R Coll Radiol), 2009. **21**(1): p. 32-8.
192. Sharma, M., E. Weiss, and J.V. Siebers, *Dose deformation-invariance in adaptive prostate radiation therapy: implication for treatment simulations*. Radiother Oncol, 2012. **105**(2): p. 207-13.
193. Sonke, J.J., M. Aznar, and C. Rasch, *Adaptive Radiotherapy for Anatomical Changes*. Semin Radiat Oncol, 2019. **29**(3): p. 245-257.
194. Riegel, A.A., et al., *Deformable image registration and interobserver variation in contour propagation for radiation therapy planning*. Journal of applied clinical medical physics, 2016. **17**(347 - 367).

195. Bodensteiner, D., *RayStation: External beam treatment planning system*. Med Dosim, 2018. **43**(2): p. 168-176.
196. Egekvist, A.G., et al., *Intra- and interobserver variability in nodule size of rectosigmoid endometriosis measured by two- and three-dimensional transvaginal sonography*. Acta Obstet Gynecol Scand, 2018. **97**(6): p. 734-743.
197. Delpon, G., et al., *Comparison of Automated Atlas-Based Segmentation Software for Postoperative Prostate Cancer Radiotherapy*. Front Oncol, 2016. **6**: p. 178.
198. Gardner, S.J., et al., *Contouring variability of human- and deformable-generated contours in radiotherapy for prostate cancer*. Phys Med Biol, 2015. **60**(11): p. 4429-47.
199. Thor, M., et al., *Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer*. Acta Oncol, 2011. **50**(6): p. 918-25.
200. Gregoire, V., et al., *Image guidance in radiation therapy for better cure of cancer*. Mol Oncol, 2020. **14**(7): p. 1470-1491.
201. Scaife, J., et al., *Random variation in rectal position during radiotherapy for prostate cancer is two to three times greater than that predicted from prostate motion*. Br J Radiol, 2014. **87**(1042): p. 20140343.
202. *The Ionising Radiation (Medical Exposure) Regulations 2017*. 2017 October 2020]; Available from: <https://www.legislation.gov.uk/ukxi/2017/1322/contents/made>.
203. Oh, S. and S. Kim, *Deformable image registration in radiation therapy*. Radiat Oncol J, 2017. **35**(2): p. 101-111.
204. Brock, K.K. and C. Deformable Registration Accuracy, *Results of a multi-institution deformable registration accuracy study (MIDRAS)*. Int J Radiat Oncol Biol Phys, 2010. **76**(2): p. 583-96.
205. Kashani, R., et al., *Objective assessment of deformable image registration in radiotherapy: a multi-institution study*. Med Phys, 2008. **35**(12): p. 5944-53.
206. Chong, I., et al., *Quantification of organ motion during chemoradiotherapy of rectal cancer using cone-beam computed tomography*. Int J Radiat Oncol Biol Phys, 2011. **81**(4): p. e431-8.
207. Davis, K.M., et al., *The association of long-term treatment-related side effects with cancer-specific and general quality of life among prostate cancer survivors*. Urology, 2014. **84**(2): p. 300-6.
208. Yamazaki, H., et al., *Transitioning from conventional radiotherapy to intensity-modulated radiotherapy for localized prostate cancer: changing focus from rectal bleeding to detailed quality of life analysis*. J Radiat Res, 2014. **55**(6): p. 1033-47.
209. Fonteyne, V., et al., *Impact of changing rectal dose volume parameters over time on late rectal and urinary toxicity after high-dose intensity-modulated radiotherapy for prostate cancer: A 10-years single centre experience*. Acta Oncol, 2015. **54**(6): p. 854-61.
210. Fonteyne, V., et al., *Rectal toxicity after intensity modulated radiotherapy for prostate cancer: which rectal dose volume constraints should we use?* Radiother Oncol, 2014. **113**(3): p. 398-403.
211. Bentzen, S.M., et al., *Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues*. Int J Radiat Oncol Biol Phys, 2010. **76**(3 Suppl): p. S3-9.

212. Faria, S., et al., *Searching for optimal dose-volume constraints to reduce rectal toxicity after hypofractionated radiotherapy for prostate cancer*. Clin Oncol (R Coll Radiol), 2010. **22**(10): p. 810-7.
213. Radiation Therapy Oncology Group. *RTOG 0126. A phase III randomized study of high dose 3D-CRT/IMRT versus standard dose 3D-CRT/IMRT in patients treated for localized prostate cancer* [21st August 2020]; Available from: www.rtog.org.
214. Radiation Therapy Oncology Group. *RTOG 0415. A phase III randomized study of hypofractionated 3D-CRT/IMRT versus conventionally fractionated 3D-CRT/IMRT in patients with favorable-risk prostate cancer*. Accessed 21st August 2020]; Available from: www.rtog.com.
215. PROFIT. *Canadian prostate fractionated irradiation trial (PROFIT)*. [cited 2020 21st August]; Available from: <http://clinicaltrials.gov/ct/show/NCT00304759?order=1/411>.
216. Peng, C., et al., *Characterizing interfraction variations and their dosimetric effects in prostate cancer radiotherapy*. Int J Radiat Oncol Biol Phys, 2011. **79**(3): p. 909-14.
217. Akino, Y., et al., *Estimation of rectal dose using daily megavoltage cone-beam computed tomography and deformable image registration*. Int J Radiat Oncol Biol Phys, 2013. **87**(3): p. 602-8.
218. Sedgwick, P., *Pearson's correlation coefficient*. Bmj, 2012. **345**(jul04 1): p. e4483-e4483.
219. Akoglu, H., *User's guide to correlation coefficients*. Turk J Emerg Med, 2018. **18**(3): p. 91-93.
220. Engels, B., et al., *Assessment of rectal distention in radiotherapy of prostate cancer using daily megavoltage CT image guidance*. Radiother Oncol, 2009. **90**(3): p. 377-81.
221. Ariyaratne, H., et al., *Image-guided radiotherapy for prostate cancer with cone beam CT: dosimetric effects of imaging frequency and PTV margin*. Radiotherapy and Oncology, 2016. **121**(1): p. 103-108.
222. van der Wielen, G.J., et al., *Deformation of prostate and seminal vesicles relative to intraprostatic fiducial markers*. Int J Radiat Oncol Biol Phys, 2008. **72**(5): p. 1604-1611 e3.
223. Yahya, N., et al., *Impact of treatment planning and delivery factors on gastrointestinal toxicity: an analysis of data from the RADAR prostate radiotherapy trial*. Radiat Oncol, 2014. **9**: p. 282.
224. Bewick, V., L. Cheek, and J. Ball, *Statistics review 13: receiver operating characteristic curves*. Crit Care, 2004. **8**(6): p. 508-12.
225. de Crevoisier, R., et al., *Daily Versus Weekly Prostate Cancer Image Guided Radiation Therapy: Phase 3 Multicenter Randomized Trial*. Int J Radiat Oncol Biol Phys, 2018. **102**(5): p. 1420-1429.
226. Wilkins, A., et al., *Hypofractionated radiotherapy versus conventionally fractionated radiotherapy for patients with intermediate-risk localised prostate cancer: 2-year patient-reported outcomes of the randomised, non-inferiority, phase 3 CHHiP trial*. The Lancet Oncology, 2015. **16**(16): p. 1605-1616.
227. Sundahl, N., et al., *Combining high dose external beam radiotherapy with a simultaneous integrated boost to the dominant intraprostatic lesion: Analysis of genito-urinary and rectal toxicity*. Radiother Oncol, 2016. **119**(3): p. 398-404.

228. Casares-Magaz, O., et al., *Spatial rectal dose/volume metrics predict patient-reported gastro-intestinal symptoms after radiotherapy for prostate cancer*. *Acta Oncol*, 2017. **56**(11): p. 1507-1513.
229. Wortel, R.C., et al., *Dose-surface maps identifying local dose-effects for acute gastrointestinal toxicity after radiotherapy for prostate cancer*. *Radiother Oncol*, 2015. **117**(3): p. 515-20.
230. Munbodh, R., et al., *Dosimetric and anatomic indicators of late rectal toxicity after high-dose intensity modulated radiation therapy for prostate cancer*. *Med Phys*, 2008. **35**(5): p. 2137-50.
231. Palorini, F., et al., *Bladder dose-surface maps and urinary toxicity: Robustness with respect to motion in assessing local dose effects*. *Phys Med*, 2016. **32**(3): p. 506-11.
232. Moulton, C.R., et al., *Spatial features of dose-surface maps from deformably-registered plans correlate with late gastrointestinal complications*. *Phys Med Biol*, 2017. **62**(10): p. 4118-4139.
233. Henderson, D.R., et al., *An Investigation of Dosimetric Correlates of Acute Toxicity in Prostate Stereotactic Body Radiotherapy: Dose to Urinary Trigone is Associated with Acute Urinary Toxicity*. *Clin Oncol (R Coll Radiol)*, 2018. **30**(9): p. 539-547.
234. Shelley, L.E.A., et al., *Associations between voxel-level accumulated dose and rectal toxicity in prostate radiotherapy*. *Physics and Imaging in Radiation Oncology*, 2020. **14**: p. 87-94.
235. Marcello, M., et al., *Relationships between rectal and perirectal doses and rectal bleeding or tenesmus in pooled voxel-based analysis of 3 randomised phase III trials*. *Radiother Oncol*, 2020. **150**: p. 281-292.
236. Buettner, F., et al., *The dose-response of the anal sphincter region--an analysis of data from the MRC RT01 trial*. *Radiother Oncol*, 2012. **103**(3): p. 347-52.
237. Fu, Y., et al., *Pelvic multi-organ segmentation on cone-beam CT for prostate adaptive radiotherapy*. *Med Phys*, 2020. **47**(8): p. 3415-3422.
238. Lei, Y., et al., *Male pelvic multi-organ segmentation aided by CBCT-based synthetic MRI*. *Phys Med Biol*, 2020. **65**(3): p. 035013.
239. Widmark, A., et al., *Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial*. *The Lancet*, 2019. **394**(10196): p. 385-395.
240. Brand, D.H., et al., *Intensity-modulated fractionated radiotherapy versus stereotactic body radiotherapy for prostate cancer (PACE-B): acute toxicity findings from an international, randomised, open-label, phase 3, non-inferiority trial*. *The Lancet Oncology*, 2019. **20**(11): p. 1531-1543.
241. Devlin, L., et al., *Dosimetric impact of organ at risk daily variation during prostate stereotactic ablative radiotherapy*. *Br J Radiol*, 2020. **93**(1108): p. 20190789.
242. Mannerberg, A., et al., *Dosimetric effects of adaptive prostate cancer radiotherapy in an MR-linac workflow*. *Radiat Oncol*, 2020. **15**(1): p. 168.
243. McVicar, N., I.A. Popescu, and E. Heath, *Techniques for adaptive prostate radiotherapy*. *Phys Med*, 2016. **32**(3): p. 492-8.
244. Gorovets, D., et al., *Prostate SBRT With Intrafraction Motion Management Using a Novel Linear Accelerator-Based MV-kV Imaging Method*. *Pract Radiat Oncol*, 2020. **10**(5): p. e388-e396.
245. Vanhanen, A., P. Poulsen, and M. Kapanen, *Dosimetric effect of intrafraction motion and different localization strategies in prostate SBRT*. *Phys Med*, 2020. **75**: p. 58-68.

Appendices

Appendix 1 – Post radiotherapy toxicity questionnaire used by clinicians

DELINEATE

__ MONTHS POST-RADIOTHERAPY FORM pg1

NB: Remember to give PRO questionnaire

Patient's Initials _____	Trial Number <input type="text"/>
Date of assessment <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <small>dd mm yy</small>	
PSA <input type="text"/> <input type="text"/> <input type="text"/> . <input type="text"/> <input type="text"/> ng/ml	Date <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <small>dd mm yy</small>
Rectal exam: 1. Clinically normal <input type="checkbox"/> 2. Suspicious <input type="checkbox"/> 3. Malignant <input type="checkbox"/>	
4. Not done <input type="checkbox"/> Please give the reason: _____	

ASSESSMENT FOR LATE SIDE EFFECTS AFTER RADIOTHERAPY		
URINARY SYMPTOMS (excluding urinary tract infections)		
Average daytime frequency <input type="checkbox"/> >2 hourly (Grade 0) <input type="checkbox"/> 2 hourly (Grade 1) <input type="checkbox"/> 1-2 hourly (no treatment) (Grade 2) <input type="checkbox"/> 1-2 hourly (simple outpatient management) (Grade 3) <input type="checkbox"/> <1 hourly (Grade 4) <input type="checkbox"/> Unknown	Nocturia <input type="checkbox"/> 0-1 times (Grade 0) <input type="checkbox"/> 2-3 times (Grade 1) <input type="checkbox"/> 4-5 times (Grade 2) <input type="checkbox"/> 6-8 times (Grade 3) <input type="checkbox"/> >8 times (Grade 4) <input type="checkbox"/> Unknown	Incontinence <input type="checkbox"/> None (Grade 0) <input type="checkbox"/> Occasional incontinence (Grade 1) <input type="checkbox"/> Frequent incontinence requiring pads (Grade 2) <input type="checkbox"/> Unknown
BOWEL SYMPTOMS		ERECTILE POTENCY
Frequency <input type="checkbox"/> 1-2 times (Grade 0) <input type="checkbox"/> 3-4 times (no medical treatment) (Grade 1) <input type="checkbox"/> 3-4 times (simple outpatient management) (Grade 2) <input type="checkbox"/> ≥ 5 times ± treatment (Grade 3) <input type="checkbox"/> Unknown	Rectal bleeding <input type="checkbox"/> None (Grade 0) <input type="checkbox"/> Occasional (no treatment) (Grade 1) <input type="checkbox"/> Moderate (simple outpatient management) (Grade 2) <input type="checkbox"/> Severe (blood transfusion, surgery) (Grade 3) <input type="checkbox"/> Unknown	<input type="checkbox"/> Normal erection (Grade 0) <input type="checkbox"/> Decreased (Grade 1) <input type="checkbox"/> Absent (Grade 2) <input type="checkbox"/> Unknown
Is the patient taking α-Blockers or Anticholinergic drugs for Bladder Symptoms? Yes <input type="checkbox"/> No <input type="checkbox"/>		

Any inpatient days related to prostate cancer since last assessment? Yes <input type="checkbox"/> No <input type="checkbox"/> If yes, number of days <input type="text"/>
Any unscheduled outpatient visits related to prostate cancer since last assessment? Yes <input type="checkbox"/> No <input type="checkbox"/> If yes, number of days <input type="text"/>
Reason for visit: 1. Assessment of Toxicity <input type="checkbox"/> 2. Treatment of toxicity <input type="checkbox"/>
3. Assessment of recurrence <input type="checkbox"/> 4. Treatment of recurrence <input type="checkbox"/>

If patient has developed clinical or biochemical evidence of progressive disease or recommenced hormonal treatment, please complete Progression Form. If patient has developed Serious Adverse Event or died, complete SAE/DEATH form.

DELINEATE

_ MONTHS POST-RADIOTHERAPY FORM pg 2

Patient's Initials _____	Trial Number <input style="width: 20px; height: 20px;" type="text"/>
LATE TOXICITY SCORING USING RTOG GRADING	
	Grade 0 Grade 1 Grade 2 Grade 3 Grade 4 Grade 5 Unknown
Diarrhoea	<input type="checkbox"/>
Rectal-anal stricture	<input type="checkbox"/>
Proctitis	<input type="checkbox"/>
Urethral stricture	<input type="checkbox"/>
Cystitis	<input type="checkbox"/>
Rectal ulcer	<input type="checkbox"/>
Haematuria	<input type="checkbox"/>
Bowel obstruction	<input type="checkbox"/>
RTOG Late Toxicity scoring system	
Grade 0 - No symptoms	
Grade 1 - Minor symptoms requiring no treatment	
Grade 2 - Symptoms responding to a simple outpatient management, lifestyle (performance status not affected)	
Grade 3 - Distressing symptoms altering patient's lifestyle (performance status). Hospitalisation for diagnosis or minor surgical intervention (such as urethral dilatation) may be required.	
Grade 4 - Major surgical intervention (such as laparotomy, colostomy, and cystectomy) or prolonged hospitalisation required.	
Grade 5 - Fatal complications	
Diarrhoea is defined as a clinical syndrome characterised by frequent loose bowel movements without associated rectal irritation (tenesmus)	
Proctitis is defined as a clinical syndrome characterised by rectal irritation or urgency (tenesmus), presence of mucous or blood in the stool and, in some patients, with frequent, sometimes loose bowel movements.	
Cystitis is defined as a syndrome characterised by irritative bladder symptoms such as frequency and dysuria. Haematuria may or may not be a part of the clinical picture of cystitis.	
Incidental / Non protocol PSAs	
PSA <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> . <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> ng/ml	Date <input style="width: 20px;" type="text"/> - <input style="width: 20px;" type="text"/> - <input style="width: 20px;" type="text"/> <small style="margin-left: 40px;">dd mm yy</small>
PSA <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> . <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> ng/ml	Date <input style="width: 20px;" type="text"/> - <input style="width: 20px;" type="text"/> - <input style="width: 20px;" type="text"/> <small style="margin-left: 40px;">dd mm yy</small>
Please record any other additional PSA results since the last trial visit.	

DELINEATE

MONTHS POST-RADIOTHERAPY FORM pg 3

Patient's Initials _____	Trial Number <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
---------------------------------	---

Rectal Questions for Gulliford Scores (IJROBP Volume 76, Number 3, 2010)

Subjective Stool Frequency (LENT SOM)
< 2 / day 2-4 / day 5-8 day >8 day Uncontrolled diarrhoea Unknown

Subjective sphincter control (LENT SOM)
None Occasional Intermittent Persistent Refractory Unknown

Management sphincter control (LENT SOM)
None Occasional Intermittent Persistent Surgical Intervention Unknown

How often has the patient had stools that were loose or liquid (no form, watery or mushy) during the last 4 weeks? (UCLA-PCI)
Never Rarely About half the time Usually Always Unknown

How often has the patient had rectal urgency (felt like they had to pass stool but did not) during the last 4 weeks? (UCLA – PCI)
Rarely / never About once a week More than once a week
About once a day More than once day Unknown

Overall, how big a problem have the patient's bowel habits been for them during the last 4 weeks? (UCLA-PCI)
No Problem Very Small Problem Small Problem
Moderate Problem Big Problem Unknown

DELINEATE

__ MONTHS POST-RADIOTHERAPY FORM pg 4

Patient's Initials _____	Trial Number <input style="width: 20px; height: 20px;" type="text"/>
---------------------------------	---

NCI CTCAE v4	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Unknown
Gastrointestinal Disorders							
Abdominal Pain	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
Constipation	<input type="checkbox"/>						
Diarrhoea	<input type="checkbox"/>						
Proctitis	<input type="checkbox"/>						
Rectal Haemorrhage	<input type="checkbox"/>						
Rectal Pain	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
Renal and Urinary Disorders							
Haematuria	<input type="checkbox"/>						
Urinary Frequency	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>
Urinary Incontinence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
Urinary Retention	<input type="checkbox"/>						
Urinary Urgency	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>
Reproductive Disorders							
Erectile Dysfunction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>
Other _____	<input type="checkbox"/>						
Other _____	<input type="checkbox"/>						
Other _____	<input type="checkbox"/>						

Confirm that PRO questionnaire has been given to patient Yes <input type="checkbox"/> No <input type="checkbox"/>
If not done please record the reason _____

Print _____
Signed _____
Date <input style="width: 20px; height: 20px;" type="text"/> <input style="width: 20px; height: 20px;" type="text"/> <input style="width: 20px; height: 20px;" type="text"/>

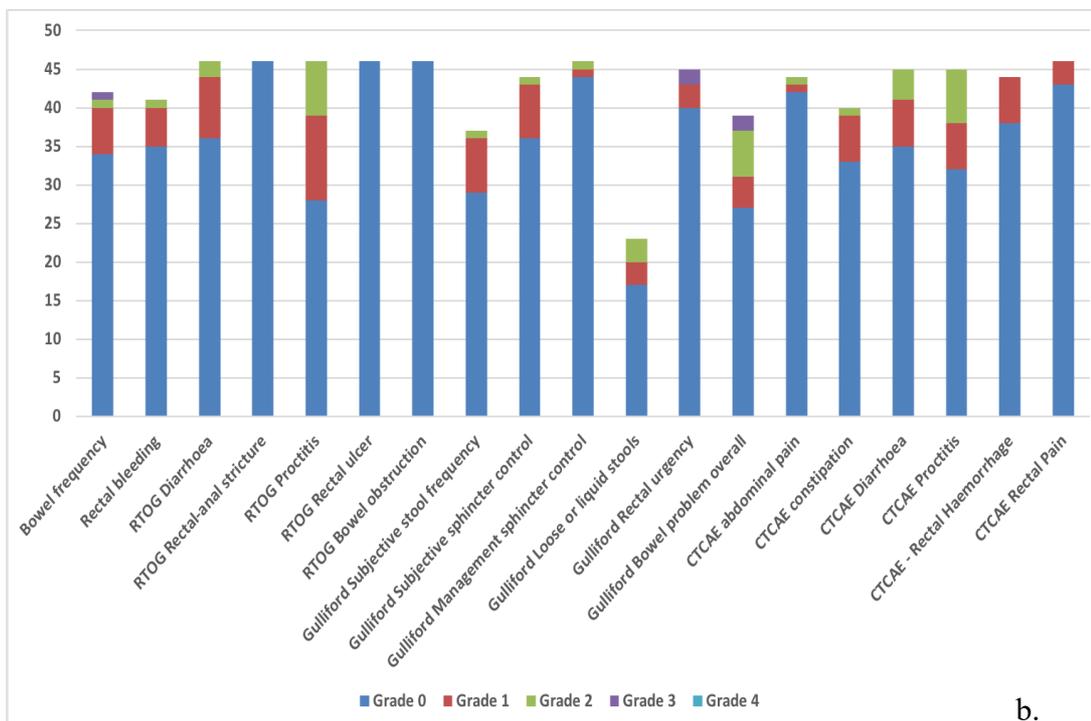
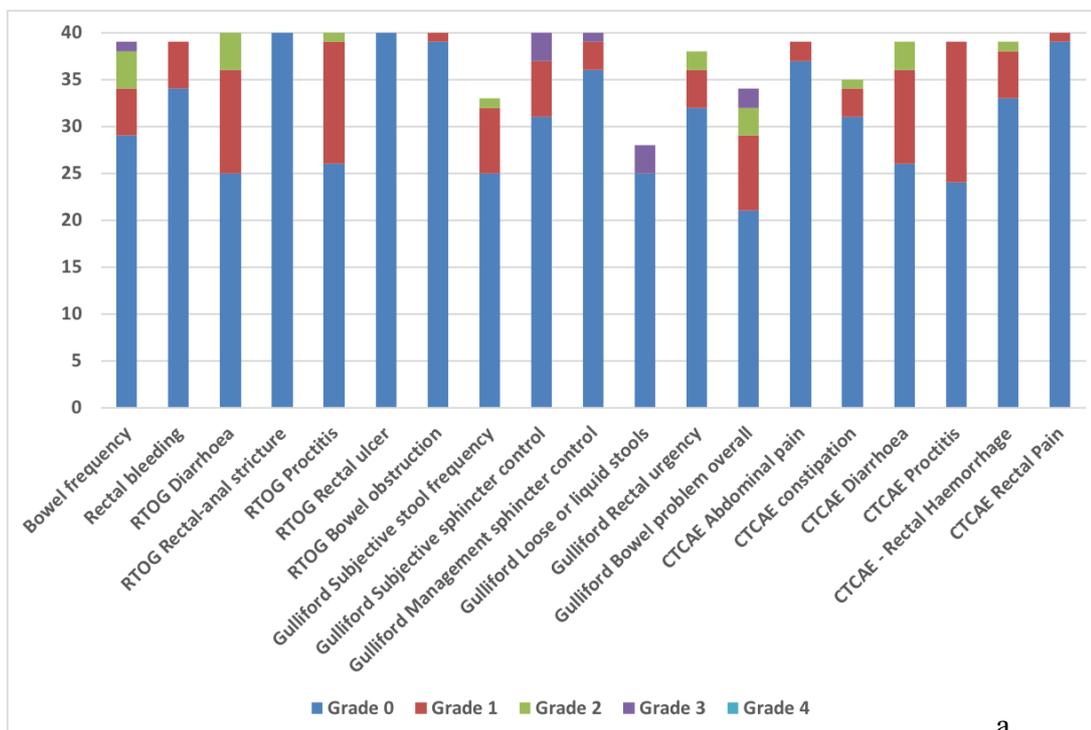
NCICTCAE Toxicity Gradings					
Grade					
Symptom	1	2	3	4	5
Gastrointestinal Disorders					
Abdominal pain	Mild pain	Moderate pain; limiting instrumental AOL	Severe pain; limiting self care AOL	.	.
Constipation	Occasional intermittent symptom; occasional use of stool softeners, laxatives or enemas	Persistent symptoms with regular use of laxatives or enemas, limiting instrumental AOL	Obstipation with manual evacuation indicated; limiting self care AOL	Life-threatening consequences; urgent intervention indicated	Death
Diarrhea	Increase of <4 stools per day over baseline; mild increase in stool output compared to baseline	Increase of 4-6 stools per day over baseline; moderate increase in stool output compared to baseline	Increase of >=7 stools per day over baseline; hospitalization indicated; severe increase in stool output compared to baseline; limiting self care AOL	Life-threatening consequences; urgent intervention indicated	Death
Proctitis	Rectal discomfort; medical intervention not indicated	Symptoms (e.g., rectal discomfort, passing blood or mucus); medical intervention indicated; limiting instrumental AOL	Severe symptoms: rectal urgency or stool incontinence; limiting self care AOL	Life-threatening consequences; urgent intervention indicated	Death
Rectal hemorrhage	Medical intervention not indicated	Mild symptoms; medical intervention or minor cauterization indicated	Transfusion, radiologic endoscopic, or elective operative intervention indicated	Life-threatening consequences; urgent intervention indicated	Death
Rectal pain	Mild pain	Moderate pain; limiting instrumental AOL	Severe pain; limiting self care AOL	.	.
Renal & Urinary Disorders					
Hematuria	Asymptomatic; clinical or diagnostic observation only; intervention not indicated	Symptomatic; urinary catheter or bladder irrigation indicated; limiting instrumental AOL	Gross hematuria; transfusion, IV medications or hospitalization indicated; elective endoscopic, radiologic or operative intervention indicated; limiting self care AOL	Life-threatening consequences; urgent intervention indicated	Death
Urinary frequency	Present	Limiting instrumental AOL; medical management indicated	.	.	.
Urinary incontinence	Occasional (e.g., with coughing, sneezing, etc.); pads not indicated	Spontaneous; pads indicated; limiting instrumental AOL	Intervention indicated (e.g., clamp, collagen injection); operative intervention indicated; limiting self care AOL	.	.
Urinary retention	Intermittent catheter placement not indicated; able to void with some residual	Placement of urinary suprapubic or intermittent catheter placement indicated; medical management indicated	Elective operative or radiologic intervention indicated; substantial loss of affected kidney function or m3ss	Life-threatening consequences; urgent intervention indicated	Death
Urinary urgency	Present	Limiting instrumental AOL; medical management indicated	.	.	.
Reproductive system disorders					
Erectile dysfunction	Occasional decrease in erectile function (frequency/rigidity of erections) but medical intervention not indicated (e.g., medication or use of mechanical device, penile pump)	Decrease in erectile function (frequency/rigidity of erections); medical intervention indicated (e.g., medication or mechanical device such as penile pump)	Decrease in erectile function (frequency/rigidity of erections) but medical intervention not helpful; placement of a permanent penile prosthesis indicated (not previously present)	.	.

Appendix 2 – Clinical characteristics by cohort for patients with dose accumulated using DSM methodology

	Standard fraction cohort (%) (n=46)	Hypofractionated cohort (%) (n=40)
<u>Age (years)</u>	70 (57 - 80)	71.5 (61 - 79)
<u>T Stage</u>		
T1a/b/c/x	18 (39%)	21 (52.5%)
T2a/b/c/x	23 (50%)	12 (30%)
T3/x	5 (10.8%)	7 (17.5%)
<u>Gleason Score</u>		
6	11 (24%)	10 (25%)
7	34 (74%)	26 (65%)
8	1 (2.2%)	4 (10%)
<u>NCCN Risk Group</u>		
Intermediate	34 (74%)	26 (65%)
High	12 (26%)	14 (35%)
<u>Diabetes</u>		
Yes	2 (4.4%)	11 (27.5%)
No	44 (95.6%)	29 (72.5%)
<u>Hypertension</u>		
Yes	19 (41.3%)	22 (55%)
No	27 (58.7%)	18 (45%)
<u>Inflammatory bowel or diverticular disease</u>		
Yes	1 (2.2%)	6 (15%)
No	45 (97.8%)	34 (85%)
<u>Pelvic surgery</u>		
Yes	3 (6.5%)	11 (27.5%)
No	43 (93.5%)	29 (72.5%)
<u>Symptomatic Haemorrhoids in the last 12 months</u>		
Yes	11 (24%)	2 (5%)
No	34 (74%)	38 (95%)
Unknown	1 (2.2%)	0
<u>Previous transureth</u>		
Yes	4 (8.7%)	3 (7.5%)
No	42 (91.3%)	37 (92.5%)
<u>Statins</u>		
Yes	17 (37%)	18 (45%)
No	29 (63%)	22 (55%)
<u>Current Smoker</u>		
Yes	3 (6.5%)	4 (10%)
No	42 (91.3%)	36 (90%)
Unknown	1 (2.2%)	0

Clinical characteristics for patients included in the DIR methodology shown by cohort

Appendix 3 – Incidence of toxicity by cohort for patients included in deformable image registration study



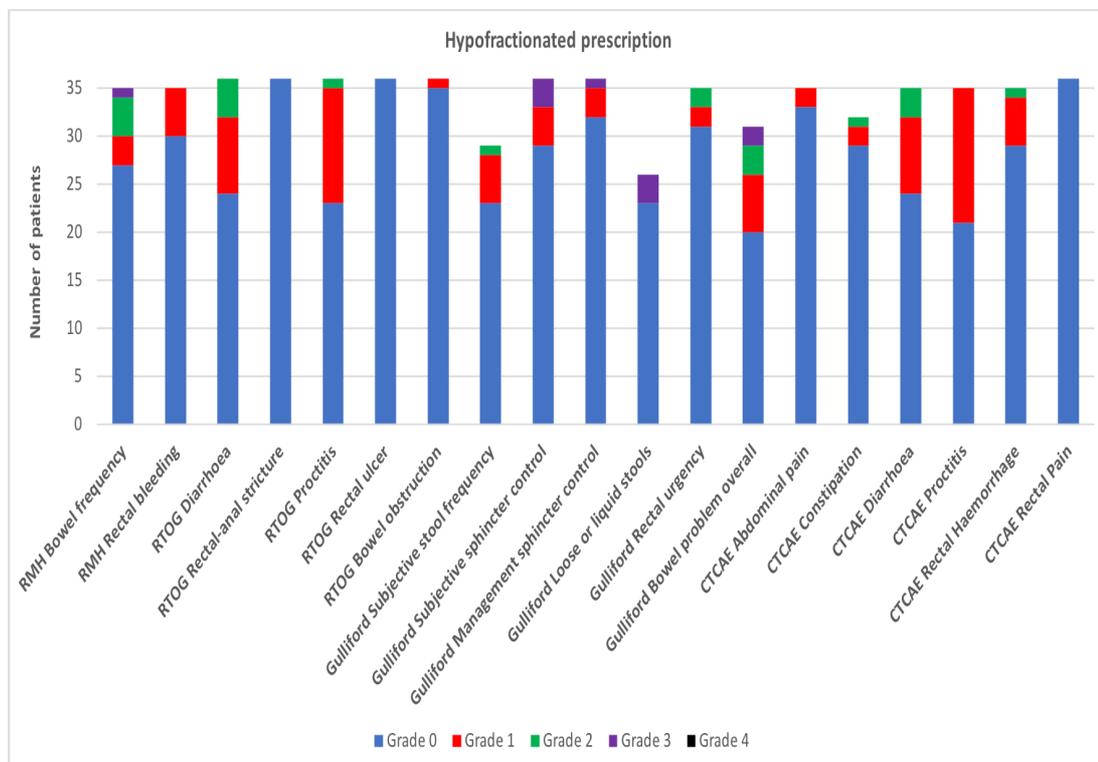
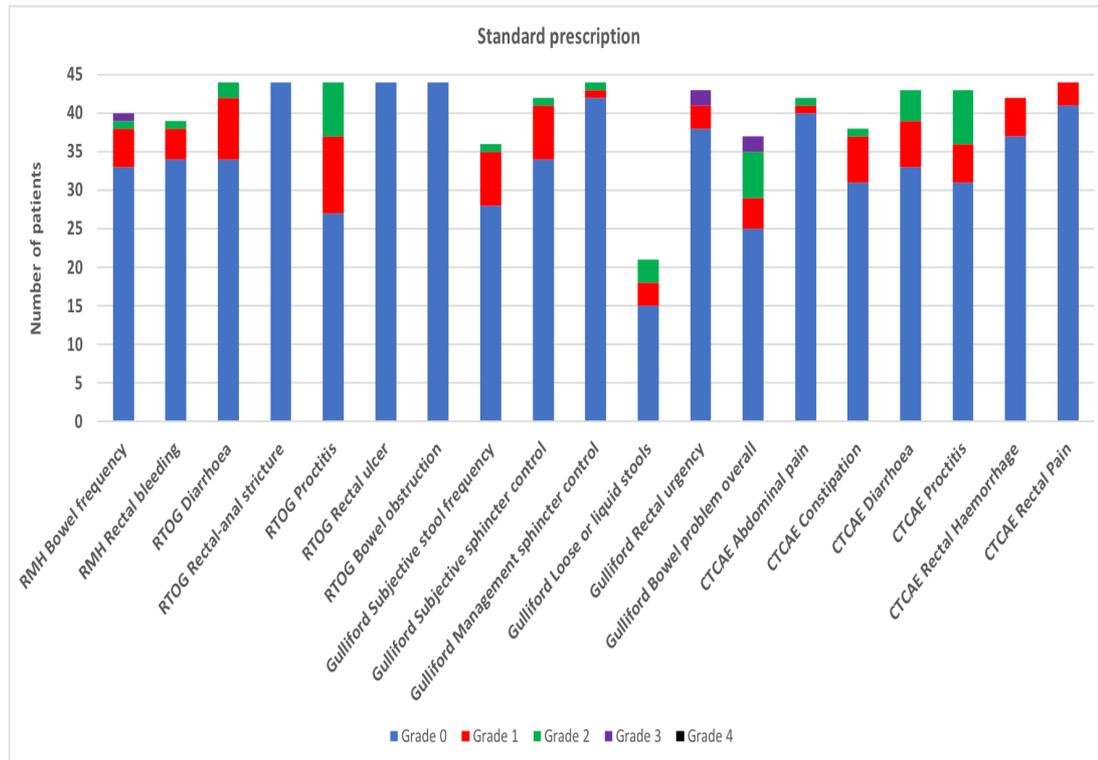
Peak toxicity for standard cohort given in a. (n=46) and hypofractionated given in b. (n=40) cohorts given as the number of patients experiencing toxicity for a given endpoint. The number of patients analysed for each endpoint varies due to exclusion of patients reporting baseline similar to toxicity being reported

Appendix 4 - Clinical characteristics by cohort for patients with dose accumulated using DSM methodology

	Standard fraction cohort (n=44)	Hypofractionated cohort (n=36)
<u>Age (years)</u>	70.2 (57-80)	70.8 (61-79)
<u>T Stage</u>		
T1a/b/c/x	17 (39)	19 (53)
T2a/b/c/x	22 (50)	10 (28)
T3/x	5 (11)	7 (19)
<u>Gleason Score</u>		
6	6 (25)	8 (22)
7	32 (73)	24 (67)
8	1 (2)	4 (11)
<u>NCCN Risk Group</u>		
Intermediate	12 (27)	14 (39)
High	32 (73)	22 (61)
<u>Diabetes</u>		
Yes	2 (5)	9 (25)
No	42 (95)	27 (75)
<u>Hypertension</u>		
Yes	19 (43)	20 (56)
No	25 (57)	16 (44)
<u>Inflammatory bowel or diverticular disease</u>		
Yes	1 (2)	5 (14)
No	43 (98)	31 (86)
<u>Pelvic surgery</u>		
Yes	3 (7)	9 (25)
No	41 (93)	27 (75)
<u>Symptomatic Haemorrhoids in the last 12 months</u>		
Yes	11 (25)	2 (6)
No	32 (73)	34 (94)
Unknown	1 (2)	0
<u>Previous transurethral resection of the prostate</u>		
Yes	4 (9)	3 (8)
No	40 (91)	33 (92)
<u>Statins</u>		
Yes	17 (39)	15 (42)
No	27 (61)	21 (58)
<u>Current Smoker</u>		
Yes	3 (7)	3 (8)
No	40 (91)	33 (92)
Unknown	1 (2)	0

Clinical characteristics of patients included in the DSM methodology (n=80), given by dose prescription cohort.

Appendix 5 Incidence of toxicity by cohort for patients included in dose surface map study



Histograms of peak toxicity for each endpoint given by cohort. Standard prescription (n= 44) and Hypofractionated prescription (n=36).