Predictors of Toxicity in Prostate Hypofractionation Clinical Trials

A thesis submitted for the degree of Doctor of Philosophy

Douglas Henry Brand

Division of Radiotherapy and Imaging
The Institute of Cancer Research
University of London, United Kingdom

I, Douglas Henry Brand, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Moderately hypofractionated external beam radiotherapy (EBRT) for non-metastatic prostate cancer is now a standard-of-care radical treatment, following multiple large randomised trials. Much research is directed towards even quicker ultrahypofractionated courses of treatment. This thesis explores the toxicity effects associated with hypofractionation.

Fraction size sensitivity (α/β ratios) for late rectal effects are estimated using data from CHHiP, a large moderate hypofractionation trial. α/β ratios are estimated by fitting equivalent dose in 2 Gy per fraction (EQD2) adjusted Lyman-Kutcher-Burman models. Although α/β ratios as low as 1.6 Gy (95% CI 0.9-2.5) for rectal bleeding G1+ are observed, none significantly improve on the commonly used α/β = 3 Gy.

Ultrahypofractionation theoretically reduces acute toxicity equivalent dose, but the treatment is accelerated. Clinician reported outcomes (CROs) for acute toxicity are explored in PACE-B, a large trial comparing ultrahypofractionated to non-ultrahypofractionated EBRT. RTOG toxicity is similar, while CTCAE gastrointestinal G2+ toxicity is worse with ultrahypofractionation.

The acute toxicity patient reported outcomes (PROs) for PACE-B are explored next. No significant differences are seen between ultrahypofractionation and non-ultrahypofractionation. In combination, this

suggests that acute toxicity should not be a reason to avoid ultrahypofractionation.

Contouring of the rectum on EBRT planning scans is important for planning and estimation of toxicity likelihood. Whether centrally reviewed rectal contours improve toxicity prediction is explored using the CHHiP dataset. No significant difference is seen for any toxicity. Additionally, no benefit to toxicity prediction is seen from use of absolute, rather than relative, volume predictors, nor by procedurally truncating the rectum.

Finally, the prediction of acute gastrointestinal and genitourinary toxicity (CRO & PRO) of ultrahypofractionation is explored by multivariate modelling of the PACE-B dataset. Consistent signals are seen to suggest worse toxicity with overall treatment time ≤1 week with ultrahypofractionation. The use of non-co-planar delivery appears protective, but only by CRO measures.

Acknowledgements

This PhD would not have been possible without the generous Clinical Research Fellowship funding from Cancer Research UK.

Amongst individuals, I would first like to thank my supervisory team of Nick van As, Alison Tree, Emma Hall & Sarah Gulliford. They have provided essential guidance across the varied research topics covered in this thesis. I must also thank Sarah Brüningk for kindly joining this team in my second year, and providing valuable insights to the modelling processes.

Outside of the supervisory team, I must pre-eminently thank David Dearnaley for the many hours he kindly spent with me in general discussion of radiotherapy research. He has been a significant influence in shaping many of the ideas presented here.

All work in the thesis derives from the CHHiP and PACE trials, the conduct of which has depended on thousands of individuals. Thanks must of course first go to the patients, on whom such clinical research is wholly contingent. I am of course also grateful to all investigators and research staff whose work has made these trials possible.

The ICR Clinical Trials and Statistics Unit has processed all non-DICOM trial data used in this thesis and this thus critical to its completion. I would specifically like to thank Vicki Hinder and Clare Griffin for their detailed review of the PACE Acute toxicity statistical analysis plan. Also, Steph

Burnett, Steph Brown and Julia Pugh for their work in rapidly getting the database ready for the acute toxicity data snapshot.

This work has also relied heavily on the use of DICOMs from these trials, for which I am extremely grateful to Anna Wilkins, Katie Fernandez, Richard Boyle, Jake Probert who contributed towards retrieval, processing and recontouring of CHHiP rectums used in this thesis. I am also grateful to Julia Murray, who retrieved many CHHiP DICOMs from the treating centres, similarly to Olivia Naismith, who retrieved many PACE-B DICOMs from the treating centres. I must thank Katie Fernandez again, along with Joanna Parker for their efforts in processing the PACE-B DICOMs.

I am of course grateful to my many fellow fellows, who made the PhD time in Orchard House an enjoyable one. Additionally, I must thank Jane Pickering for the many PhD administration problems she has helped me through, as well as Meg Messick, for her tireless organisation of supervisory meetings.

Finally, I am indebted to my wife Sarah, for her love and support at all times during this research.

Contents

List of Figures

List of Tables

List of Abbreviations

3DCRT = 3-Dimensional Conformal Radiotherapy

5-ARI = 5-alpha reductase inhibitor

ADT = Androgen Deprivation Therapy

Alt. = Alternate

ASTRO = American Society for Radiation Oncology

AUC = Area Under Curve

bPFS = Biochemical Progression Free Survival

BED = Biologically Effective Dose

C. = Centre

CBCT = Cone Beam Computed Tomography

CERR = Computational Environment for Radiotherapy Research

CFRT = Conventionally Fractionated RT

CI = Confidence Interval

CK = CyberKnife

CLE = Consequential Late Effect

CNN = Convolutional Neural Network

CRO = Clinician Reported Outcome

CT = Computerised Tomography

CTCAE = Common Toxicity Criteria Adverse Events

CTV = Clinical Target Volume

CV = Cross-Validation

DICOM = Digital Imaging and Communications in Medicine

DMF = Dose Modifying Factor

DSC = Dice Similarity Coefficient

DVH = Dose-Volume Histogram

DNA = Deoxyribonucleic Acid

Edit. = Edited

EBRT = External Beam Radiotherapy

EPIC = Expanded Prostate Cancer Index Composite

EPIC-26 = Expanded Prostate Cancer Index Composite (26 Question)

EORTC = European Organisation for Research and Treatment of Cancer

EORTC QLQ-C30 = EORTC core quality of life questionnaire

EQ-5D = EuroQol-5D

EQD2 = Equivalent Dose in 2 Gy Fractions

FACT-G = Functional Assessment of Cancer Therapy - General

FFF = Flattening Filter Free

Fidx = Fiducials

Fr = Fraction

Freq. = Frequency

G1/2/3 = Grade 1/2/3

G1/2/3+ = Grade 1/2/3 or above

GI = Gastrointestinal

GU = Genitourinary

HDRBT = High Dose Rate Brachytherapy

HES = Hospital Episodes Statistics

HSCL-25 = Hopkins Symptom Checklist 25

HIFU = High Intensity Focussed Ultrasound

IBD = Inflammatory Bowel Disease

IBDQ = Inflammatory Bowel Disease Questionnaire

IDMC = Independent Data Monitoring Committee

IGRT = Image-Guided Radiotherapy

IIEF-5 = International Index of Erectile Function 5-question

IMRT= Intensity Modulated Radiotherapy

ICR-CTSU = Institute of Cancer Research Clinical Trials and Statistics Unit

Intra-kV = Intra-fractional kV

IPSS = International Prostate Symptom Score

IQR = Interquartile Range

IR = Intermediate Risk

LASSO = Least Absolute Shrinkage and Selection Operator

LDRBT = Low Dose Rate Brachytherapy

LENTSOM = Late Effects Normal Tissue: Subjective, Objective & Management

LKB = Lyman-Kutcher Burman

LKB-NoEQD2 = LKB Model Without EQD2 Correction

LKB-EQD2 = EQD2-corrected Lyman-Kutcher Burman Model

LKB-EQD2-DMF = Dose-Modifying-Factor & EQD2-corrected Lyman-Kutcher Burman Model

LINAC= Linear Accelerator

LQ = Linear-Quadratic

LR = Low-risk

MCID = Minimal Clinically Important Difference

Med. = Median

MHRT = Moderately Hypofractionated Radiotherapy

MLC = Multi-leaf Collimator

MRI = Magnetic Resonance Imaging

MVar = Multivariate

NCCN = National Comprehensive Cancer Network

NHS = National Health Service

nmPCa = Non-metastatic Prostate Cancer

NN = Neural Network

NPV = Negative Predictive Value

NTCP = Normal Tissue Complication Probability

NS = Not Stated.

OAR = Organ at Risk

OOB = Out of Bag

OR = Odds Ratio

Orig. = Original

OS = Overall Survival

PCa = Prostate Cancer

PCSS = Prostate Cancer Symptom Scale

PCSSurv = Prostate Cancer Specific Survival

Post. = Posterior

PPV = Positive Predictive Value

Pros. = Prostate

PS = Performance Status

PSA = Prostate Specific Antigen

PRO = Patient Reported Outcome

PTV = Planning Target Volume

QA = Quality Assurance

QUANTEC = Quantitative Analyses of Normal Tissue Effects in the Clinic

QoL = Quality of Life

RCT = Randomised Controlled Trial

REC = Research Ethics Committee

RMH = Royal Marsden Hospital

RMSE = Root Mean Squared Error

ROC = Receiver Operating Characteristic

RR = Relative Risk

RT = Radiotherapy

RTDS = Radiotherapy Dataset

RTOG = Radiation Therapy Oncology Group

SAP = Statistical Analysis Plan

SBRT = Stereotactic Body Radiotherapy

SD = Standard Deviation

Sens. = Sensitivity

SF-36 = Short Form 36 Question

Spec. = Specificity

STAPLE = Simultaneous Truth And Performance Level Estimation

Subj. = Subjective

SV = Seminal Vesicles

SVM = Support Vector Machine

TCP = Tumour Control Probability

TD50 = Toxic Dose 50%

TURP = Trans-urethral Resection of Prostate

UCLA-PCI = University of California Los Angeles Prostate Cancer Index

UI = Urinary Incontinence

UK = United Kingdom

UO = Urinary Obstructive

VMAT = Volumetric Modulated Arc Therapy

WHO = World Health Organisation

VODCA = Visualisation and Organisation of Data for Cancer Analysis

Chapter 1. Thesis Outline

1.1 Thesis Approach to Background Material

This section will briefly describe the research motivation and questions. The viewpoint presented is as seen at the outset of the PhD study period (October 2017). As a brief summary, the prose is deliberately presented without extensive discussion of the references cited. The literature review (**Chapter 2**) will give a complete, fully referenced, background. The literature review will also discuss some references published subsequent to October 2017, where this will not influence the research narrative. New literature, published subsequent to October 2017, which alters the conclusions of the data chapters, will be discussed in the relevant data chapter discussion section.

1.2 Thesis Description

External beam radiotherapy (EBRT) is a common treatment modality used in the curative treatment of non-metastatic prostate cancer (nmPCa). This was historically delivered over long courses of small daily radiotherapy (RT) treatments (fractions), at a dose of 1.8-2.0 Gy. However, between 2016-2017, four phase III trials of moderate hypofractionation were published. The non-inferiority disease control results from three of these trials (CHHiP [1], RTOG-0415 [2] & PROFIT [3]) prompted a paradigm shift, with moderate hypofractionation (60 Gy in 20 fractions) becoming the UK standard-of-care. The fourth trial (HYPRO [4]) showed unexpectedly high acute and late toxicity.

These trials predicted the expected toxicity, and tumour response, to hypofractionation using the linear-quadratic (LQ) model [5]. A model parameter, the α/β ratio, governs the ratio of linear and quadratic response to an increase in fraction size. A smaller α/β ratio means a bigger response (tumour control or toxicity) to hypofractionation. The moderate hypofractionation trials were motivated by contemporaneous new evidence of a low α/β ratio in prostate cancer (PCa) [6]. Assuming normal tissues (rectum and bladder) had higher α/β ratios, a therapeutic gain could be anticipated. However, very limited human data on the α/β ratios for normal tissue toxicity was available: a 2004 study suggesting late rectal α/β = 5 Gy, by whole trial analysis [7]. Interestingly, the trial with highest late rectal toxicity (HYPRO) also used the highest late rectal α/β estimate (4-6 Gy vs 3 Gy in the three other trials), a choice likely motivated by animal data [8]. In **Chapter 3**, this thesis will examine late rectal toxicity data from the CHHiP trial, to better define the human late rectal α/β ratio. This will include examination of different late rectal toxicity endpoints (bleeding, frequency etc.), since their response to hypofractionation may differ.

Given the success of moderate hypofractionation studies, it seems logical to consider ultrahypofractionation (≥5 Gy per fraction). Assuming continued LQ-model validity at higher doses per fraction, the lower prostate α/β ratio vs normal tissue α/β ratios, might allow more therapeutic gain. Two relevant trials in the space are: i) The HYPO-RT-PC trial (ISRCTN45905321), initiated 2005, comparing 42.7 Gy in 7 fractions regimen to conventional 78 Gy in 39

fractions; ii) The PACE-B trial (NCT01584258), initiated 2012, comparing 36.25 Gy in 5 fractions to either 78 Gy in 39 fractions or 62 Gy in 20 fractions. A key consideration in ultrahypofractionation is the concomitant acceleration of treatment (reduction in days from first to last fraction), a factor independently shown to worsen acute toxicity [9]. Detailed analysis of the clinician reported outcomes (CROs) for acute toxicity in the PACE-B trial is therefore undertaken in **Chapter 4**. Differences between CROs and patient reported outcomes (PROs) have been demonstrated in patients receiving hypofractionationated EBRT for nmPCa [10]. A further detailed analysis of the acute toxicity PROs is therefore conducted in **Chapter 5**. A potential signal for lower stereotactic body radiotherapy (SBRT) acute genitourinary (GU) toxicity in patients receiving Cyberknife (CK) is seen, which is explored later in **Chapter 7**.

For prostate EBRT, the planning is optimised by achieving adequate dose to the tumour target, while keeping normal tissue doses below tolerance. For the rectal contours from the CHHiP trial, which were modelled in **Chapter 3**, each contour was checked against the protocol definition and edited where necessary. Such central review work, which was also undertaken for the prior UK-wide radiotherapy study (MRC RT-01) [11], is time-consuming and of uncertain benefit to toxicity modelling. The effect of retrospective centralised review on the rectal morphology, dose and relation to toxicity is examined in **Chapter 6**. Additionally, it is worth remembering that rectal contouring for EBRT planning exerts a person-time cost factor. Other groups have previously suggested the use of smaller, more reproducible rectal

definitions [12], based on distance from the prostate, although on small sample sizes. Whether these definitions, or the use of absolute volumes, might improve toxicity prediction in the large CHHiP trial cohort is also investigated in **Chapter 6**.

In the prediction of toxicity following radiotherapy, dosimetry alone is a relatively weak predictor [13]. It is therefore useful to combine dosimetric information with other potential predictors in multivariate predication frameworks. **Chapter 7** will leverage the PACE-B dataset to construct multivariate models examining clinical and dosimetric risk factors for acute toxicity after prostate EBRT, including the influence of Cyberknife delivery as suggested by **Chapter 4**.

Finally, overarching conclusions to the thesis, along with suggestions for future research will be presented in **Chapter 8**.

Chapter 2. Literature Review

2.1 Publications Related to This Chapter

The Linear-Quadratic Model and Implications for Fractionation.

D H Brand & J R Yarnold.

Clin Oncol (R Coll Radiol). 2019 Oct;31(10):673-677.

DOI: 10.1016/j.clon.2019.06.007

[Educational review/editorial]

2.2 The Prostate Cancer Treatment Landscape

2.2.1 Prostate Cancer Incidence and Survival in the UK

PCa is the most common United Kingdom (UK) male non-cutaneous malignancy, with ~50,000 new cases in 2017 [14] . PCa survival in the UK has seen enormous progress since the 1970s, with a 10-year survival rate of 84% in 2010-2011, compared to 25% in 1970-1971 [15]. Earlier diagnosis through prostate specific antigen (PSA) testing, has meant more patients treated with potentially curative intent, although also the overdiagnosis of non-life threatening PCa [16].

2.2.2 Prostate Cancer Risk Stratification

Staging and risk stratification for PCa should occur at the time of diagnosis [17], commonly with the National Comprehensive Cancer Network (NCCN) risk stratification [18]. This requires TNM staging, Gleason pathology score and PSA (before any hormonal therapy). This thesis will concern itself with

patients that have nmPCa suitable for radical (curative) treatment, although active surveillance has a role in lower risk disease [19]. Metastatic disease treatment and progress in systemic therapies are outside the scope of this thesis.

2.2.3 Radical Modalities for Non-Metastatic Prostate Cancer

Radical options for nmPCa include: EBRT, low dose rate brachytherapy (LDRBT), high dose rate brachytherapy (HDRBT) and radical prostatectomy (surgery). Focal ablative therapies, such as cryotherapy and high intensity focussed ultrasound (HIFU), are the subject of clinical trials [20].

With similar prostate cancer survival[1], based on the large randomised ProtecT trial [19], patient choice between EBRT and surgery may be guided by the differential toxicity profiles and delivery logistics of each modality. For example, surgery causes worse incontinence and sexual function, but improves nocturia; while EBRT results in worse bloody stools [21].

LDRBT as a boost to EBRT has strong evidence from the ASCENDE-RT randomised controlled trial (RCT), which demonstrated higher rates of biochemical progression with EBRT boost vs LDRBT boost after whole pelvis EBRT (HR 2.04, $p<0.001$) [22,23]. Unfortunately, grade 3+ (G3+) GU events were much worse with LDRBT boost (18.4% vs 5.2%, $p<0.001$). The

[1] Prostate-cancer specific survival >98% at 10 years.

randomised evidence for LDRBT monotherapy is very limited [24], although it has guideline level established clinical usage in the UK [20].

The evidence base for HDRBT monotherapy largely comes from prospective cohort studies, which have shown good outcomes for nmPCa with 4-6 fractions [25]. More recent interest has focussed on reducing the number of fractions. Morton and colleagues have reported randomised data examining HDRBT 19 Gy in 1 fraction versus 27 Gy in 2 fractions as monotherapy for low/intermediate risk nmPCa [26]. The 5-year biochemical progression free survival (bPFS) was significantly worse with 1 fraction versus 2 fractions (74.5% vs 97.3%, p=0.002), a finding that calls into question the validity of the LQ-model at very high doses per fraction (Discussed in **Section 2.3.3.4.3**).

HDRBT can also act as a boost to EBRT. An RCT comparing EBRT (55Gy / 20#) to EBRT + HDRBT boost (37.5 Gy / 15# + 17 Gy / 2#) showed better relapse free survival with boost [27]; however the control arm is clearly subtherapeutic following the results of CHHiP [1] (see **Section 2.3.3.5.1**).

The thesis will now focus on EBRT as the radical modality of interest.

2.3 External Beam Radiotherapy for Prostate Cancer

2.3.1 External Beam Radiotherapy Overview

EBRT is the most common radical treatment modality for UK men with nmPCa[2]. Most patients will receive ADT neoadjuvantly, concurrently and/or adjuvantly, although detailed discussion of ADT is outside the scope of this thesis. Most UK nmPCa EBRT is now moderately hypofractionated: typically 60 Gy in 20 fractions [28]. The evidence base behind this fractionation choice will be examined.

2.3.2 Advances in Prostate External Beam Radiotherapy

2.3.2.1 Historical Radiotherapy for Prostate Cancer

Recognisable radical EBRT doses of 60-70 Gy in 1.8-2 Gy per fraction were being delivered as early as the 1960s [29]. Prostate EBRT of the early 1980s was generally non-conformal, with x-ray simulation and bony landmarks being used for simulation. Delivery techniques were crude: anterior/posterior field, three or four field “bricks” or non-conformal arc rotations [30]. This can be considered the “pre-dose-escalation” era. Major advances in technology and techniques have permitted dramatic changes in the decades since.

2.3.2.2 Conformal Dosimetry

Development of computerised tomography (CT) machines permitted three-dimensional (3D) visualisation of pelvic anatomy. An early computerised

[2] The 2019 report of the National Prostate Cancer Audit identified 14000 men having radical EBRT versus 7000 undergoing RP [257]

treatment planning system with delineated targets was described in 1983 [31]. The coincident development of multileaf collimators (MLCs) during the 1980s [32] made conformal delivery far easier (**Figure 1**); obviating the need for customised lead blocks. A UK RCT of conformal vs non-conformal EBRT (n=225), both 64 Gy in 32 fractions, for nmPCa, demonstrated reduced late G2+ Radiation Therapy Oncology Group (RTOG) rectal toxicity (5% vs 15%, p=0.01)[3], a dose-limiting toxicity for non-conformal radiotherapy [33].

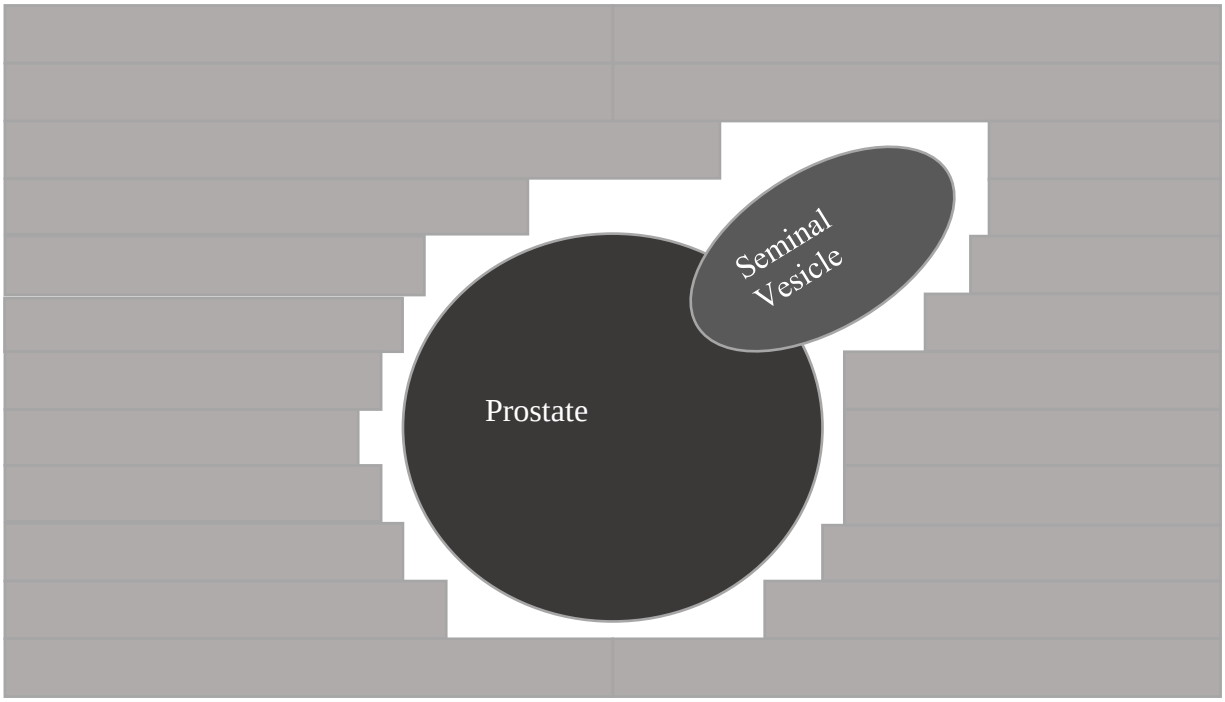


Figure 1. Multi-leaf Collimation for Conformality
Lateral beam's eye view of a prostate and SVs planning target volume. The multi-leaf collimators are used to conform the beam shape to the target.

2.3.2.3 Dose Escalation

The reduced toxicity of 3D conformal radiotherapy (3DCRT) suggested the possibility of increasing dose, without increasing side effects beyond the 2-dimensional era (isotoxicity). Multiple dose-escalation era RCTs tested the

[3] Tumour control and bladder toxicity were not statistically significantly different.

hypothesis that dose-escalated EBRT would improve local control of nmPCa, without excess side effects. Meta-analysis (6 trials, n=2822) demonstrated that, compared to standard dose, dose-escalation resulted in reduced biochemical progression rates (24.7 % vs 34.0 %, odds ratio (OR) 0.61, 95 % confidence interval (CI) 0.51-0.74), but no improvement in overall survival (OS) or Prostate Cancer Specific Survival (PCSSurv[4]) [34]. Dose escalation worsened late toxicity (G2+) for both gastrointestinal (GI) (28.0% vs 18.6%) and GU (22.6% vs 19.5%) endpoints. Despite a failure to translate improved bPFS into improved OS, dose-escalated EBRT is now standard-of-care [20].

2.3.2.4 Intensity Modulated Radiotherapy

Intensity modulated radiotherapy (IMRT) describes the delivery of beams with heterogenous fluence, permitting generation of complex isodose shapes [35], improving conformality beyond 3DCRT. The technology has progressed from field-in-field (step-and-shoot), through sliding window, up to modern volumetric modulated arc therapy (VMAT) with dynamic MLCs. VMAT, continually delivering from any co-planar angle, permits complex distributions (improving conformality) and rapid delivery [36].

Many studies have examined IMRT in nmPCa. A 2016 meta-analysis (n=9556, 23 trials) showed that IMRT, compared to 3DCRT, was associated with better G2+ RTOG GI toxicity, for both acute (relative risk (RR) 0.59, 95 % CI = 0.44-0.78) and late (RR 0.54, 95 % CI = 0.38-0.78) reactions, plus

[4] PCSSurv used as acronym due to a quality of life instrument described later, also called PCSS.

improved biochemical control (RR 1.17, 95 % CI = 1.08-1.27) [37]. However acute G2+ RTOG GU toxicity was worse (RR 1.08, 95 % CI = 1.00-1.17). There was no difference in OS. IMRT delivery is recommended for all prostate EBRT in UK national guidance [20].

2.3.2.5 Image Guided Radiotherapy

A key concern in EBRT is the potential for geographic miss, i.e. missing the tumour and under-dosing. Although pre-treatment planning CTs had enabled accurate clinical target volume (CTV) delineation, such accuracy is only relevant at the time of the CT capture. By 1991, CTV motion of ±2cm relative to bony set-up anatomy (termed inter-fractional motion) was recognised by comparison of pre-treatment CT to simulator plain films with rectum and bladder opacified[5] [38]. Due to this CTV positional uncertainty, large CTV to planning target volume (PTV) expansion margins were applied, to prevent geographic miss. However, large CTV to PTV margins for the prostate are problematic. The key organs at risk (OARs) (bladder, rectum, bowel and penile bulb) sit in close approximation to the prostate. In such circumstance, large PTV margins may cause portions of the OARs to receive full prescription dose, making toxicity more likely.

This issue has been addressed through the development of image-guided radiotherapy (IGRT), allowing reduction of PTV margins through improved

[5] Neatly, this study also suggested the cause as differential filling by bladder and rectum, an issue still addressed in clinical practice through many centres recommending bowel and bladder preparation for prostate radiotherapy [258].

certainty in daily prostate positioning. Techniques have developed from megavoltage port films [39] through to real-time magnetic resonance imaging (MRI) [40]. In depth discussion of such methods is not a focus of this thesis, but is worth noting that without IGRT, the progression from moderate hypofractionation to ultrahypofractionation (discussed below) would have been very difficult. With (e.g.) 5 fractions, the consequence of geographic miss would be great; IGRT permits such delivery with confidence in prostate position, without the need for excessively large margins.

2.3.2.6 Stereotactic Body Radiotherapy

SBRT can be regarded in some respects as a fusion of the IMRT and IGRT processes. Tumour positioning must be highly certain in order to permit the extremely conformal radiotherapy doses defining SBRT. The first stereotactic radiotherapy system, for intra-cranial delivery, was developed in the 1970s[6]. Stereotactic intra-cranial radiotherapy is characterised by rigid immobilisation, with large fractions, commonly in single treatments, and highly conformal dose distribution [41]. Extension of this concept of ultrahypofractionated radiotherapy, with high conformality and steep dose gradients, to non-cranial sites, termed SBRT, was initially reported in the 1990s, across a range of body sites [42]. Platforms for technical delivery of SBRT in the prostate setting will be briefly considered.

[6] Lars Leksell at the Karolinska institute, Sweden

2.3.2.6.1 Co-Planar SBRT Delivery

Conventional linear accelerators (LINACs), with isocentric set-up, have a single rotational plane about fixed central point. SBRT can be delivered through step-and-shoot IMRT or, more commonly, VMAT. These machines have high dose rates, with modern flattening filter free (FFF) modes delivering dose rates around 1400 monitor units per minute at 6 MV [43]. With FFF delivery in combination with VMAT, prostate SBRT fractions can be delivered in 2 minutes or less [44]. This speed is important, since it minimises the probability of significant intra-fractional prostate motion, which doubles in standard deviation (SD) over four minutes versus two minutes [45]. A potential drawback of the conventional LINAC is that accuracy of delivery may be limited by lack of real-time intra-fractional IGRT. IGRT methods chosen by centres in PACE-B are analysed in **Chapter 5**.

2.3.2.6.2 Non Co-Planar Delivery

Radiotherapy treatment using beam angles outside those in a single 360-degree rotational arc is non-co-planar delivery. This increases the spatial flexibility for dose conformation by the treatment planning software [46]. Most commonly, such treatment is delivered with a dedicated machine, such as the CyberKnife (Accuray Incorporated, Sunnyvale, CA, USA) [47]. The CK houses a small LINAC on a 6-axis robotic arm, permitting non-co-planar beams from a large number of nodal positions. Additionally, real-time IGRT is provided by orthogonal kV imaging, matched to fiducial markers. Real-time positional correction is there possible, based on shifts in fiducial position.

2.3.2.6.3 Comparison of IMRT/VMAT and Cyberknife Methods

In 2003, King *et al* provided early evidence that CK may improve on an IMRT prostate plan, by reduced bladder and rectal dose-volume histogram (DVH) dose parameters [48]. CK and modern VMAT produce comparable plan distributions, but the delivery of VMAT is typically significantly quicker [49]. To my knowledge, comparative evidence for toxicity between the techniques does not exist. This will be investigated in **Chapter 7**.

2.3.3 Hypofractionated EBRT For Prostate Cancer

Having discussed the technological developments in prostate EBRT, our focus will turn towards dose-fractionation developments. In recent times, trials of EBRT as a primary treatment for nmPCa have trended towards hypofractionated schedules. The rationale and evidence underlying this is a core focus of this thesis, meriting a detailed examination of the evidence.

2.3.3.1 The Linear Quadratic Model

Since the early 20th century, radiotherapists have recognised that a fractionated course of radiation permits a higher total dose, with manageable increases in toxicity [50]. A seminal observation was that by Strandqvist in 1944, recognising that, when irradiating skin tumours, both tumour control and normal skin erythema followed a power law in relation to fractional dose [51]. The now dominant theory explaining this relationship is the LQ-model, first proposed in 1976, based on the fit of mouse skin reactions to fraction size [52]. It is commonly expressed in a form relevant for cell survival assays:

$$Surviving\ Fraction = e^{-(\alpha.D+\beta.D^2)} \quad (1)$$

Where:

α = Linear cell kill coefficient

β = Quadratic cell kill coefficient

D = Total dose in Gy (single fraction)

The component parts of this relationship will be considered in more detail.

2.3.3.2 Alphas and Betas in the LQ-model

For increasing dose, α describes linear change and β describes quadratic change, in cell kill. The nature of alpha-type damage has been characterised as “single-hit” lethality damage [53]. This can be thought of as a lethal deoxyribonucleic acid (DNA) insult, typically a double strand break (DSB), induced by a single photon track. More densely ionising radiation (e.g. carbon ions) exhibit greater alpha-type cell kill [54]. Beta-type cell kill is generally typified as “multi-hit” lethality; coming from the interaction of multiple photon/particle tracks [53]. For example, multiple non-lethal single-strand DNA breaks interacting to form a lethal DSB or complex DNA lesion.

The α/β ratio is effectively the inverse relative fraction size sensitivity of a tumour/tissue. Traditionally tumour control and acute normal tissue responses are considered to exhibit high α/β ratios (≈10 Gy), whilst late normal tissue responses exhibit low α/β ratios (≈3-5 Gy) [53]. However, at the turn of the millennium, this orthodoxy was challenged for PCa by two seminal reports. Firstly, Brenner & Hall used Poisson statistics to model stem cell survival, in two stages [55]. First estimating α from a curve fit to LDRBT dose-failure data, before using this value to estimate β from EBRT dose-

failure data, estimating a PCa α/β ratio of 1.5 Gy (95 % CI 0.8-2.2 Gy). Fowler *et al* corroborated this using LDRBT and EBRT dose-failure data, fitting α, β and sublethal repair half-life. They estimated a PCa α/β ratio of 1.49 Gy (95% CI 1.25–1.76 Gy) [6]. The cytotoxic effect of α/β ratio variation is examined in **Figure 2,** from a cell survival assay perspective.

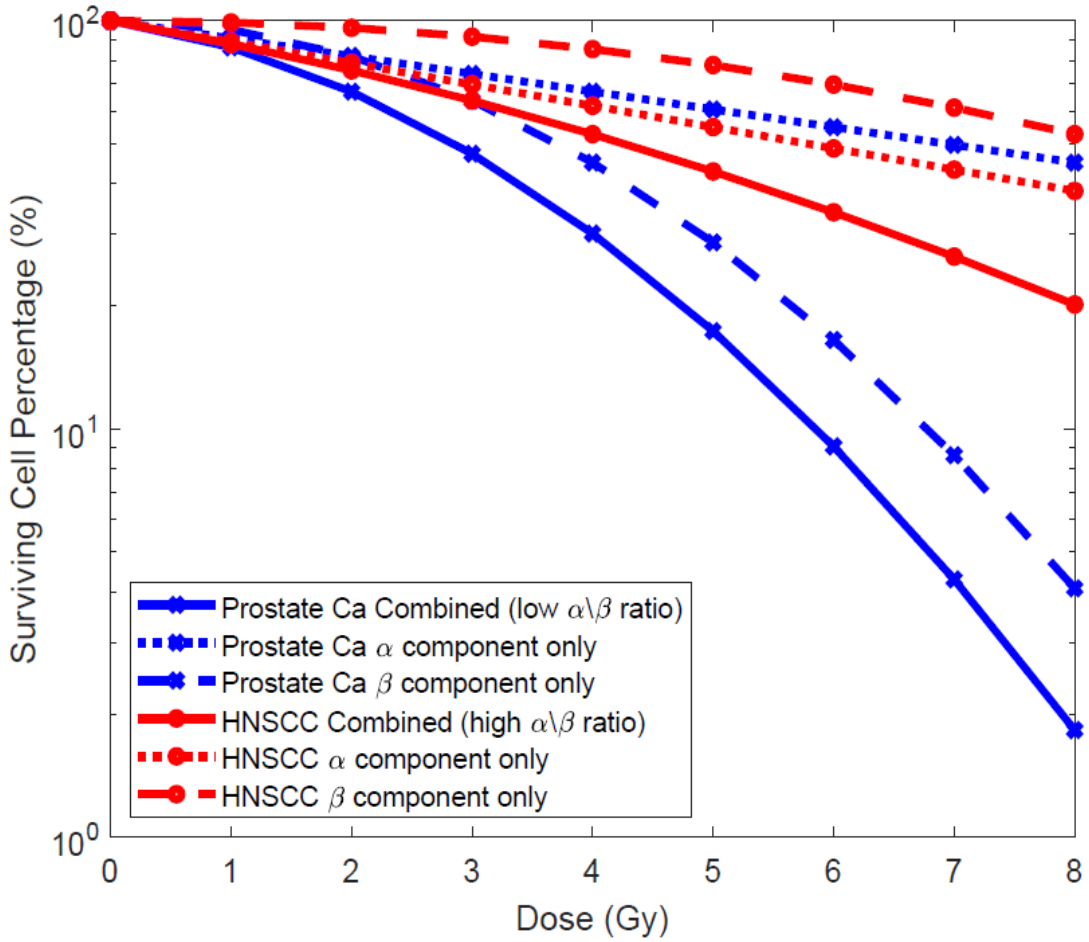


Figure 2. Cell Survival Curves For Tumours with High and Low α/β Ratios

Simulated cell survival curves demonstrating treatment with single fraction radiotherapy to prostate cancer (low α/β) and Head and Neck Squamous Cell Carcinoma (HNSCC) (high α/β) in-vitro. The "single hit lethality" alpha damage and "multi-hit lethality" beta damage are seen in respectively fine and semi-fine dash. The continuous lines show the sum of these processes. It can be seen that increasing dose per fraction results in a stronger effect on cell killing for prostate cancer, with a lower α/β ratio. Values used in simulation were: prostate cancer $\alpha = 0.1\ Gy^{-1}$ $\beta = 0.05\ Gy^{-1}$ HNSCC $\alpha = 0.12\ Gy^{-1}$ $\beta = 0.01\ Gy^{-1}$ [56]. Abbreviations: HNSCC = Head and neck squamous cell carcinoma.
Modified from article by self – thesis publishing rights retained [57].

2.3.3.3 The Biologically Effective Dose

Establishing that higher doses per fraction might kill PCa cells more efficiently is only half the story; the effect on normal tissues must be simultaneously considered. Comparison of dose-fraction regimens has been undertaken for thirty years through Fowler's biologically effective dose (BED), derived from the LQ-model [5,58]:

$$BED = N \cdot d \cdot \frac{1 + d}{\frac{\alpha}{\beta}} - K \cdot (T - T_k) \tag{2}$$

Where:
N = total number of fractions
d = dose per fraction
T = time of delivery (days)
T_k = kickoff day for repopulation (days)
K = daily BED equivalent repopulation ($Gy.days^{-1}$)

Derived from the LQ-model, for a given tissue, this allows a dose-fraction regimen to be reduced to a single number, the BED. The existence of a time factor for prostate tumour and related OARs is not universally accepted, so it is common to use only the left side terms. An alternative expression of BED is equivalent dose in 2 Gy fractions (EQD2), a form of units that many find easier to conceptualise:

$$EQD2 = N.d.\frac{(d + \alpha/\beta)}{(2 + \alpha/\beta)} \tag{3}$$

EQD2 will preferentially be used throughout this thesis.

2.3.3.4 Normal Tissue Complication Probability and Therapeutic Gain

With the putative low α/β (1.5 Gy) for PCa being less than estimates of α/β ratios for late normal tissue [6], a compelling rationale existed for hypofractionation, since the EQD2 will increase more for tumour than normal tissue. This would widen the therapeutic ratio, conceptually the difference between tumour control probability (TCP) and normal tissue complication probability (NTCP) for a given dose-fractionation schedule. TCP, the likelihood of local control following radiation, can be estimated by a number of mathematical frameworks, such as the Poisson model.

NTCP models, of relevance to this thesis, allow estimation of the probability of toxicity from the input of any combination of clinical, dosimetric or other parameters. These will be further discussed later in **Section 2.5**, but the Lyman-Kutcher-Burman (LKB) model [59] will be covered now, as an NTCP model with particular relevance to this thesis.

2.3.3.4.1 Lyman Kutcher-Burman Model

The LKB model allows the prediction of NTCP from dosimetric information. It was proposed by Kutcher and Burman [59], using effective volume methodology as an extension to Lyman's histogram reduction technique [60]. This effective volume method transforms a non-uniform DVH into a uniform one with equal NTCP as the original. The reduction considers the relative seriality of the organ involved, fitted as a parameter *n*. A small *n* implies a more serial organ, where loss of any tissue functional subunit may result in toxicity expression. The spinal cord is a typical example of a serial organ,

while the lung parenchyma would be an example of parallel architecture (**Figure 3**) [53].

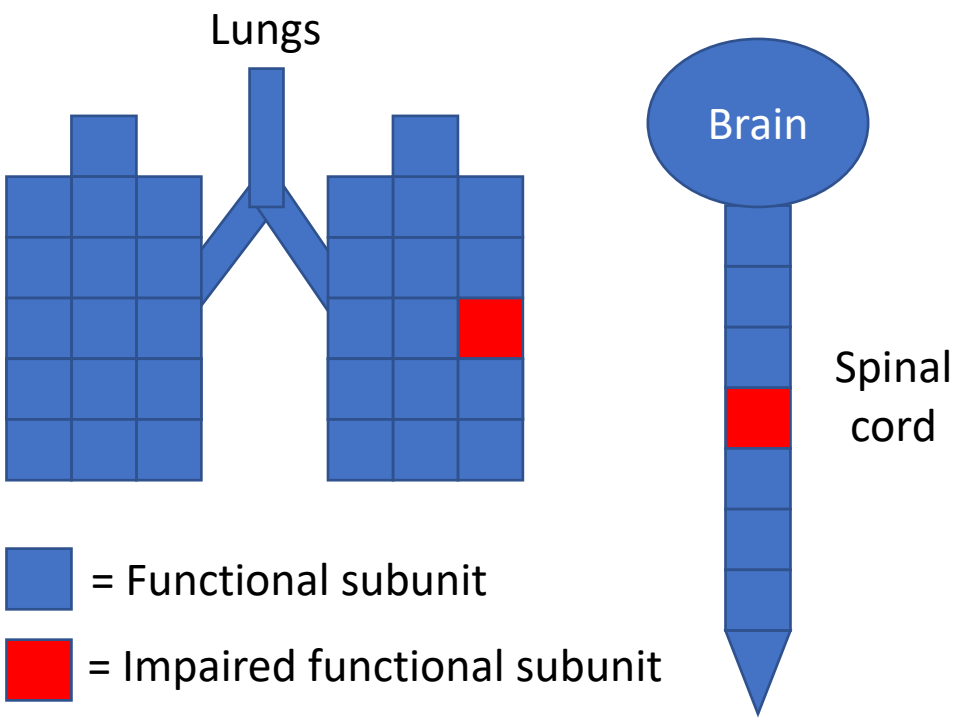


Figure 3. Example of Parallel Versus Serial Organ Architecture

The parallel organ (left, lungs) has a single functional subunit impaired, with only a proportional loss of function. The serial organ (right, spinal cord) also has a single functional subunit impaired, however here the function below this lesion will be lost.

The resultant effective dose value (termed effective dose, DEff) can be compared against fitted values for the tolerance dose for 50% toxicity (TD50)[7], based on a dose response slope of varying steepness (with steepness controlled by parameter *m*). Effects on the NTCP curve of *m* and TD50 adjustment are shown in **Figure 4**.

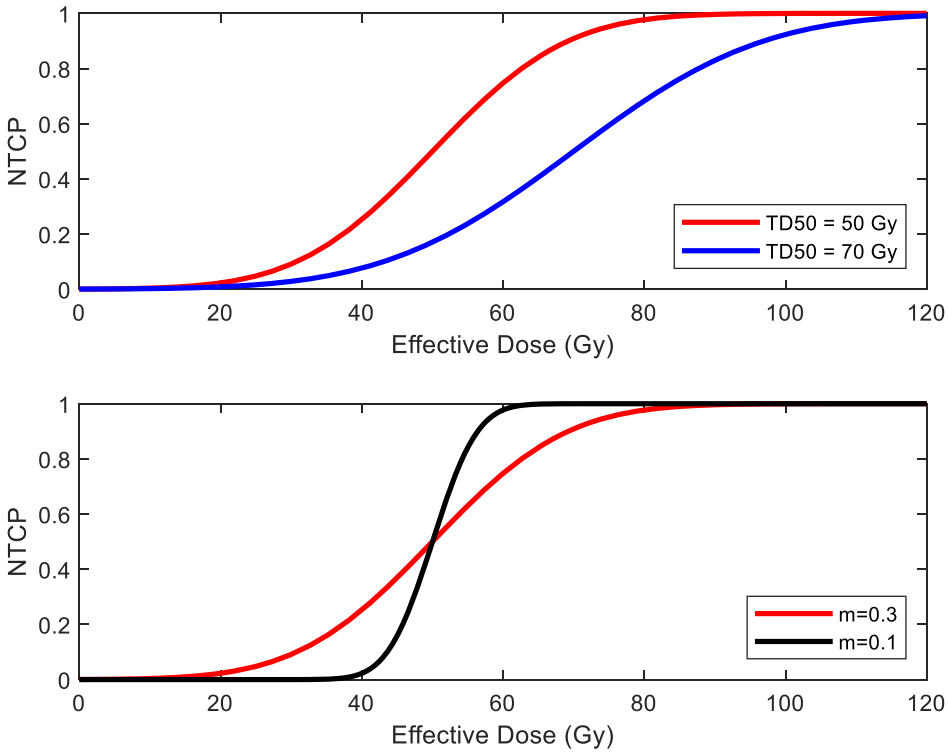


Figure 4. NTCP Curve Changes with Changes in LKB Model TD50 and m
LKB model parameter change examples are demonstrated. The upper plot shows the effect on the NTCP curve of increasing TD50 from 50 Gy (red) to 70 Gy (blue). It is important to note that by constraining at the origin (0,0), the alteration to TD50 also influences the steepness of the graph. The lower plot shows the effect on NTCP curve steepness of decreasing *m* from 0.3 (red) to 0.1 (black).

[7] The dose at which a 50% toxicity rate is expected.

Although it discards some information, the LKB methodology has been a popular way of handling DVH data, with parameter fits for various toxicities continuing to be published in first tier radiotherapy journals to the current day [61]. Two more recent extensions of the LKB model permit EQD2 correction [62] and also use of dose modifying factors (DMFs) [63]. The combination of these techniques in **Chapter 3** will permit estimation of late rectal α/β ratios for individual endpoints, while also checking sensitivity of these estimates to inclusion of various DMFs. The mathematical formulation of the LKB model will also be considered in **Chapter 3**.

2.3.3.4.2 Therapeutic Gain

Having introduced the concept of NTCP curves, **Figure 5** demonstrates graphically the concept of therapeutic gain. The alteration in fractionation regimen shifts the TCP curve relatively more than the NTCP. This might be expected for a more hypofractionated regimen if the tumour α/β ratio was lower than the corresponding late normal tissue α/β ratio. Of course, this is a simplification, since although the target (prostate) typically receives a fairly uniform dose, the OARs will receive highly heterogeneous doses (issue discussed with modelling in **Section 2.5.2.3.3**).

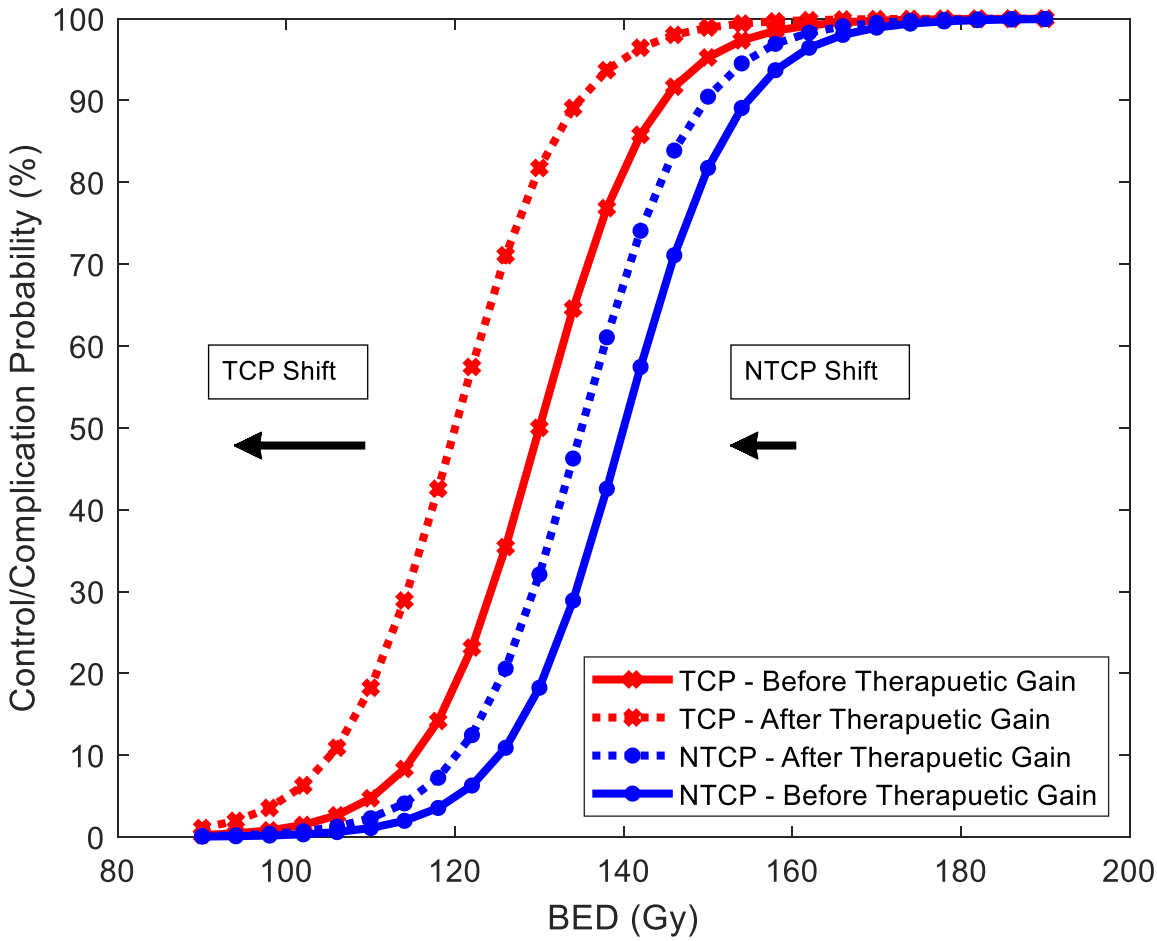


Figure 5. Stylised Therapeutic Gain Example

Stylised curves to demonstrate the concept of therapeutic gain. In this example, an effector (e.g. radiosensitiser) has increased the radiosensitivity of the tumour more than the normal tissue. Hence a wider gap between the two curves is produced and therapeutic gain is effected. In the case of fractionation changes, the curves would remain in the same position, but the BED delivered to tumour and normal tissue would alter depending on their respective α/β ratio.

Modified from article by self – thesis publishing rights retained [57].

2.3.3.4.3 The LQ-Model at Fractionation Extremes

The LQ-model has proven robust, still in regular use for comparison of dose-fractionation regimens, decades after its development [5]. However, it is worth noting that its reliability at the extremes of fraction size is uncertain. Low-dose hypersensitivity (which might occur during LDRBT) has been shown, with unexpectedly high cell kill at very low doses (<0.1 Gy), a possible function of cell cycle stage radiosensitivity [64].

At higher doses per fraction, (>6 Gy), different target mechanisms have been proposed, beyond the prototypical direct and indirect radical damage to tumour DNA. These include endothelial cell damage, ceramide-dependent apoptosis or immune-cell mediated death [65]. Whilst it has been questioned whether the LQ-model alone can predict dose-response at high doses per fraction. Brenner has summarised that in-vitro[8] and in-vivo animal data remain well fitted to the LQ-model up to 10 Gy per fraction, and probably up to 15 Gy [66]. The potential danger of assuming further extrapolation is illustrated by the failure of the LQ-model for prediction of tumour control with single fraction (19 Gy) HDRBT monotherapy[9] [67]. However, for lung SBRT (all <15 Gy / fraction), escalation in biologically effective dose, as, predicted by the LQ-model, appears to account for the increased tumour control observed [68].

[8] DNA flow cytometry and colony survival data
[9] Morton Phase II HDR monotherapy data: 19 Gy in 1 Fr vs 27 Gy in 2 fractions. 5-year bPFS 74% in single fraction arm versus 95% in the 2-fraction arm (p=0.001).

2.3.3.5 The Moderate Hypofractionation Era

2.3.3.5.1 Trials of Moderate Hypofractionation

This strong radiobiological rationale led to multiple phase II/III moderate hypofractionation trials; summarised in **Table 1**. Given the fewer patient visits required, most trials were designed to examine non-inferiority of moderate hypofractionation to conventional fractionation.

Meta-analysis (n=8146, 9 trials) comparing hypofractionated radiotherapy to conventional, found no significant differences in bPFS, OS, PCSSurv; nor acute GU, late GU and late GI toxicity [69]. Hypofractionation was associated with significantly worse acute GI toxicity (RR 1.47, $p<0.001$). A more tightly focussed meta-analysis, only including the large phase III non-inferiority studies (CHHiP 60 Gy vs 74 Gy, PROFIT and RTOG-0415) [70] also suggested hypofractionation worsened acute G2+ GI toxicity, but noted statistically significantly improved disease free survival (HR 0.87, 95% CI 0.76-0.99).

Table 1. Phase II/III Moderate Hypofractionation Randomised Trials

Trial & Countries	N (Trial Phase)	NCCN Risk Group	Test Arms (Gy/#) (timing)	Control Arms (Gy/#) (timing)	Test Fraction (Gy)	Hormone Therapy	Median F/U at time of survival analysis (Year published)	Disease Free Survival Test v Control	G2+ Late GI Tox. Cumulative Test vs Control	G2+ Late GU Tox. Cumulative Test vs Control
CHHiP [1] UK / Swiss/ New Zealand /Ireland	3216 (III)	Low to High	60 / 20 daily 57 / 19 daily	74 / 37 daily	3 Gy	Optional for low risk patients; 3-6 months	62.4 mon. 2016	90.6% v 88.3% HR 0·84 Non-inferior 85.9% v 88.3% HR 1.2 Not non-inferior	11.9% v 13.7% Non-Significant 11.3% v 13.7% Non-Significant	11.7% v 9.1% Non-Significant 6.6% v 9.1% Non-Significant
PROFIT [3] Canada	1206 (III)	Intermediate	60 / 20 daily	78 / 39 daily	3 Gy	No	72 mon. 2017	85.0% v 85.0% HR 0.96 Non-inferior	8.9% v 14.0% RR 0.63*	22.2% v 22.0% Non-Significant
RTOG 0415 [2] USA	1115 (III)	Low	70 / 28 daily	73.8 / 41 daily	2.5 Gy	No	70 mon. 2016	86.3% v 85.3% HR 0.85 Non-inferior	22.4% v 14.0% RR 1.55-1.59	29.7% v 22.8% RR 1.31-1.56
HYPRO [4] Netherlands	820 (III)	Intermediate & High	64.6 / 19 3# per week	78 / 39 daily	3.4 Gy	67% yes; Ave. 32 mon.	60 mon. 2016	80.5% v 77.1% HR 0·86, Not superior	21.9% vs 17.7 %, **HR 1.19** – not non-inferior	41.3% v 39.0% **HR 1.16** – not non-inferior**
Fox-Chase [71] USA	303 (II)	Low - High	70.2 / 26 daily	76 / 38 daily	2.7 Gy	50%; 4 or 24 mon.	68.4 mon. 2013	76.7% v 78.6% HR not reported Not superior	18.1% vs 22.5% Non-Significant	21.5% vs 13.4% Non-Significant
Adelaide Trial [72] Australia	217 (II)	Low-Intermediate	55 / 20 Daily	64 / 32 Daily	2.75 Gy	No	90 mon. 2011	53% vs 34% HR 0.65	GI toxicity not significantly different	GU toxicity not significantly different
MD Anderson [73] USA	206 (II)	Low-Int (1% high)	72 / 30 Daily	75.6 / 42 Daily	2.4 Gy	24%	100 mon. 2016	89.3% v 84.6% HR not reported p=0.034	10% vs 5.1% Non-Significant	15.8% v 16.5% Non-Significant
Roma Trial [74] Italy	168 (II)	High	62 / 20 Over 5 weeks	80 / 40# daily	3.1 Gy	All, 9 months	108 mon. 2017	72% v 65% HR not reported Non-Significant	13.5% v 15.4% Non-Significant	14% v 21% Non-Significant

Statistically significant results in **bold**. Abbreviations: n = number of patients; # = fractions; F/U = follow up; v = versus; GI = gastrointestinal; GU = genitourinary; G2+ = grade 2 or more; Tox. = toxicity; mon. = months; Ave. = average; NS = not statistically significant; RR = relative risk

* = Value not reported in paper (only as significant), but calculated from reported values

** = Note also for HYPRO that late G3+ GU significantly worse for test arm (19.0%) versus control (12.9%), p=0.021

2.3.3.5.2 Excess Toxicity in the HYPRO Trial

Unlike the three other phase III trials, the Dutch HYPRO trial may have underestimated the likely normal tissue EQD2 of the hypofractionated arm, resulting in higher late G3+ GU toxicity for the 64.6 Gy / 19 # arm (3.4 Gy per fraction) versus the conventional 78 Gy / 39 # arm (19% vs 12.9%, p=0.021) [75]. Despite superiority design, disease control was unfortunately not significantly improved, for hypofractionation vs conventional EBRT (80.5% v 77.1% at 5 years, HR 0.85, 95% CI 0.63-1.16, p=0·36).

The assumptions made in the design of each of the major moderate hypofractionation trials are therefore of interest. The postulated advantage of each trial's test dose-fractionation regimen was justified through the putative α/β ratio of PCa being low, however the exact assumptions of α/β ratio varied, as did those for late normal tissue side effects. The assumptions made in each trial are shown in **Table 2**.

Table 2. Hypofractionation Trial Design α/β Ratio Assumptions

Note the higher α/β ratio assumed for late reactions in HYPRO (bold).

Trial	Prostate Assumptions			Rectum Assumptions			Bladder Assumptions		
	α/β Ratio (Gy)	Test EQD2 (Gy)	Control EQD2 (Gy)	α/β Ratio (Gy)	Test EQD2 (Gy)	Control EQD2 (Gy)	α/β Ratio (Gy)	Test EQD2 (Gy)	Control EQD2 (Gy)
HYPRO	1.5	90.4	78	**4 - 6**	79.7 - 76	78	**4 - 6**	79.7 - 76	78
CHHiP 57Gy	1.5 - 2.5	73.3 – 69.7	74	3	68.4	74	3	68.4	74
CHHiP 60Gy	1.5 - 2.5	77.1 – 73.3	74	3	72.0	74	3	72.0	74
PROFIT	1 - 3	80 - 72	78	3 - 5	72 - 68.6	78	3 - 5	72 - 68.6	78
RTOG 0415	3	77.0	70.8	3	77	70.8	3	77	70.8

It can be seen that the HYPRO trial was unique in assuming a high α/β ratio for the late normal tissues. Their worst-case scenario (late normal α/β = 4 Gy) would result in a slight dose escalation EQD2 79.7 $Gy_{\alpha/\beta=4}$ compared to the control 78 Gy. However, both GI and GU toxicity were significantly worse, an outcome that might be expected if the true α/β was <4 Gy, since the hypofractionated arm EQD2 would be >80 Gy.

2.3.4 Normal Tissue α/β Ratios in Prostate Radiotherapy

2.3.4.1 Overview of Fraction Size Sensitivity for Normal Tissues

As outlined above, the unexpectedly toxic results of the HYPRO trial were most likely the result of EQD2 underestimation for late normal tissues, caused by higher late normal tissue α/β ratio estimates than the other moderate hypofractionation trials. This section will examine evidence for the fraction size sensitivity of normal tissues, with a key focus on GI toxicity.

2.3.4.2 Acute and Late Side Effects of Radiation Treatment

Normal tissue radiation responses can be divided into early/acute reactions (defined here as ≤ 3 months from end of radiotherapy) and late reactions (defined here as >3 months). Acute reactions most often occur in those tissues with a significant component of replicating cells, for example the gastrointestinal tract (e.g. mucositis, diarrhoea) and skin (e.g. dry or moist erythema) [53]. The radiation insult temporarily halts stem cell division, slowly denuding the tissue surface through loss of non-replicating cells. This is followed by recovery within a few weeks of radiation completion; once stem cell division recommences and the replicative layer is repopulated. The

pathogenesis of late reactions are less clearly understood, with different mechanisms such as vascular injury, fibrotic repair and angiogenesis likely contributing to different side effects [53].

2.3.4.3 Animal Estimates of Rectal α/β Ratio

Early estimation of α/β ratios tended to be derived from dose-fractionation experiments in animal models. The applicability of such information to human patients is questionable. We shall briefly look at an example of a mouse study and then some estimates produced for rectal endpoints.

2.3.4.3.1 An Example of Mouse Model Rectal α/β Ratio Estimation

In 1988, Van der Kogel *et al* used 250kVp X-rays to irradiate a 2.5cm sup-inf single field of lower mouse abdomens (encompassing rectum in 90% isodose) [76]. Doses varied to fit dose-response curves, with 1,4 or 10 fractions. Although they reported a late rectal α/β ratio of 6.5 Gy, this was based upon fitting just 3 different dose-per-fraction isoeffective data points, with an endpoint of death at 250 days.

2.3.4.3.2 The Limited Generalisability of Animal Data to Humans

Fortunately for prostate EBRT patients, we can readily avoid irradiating OARs to doses causing death by 250 days. Few mouse models have had recognisable endpoints for humans, for example Dewit *et al* estimating rectal stenosis α/β ratio of 4.4 Gy (95% CI 1.6-7.7) or anal discharge α/β ratio 5.3 Gy (95% CI 3.2-7.9) [77]. Gasinska *et al* (1993) estimated a range of early and late rectal α/β ratios (lethality, weight loss), with late estimates at 6.4 Gy

or above; interestingly the late α/β ratio for short faeces (perhaps analogous to a reduced rectal capacity in humans) was lower at 1.4 Gy [78].

2.3.4.3.3 Dangers of Animal Estimate Usage

However, as a whole, the animal literature lacks key human endpoints of interest, (e.g. rectal bleeding), targets much higher event rates than humans and is hampered by crude irradiation techniques. It is likely that the high α/β ratio (4-6 Gy) chosen for late rectal side effects in HYPRO, was derived from a contemporary paper by Fowler [8]. He analysed possible hypofractionated regimens, generally assuming late rectal α/β = 3 Gy, but in the discussion suggested that an α/β ratio of 4-6 Gy might be plausible. This was based on selected late rectal α/β estimates from six animal experiments fitted to short faeces [79], rectal stenosis/obstruction [77,78,80,81], late lethality [78,79,82]. The excess toxicity of the HYPRO trial highlights the risk of extrapolating animal data, based predominantly on endpoints rarely seen in humans.

2.3.4.4 Late Rectal Toxicity α/β Ratio Estimates in Humans

Estimates of human late rectal α/β ratio(s) are few; three published studies have examined data from PCa EBRT trials, plus one abstract-only study. All of them examine rectal toxicity as a whole, rather than specific endpoints, such as rectal bleeding, proctitis etc.

2.3.4.4.1 Full Length Article Studies of Human Late Rectal α/β Ratio

Brenner estimated late G2+ RTOG α/β ratio=5.4 Gy, by fitting a curve to toxicity outcomes of eight dose-fractionation regimens, from four USA/Japan

PCa EBRT trials[10], with dose per fraction 1.8 – 3.0 Gy [7]. A key critique is that almost all patients received 1.8-2 Gy per fraction (2254 patients [83–85]), with only 52 receiving 3 Gy per fraction [86]. This lack of heterogeneity in dose per fraction will hamper modelling of the α/β ratio estimate when fitting. Another critique is that, in analysing all patients from one arm as a single datapoint, the relationships between individual DVHs and subsequent toxicities is discarded (dose heterogeneity discussed in **Section 2.5.2.3.3**).

Marzi examined 162 patients in the Roma hypofractionation trial (recall **Table 1**), with 13% RTOG G2+ late GI toxicity [62]. An LKB model of the rectal wall (i.e. hollow viscera) DVH was corrected by EQD2 on a per-dose-bin basis and fitted the LKB model by maximum likelihood. They estimated RTOG G2+ late GI toxicity to have an α/β ratio of 2.3 Gy (95 % CI 1.1 – 5.6 Gy). However, *m* and *n* were fixed to 0.15 and 0.12 respectively, based on prior work by Burman [87]. The reduction of variance from fixing these means confidence intervals are artificially narrow, an approach likely taken due to the low number of patients.

A third estimate of rectal α/β ratio comes from work by Tucker *et al* on 509 patients from RTOG 94-06 [88], using a similar LKB model with EQD2-corrected dose bins. 15% (77/509) experienced G2+ RTOG late rectal toxicity. The parameters were modelled with a generalised version of the LKB, incorporating the toxicity data in a time-to-event manner; however, this

[10] European/Australian data was excluded on grounds of differing diabetes rate. The influence of diabetes on rectal α/β will be investigated in Chapter 3

did not make a difference to other parameter estimates. Wide confidence intervals were found: α/β 4.8 Gy (**68 % CI** 0.6 – 46 Gy) estimated for late RTOG G2+ GI toxicity. This is likely in part caused by limited prescribed dose heterogeneity (1.8-2 Gy). No statistically significant improvement in model performance was demonstrated between standard and EQD2-corrected LKB models.

2.3.4.4.2 Abstract-Only Studies of Human Late Rectal α/β Ratio

Zhu and colleagues published an abstract estimating late rectal (G2+ Late Effects Normal Tissue: Subjective, Objective & Management (LENTSOM)) α/β ratio, amongst 213 patients receiving either 70 Gy in 35 fractions versus 70 Gy in 28 fractions [89]. The Marzi EQD2-corrected LKB model methodology [62] was utilised. They estimated the late rectal α/β ratio was 7.17 Gy (95% CI 5.21 – 9.13 Gy). Full length publication is awaited.

2.3.4.4.3 Poor Agreement Amongst Human Studies

In summary, estimates range from 2.3 to 7.7 Gy for human late rectal α/β ratio, with confidence intervals very wide or artificially constrained. As can be seen from the HYPRO trial, small differences in α/β ratio estimation may be responsible for large differences in toxicity. More accurate estimates are pursued in **Chapter 3.**

2.3.5 Ultrahypofractionation for Prostate Cancer

The moderate hypofractionation era trials were a resounding success, vindicating shorter, cheaper regimens with similar efficacy. The logical extension from this is ultrahypofractionated radiotherapy (≥5 Gy/fraction)

2.3.5.1 Early Experience with Ultrahypofractionation

Ultrahypofractionated EBRT for nmPCa has been delivered since the 1960s. In 1990, Lloyd-Davies *et al* reported a series of 189 patients with T1-4NXMX PCa, treated between 1966–1984 with 36 Gy in 6 fractions over 3 weeks [90]. Late side effects were noted as surprisingly few, given LQ-model expectations, although detailed quantification was not provided [91].

2.3.5.2 Efficacy of Modern Ultrahypofractionation

The first report of ultrahypofractionated SBRT for PCa is a phase I/II study in 40 low risk patients, by Madsen *et al* in 2007[11] [92]. Phase II data with a dose regimen similar to the PACE-B trial (36.25 Gy in 5 fractions) was published in 2009 [93]. Pooled individual patient data level analysis of 2142 patients treated with SBRT has shown excellent 7-year biochemical relapse rates for both low (4.5%) and intermediate (10.2%) risk disease [94]. Results from the Phase III HYPO-RT-PC trial [95] will be discussed in **Chapter 4**.

[11] 33.5 Gy in 5 fractions daily, with 100% of the CTV receiving 90% dose (i.e. D90 = 100%), achieving 48-month bPFS of 90%

2.3.5.3 Acute Toxicity with Hypofractionation

With the promising efficacy signals from ultrahypofractionation, it is worth considering the acute toxicity associated with such an accelerated regimen. It is commonly understood that acute toxicity reactions have a higher α/β value than late toxicity reactions, for example α/β = 10 Gy [53]. Therefore, if the therapeutic ratio for late reactions would be improved by hypofractionation (PCa α/β ratio < late α/β ratio) then it might be anticipated that such therapeutic gain would be even greater for acute reactions (PCa α/β ratio << acute α/β ratio).

The acute toxicity seen for acute GI and GU toxicity in the moderate hypofractionation phase III trials is shown in **Table 3**, along with the $EQD2_{\alpha/\beta=10Gy}$ for each arm. The hypofractionated arm $EQD2_{\alpha/\beta=10Gy}$ was lower than control in CHHiP-60Gy, CHHiP-57Gy, PROFIT and HYPRO. However, in all cases the G2+ acute GI toxicity was worse than the conventional arm. RTOG-0415 measured acute toxicity using CTCAE rather than RTOG methodology, which may potentially have influenced frequency of events.

Table 3. Acute Toxicity in the Moderate Hypofractionation Trials

The worst toxicity experienced in each arm, with $EQD2_{\alpha/\beta=10Gy}$. These are compared for hypofractionated minus control arm on the right side of the table. For CHHiP, each hypofractionated arm has been compared separately. Note that although $EQD2_{\alpha/\beta=10Gy}$ is lower for CHHiP, PROFIT and HYPRO, the G2+ GI toxicity is increased and GU G2+ toxicity is similar or slightly worse (bold text).

Novel Abbreviations: Fr = fractions.

Trial	Hypofractionated Arm							Control Arm							Hypofractionated Minus Control Difference				
	Dose (Gy)	Fr	EQD2 (Gy_{10})	GI G2+ (%)	GI G3+ (%)	GU G2+ (%)	GU G3+ (%)	Dose (Gy)	Fr	EQD2 (Gy_{10})	GI G2+ (%)	GI G3+ (%)	GU G2+ (%)	GU G3+ (%)	EQD2 Diff (Gy_{10})	GI G2+ Diff (%)	GI G3+ Diff (%)	GU G2+ Diff (%)	GU G3+ Diff (%)
CHHiP 60	60	20	65	38	N/R	49	N/R	74	37	74	25	N/R	46	N/R	**-9**	**13**	N/A	**3**	N/A
CHHiP 57	57	19	61.8	38	N/R	46	N/R	74	37	74	25	N/R	46	N/R	**-12.3**	**13**	N/A	**0**	N/A
PROFIT	60	20	65	16.7	0.7	30.9	3.9	78	39	78	10.5	0.5	31	4	**-13**	**6.2**	0.2	**-0.1**	-0.1
RTOG-0415	70	28	72.9	10.7	0.8	27	3.3	73.8	41	72.6	10.3	0.6	27.1	2.4	0.3	0.4	0.2	-0.1	0.9
HYPRO	64.6	19	72.1	42	5.7	60.5	20.3	78	39	78	31.2	4.6	57.8	17.6	**-5.9**	**10.8**	1.1	**2.7**	2.7

2.3.5.3.1 The Role of Overall Treatment Time in Acute Toxicity

This discrepancy may be explained by the impact of overall treatment time (days) on acute toxicity. The head and neck cancer DAHANCA-6 trial neatly demonstrates this [9]. In that trial, the same dose and fractionation delivered as 6 fractions/week, rather than 5, caused significantly increased acute mucositis (as well as improved locoregional control). For acute epithelial reactions, radiotherapy impairs the replicating compartment of an epithelial layer – as the non-dividing layers slough, mucositis will occur [53]. However, with longer overall treatment time, repopulation can occur, which compensates for some dose per day. Split course radiotherapy improves acute toxicity in head and neck cancer, but worsens tumour control due to simultaneous tumour repopulation [96].

2.3.5.3.2 Acute Toxicity with Ultrahypofractionation

The expected effect of ultrahypofractionation on acute toxicity is thus governed by two competing factors: reducing $EQD2_{\alpha/\beta=10Gy}$ versus accelerating treatment time. **Table 4** summarises the acute toxicity in prospectively collected phase I/II SBRT studies (n=1775). Wide variation in the proportions experiencing GI/GU G2+ toxicity can be appreciated, with variation in scales (RTOG vs Common Toxicity Criteria Adverse Events (CTCAE)) making meta-analysis difficult.

Table 4 Caption [*Table found overleaf*]
* Grade percentages estimated from figures in paper
Novel Abbreviations: Alt. = alternate; C. = centres; fidx = fiducials; CBCT = cone beam CT; HR = high risk; IR = intermediate risk; Intra-kV = intra-fractional kV; LR = low risk; NS = Not Stated; Post. = Posterior; Pros. = prostate; SV = Seminal Vesicles;

Table 4. Acute Toxicity After Prostate SBRT: Low-Intermediate Risk Patients

Trial	C.	n	Risk	ADT	CTV	Margins (mm)	Plan Extras	Machine/IGRT	PTV Dose (Gy)	Fr	Frequency	Score	Acute GI		Acute GU	
													G2 (%)	G3/4 (%)	G2 (%)	G3/4 (%)
Meier [97] NCT00643994	21	309	LR-IR	No	Pros. Only	5 (3 Post.)	MRI	CK fidx + Intra-kV	36.25	5	Daily/ Alt Days	CTCAE	8.1	0	26	0
Katz [98]	1	304	LR-HR	18.8%	LR - Pros. Only IR - Pros. + Prox SV	3-5 Some 8	± MRI	CK fidx + Intra-kV	35 (n=50) 36.25 (n=254)	5	Daily	RTOG	4 3.6	0 0	4 4.7	0 0
Fuller [99] NCT00643617	7	259	LR-IR	No	LR - Pros. Only IR - Pros. + 1cm SV	2 (0 Post.)	Foley-CT ± MRI	CK fidx + Intra-kV	38	4	Daily	CTCAE	6.9	0	35.1	1.1
Quon [100] NCT01423474	3	152	LR-Low IR	<5%	Pros. Only	0·3	No	LINAC IMRT fidx + kV/CBCT	38	5	Alt Days **Or** Weekly	RTOG	18·4 10·8	0 0	31·6 33·8	1.3 2.7
Zelefsky [101]	1	136	LR-IR	No	Pros. + SV	5 (3 Post.)	No	LINAC IMRT Calypso or fidx	32.5-40	5	Alt Days	CTCAE	4.4	0	16.2	0
Mantz [102]	1	102	LR	No	Pros. Only	2	No	LINAC IMRT Calpyso	40	5	Alt Days	CTCAE	?	?	?	2 Pts
Hannan [103]	5	91	LR-Low IR	16.5%	Pros. Only	3	± MRI	LINAC IMRT Calypso or fidx	45-50	5	Alt Days	CTCAE	17	0	20	0
Loblaw [104] NCT01578902	1	84	LR	1%	Pros. Only	4	No	LINAC IMRT fidx + kV	33.25	5	Weekly	CTCAE	10	0	19	1
Jackson [105] NCT01288534	5	66	LR-Low IR	No	Pros. Only	3	MRI or Foley-CT	LINAC IMRT Calypso	37	5	Every 3 Days	CTCAE	4	0	23	0
Boyer [106] NCT00941915	3	60	LR-Low IR	No	Pros. Only	5 (3 Post.)	MRI	LINAC IMRT Various IGRT	37	5	Alt Days	CTCAE	5	0	25	0
McBride [107]	4	45	LR	No	Pros. Only	5 (3 Post.)	± MRI ± Foley-CT	CK fidx + Intra-kV	36.25-37.5	5	Max 10 days	CTCAE	7	0	19	2.2
Bolzicco [108]	1	45	LR-IR	37%	Pros. Only	5 (3 Post.)	Catheter	CK fidx + Intra-kV	35	5	Daily	RTOG	24.4	0	11.1	0
Alongi [44]	NS	42	LR-IR	Some	NS	NS	NS	LINAC IMRT CBCT±fidx	35 (LR) 37.5 (IR)	5	Daily	CTCAE	5	0	13	0
Alongi [109]	NS	40	LR-IR	Some	LR - Pros. Only IR - Pros+1/3SV	3-5	MRI	LINAC IMRT CBCT±fidx	35	5	Alt Days	CTCAE	10	0	40	0
Madsen [92]	1	40	LR	NS	Pros. Only	4-5	MRI	3DCRT fidx + kV	33.5	5	Daily	CTCAE	13	0	20.5	2.5

Phase III comparative evidence is therefore highly desirable. At the time of commencing this thesis, no phase III data had been published on the comparative acute toxicity with SBRT. **Chapter 4 & Chapter 5** therefore set out to provide detailed analysis of acute toxicity occurring within the PACE-B trial. Results from the recently reported HYPO-RT-PC trial [110] are evaluated in the discussion sections of those chapters.

2.3.5.3.3 Consequential Late Effects

It is worth briefly touching on consequential late effects (CLEs). These are late reactions arising as a consequence of severe acute reactions, therefore being sensitive to total dose and overall treatment time, but relatively insensitive to fraction size [111]. This contrasts with typical late effects, which are strongly dependent on fraction size, reflecting the lower α/β ratio for typical late effects relative to acute reactions, the precursor of CLEs. Increased acute reactions with ultrahypofractionation may thus portend increased late reactions in the form of CLEs.

2.4 Endpoints in Prostate Radiotherapy

The discussion will now turn to the endpoints measured in prostate EBRT trials. For any oncology RCT comparing a new treatment to standard-of-care, endpoints/measures must be selected to answer three key questions [112]:

1) Does it improve overall survival?
2) Does it improve patient quality of life?
3) Is it cheaper?

With these three metrics, a comparative assessment of cost-effectiveness can be performed[12] [113]. However, for nmPCa treatments, demonstrating an OS benefit (or reliable surrogate) is challenging.

2.4.1 Hypofractionation Trial Efficacy Endpoints

The ICECAP study demonstrated that PSA-based metrics (e.g. bPFS) are not robust OS surrogates [114]. This is of importance given the positive phase III moderate hypofractionation studies showed non-inferiority of bPFS [1–3]. Similarly, for ultrahypofractionation, non-inferiority of bPFS is the primary endpoint for HYPO-RT-PC and PACE-B. Given the non-inferioty designs and difficulties of bPFS:OS correlation, adoption of these technologies is therefore more dependent on patient side effects and treatment cost. Assaying side effects of treatment on the patient is usually as toxicity (CROs) or quality of life (PROs).

[12] Typically using the quality adjusted life year

2.4.2 Clinician Reported Outcome Metrics

Assessment of toxicity severity is made challenging by the absence of a measurable unit; a situation which has leant itself to ordinal scales. As early as the 1950s, a four-level grading system for acute skin toxicity was reported[13] [115]. Unfortunately, progress over subsequent decades has not led to a single universally accepted scale for grading radiotherapy toxicities.

2.4.2.1 The WHO Handbook

Development of comprehensive ordinal scales for oncological toxicities began with the World Health Organisation (WHO) Handbook for Reporting Results of Cancer Treatment, 1979 [116]. It included grade 0-4 ordinal scales for several acute toxicities, although principally for systemic therapy toxicities (haematological, renal etc). Late effects reporting was encouraged, with a template for collection, but without ordinal scales for specific late toxicities.

2.4.2.2 Common Toxicity Criteria Adverse Events

The WHO scales were gradually superseded by the 1988 Common Toxicity Criteria (referred to as Common Toxicity Criteria Adverse Events (CTCAE) hereafter) [117]. The CTCAE provided ordinal CRO scales for a wider range of oncological toxicities. Unfortunately, like the WHO handbook, CTCAE was principally designed to assess systemic therapy toxicities, meaning radiotherapy practitioners did not feel that it was sufficient as a codified assessment of late effects [118].

[13] Dry erythema, dry desquamation, moist desquamation, necrosis

2.4.2.3 The RTOG Radiotherapy Toxicity Scales

The omission of late effects in CTCAE v1.0 was unfortunate, since during the 1980s, members of RTOG had developed scales for the reporting of acute and late radiotherapy effects. These were distributed with the RTOG trials and endorsed by the European Organisation for Research and Treatment of Cancer (EORTC) (although the scale will be referred to as RTOG scale hereafter), but not published in comprehensive form until 1995 [119,120]. The RTOG scales differed from CTCAE in giving a grade for an organ system, rather than on a per symptom/syndrome basis. Wide practitioner variety has long been noted in the reporting of this scale [120].

2.4.2.4 LENTSOM

While acute toxicity of radiotherapy could be addressed by either CTCAE or RTOG scales, by 1995 a further RTOG/EORTC late effects conference was convened to try and develop new universal scales for late toxicities [118]. Known as LENTSOM[14], these new scales aimed to reduce subjectivity in the reporting of late effects [121], including subjective, objective and management scores for each organ system.

2.4.2.5 Current Practice in Late Toxicity Reporting

The story of major convergences in radiotherapy toxicity reporting ceases with the 1995 LENTSOM conference. There remains substantial heterogeneity in the usage of scales to report toxicity, with RTOG and

[14] Recall abbreviation as Late Effects Normal Tissues – Subjective, Objective, Management

CTCAE used for acute toxicity and RTOG, LENTSOM and CTCAE for late toxicity [122]. Analysis of publications over five years (2010-2015) has shown similar use of RTOG and CTCAE (40-50% each, with a slow trend towards CTCAE), but only 5% usage of LENTSOM.

2.4.2.6 Clinician Reported Metrics in Hypofractionation Trials

Table 5 shows the heterogeneity of CRO scales across phase III hypofractionation trials. Both trials analysed in this thesis (CHHiP & PACE-B) use multiple CRO scales. CHHiP includes the Royal Marsden Hospital (RMH) scale; an ordinal grading system developed locally. This presents challenges in terms of multiplicity of testing, with data from each scale requiring analysis, increasing the potential for type I statistical errors. Overall, the RTOG scale is most consistently chosen for acute and late toxicity[15]. These considerations contributed to primary endpoint definition in **Chapter 4**.

Table 5. CRO Scales Used in Phase III Hypofractionation Studies

Clinician Reported Outcomes	CHHiP	RTOG-0415	PROFIT	HYPRO	PACE-B	HYPO-RT-PC
Acute						
RTOG	Yes		Yes	Yes	Yes	Yes
CTCAE		Yes			Yes	
Late						
CTCAE		Yes			Yes	
LENTSOM	Yes					
RMH	Yes					
RTOG	Yes		Yes	Yes	Yes	Yes

[15] Omitted only by the eponymous trial, RTOG-0415.

2.4.3 Patient Reported Metrics

Beyond CROs, another measure of the side effects from therapy are PROs; often termed QoL data. Many PRO scales exist, so this discussion will focus on those used in phase III hypofractionation trials.

2.4.3.1 Generic Health Quality of Life Questionnaires

Some PRO tools assess general QoL. Two cancer specific examples are: the EORTC core QoL questionnaire (EORTC QLQ-C30)[16] [123]; the Functional Assessment of Cancer Therapy - General (FACT-G) [124]. The Short Form 36 (SF-36)[17] is a commonly used generic QoL instrument, developed for non-cancer-specific usage [125,126].

From a health economic perspective, the most important generic health questionnaire is the EuroQol-5D (EQ-5D)[18] [127]. Since 2008, this is has been the gold-standard tool for quality adjusted life years calculation by NICE during UK health economic appraisals [128]. Other generic QoL tools need to be mapped to EQ-5D for such appraisals [129].

2.4.3.2 Prostate Cancer Specific Questionnaires

Some extensions to the generic cancer questionnaires cover PCa specific issues: e.g. QLQ-PR25, supplementing QLQ-C30 with urinary, bowel,

[16] Derived from the original EORTC QLQ-36.
[17] Sometime used in the abbreviated 12 questions SF-12 format.
[18] Specifically, the original questionnaire which is now called the EQ-5D-3L.

hormonal and sexual change questions [130]. FACT-P is a similar 12-question extension to the FACT-G questionnaire [131].

The University of California Los Angeles Prostate Cancer Index (UCLA-PCI) supplements the SF-36 with 20 PCa specific questions [132]. Although the prostate section is non-proprietary, the SF-36 component is subject to copyright. The Expanded Prostate Cancer Index Composite (EPIC) is a 50-question, validated, free-to-use expansion of the 20 PCa questions in the UCLA-PCI, aimed at urinary, bowel, sexual and hormonal domains [133]. Compared to UCLA-PCI, it provides more detailed interrogation of certain symptoms such as haematuria, urinary obstruction and hormonal symptoms. An abbreviated 26-question version (EPIC-26) has been validated for nmPCa, making it a convenient option in EBRT studies [134].

The Prostate Cancer Symptom Scale (PCSS) is an 18-question instrument developed specifically for the PCa EBRT setting, but with, to my knowledge, no usage in trials outside Scandinavia [135].

2.4.3.3 Benign Disease Questionnaires Used in EBRT Trials

QoL tools for benign prostatic disease have been used as PRO instruments in PCa trials. The International Prostate Symptom Score (IPSS) [136], is a common tool used to screen men for symptoms of benign prostatic hypertrophy. In 1998, Terk *et al* identified the pre-treatment IPSS as a predictor of urinary retention after LDRBT [137]. Since both SBRT and brachytherapy may be heterogeneously dosed, there has been interest in

IPSS as a measure of potential obstructive symptoms after SBRT. Early SBRT data for UK nmPCa patients demonstrated median IPSS increase from baseline 6 to 11 at 1-3 weeks post-treatment [138].

Tools for benign bowel disease have been investigated as instruments to report GI symptoms after pelvic radiotherapy. Both the Inflammatory Bowel Disease Questionnaire (IBDQ) and the Vaizey incontinence score have been shown to agree with LENTSOM and acute RTOG scoring [139,140].

Likewise, tools for male erectile dysfunction are also of interest in assessment of prostate EBRT side effects. The International Index of Erectile Function 5-question (IIEF-5) QoL instrument was developed from the larger 15-question scale, with 98% sensitivity (88% specificity) for erectile dysfunction [141]. It shows strong correlation with more complicated instruments for erectile dysfunction after brachytherapy [142].

2.4.3.4 Usage of PRO Data in Hypofractionation Trials

As might be expected from the wide range of tools described above, there is little concordance on QoL measurements between phase III hypofractionation trials (**Table 6**). No tool has been used in every study, with efforts to achieve cross-trial alignment by CHHiP resulting in intra-trial differences in QoL metrics utilised [143]. This limits utility of PRO instruments in this trial for modelling, since only a subset of the patients who consented to the QoL substudy will be available. This factor influenced the choice of the more consistent CROs for modelling in **Chapter 3**.

Table 6. Patient Reported Outcome Scales in Major Hypofractionation Studies

Hopkins Symptom Checklist 25 (HSCL-25) specifically assesses anxiety/depression. Other scales discussed in main text.

PRO Instruments	CHHiP	RTOG-0415	PROFIT	HYPRO	PACE-B	HYPO-RT-PC
Generic Instruments						
EORTC QLQ-C30						Yes
EQ-5D		Yes				
FACT-G	Until 2009					
SF-12	From 2009		Yes			
SF-36	Until 2009					
Prostate Cancer						
EORTC QLQ-P25				Yes		
EPIC-50	From 2009 (GI & GU)	Yes	Yes			
EPIC-26	From 2009 (sexual & hormonal)				Yes	
FACT-P	Until 2009					
PCSS						Yes
UCLA-PCI	Until 2009					
Other Questionnaires						
IIEF-5					Yes	Yes
IPSS					Yes	
Vaizey					Yes	
Mental Health						
HSCL-25		Yes				

2.5 Modelling of Toxicity Endpoints

2.5.1 Predictive Models

Predictive models attempt to describe a mathematical relationship between certain variables (predictors), and an outcome of interest which they might influence. For the modelling of radiotherapy toxicity (the outcome), predictors might be patient-related (e.g. age, comorbid status etc) or treatment-related (radiotherapy dose, fractionation, concomitant medications).

2.5.2 Modelling Radiotherapy Toxicity

This section will consider components of the modelling process and discuss some of the issues most pertinent to the modelling of radiotherapy toxicity, which has some unique challenges relative to other cancer therapies.

2.5.2.1 Outcome Variable

Outcome variable choice is important. Firstly, the outcome variable data type influences the types of predictive model that may be considered:

- Binary/categorical (e.g. G2+ RTOG toxicity) → Classification models
- Discrete/continuous (e.g. Worst IPSS score) → Regression models

It is generally better to avoid converting numerical outcomes (e.g. blood sugar level) into binary variables (blood sugar: <7 mmol/L or ≥7 mmol/L), since this discards information regarding the quantitative differences between patients, that a regression model can utilise[19] [144,145]. For some

[19] Additionally, an arbitrary cut-point may act to bias the effect direction seen.

toxicity data, such as RTOG toxicity score, the data is available in ordinal format (0 / 1 / 2 / 3), which has some additional information beyond nominal classification, but is not numerical. In the prostate EBRT field, it is common to model G2+ toxicity [62,88]. This avoids a more complicated model (e.g. ordinal logistic regression) when there is limited interest in G1 toxicity. Additional it avoids the problems of modelling very small numbers of G3 or G4 events; difficult even in a binary logistic regression [146].

2.5.2.2 Model Choice

Having selected an outcome, model choice will be further determined by what degree of interpretability is desired. The outright best models for prediction tend to lose human interpretability [147]. A good example being neural networks, where, after training, the thousands of neurone weights between the hidden layers make it challenging to understand the key features contributing to model prediction [148].

For radiotherapy toxicity models, interpretability is important. Factors such as an odds ratio (or coefficient) for each input parameter allows clinicians to consider the relative weights of predictors. These can help to guide clinical practice – e.g. consenting patients with more risk factors to a higher risk of complications. For this reason, multivariate models with good interpretability (e.g.) multivariate logistic/linear regression are pursued in **Chapter 7**.

2.5.2.3 Handling of Radiotherapy Predictors

The inclusion of non-radiotherapy predictors in a model can follow standard pre-processing techniques [149]. However, radiotherapy dosimetry presents some unique challenges.

2.5.2.3.1 Definitions of an Organ at Risk and Dose Received

For a radiotherapy toxicity model, the dose received by a normal tissue related to that endpoint is an important predictor. However, unlike the CTV, this dose is heterogenous in nature. Clinically, this is represented as a DVH, showing either the percentage (relative DVH), or volume (absolute DVH), of an OAR receiving a given dose.

The relative seriality of the OAR[20] (recall **Figure 3**) guides OAR volume definition and dosimetry of interest. For highly serial OARs (e.g. spinal cord), maximum dose is most useful, since if below tolerance, the anticipated risk of toxicity is low. For more parallel OARs, such as the lung parenchyma, it is important to contour both whole lungs, allowing calculation of the proportion exceeding a dose which might impair function [150]. Here a percentage dose-volume constraint will be useful, to ensure that a functionally relevant proportion of OAR is not irradiated beyond tolerance.

[20] A serial organ being one impaired by damage to any part of its structure, while a parallel one experiences independent loss of functional units, with a minimum number of units to maintain organ function [53]. In many cases it is a mix of the two.

This simple description of fully serial and parallel OARs is complicated by the fact that OARs may display some seriality and some parallelism [53]. An OAR such as the rectum may even have different endpoints (e.g. bleeding vs proctitis) with different degrees of seriality.

It is unfortunate that considerable spatial information is discarded in the DVH. Attempts have been made to address this, such as the dose-surface maps [151,152], spatial dose metrics [153], however, their utility has not yet been sufficiently proven to enter routine clinical practice.

2.5.2.3.2 Challenges for Prostate OARs

Both bladder (for GU toxicity) and rectum (for GI toxicity) are hollow viscera, with multi-layered functional walls surrounding a non-functional interior of excreta. The muscular walls may expand or contract dependent on the amount of excreta present, changing the volume of the organ but not changing the volume of functional wall tissue.

For this reason, there exists heterogeneity in the practice of OAR delineation. The rectum is commonly contoured as a solid organ, with resultant dose-volume constraints [154], but other groups contour as a hollow viscera (wall), necessitating different DVH constraints [155]. This thesis will concern itself with the rectum as a solid structure, with dose information extracted as a DVH[21]; this being the methodology most familiar to UK practitioners from the

[21] As opposed to the dose-surface histogram, which analogises the rectal wall structure.

RT-01, CHHiP and PACE trials. Practice in each centre may vary for the superior and inferior extent of the rectum, potentially altering the dose metrics produced and hence toxicity prediction. The influence of rectal contouring on toxicity prediction is investigated in **Chapter 6**.

2.5.2.3.3 The Dose Received by a Normal Tissue

It is important to note that we are considering the planned dose to the rectum[22], which is only in itself a surrogate of the true delivered dose [156]. Techniques such as MRI-guided RT might permit routine OAR dose accumulation mapping, although this is not yet in clinical practice.

However, accepting the planned solid rectum DVH as our dosimetric predictor source, further challenges arise. It is common to use the values of dose bins as predictors, however there is severe multicollinearity[23] between each dose bin. A traditional approach to this issue is DVH reduction through models such as the LKB model, described earlier in **2.3.3.4.1**. Other approaches to this issue will now be considered.

2.5.2.3.4 Non-LKB Model Methods to Counter Dose Bin Correlation

More recently, other approaches to the DVH multicollinearity issue have been explored. Principal component analysis allows the decomposition of a

[22] I.e. that calculated from the pre-treatment planning CT scan.
[23] Multicollinearity describes situation of multiple predictive variables being highly correlated. When undertaking variable selection this may result in instability, since any correlated predictor may be arbitrarily chosen.

set of correlated numerical variables, into a smaller number of principle components[24], which are fully decorrelated (**Figure 6**) [157].

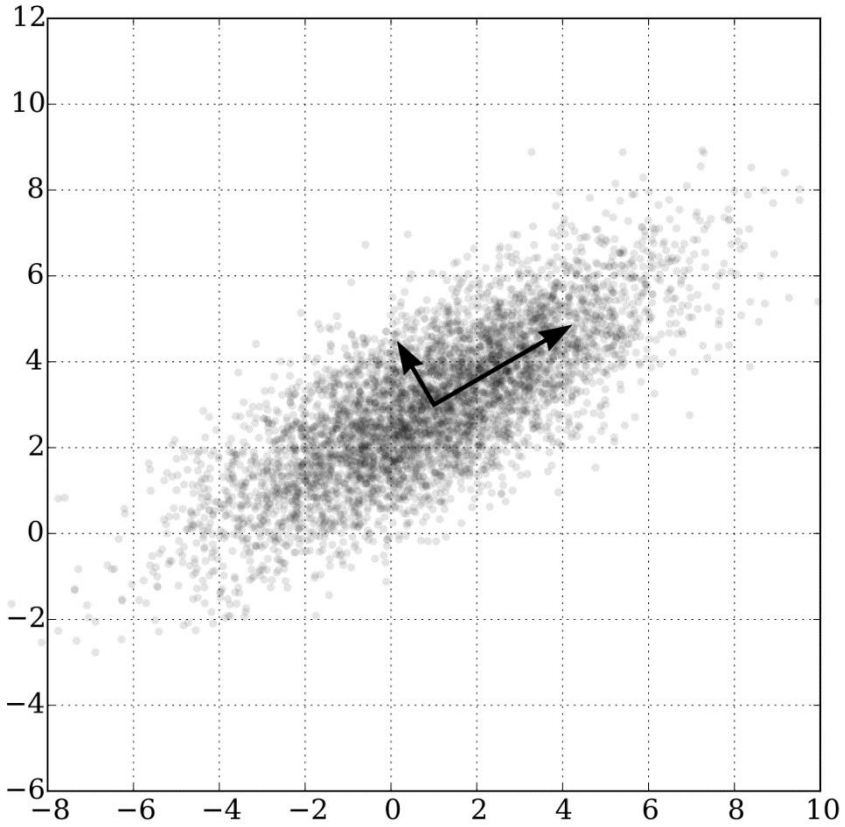


Figure 6. Example of Principle Components

Eigenvectors (shown as arrows) for the first and second principle components of two highly correlated variables. The two eigenvectors can be seen to be fully decorrelated. https://en.wikipedia.org/wiki/Principal_component_analysis#/media/File:GaussianScatterPCA.svg. User: Nicoguaro. Reused based on CCBY 4.0 licence

This methodology has been applied in a radiotherapy setting [158], but suffers from difficult interpretability when making inference with the principal components. Another approach is functional data analysis, where the patient DVH curve is treated as a single function; eliminating multicollinearity [159]. This has not yet had widespread adoption.

[24] Each principal component being expressed as a formulation of the original variables

Variable selection methods can also reduce multicollinearity. For example, manual inspection of the correlation matrix before modelling would enable the selection of less correlated dose bins. This would likely be those further apart on the DVH; i.e. V20Gy and V70 Gy, rather than V69 Gy and V70Gy.

Procedural variable selection can occur through methods such as stepwise selection or least absolute shrinkage and selection operator (LASSO) L1-regularisation [147]. Stepwise methods can be operated forward (adding variables), backward (removing variables) or a hybrid system of the two. The methods usually use p-values as a cut-off for selection [160]. LASSO regularisation applies a penalty based on the sum of the absolute coefficient values, multiplied by a hyperparameter lambda (chosen by cross-validation). This induces coefficients to be set to zero, effectively removing predictors from the model. Variable selection for a radiotherapy toxicity model will be undertaken in **Chapter 7**.

2.6 Brief Summary

Radical EBRT is an important treatment modality for nmPCa. Moderate hypofractionation is now standard-of-care, with ultrahypofractionation in trials; both on the basis of non-inferior disease control. Patient effects associated with hypofractionation are therefore important, both from CRO and PRO perspective. This thesis seeks to examine the effects of hypofractionation on CRO and PRO endpoints, through a variety of modelling and analysis techniques. A fuller summary can be recalled from **Chapter 1**.

Chapter 3. Estimates of α/β Ratios for Late Rectal Toxicities

3.1 Publications Related to Chapter

Estimates of Alpha/Beta (α/β) Ratios for Individual Late Rectal Toxicity Endpoints: An Analysis of the CHHiP trial

Douglas H. Brand, Sarah C. Brüningk, Anna Wilkins, Katie Fernandez, Olivia Naismith, Annie Gao, Isabel Syndikus, David P. Dearnaley, Alison C. Tree, Nicholas van As, Emma Hall*, Sarah Gulliford*, On behalf of the CHHiP Trial Management Group.

Int. Journal of Radiation Oncology, Biology, Physics. 2020.

https://doi.org/10.1016/j.ijrobp.2020.12.041

[Full Length Research Article, in-press, PMID: 33412260]

3.2 Introduction

Rectal toxicity endpoints are important late side effects of prostate EBRT. Models have been produced for many common individual rectal endpoints such as bleeding, proctitis, stool frequency and faecal incontinence [11,152,161–166]. These models incorporate DVH derived values as dosimetric predictors. In the hypofractionation era, researchers have adjusted the rectal dose bins using the LQ-model [5], describing normal tissue fraction sensitivity by means of the α/β ratio. Commonly, a late rectal α/β = 3 Gy is assumed [167,168], to convert to EQD2 and enable comparison with 2 Gy per fraction treatments [5]. Similarly, EQD2 correction

has been used when summating brachytherapy and EBRT doses, with late rectal α/β = 3 – 5.4 Gy [169–171].

These EQD2-corrected comparisons of regimens are dependent on an accurate estimate of the α/β ratio. Researchers have previously provided human estimates for the α/β ratio of overall late rectal toxicity in the range 2.7 – 7 Gy [7,62,88,89] (**Section 2.3.4.4**). However, individual rectal toxicity endpoints (bleeding, stool frequency etc.) are likely driven by different upstream pathophysiological processes [172], and may thus have distinct sensitivity to fraction size, as manifest by the α/β ratio. Although estimates have been produced for individual central nervous system endpoints [173], to my knowledge, such estimates have not previously been made for pelvic normal tissues.

In this chapter, using data from the CHHiP trial, I estimate α/β ratios for individual rectal toxicity endpoints: bleeding, stool frequency, proctitis, sphincter control and stricture/ulcer. I also test if such α/β ratio estimates are influenced by inclusion of other predictive clinical factors: age, diabetes, hypertension, inflammatory bowel or diverticular disease (IBD/diverticular), and haemorrhoids.

3.2.1 Hypotheses

1) Individual rectal endpoints, e.g. bleeding, stool frequency etc, may exhibit different fraction size sensitivity, manifest by the α/β ratios.

2) These individual α/β ratios may be modelled through analysis of dosimetry from a phase III hypofractionation trial.

3) The estimates of α/β ratio might be influenced by inclusion of clinical parameters in the model (e.g. diabetes, inflammatory bowel disease / diverticular disease).

4) Findings may be supported by whole-trial level meta-analysis of the EQD2 doses and rectal toxicity in the phase III hypofractionation trials.

3.3 Methods

3.3.1 The CHHiP Trial

The CHHiP trial (ISRCTN97182923) has previously been described in detail [1,25,26]. Briefly, 3216 men were recruited, all with histologically confirmed T1b –T3aN0M0 prostate adenocarcinoma, PSA ≤40 ng/mL and risk of lymph node involvement <30%. Open-label randomisation was 1:1:1 between conventional (74 Gy in 37 fractions (Fr) over 7.4 weeks), higher dose hypofractionated (60 Gy in 20 Fr over 4 weeks) or lower dose hypofractionated (57 Gy in 19 Fr over 3.8 weeks) EBRT. The primary endpoint of biochemical or clinical failure was met, with non-inferiority of the 60 Gy in 20 fraction regimen confirmed [1]. Ethics approval has been described previously [1]. The Institute of Cancer Research Clinical Trials and Statistics Unit (ICR-CTSU, London, UK) coordinated the study and managed the trial data used in this analysis.

3.3.2 Patient Selection

CHHiP patients who had received all fractions of one of the protocol radiotherapy regimens were eligible for inclusion in this rectal α/β ratio sub-study. Those without centrally available Digital Imaging and Communications in Medicine (DICOM) data[25] of CT, structures and dose cube were excluded.

[25] An inter-operative standard for radiotherapy data.

3.3.3 Rectal Contouring and Dose-Volume-Histogram Generation

The CHHiP trial protocol recommended an empty rectum. Contouring for the rectum, as a solid structure, was "*from the anus (usually at the level of the ischial tuberosities or 1 cm below the lower margin of the PTV whichever is more inferior) to the recto-sigmoid junction*" [1].

Centres were encouraged to submit all CHHiP radiotherapy plan files to the ICR-CTSU, for central analysis; I did not personally retrieve these. For those patients with appropriate files received (CT, structures, dose), recontouring to protocol was performed by one of five observers: myself (n=448), Anna Wilkins (n=791), Jake Probert (n=109), Katie Fernandez (n=903), Richard Boyle (n=161). The non-clinicians (JP, KF, RB) were trained in the contouring of the rectum to the CHHiP protocol and overseen by myself and Anna Wilkins.

For each patient, the patient treatment plan files from external centre were opened in VODCA (version 5.4.1, MSS Medical Software Solutions GmbH, Hagendorn, Switzerland), a research treatment planning system, and converted to DICOM if in another format (e.g. RTOG files). The presence of all CT slices, structures and dose cube were checked manually. Separate dose cubes (e.g. two-phase treatment) were summed into a single dose cube. Using VODCA, the rectum was edited, where required, to meet the CHHiP protocol definition: particular attention being paid to the superior and inferior extent. DVH data was then re-calculated and the resulting DICOM file saved. For this chapter, I wrote scripts in MATLAB (v2018b, Mathworks, MA,

USA) to extract the checked rectal differential DVH data into tables for modelling.

3.3.4 Endpoints for Modelling

The CHHiP trial collected bowel toxicity information in the form of both CROs [1] and PROs [143]. I chose CROs for modelling, since the PRO measures changed during the trial (recall **2.4.3.4**), which would reduce the patient numbers per model. The CROs collected were RTOG late rectal toxicity [119], the RMH scale [174] and LENTSOM [121]. In the trial follow-up forms, rather than an overall RTOG score, possible contributory components were requested separately: bowel obstruction, diarrhoea, proctitis, rectal-anal stricture, rectal ulcer. Only RMH and LENTSOM were collected at registration (baseline) and pre-radiotherapy (pre-RT). All scales were collected for late rectal toxicity at 6-, 12-, 18-, 24-, 36-, 48- & 60-months follow-up after the start of RT. This data is managed by ICR-CTSU and I was provided with the 5-year follow-up database [1] for this analysis.

I merged RTOG, RMH and LENTSOM into new amalgamated endpoints representing underlying separate symptomatic issues, using similar methodology to previously described by Gulliford *et al* [175]. Grading was simplified to: grade 0 for no toxicity; grade 1 for toxicity not needing intervention; grade 2 for any toxicity requiring intervention. The scores were dichotomised to consider: grade 0 vs grade 1 or above (G1+ comparison); grade 0 and grade 1 vs grade 2 or above (G2+ comparison). For bowel pain, sphincter control and stricture/ulcer, grade 2 or above events were rare

(<5%), so only a G1+ comparison was performed. No attempt was made to amalgamate endpoints to generate G3+ models, both due to the rarity of G3+ events and the difficulty of unifying such events between scales.

For each endpoint, patients were excluded if any relevant toxicity (G1+) was reported at baseline or pre-RT assessments; or if both assessments were missing. This was to avoid those with pre-existent symptoms registering as having treatment-induced toxicity events during follow-up. Patients were further excluded for an endpoint if they were missing the relevant follow-up data at more than 3 of the 7 (>50%) late toxicity assessments. Toxicity events were scored for any relevant toxicity of sufficient grade at any time point (i.e. worst toxicity). A full overview of the endpoint generation process is provided in **Table 7**. I personally wrote and debugged code in Stata (version 15, Statacorp, TX, USA) for the endpoint amalgamation process, along with extraction of all clinical data for the modelling stages.

Table 7. Explanation of Endpoint Generation

The amalgamated individual endpoints are listed, along with subdomain scores that would generate an event score in the amalgamated endpoint. Exclusion criteria are explained. The baseline score is generated as the WORST score of the baseline assessment and the pre-RT assessment. Patients would not be assigned a baseline score without the relevant endpoint being scored at one or both of those visits (and thus would be excluded from that endpoint). Only RMH and LENTSOM were collected at those timepoints, so RTOG scores are not considered in the adjudication of zero baseline toxicity.
New abbreviations: freq = frequency; subj = subjective.

Modelling Endpoint	Exclude Unless Grade 0 in ALL of these baseline scores	Exclude if Missing >50% Follow-up Scores for ANY of:	Score Positive Toxicity if ANY of these toxicities recorded ≥6 months
Bleeding G1+	RMH Rectal bleeding	RMH Rectal bleeding	RMH Rectal bleeding G1+
	LENTSOM Objective bleeding	LENTSOM Objective bleeding	LENTSOM Objective bleeding G1+
	LENTSOM Manage bleeding	LENTSOM Manage bleeding	LENTSOM Manage bleeding G1+
Bleeding G2+	RMH Rectal bleeding	RMH Rectal bleeding	RMH Rectal bleeding G2+
	LENTSOM Objective bleeding	LENTSOM Objective bleeding	LENTSOM Objective bleeding G2+
	LENTSOM Manage bleeding	LENTSOM Manage bleeding	LENTSOM Manage bleeding G1+
Frequency G1+	RMH Bowel frequency	RTOG Diarrhoea	RTOG Diarrhoea G1+
	LENTSOM Subj stool frequency	RMH Bowel frequency	RMH Bowel frequency G1+
	LENTSOM Manage tenesmus/stool freq	LENTSOM Subj stool frequency	LENTSOM Subj stool frequency G1+
		LENTSOM Manage tenesmus/stool freq	LENTSOM Manage stool freq. G1+
Frequency G2+	RMH Bowel frequency	RTOG Diarrhoea	RTOG Diarrhoea G2+
	LENTSOM Subj stool frequency	RMH Bowel frequency	RMH Bowel frequency G2+
	LENTSOM Manage tenesmus/stool freq	LENTSOM Subj stool frequency	LENTSOM Subj stool frequency G2+
		LENTSOM Manage tenesmus/stool freq	LENTSOM Manage stool freq. G1+
Pain G1+	LENTSOM Subj pain	LENTSOM Subj pain	LENTSOM Subj pain G1+
	LENTSOM Manage pain	LENTSOM Manage pain	LENTSOM Manage pain G1+
Proctitis G1+	LENTSOM Subj tenesmus	RTOG Proctitis	RTOG Proctitis G1+
	LENTSOM Subj mucosal loss	LENTSOM Subj tenesmus	LENTSOM Subj tenesmus G1+
		LENTSOM Subj mucosal loss	LENTSOM Subj mucosal loss G1+
Proctitis G2+	LENTSOM Subj tenesmus	RTOG Proctitis	RTOG Proctitis G2+
	LENTSOM Subj mucosal loss	LENTSOM Subj tenesmus	LENTSOM Subj tenesmus G2+
		LENTSOM Subj mucosal loss	LENTSOM Subj mucosal loss G2+
Continued overleaf…			

Table 7 continued…			
Modelling Endpoint	**Exclude Unless Grade 0 in ALL of these baseline scores**	**Exclude if Missing >50% Follow-up Scores for ANY of:**	**Score Positive Toxicity if ANY of these toxicities recorded ≥6 months**
Sphincter Control G1+	LENTSOM Subj sphincter control	LENTSOM Subj sphincter control	LENTSOM Subj sphincter control G1+
	LENTSOM Manage sphincter control	LENTSOM Manage sphincter control	LENTSOM Manage sphincter control G1+
Stricture/Ulcer G1+	LENTSOM Objective ulceration	LENTSOM Objective ulceration	RTOG bowel obstruction G1+
	LENTSOM Objective stricture	LENTSOM Objective stricture	RTOG rectal-anal stricture G1+
	LENTSOM Manage ulceration	LENTSOM Manage ulceration	RTOG rectal ulcer G1+
	LENTSOM Manage stricture	LENTSOM Manage stricture	LENTSOM Objective ulceration G1+
			LENTSOM Objective stricture G1+
			LENTSOM Manage ulceration G1+
			LENTSOM Manage stricture G1+

3.3.5 Generalised Lyman-Kutcher-Burman Model

A generalised LKB model has been previously described for rectal α/β ratio estimation [88]. Dose modifying factors were incorporated as modulators of each individual patient's effective dose parameter (D_{Eff}), per prior work by Tucker *et al* [176]. The model is expressed as a definite integral:

$$NTCP = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{t} e^{-0.5 \cdot x^2}\, dx \tag{4}$$

Where *NTCP* is the normal tissue complication probability. Furthermore:

$$t = \frac{D_{Eff} \cdot e^{\delta \cdot DMF} - TD_{50}}{m \cdot TD_{50}} \tag{5}$$

Here, TD_{50} represents the tolerance dose for 50% toxicity, at the median (steepest) part of the NTCP dose response curve. *m* is a parameter inversely controlling the steepness at TD_{50}. *DMF* is the dose modifying factor corresponding to either: ones and zeros for binary risk factors, or a positive integer for age. *δ* is the dose modifying coefficient, used to adjust TD_{50} in the presence of the risk factor specified by *DMF*. For binary DMFs, the co-efficient is for presence of risk factor, for numerical DMFs (age only) it is evaluated on a per-unit basis. Note that a DMF covariate of zero will result in no change to the effective dose (D_{Eff}), which is defined by:

$$D_{Eff} = \left(\sum_{i=1}^{w} (EQD2_i)^{\frac{1}{n}} \cdot v_i \right)^n \tag{6}$$

Where *n* represents the relative seriality of a tissue endpoint dose response: values towards 0 being more serial and towards 1 being more parallel [177]. *w* is the number of dose bins, iterated by dose bin *i*. v_i is the relative volume

of an organ present in the dose bin *i*. $EQD2_i$, is the EQD2 for dose bin *i*, which is given by:

$$EQD2_i = D_i \cdot \left(\frac{d_i + \alpha/\beta}{2\ Gy + \alpha/\beta}\right) \tag{7}$$

Where D_i is the total dose in Gy, to a given DVH dose bin *i*. d_i is the dose in Gy per fraction, to a given dose bin (i.e. D_i divided by number of fractions). *α/β* (Gy) is the theoretical single fraction dose giving equal contribution for linear (*α*) and quadratic (*β*) components of the LQ formula [5].

This model is termed LKB-EQD2, or LKB-EQD2-DMF with the inclusion of a DMF in **Equation 5.** The LKB-NoEQD2 model without EQD2 correction uses **Equations 4 & 5** (without DMF inclusion), but substitutes physical dose bin dose for $EQD2_i$ in **Equation 6**. This LKB-NoEQD2 model was fitted separately for 2 Gy per fraction patients (74 Gy in 37 Fr) and 3 Gy per fraction patients (60 Gy in 20 Fr and 57 Gy in 19 Fr).

3.3.5.1 Initial Grid Search

MATLAB (v2018b) was used for modelling in this chapter. I personally wrote and debugged all code. For each model, initial fitting was done using the grid search method, as previously described [11]. Each unknown parameter was searched on a grid with dimensionality equal to number of fit parameters. LKB-EQD2 models with fixed α/β were also produced, using the same parameter grid as those with fitted α/β, but fixing the α/β to either 3 Gy or 4.8 Gy, as per prior literature estimates [62,88].

I assessed model performance in two ways. First the naïve performance was assessed by calculating a log likelihood sum. Better model performance will produce a less negative log likelihood sum. It was calculated as:

$$Likelihood = f(toxicity) = \begin{cases} NTCP & toxicity = 1 \\ 1 - NTCP & toxicity = 0 \end{cases} \quad (8)$$

$$Performance = Summed\ Log\ Likelihood = \sum_{j=1}^{c} \ln Likelihood_j \quad (9)$$

Where: *c* = number of patients (with *j* as iterator through such patients).

The model parameter values generating the ten least negative performance metrics were recorded at the end of the grid search. The best (least negative) of these was noted as the naïve model performance, for later use in **Equation 11**.

The second action at each grid step was to assess performance of 2000 bootstraps, drawn with replacement, with unique bootstraps for each endpoint. The bootstrap performance was also assessed by **Equation 9**. At the end of the grid search, the parameters giving the ten least negative performance metrics for each bootstrap were recorded so that these could be used later, for out-of-the-bag (OOB) prediction, in **Equation 10** [178].

3.3.5.2 Second Stage Search

To account for the known sensitivity of fitting algorithms to initial starting parameters and hence to improve model performance [179], a secondary optimisation search for parameter values was undertaken. For this, the

values of *n*, *m*, *TD50*, *α/β* and *DMFs* producing the ten best performance metrics (by **Equation 9**) were used as the initial parameters in a constrained Nelder-Mead simplex algorithm search [180], utilising a bounded version of *fminsearch* (fminsearchbnd, v 1.4.0.0)[26] [181], to see if further improvement in performance could be found. I.e. for each endpoint; 1 naïve model and 2000 bootstraps with 10 searches = 20010 algorithm searches. This algorithm was run with constraints: *n* = 0.01-10; *m* = 0.01-10; *TD50* = 0.01-1000 Gy. Where freely fitted, α/β varied between 0.001 to 1000 Gy. The DMF covariate was varied between -10 to 10, which, when raised to the natural base *e*, searches a dose multiplier range of 4.54×10^{-4} – 22026. This wide bounding of all fit parameters was chosen to prevent bootstrap distributions being inappropriately constrained, which would bias the coverage of the non-parametric 95% confidence interval. For the naïve likelihood and each bootstrap, the final best model parameters were those resulting in best performance (by **Equation 9**) from any of the grid search positions or any of the subsequent ten Nelder-Mead simplex algorithm searches.

3.3.5.3 Estimating Test Performance and Model Comparison

A model comprising more free parameters is always likely to improve naïve likelihood performance, but this can be due to overfitting [147]. To address this difficulty, the 632 bootstrap estimator was used as an unbiased estimator of test performance [182]. It balances out the over-optimistic naïve likelihood

[26] The search is constrained by sine transformation of the searched parameter.

(fitted on the population) against the negatively biased out-of-the-bag bootstrap estimate. The 632 estimator was preferred over the 632+ estimator, due to faster calculation and the low risk of near-perfect prediction with a relatively simple model [178]. The first step calculated the out-of-the-bag performance for the model:

$$OOB\ performance = \sum_{j=1}^{c}\left(\frac{1}{z} \times \sum_{boot=1}^{z} \ln \widehat{likelihood}_{p,boot}\right) \qquad (10)$$

Where *c* is the total number of patients (iterated by *j*), and *z* is the number of bootstraps not containing patient (iterated by *boot*). The predicted likelihood is derived by inserting the predicted NTCP into **Equation 8**. The 632 estimator was then calculated [178]:

$$632\ Estimator = 0.368 \cdot Naive\ Performance + 0.632 \cdot OOB\ Performance \qquad (11)$$

Models were compared by means of the likelihood ratio test of the 632 estimators. Firstly, comparing whether the LKB-EQD2 model with free fitted α/β ratio had significantly better 632 estimator than the model with the α/β ratio fixed at two reported literature values: α/β = 3 Gy [62] or 4.8 Gy [88]. Secondly, examining for significant improvement from LKB-EQD2 to LKB-EQD2-DMF, sequentially tested with each of the DMFs. Tests were only planned where log likelihood improvement occurred; with approximately 50 tests anticipated, a Bonferroni-corrected p-value significance level of 0.001 was used [183]. Parameter estimates were obtained at the 50th centile of the bootstrap distribution. 95% bootstrap CIs for the optimum model parameter values were obtained as the 2.5th and 97.5th centiles of the corresponding

parameter values producing the best summed log likelihood performance metric for each bootstrap.

3.3.6 Graphical Outputs of Calibration

Model calibration was fitted as a logistic regression of predicted NTCP values for each patient as single predictor against observed binary outcomes (toxicity/no toxicity). The fitted model was then displayed graphically against ideal (perfect) prediction; termed the calibration curve. Furthermore, binned calibration plots were examined, with patients grouped into deciles of predicted risk: average bin NTCP plotted against observed bin toxicity proportion. I personally wrote and debugged all code for this in MATLAB.

3.3.7 Whole-Trial Level Meta-Analysis

A simple model was fitted to the whole trial G2+ rectal toxicity data for the phase III hypofractionation trials. First, calculating the percentage difference in late G2+ rectal toxicity between each moderate hypofractionated (test) and conventional (control) arm. Then, for a given α/β ratio, calculating the difference between test and control rectal EQD2 for each trial. Taking advantage of the similar scales of units, fitting the α/β ratio to minimise the sum of squared differences between the EQD2 differences and toxicity differences:

$$\min_{\substack{\alpha \\ \beta \in \mathbb{R}}} \left(\sum_{t=1}^{6} \left(\left(EQD2_{Test,t} - EQD2_{Control,t} \right) - \left(GITox_{Test,t} - GITox_{Control,t} \right) \right)^2 \right) \quad (12)$$

Where:

EQD2 is as defined in equation 3 earlier. t is the trial (one of CHHiP 57Gy, CHHiP 60Gy, PROFIT, RTOG 0415, HYPRO, HYPO-RT-PC). GITox is the trial reported G2+ GI toxicity rate (cumulative expressed as a %). Test refers to the investigation dose-fractionation regimen, control to the control regimen.

Normalisation/standardisation of data was not undertaken, given the limited number of data points. I personally wrote and debugged all code for this in MATLAB.

3.4 Results

3.4.1 Included Patients

A total of 2215/3216 patients were included. **Figure 7** is a CONSORT-style flow diagram accounting for all patients that were randomised into the CHHiP study and their reasons for non-inclusion in this analysis.

Figure 7. CONSORT-Style Patient Flow Diagram

Showing any reasons for exclusion of all patients randomised into the CHHiP trial.

3.4.2 Baseline Patient Data

Key relevant baseline and treatment characteristics for the included patients are shown in **Table 8**, which are similar to those in the CHHiP trial as a whole. This indicates that patients in this study are representative of the whole trial cohort.

Table 8. Baseline characteristics for patients included in this study

Characteristic	This Study		Whole CHHiP Trial	
	No.	%	No.	%
Age	69 yrs median	44-85 (range)	69 yrs median	44-85 (range)
Arm				
57Gy/19f	755	34%	1077	33%
60Gy/20f	753	34%	1074	33%
74Gy/37f	707	32%	1065	33%
NCCN Risk Group				
Low risk	308	14%	484	15%
Intermediate risk	1655	75%	2347	73%
High risk	252	11%	385	12%
Gleason score				
≤6	750	34%	1122	35%
7	1399	63%	1995	62%
8	66	3%	99	3%
Clinical T Stage				
T1	851	38%	1170	36%
T2	1196	54%	1766	55%
T3	167	8%	277	9%
Missing	1	<1%	3	<1%
Pre-ADT PSA				
<10 ng/mL	1082	49%	1567	49%
10-20 ng/mL	1006	45%	1415	44%
≥20 ng/mL	127	6%	208	6%
Missing	0	0%	26	<1%
Comorbidities				
Diabetes	227	10%	342	11%
Hypertension	874	40%	1276	40%
IBD or diverticular disease	85	4%	124	4%
Pelvic Surgery	162	7%	252	8%
Symptomatic Haemorrhoids	153	7%	209	6%
Total	**2215**	**100%**	**3216**	**100%**

3.4.3 Endpoint Data Summary

A summary of the number of patients meeting requirements (≥50% follow-up form completion) for each endpoint modelled are shown in **Table 9**, with the proportion of patients expressing toxicity ranging from 3.6% for stricture/ulcer G1+ (79/2206) to 38.1% for stool frequency G1+ (771/2025). Numbers of excluded patients were generally low, ranging from 9 (<1%) excluded from stricture/ulcer G1+, to 209 (9.4%) excluded from bleeding G2+. It should be noted that more patients were excluded from G2+ endpoints because any toxicity in follow-up allows scoring (and there are more G1 toxicities), while all composite endpoints must be non-missing and recorded as G0 to score a grade zero. Hence if a patient is missing some follow-up endpoints (e.g. RTOG) but has a G1 toxicity in another (e.g. LENTSOM) they will be recorded as a toxicity in G1+ endpoint, but missing in the G2+ endpoint.

Table 9. Summary of Patient Numbers in Each Modelling Endpoint

Patients excluded for any of: missing baseline data; baseline toxicity above grade 0; missing >50% of follow-up forms. Presented percentages are calculated without the inclusion of patients excluded for each endpoint, so that event rates in modelled patients can be seen.

Rectal Endpoints & Grades of Interest	Dose-Fractionation Regimen						Total	
	57 Gy in 19 fractions		60 Gy in 20 fractions		74 Gy in 37 fractions			
	No.	%	No.	%	No.	%	No.	%
Bleeding G1+								
No	479	70.5%	434	63.4%	434	67.4%	1347	67.1%
Yes	200	29.5%	251	36.6%	210	32.6%	661	32.9%
Excluded	73		67		67		207	
Bleeding G2+								
No	590	86.9%	575	83.9%	549	85.5%	1714	85.4%
Yes	89	13.1%	110	16.1%	93	14.5%	292	14.6%
Excluded	73		67		69		209	
Frequency G1+								
No	437	62.8%	428	62.4%	389	60.5%	1254	61.9%
Yes	259	37.2%	258	37.6%	254	39.5%	771	38.1%
Excluded	56		66		68		190	
Frequency G2+								
No	611	87.9%	587	85.8%	545	84.9%	1743	86.2%
Yes	84	12.1%	97	14.2%	97	15.1%	278	13.8%
Excluded	57		68		69		194	
Pain G1+								
No	686	93.1%	671	90.1%	638	90.8%	1995	91.3%
Yes	51	6.9%	74	9.9%	65	9.2%	190	8.7%
Excluded	15		7		8		30	
Proctitis G1+								
No	509	69.3%	449	62.2%	433	62.7%	1391	64.8%
Yes	225	30.7%	273	37.8%	258	37.3%	756	35.2%
Missing	18		30		20		68	
Proctitis G2+								
No	666	90.9%	641	88.8%	607	87.8%	1914	89.2%
Yes	67	9.1%	81	11.2%	84	12.2%	232	10.8%
Excluded	19		30		20		69	
Sphincter Control G1+								
No	680	91.0%	664	88.7%	615	87.5%	1959	89.1%
Yes	67	9.0%	85	11.3%	88	12.5%	240	10.9%
Excluded	5		3		8		16	
Stricture/Ulcer G1+								
No	732	97.5%	719	95.9%	676	95.9%	2127	96.4%
Yes	19	2.5%	31	4.1%	29	4.1%	79	3.6%
Excluded	1		2		6		9	
Total	**752**	**100%**	**752**	**100%**	**711**	**100%**	**2215**	**100%**

3.4.4 LKB-NoEQD2 Model Fits with Fitted and Fixed α/β Ratio

Table 10 shows parameter estimates of *n*, *m* and *TD50* for fits of the LKB-NoEQD2 model to two groups: 74 Gy only; 57 Gy and 60 Gy combined. Each endpoint is presented separately. The low *n* values generally seen are indicative of the generally serial nature of responses. Wide confidence intervals are noted for rarer endpoints; pain, sphincter control and stricture/ulcer. Lower *m* values are typical of a steeper dose response relationship (at a trial level). *TD50* values were observed to be higher for all G2+ endpoints compared to matching G1+ endpoints.

Table 10. LKB Model fits (No EQD2 Correction)

First two sections show LKB-NoEQD2 model fitted for each endpoint to the conventionally fractionated (74Gy) patients and the hypofractionated (57 & 60 Gy) patients.

Model	Pts	n (95% CI)	m (95% CI)	TD50 (95% CI) [Gy]	632 Likelihood
LKB-NoEQD2 (74Gy Pts)					
Bleeding G1+	644	0.26 (0.01-1.12)	0.33 (0.09-0.68)	61.5 (54.5-74.0)	-401.8
Bleeding G2+	642	0.13 (0.01-0.42)	0.21 (0.06-0.43)	74.0 (67.2-96.6)	-262.6
Frequency G1+	643	0.17 (0.01-0.53)	0.30 (0.09-0.76)	60.8 (53.7-72.8)	-427.7
Frequency G2+	642	0.11 (0.03-0.69)	0.20 (0.09-0.49)	73.8 (66.2-98.6)	-269.9
Pain G1+	703	0.24 (0.01-3.15)	0.33 (0.15-0.61)	92.7 (72.2-271.6)	-216.5
Proctitis G1+	691	0.10 (0.01-0.18)	0.22 (0.08-0.50)	64.9 (60.8-73.7)	-452.2
Proctitis G2+	691	0.05 (0.01-0.14)	0.14 (0.06-0.44)	78.0 (71.6-111.6)	-254.3
Sphincter Control G1+	703	0.19 (0.09-3.30)	0.29 (0.16-0.63)	81.7 (68.5-185.3)	-263.8
Stricture/Ulcer G1+	705	0.28 (0.01-5.79)	0.16 (0.05-0.31)	74.4 (66.2-92.8)	-117.6
LKB-NoEQD2 (57Gy/60Gy Pts)					
Bleeding G1+	1364	0.13 (0.07-0.20)	0.22 (0.15-0.31)	50.7 (48.2-53.8)	-845.9
Bleeding G2+	1364	0.11 (0.01-0.28)	0.22 (0.13-0.40)	61.7 (56.3-74.2)	-560.6
Frequency G1+	1382	0.20 (0.12-0.33)	0.47 (0.30-0.89)	50.5 (46.8-59.2)	-908.2
Frequency G2+	1379	0.26 (0.02-0.73)	0.33 (0.20-0.53)	64.9 (56.5-94.4)	-531.9
Pain G1+	1482	0.02 (0.01-9.99)	0.37 (0.16-0.69)	105.4 (69.5-619.1)	-429.8
Proctitis G1+	1456	0.09 (0.01-0.17)	0.34 (0.18-0.70)	56.5 (52.0-67.8)	-931.3
Proctitis G2+	1455	0.12 (0.01-4.16)	0.28 (0.15-0.58)	73.8 (61.6-153.8)	-477.8
Sphincter Control G1+	1496	0.17 (0.09-0.29)	0.26 (0.17-0.43)	65.8 (58.0-93.9)	-486.6
Stricture/Ulcer G1+	1501	0.17 (0.01-0.47)	0.20 (0.09-0.35)	72.3 (60.6-113.6)	-217.4

3.4.5 LKB-EQD2 Model Fits

Table 11 shows LKB-EQD2 model fits for all patients combined, including α/β ratio estimates. α/β ratio estimates for most endpoints were below 3 Gy, with the upper bound of the 95% CI for rectal bleeding G1+ being less than 3 Gy. The 95% CI for Pain G1+ was extremely wide (α/β 0.0 to 840 Gy), suggesting a poor fit for this endpoint.

Table 11 also shows fits for the LKB-EQD2 model, with α/β ratio fixed at 3 Gy and 4.8 Gy. The p-values for likelihood ratio test comparison between the LKB-EQD2 model (unfixed α/β) and the two fixed α/β models are shown. In many cases, the less flexible model (LKB-EQD2 with fixed α/β ratio) had a better fit (by 632 estimator), implying overfitting, making likelihood ratio testing inappropriate. The LKB-EQD2 model with fitted α/β ratio was significantly better than the model with fixed α/β 4.8 Gy for rectal bleeding G1+ ($p = 0.00032$). Other comparisons, where LKB-EQD2 models with fitted α/β ratio were better, did not meet adjusted significance threshold.

Table 11. LKB-EQD2 Fits With Free and Fixed Alpha/Beta Ratio (OVERLEAF)
Showing the LKB-EQD2 model fitted with a varying α/β ratio, then fixed to α/β = 3 Gy and α/β = 4.8 Gy. p-values are from likelihood ratio tests between an endpoint 632 likelihood in the fixed α/β LKB-EQD2 models and same endpoint 632 likelihood in the unfixed LKB-EQD2 model. Note that “Better Fit” implies that the simpler fixed α/β ratio model has better (less negative) 632 estimator than the more complex model (varying α/β ratio), implying the more complex model is overfitted and making likelihood ratio testing inappropriate. Bold p-values are significant at adjusted $p<0.001$. (*Table overleaf*)

Model	Pts	n (95% CI)	m (95% CI)	TD50 (95% CI) [Gy]	α/β Ratio [Gy]	632 Likelihood	p-value vs LKB-EQD2
LKB-EQD2 (All Pts). Free fitted α/β ratio.							
Bleeding G1+	2008	0.21 (0.08-0.34)	0.33 (0.20-0.47)	58.8 (54.2-66.0)	1.6 (0.9-2.5)	-1248.1	N/A
Bleeding G2+	2006	0.16 (0.01-0.34)	0.27 (0.14-0.42)	75.8 (68.2-88.6)	1.7 (0.7-3.0)	-822.6	N/A
Frequency G1+	2025	0.27 (0.17-0.44)	0.55 (0.39-0.86)	56.0 (51.4-62.3)	2.3 (0.9-5.3)	-1334.7	N/A
Frequency G2+	2021	0.31 (0.10-0.71)	0.36 (0.23-0.52)	75.7 (66.2-96.8)	2.7 (0.9-8.5)	-801.3	N/A
Pain G1+	2185	0.15 (0.01-9.89)	0.48 (0.21-0.68)	139.7 (88.7-499.1)	3.6 (0.0-839.6)	-647.4	N/A
Proctitis G1+	2147	0.14 (0.02-0.22)	0.42 (0.22-0.68)	63.6 (58.7-75.5)	2.7 (1.5-5.4)	-1384.1	N/A
Proctitis G2+	2146	0.11 (0.01-0.25)	0.30 (0.17-0.51)	87.8 (75.2-137.0)	2.7 (1.3-15.1)	-731.9	N/A
Sphincter Control G1+	2199	0.23 (0.15-0.38)	0.32 (0.24-0.45)	79.3 (69.8-103.3)	3.1 (1.4-9.1)	-749.7	N/A
Stricture/Ulcer G1+	2206	0.31 (0.01-0.74)	0.25 (0.10-0.34)	83.8 (71.5-110.3)	2.5 (0.9-8.2)	-335.1	N/A
LKB-EQD2 (all Pts). Fixed α/β = 3 Gy							
Bleeding G1+	2008	0.23 (0.15-0.35)	0.37 (0.28-0.51)	57.3 (53.5-61.8)	3.0 (3.0-3.0)	-1250.2	0.042
Bleeding G2+	2006	0.19 (0.03-0.36)	0.32 (0.21-0.46)	75.8 (67.8-92.3)	3.0 (3.0-3.0)	-822.9	0.49
Frequency G1+	2025	0.27 (0.17-0.42)	0.56 (0.40-0.86)	55.7 (51.5-62.2)	3.0 (3.0-3.0)	-1334	Better fit
Frequency G2+	2021	0.31 (0.10-0.71)	0.36 (0.25-0.52)	75.8 (66.3-97.4)	3.0 (3.0-3.0)	-800.3	Better fit
Pain G1+	2185	0.17 (0.01-9.98)	0.49 (0.24-0.70)	142.6 (89.4-701.6)	3.0 (3.0-3.0)	-646.6	Better fit
Proctitis G1+	2147	0.14 (0.02-0.22)	0.43 (0.25-0.68)	63.4 (58.6-75.6)	3.0 (3.0-3.0)	-1383.2	Better fit
Proctitis G2+	2146	0.12 (0.01-0.25)	0.30 (0.18-0.51)	88.1 (75.3-136.5)	3.0 (3.0-3.0)	-730.8	Better fit
Sphincter Control G1+	2199	0.24 (0.15-0.38)	0.32 (0.24-0.45)	79.1 (69.9-103.4)	3.0 (3.0-3.0)	-748.7	Better fit
Stricture/Ulcer G1+	2206	0.32 (0.01-0.74)	0.25 (0.13-0.35)	84.4 (71.7-115.0)	3.0 (3.0-3.0)	-334.2	Better fit
LKB-EQD2 (all Pts). Fixed α/β = 4.8 Gy							
Bleeding G1+	2008	0.28 (0.20-0.42)	0.46 (0.36-0.63)	57.0 (53.1-62.5)	4.8 (4.8-4.8)	-1254.6	**0.00032**
Bleeding G2+	2006	0.24 (0.14-0.46)	0.39 (0.30-0.54)	80.0 (69.5-105.9)	4.8 (4.8-4.8)	-824.9	0.032
Frequency G1+	2025	0.29 (0.19-0.45)	0.63 (0.46-0.96)	55.6 (51.2-63.0)	4.8 (4.8-4.8)	-1335.2	0.34
Frequency G2+	2021	0.34 (0.16-0.75)	0.40 (0.30-0.54)	77.5 (67.0-103.5)	4.8 (4.8-4.8)	-800.7	Better fit
Pain G1+	2185	0.21 (0.01-9.97)	0.52 (0.30-0.70)	152.5 (93.6-745.7)	4.8 (4.8-4.8)	-646.4	Better fit
Proctitis G1+	2147	0.16 (0.09-0.24)	0.52 (0.38-0.81)	63.3 (58.2-74.1)	4.8 (4.8-4.8)	-1383.8	Better fit
Proctitis G2+	2146	0.14 (0.02-0.27)	0.36 (0.25-0.54)	93.4 (77.7-148.5)	4.8 (4.8-4.8)	-731	Better fit
Sphincter Control G1+	2199	0.24 (0.16-0.38)	0.34 (0.27-0.47)	81.3 (71.1-106.6)	4.8 (4.8-4.8)	-749	Better fit
Stricture/Ulcer G1+	2206	0.36 (0.15-0.84)	0.28 (0.21-0.37)	87.5 (73.2-127.9)	4.8 (4.8-4.8)	-334.2	Better fit

Table 11: *caption on preceding page.*

3.4.6 LKB-EQD2-DMF Model Fits for Dose Modifying Factors

The effect on model parameters of sequential inclusion of each DMF is reported in **Table 12**. For each endpoint, the LKB-EQD2 model results without inclusion of DMF are reproduced in the first row for ease of comparison. Where the goodness of fit (as assessed by 632 estimator) was improved with inclusion of DMF, p-values for likelihood ratio testing of the LKB-EQD2-DMF model against the LKB-EQD2 model are presented. Only two LKB-EQD2-DMF models improved on LKB-EQD2, by adjusted significance: IBD/Diverticular disease for both stool frequency G2+ (DMF=1.37, 95% CI 1.13 – 1.82, p=0.00041) and proctitis G1+ (DMF=1.27, 95% CI 1.10 – 1.58, p=0.00046). In both of these cases, α/β ratio estimates of the LKB-EQD2 vs LKB-EQD2-DMF fits did not differ by a clinically relevant margin: stool frequency G2+ (2.7 Gy vs 2.5 Gy), proctitis G1+ (2.7 Gy vs 2.6 Gy).

Although inclusion of other DMFs did not meet adjusted significance for model fit improvement, it can be seen in **Table 12** that any differences between LKB-EQD2-DMF model and LKB-EQD2 model α/β ratio estimates are not clinically relevant.

Table 12. Effects of Including Dose Modifying Factors on Model Parameters

Model fits for the sequential inclusion of each dose modifying factor. The 632 likelihood shows model performance and is directly comparable within endpoints (where n is the same). Each DMF model is compared against the LKB-EQD2 (No DMF) model for the same endpoint by likelihood ratio test. Note that "Worse Fit" implies that the more complicated LKB-EQD2-DMF has a worse 632 estimator fit than the simpler LKB-EQD2 (No DMF) model, implying overfitting and making likelihood ratio testing inappropriate. Bold p-values are significant at adjusted p<0.001.

Rectal Endpoints & Dose Modifying Factors	Pts	n covariate	m covariate	TD50 covariate (Gy_{EQD2})	α/β ratio (Gy)	Dose modifying factor covariate	632 Likelihood	Likelihood ratio p-value
Bleeding G1+								
LKB-EQD2 (No DMF)	2008	0.21 (0.08-0.34)	0.33 (0.20-0.47)	58.8 (54.2-66.0)	1.6 (0.9-2.5)	N/A	-1248.1	N/A
Age (years)	2008	0.21 (0.08-0.35)	0.33 (0.21-0.47)	51.0 (36.0-68.9)	1.6 (0.9-2.5)	0.9976 (0.9937-1.0016)	-1248.3	Worse Fit
Diabetes Y/N	2008	0.20 (0.08-0.34)	0.32 (0.20-0.47)	58.6 (54.0-66.1)	1.6 (0.9-2.5)	0.96 (0.87-1.03)	-1248.3	Worse Fit
Haemorrhoids Y/N	2008	0.21 (0.09-0.35)	0.33 (0.21-0.47)	58.9 (54.3-66.1)	1.6 (0.9-2.5)	1.07 (0.96-1.20)	-1248.3	Worse Fit
Hypertension Y/N	2008	0.21 (0.09-0.35)	0.33 (0.21-0.47)	58.4 (53.7-65.8)	1.6 (0.9-2.5)	0.98 (0.93-1.03)	-1248.8	Worse Fit
IBD/Diverticular Y/N	2008	0.21 (0.10-0.35)	0.33 (0.21-0.46)	58.9 (54.3-65.0)	1.6 (0.9-2.5)	1.13 (1.01-1.30)	-1246.8	0.11
Pelvic Surgery Y/N	2008	0.20 (0.08-0.34)	0.33 (0.21-0.47)	59.3 (54.5-66.7)	1.6 (0.9-2.5)	1.08 (1.00-1.18)	-1247.3	0.21
Bleeding G2+								
LKB-EQD2 (No DMF)	2006	0.16 (0.01-0.34)	0.27 (0.14-0.42)	75.8 (68.2-88.6)	1.7 (0.7-3.0)	N/A	-822.6	N/A
Age (years)	2006	0.16 (0.01-0.36)	0.27 (0.14-0.44)	81.0 (57.0-124.3)	1.7 (0.7-3.0)	1.0004 (0.9956-1.0055)	-823.5	Worse Fit
Diabetes Y/N	2006	0.16 (0.01-0.35)	0.27 (0.14-0.42)	75.4 (67.7-88.5)	1.7 (0.7-3.0)	0.94 (0.80-1.03)	-822.6	0.91
Haemorrhoids Y/N	2006	0.16 (0.01-0.34)	0.27 (0.14-0.42)	76.1 (68.2-89.6)	1.7 (0.7-3.1)	1.11 (0.99-1.33)	-821.9	0.21
Hypertension Y/N	2006	0.16 (0.01-0.33)	0.27 (0.14-0.42)	74.6 (66.9-87.5)	1.7 (0.7-3.0)	0.96 (0.89-1.01)	-822.2	0.36
IBD/Diverticular Y/N	2006	0.17 (0.01-0.36)	0.28 (0.14-0.42)	75.9 (68.3-90.1)	1.7 (0.7-3.0)	1.17 (1.03-1.44)	-820.2	0.026
Pelvic Surgery Y/N	2006	0.16 (0.01-0.35)	0.27 (0.14-0.42)	76.2 (68.3-89.3)	1.7 (0.7-3.1)	1.04 (0.94-1.16)	-823.2	Worse Fit
Continued Overleaf…								

Table 12 continued…	**Pts**	**n covariate**	**m covariate**	**TD50 covariate (Gy_{EQD2})**	**α/β ratio (Gy)**	**Dose modifying factor covariate**	**632 Likelihood**	**Likelihood ratio p-value**
Stool Frequency G1+								
LKB-EQD2 (No DMF)	2025	0.27 (0.17-0.44)	0.55 (0.39-0.86)	56.0 (51.4-62.3)	2.3 (0.9-5.3)	N/A	-1334.7	N/A
Age (years)	2025	0.27 (0.17-0.44)	0.54 (0.39-0.81)	38.8 (30.0-57.9)	2.3 (0.9-5.3)	0.9942 (0.9903-1.0003)	-1334	0.25
Diabetes Y/N	2025	0.27 (0.17-0.43)	0.55 (0.39-0.83)	56.6 (51.7-63.3)	2.3 (0.9-5.3)	1.09 (0.97-1.25)	-1334.5	0.52
Haemorrhoids Y/N	2025	0.28 (0.17-0.45)	0.56 (0.40-0.88)	56.8 (51.9-63.3)	2.2 (0.8-5.1)	1.21 (1.06-1.48)	-1331.8	0.016
Hypertension Y/N	2025	0.27 (0.17-0.44)	0.55 (0.39-0.86)	55.6 (50.9-62.4)	2.2 (0.8-5.2)	0.98 (0.89-1.06)	-1335.5	Worse Fit
IBD/Diverticular Y/N	2025	0.27 (0.17-0.44)	0.55 (0.39-0.84)	56.4 (51.4-62.9)	2.3 (0.9-5.5)	1.19 (1.00-1.47)	-1334	0.23
Pelvic Surgery Y/N	2025	0.26 (0.16-0.42)	0.56 (0.40-0.85)	56.8 (51.8-63.7)	2.3 (1.0-5.6)	1.13 (0.99-1.33)	-1334.1	0.28
Stool Frequency G2+								
LKB-EQD2 (No DMF)	2021	0.31 (0.10-0.71)	0.36 (0.23-0.52)	75.7 (66.2-96.8)	2.7 (0.9-8.5)	N/A	-801.3	N/A
Age (years)	2021	0.31 (0.11-0.73)	0.35 (0.24-0.50)	54.4 (30.0-90.0)	2.7 (0.9-8.2)	0.9947 (0.9852-1.0026)	-801.4	Worse Fit
Diabetes Y/N	2021	0.31 (0.10-0.70)	0.36 (0.24-0.51)	75.7 (66.2-93.9)	2.6 (0.9-8.7)	1.02 (0.86-1.17)	-802.2	Worse Fit
Haemorrhoids Y/N	2021	0.31 (0.10-0.71)	0.36 (0.24-0.51)	76.6 (66.6-95.0)	2.7 (1.0-8.9)	1.15 (0.98-1.40)	-800.6	0.22
Hypertension Y/N	2021	0.31 (0.10-0.73)	0.36 (0.23-0.51)	75.2 (65.7-91.8)	2.6 (0.9-8.2)	0.97 (0.86-1.07)	-802.1	Worse Fit
IBD/Diverticular Y/N	2021	0.31 (0.10-0.68)	0.36 (0.23-0.50)	76.2 (66.5-95.3)	2.5 (0.8-7.1)	1.37 (1.13-1.82)	-795.1	**0.00041**
Pelvic Surgery Y/N	2021	0.31 (0.09-0.73)	0.36 (0.24-0.51)	76.7 (66.6-96.3)	2.7 (1.0-9.8)	1.11 (0.95-1.33)	-801.2	0.71
Bowel Pain G1+								
LKB-EQD2 (No DMF)	2185	0.15 (0.01-9.89)	0.48 (0.21-0.68)	139.7 (88.7-499.1)	3.6 (0.0-839.6)	N/A	-647.4	N/A
Age (years)	2185	0.15 (0.01-1.74)	0.50 (0.25-0.74)	87.0 (42.0-179.4)	5.0 (0.2-39.4)	0.9911 (0.4328-1.0064)	-647.9	Worse Fit
Diabetes Y/N	2185	0.16 (0.01-9.79)	0.48 (0.21-0.68)	138.0 (88.0-522.4)	3.7 (0.0-838.7)	0.95 (0.05-1.83)	-648.3	Worse Fit
Haemorrhoids Y/N	2185	0.16 (0.01-9.89)	0.48 (0.21-0.69)	142.5 (88.8-606.3)	3.9 (0.0-921.1)	1.26 (0.85-4.47)	-647.2	0.54
Hypertension Y/N	2185	0.14 (0.01-9.97)	0.46 (0.21-0.68)	137.5 (89.0-591.5)	3.5 (0.0-951.1)	1.04 (0.69-2.07)	-648.2	Worse Fit
IBD/Diverticular Y/N	2185	0.31 (0.01-9.95)	0.52 (0.21-0.70)	151.1 (89.6-867.0)	3.3 (0.0-942.9)	1.79 (1.07-13.76)	-644.2	0.011
Pelvic Surgery Y/N	2185	0.19 (0.01-9.90)	0.49 (0.21-0.69)	142.2 (88.8-647.6)	4.1 (0.0-945.8)	1.06 (0.31-3.28)	-648.2	Worse Fit
Continued Overleaf…								

Table 12 continued… **Proctitis G1+**	**Pts**	**n covariate**	**m covariate**	**TD50 covariate (Gy_{EQD2})**	**α/β ratio (Gy)**	**Dose modifying factor covariate**	**632 Likelihood**	**Likelihood ratio p-value**
LKB-EQD2 (No DMF)	2147	0.14 (0.02-0.22)	0.42 (0.22-0.68)	63.6 (58.7-75.5)	2.7 (1.5-5.4)	N/A	-1384.1	N/A
Age (years)	2147	0.14 (0.02-0.22)	0.42 (0.22-0.68)	54.2 (36.0-79.8)	2.7 (1.5-5.4)	0.9975 (0.9912-1.0030)	-1384.6	Worse Fit
Diabetes Y/N	2147	0.14 (0.02-0.23)	0.42 (0.21-0.68)	62.8 (57.8-74.2)	2.6 (1.5-5.3)	0.84 (0.65-0.94)	-1379	0.0013
Haemorrhoids Y/N	2147	0.14 (0.02-0.22)	0.43 (0.22-0.69)	64.1 (59.3-75.2)	2.7 (1.6-6.0)	1.12 (1.01-1.32)	-1382.6	0.081
Hypertension Y/N	2147	0.14 (0.02-0.21)	0.42 (0.21-0.68)	62.9 (57.8-74.4)	2.6 (1.5-5.2)	0.97 (0.90-1.02)	-1384.3	Worse Fit
IBD/Diverticular Y/N	2147	0.14 (0.02-0.22)	0.43 (0.22-0.68)	64.0 (59.3-75.1)	2.6 (1.5-5.4)	1.27 (1.10-1.58)	-1378	**0.00046**
Pelvic Surgery Y/N	2147	0.14 (0.02-0.21)	0.43 (0.23-0.70)	65.1 (59.6-76.6)	2.7 (1.6-6.2)	1.15 (1.04-1.38)	-1381	0.012
Proctitis G2+								
LKB-EQD2 (No DMF)	2146	0.11 (0.01-0.25)	0.30 (0.17-0.51)	87.8 (75.2-137.0)	2.7 (1.3-15.1)	N/A	-731.9	N/A
Age (years)	2146	0.12 (0.02-0.26)	0.30 (0.16-0.49)	90.1 (75.0-252.8)	2.7 (1.2-9.0)	1.0021 (0.9966-1.0129)	-732	Worse Fit
Diabetes Y/N	2146	0.11 (0.01-0.26)	0.30 (0.17-0.50)	86.9 (74.7-131.7)	2.7 (1.3-12.6)	0.90 (0.62-1.01)	-731.4	0.31
Haemorrhoids Y/N	2146	0.11 (0.01-0.27)	0.30 (0.17-0.51)	88.1 (75.3-136.6)	2.7 (1.3-14.6)	1.06 (0.92-1.31)	-732.2	Worse Fit
Hypertension Y/N	2146	0.11 (0.01-0.29)	0.30 (0.17-0.49)	86.7 (74.6-125.7)	2.6 (1.2-9.4)	0.96 (0.84-1.03)	-732.2	Worse Fit
IBD/Diverticular Y/N	2146	0.11 (0.01-0.26)	0.30 (0.17-0.51)	88.8 (75.5-138.6)	2.6 (1.2-11.2)	1.22 (1.04-1.69)	-728.9	0.015
Pelvic Surgery Y/N	2146	0.11 (0.01-0.28)	0.30 (0.17-0.51)	89.2 (75.9-142.2)	2.8 (1.3-14.7)	1.11 (0.99-1.42)	-730.9	0.16
Sphincter Control G1+								
LKB-EQD2 (No DMF)	2199	0.23 (0.15-0.38)	0.32 (0.24-0.45)	79.3 (69.8-103.3)	3.1 (1.4-9.1)	N/A	-749.7	N/A
Age (years)	2199	0.24 (0.15-0.38)	0.34 (0.24-0.45)	90.0 (63.0-186.1)	3.0 (1.4-8.5)	1.0024 (0.9968-1.0102)	-750	Worse Fit
Diabetes Y/N	2199	0.24 (0.15-0.39)	0.32 (0.24-0.45)	78.8 (69.4-99.7)	3.1 (1.4-9.4)	0.93 (0.73-1.06)	-750.2	Worse Fit
Haemorrhoids Y/N	2199	0.24 (0.15-0.38)	0.32 (0.24-0.44)	80.3 (70.2-104.1)	3.2 (1.5-10.2)	1.15 (1.00-1.37)	-748.5	0.14
Hypertension Y/N	2199	0.24 (0.15-0.38)	0.32 (0.24-0.42)	79.4 (69.7-95.4)	3.1 (1.4-8.8)	1.01 (0.93-1.10)	-750.5	Worse Fit
IBD/Diverticular Y/N	2199	0.24 (0.15-0.40)	0.33 (0.24-0.45)	80.6 (70.2-104.0)	3.1 (1.4-8.8)	1.29 (1.10-1.64)	-745.3	0.0032
Pelvic Surgery Y/N	2199	0.24 (0.15-0.39)	0.33 (0.24-0.45)	80.5 (70.2-103.6)	3.2 (1.4-10.2)	1.11 (0.96-1.30)	-749.4	0.48
Stricture/Ulcer G1+								
LKB-EQD2 (No DMF)	2206	0.31 (0.01-0.74)	0.25 (0.10-0.34)	83.8 (71.5-110.3)	2.5 (0.9-8.2)	N/A	-335.1	N/A
Age (years)	2206	0.28 (0.01-0.63)	0.25 (0.15-0.31)	136.4 (78.7-343.7)	2.4 (0.9-6.7)	1.0071 (0.9990-1.0184)	-333.9	0.12
Diabetes Y/N	2206	0.31 (0.01-0.74)	0.25 (0.11-0.34)	83.6 (71.4-110.0)	2.5 (0.9-8.1)	0.97 (0.74-1.12)	-336.1	Worse fit
Haemorrhoids Y/N	2206	0.31 (0.01-0.75)	0.25 (0.11-0.34)	83.8 (71.6-109.4)	2.5 (0.9-8.2)	1.04 (0.84-1.23)	-336	Worse fit
Hypertension Y/N	2206	0.31 (0.01-0.74)	0.24 (0.11-0.33)	84.5 (71.9-108.3)	2.5 (0.9-7.5)	1.03 (0.93-1.13)	-335.8	Worse fit
IBD/Diverticular Y/N	2206	0.32 (0.01-0.76)	0.25 (0.12-0.35)	84.1 (71.6-112.7)	2.5 (0.9-8.5)	1.05 (0.73-1.33)	-336.3	Worse fit
Pelvic Surgery Y/N	2206	0.32 (0.01-0.76)	0.25 (0.11-0.35)	85.0 (72.0-113.8)	2.6 (1.0-9.4)	1.08 (0.91-1.30)	-335.5	Worse fit

3.4.7 Calibration Curves for the LKB-EQD2 Models

Calibration curves and binned calibration plots are presented for the LKB-EQD2 model (i.e. including the α/β ratio) fitted to rectal bleeding G1+ in **Figure 8 & Figure 9**. This is an example of a better calibrated endpoint, where the calibration curve can be seen to sit close to the ideal line in the working range of probable toxicity frequency.

An example of a poorly calibrated endpoint, stricture/ulcer G1+, is shown in **Figure 10 & Figure 11**. Here the bunched up and disordered calibration bins, are readily apparent, along with the deviation of the calibration curve from the line of ideal fit.

Figure 8. Rectal Bleeding G1+ Calibration Curve: LKB-EQD2 Model

Note the zoomed in axes.

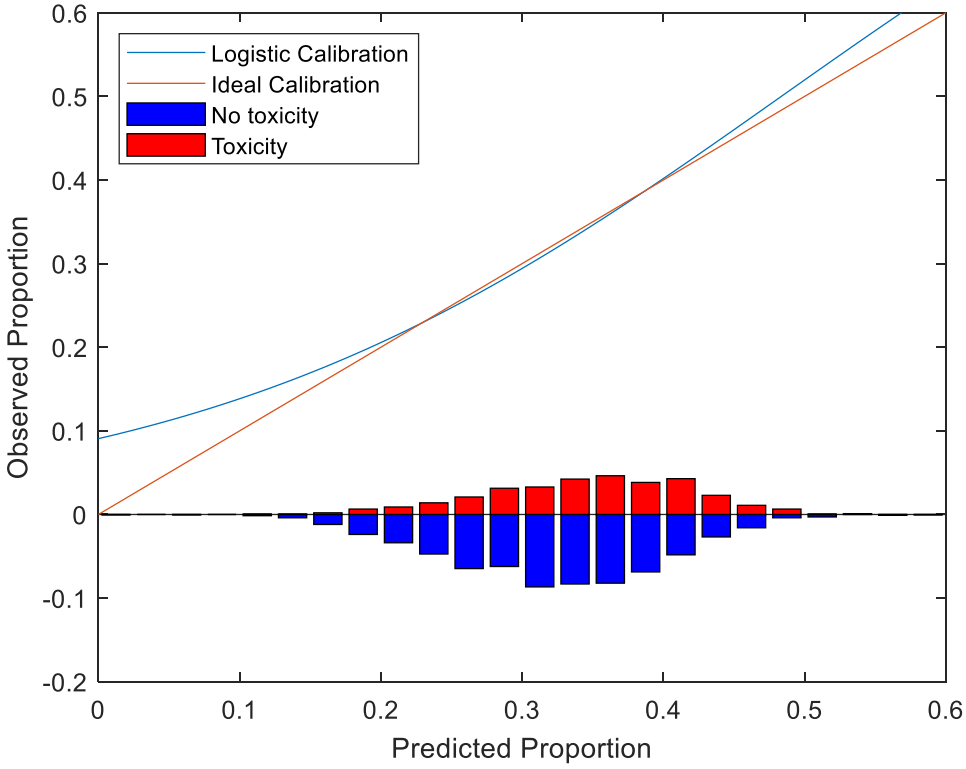


Figure 9. Rectal Bleeding G1+ Binned Calibration Plot: LKB-EQD2 Model

Note the zoomed in axes.

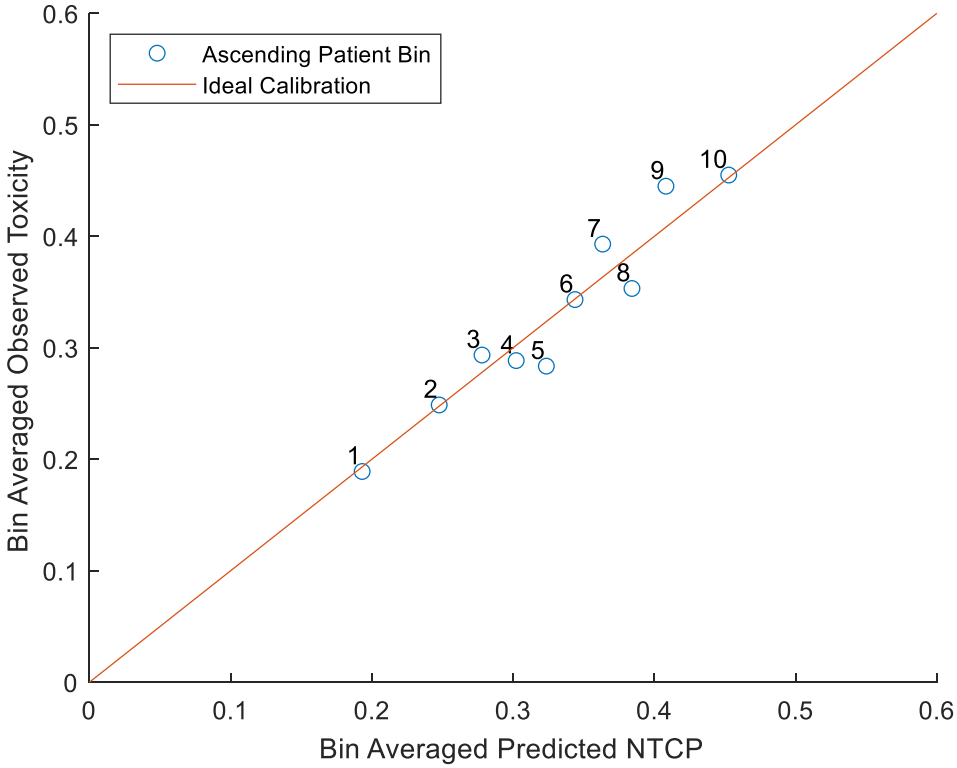


Figure 10. Stricture/Ulcer G1+ Calibration Curve: LKB-EQD2 Model

Note the zoomed in axes.

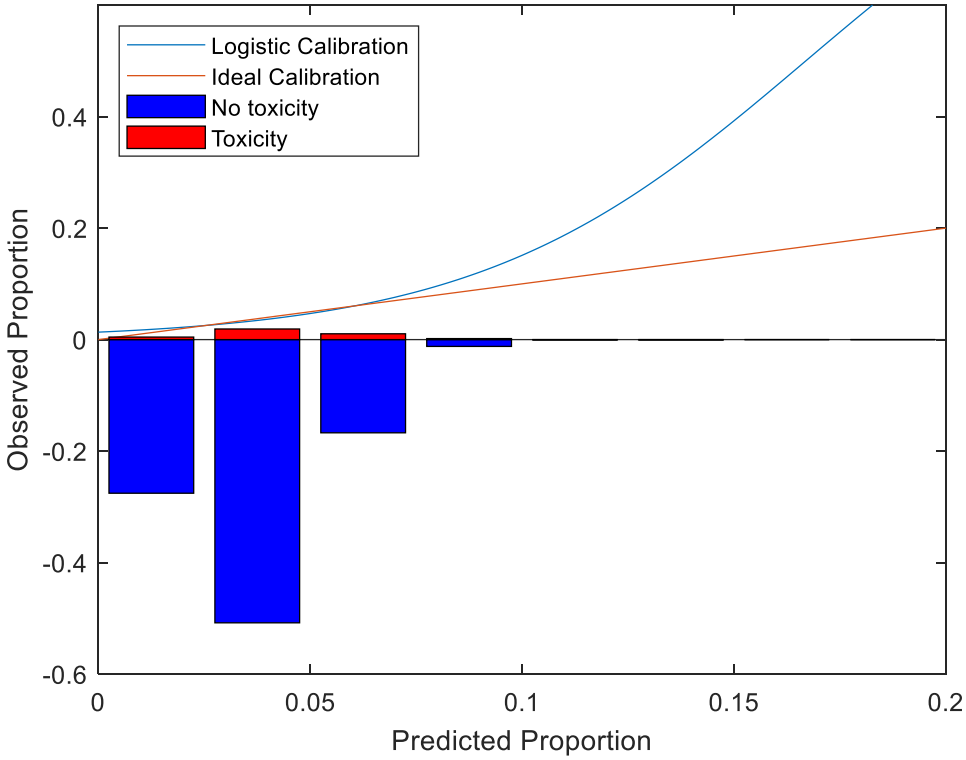
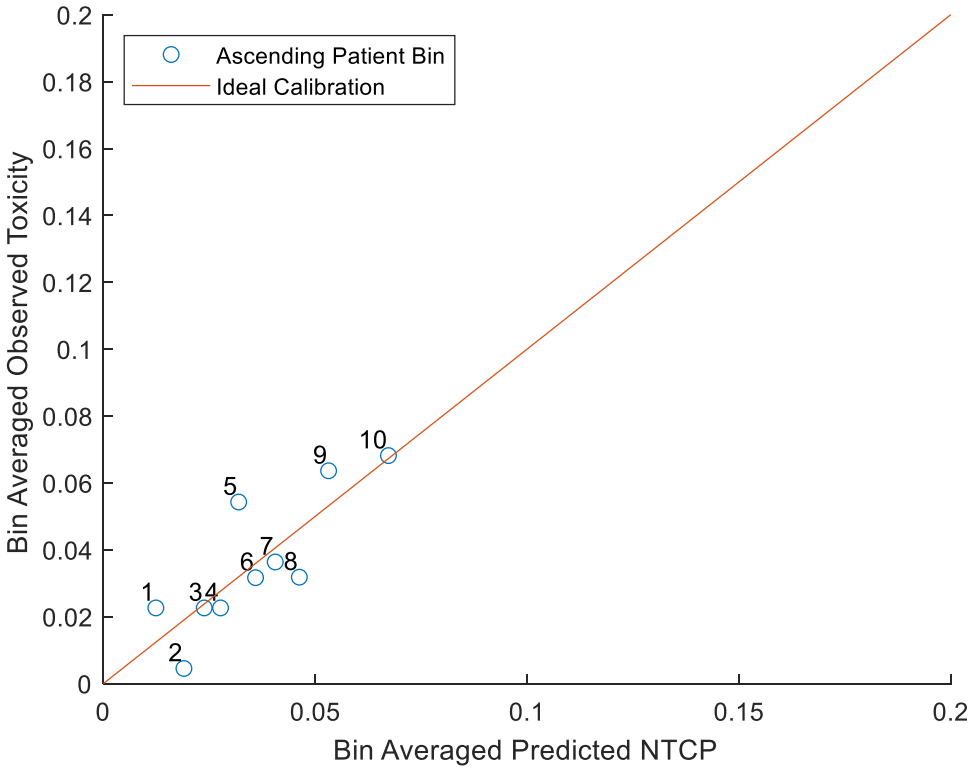


Figure 11. Stricture/Ulcer G1+ Binned Calibration Plot: LKB-EQD2 Model

Note the zoomed in axes.



3.4.8 Calibration of LKB-EQD2-DMF Models

Inclusion of IBD/Diverticular disease in an LKB-EQD2-DMF model significantly improved over the LKB-EQD2 model, for two endpoints. The binned calibration plots are compared between LKB-EQD2 and LKB-EQD2-DMF models for these endpoints: stool frequency G2+ (**Figure 12 & Figure 13**) and proctitis G1+ (**Figure 14 & Figure 15**). In both cases, it can be seen that the inclusion of a DMF has predominantly acted to increase separation of the highest risk bins.

Figure 12. Stool Frequency G2+ Binned Calibration Plot: LKB-EQD2 Model

Note the zoomed in axes.

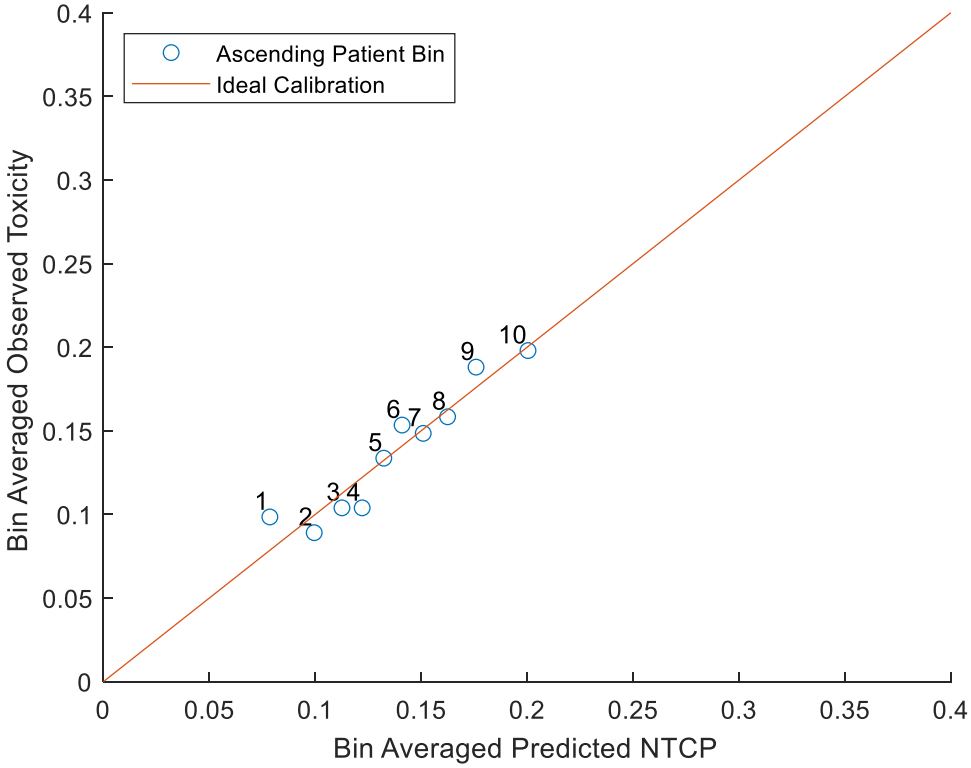


Figure 13. Stool Frequency G2+ Binned Calibration Plot: LKB-EQD2-DMF Model. DMF = IBD/Diverticular

Note the zoomed in axes.

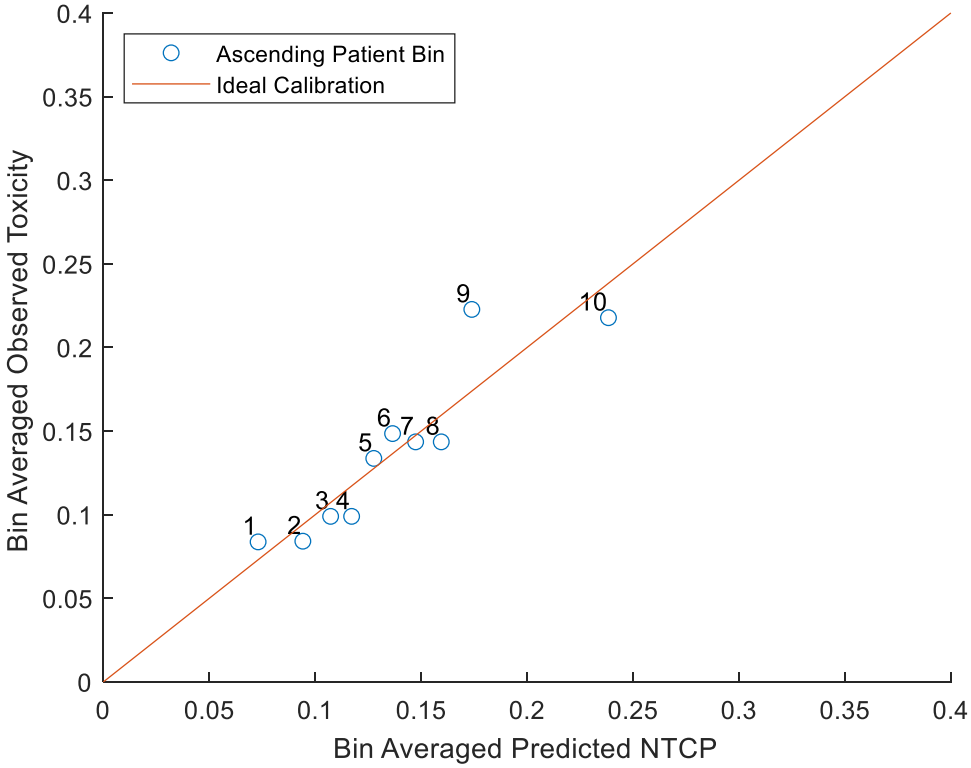


Figure 14. Proctitis G1+ Binned Calibration Plot: LKB-EQD2 Model

Note the zoomed in axes.

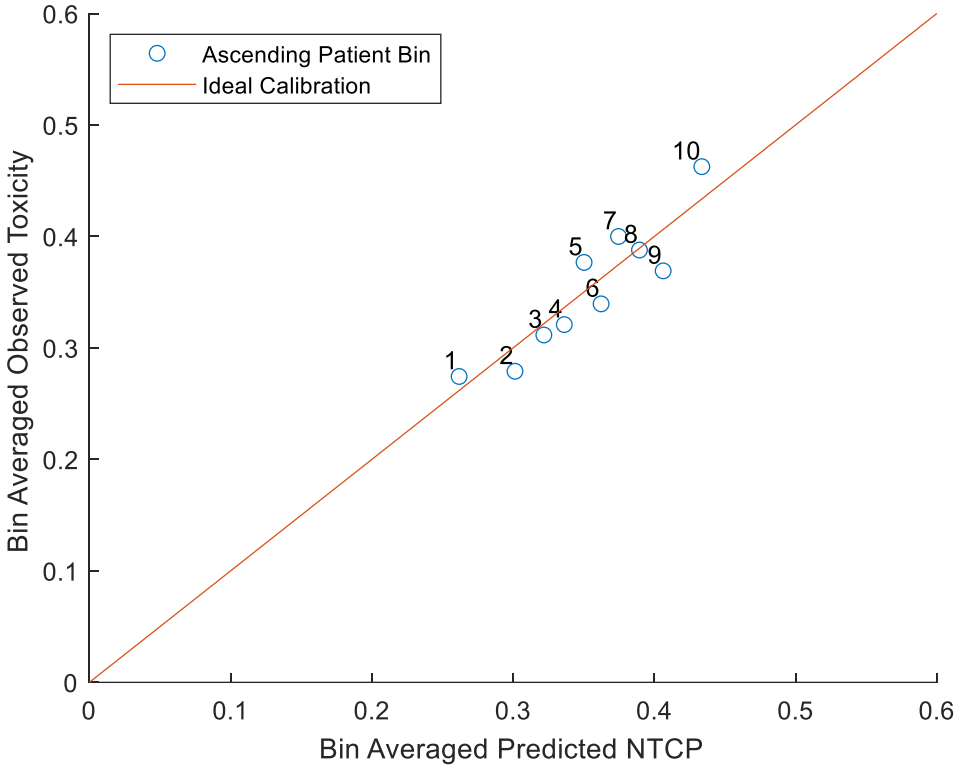
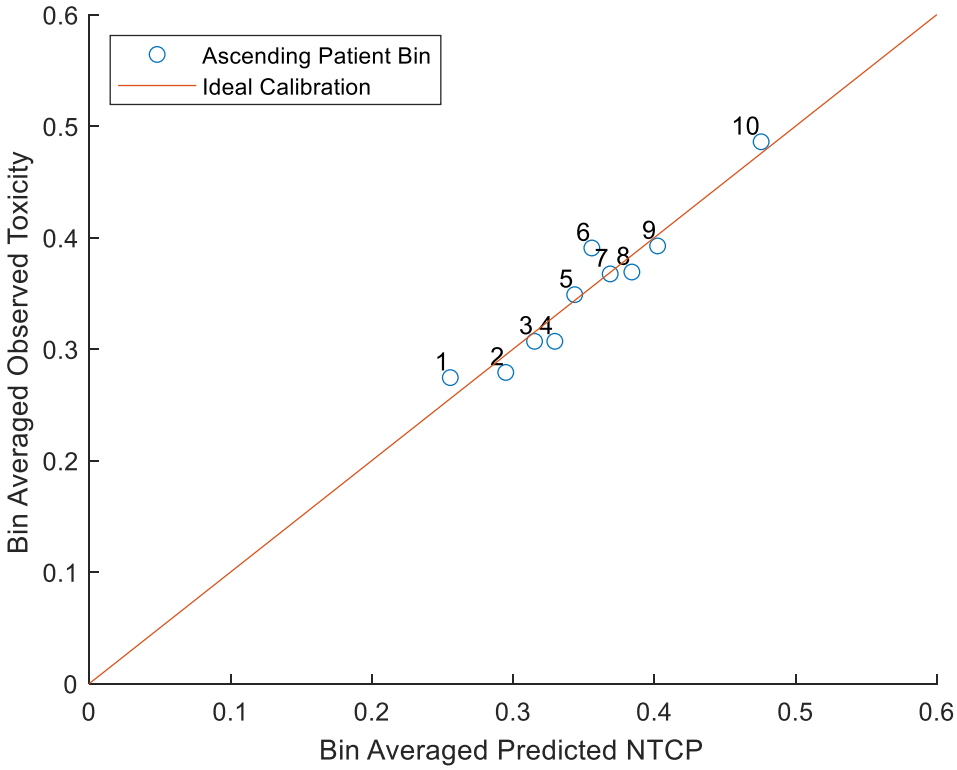


Figure 15. Proctitis G1+ Binned Calibration Plot: LKB-EQD2-DMF Model

DMF = IBD/Diverticular

Note the zoomed in axes.



3.4.9 Pooling of the α/β Ratio Estimates

One overall late rectal α/β ratio for use in the comparison of expected late rectal side effects between differing dose-fractionation schedules is desirable. The frequency weighted average for modelled late rectal G1+ events (excluding pain re poor fit) was α/β = 2.4 Gy and the equivalent for G2+ events was α/β = 2.3 Gy. Unfortunately, no transformation was found to normalise the highly positively skewed bootstrapped α/β ratio 95% CIs, meaning pooling standard errors for a unified 95% CI is not appropriate [184]. Caution is of course advised, in the application of any single figure, since as demonstrated, the true fraction size sensitivity may differ between endpoints. The calculation of these estimates is shown in **Table 13.**

Table 13. Calculation of Pooled Rectal Late α/β Ratios

Calculation for the pooled averaged late rectal α/β ratio for those more common endpoints fitted best during modelling. Weighting is by the frequency of side effect occurrence seen in patients modelled (per **Table 9**).

Late Rectal Endpoints	Frequency	Weights	α/β Ratio (Gy)
Grade 1+ Endpoints			
Bleeding G1+	0.329	9.139	1.58
Frequency G1+	0.381	10.583	2.26
Pain G1+	0.087	2.417	3.64
Proctitis G1+	0.352	9.778	2.65
Sphincter Control G1+	0.109	3.028	3.09
Stricture/Ulcer G1+	0.036	1	2.49
Grade 1+ weighted average			**2.36**
Grade 2+ Endpoints			
Bleeding G2+	0.146	1.352	1.71
Frequency G2+	0.138	1.278	2.66
Proctitis G2+	0.108	1	2.70
Grade 2 weighted average			**2.32**

3.4.10 Whole-Trial Level Meta-Analysis Results

Fitting an α/β value to minimise the squared difference between EQD2 and G2+ toxicity rate, the best value is late rectal toxicity α/β ratio = 2.8 Gy (fit this this value shown in **Table 14**). However, plotting fit quality across the possible α/β ratios, a value of 3 Gy is very similar (**Figure 16**).

Table 14. Best Fit for Rectal α/β Ratio in Moderate Hypofractionation Trials

Produced by fitting rectal α/β ratio, to minimise the sum of squared differences between each trial toxicity percentage difference (test minus control) and EQD2 difference (test minus control).

GI Toxicity α/β = 2.83 Gy	Test		Control		Differences		(EQD2 difference - Toxicity difference)^2
	EQD2 (Gy)	Toxicity G2+ (%)	EQD2 (Gy)	Toxicity G2+ (%)	EQD2 (Gy)	Toxicity (%)	
CHHiP 57Gy	68.8	11.3	74.0	13.7	-5.2	-2.4	7.83
CHHiP 60Gy	72.4	11.9	74.0	13.7	-1.6	-1.8	0.05
PROFIT	72.4	8.9	78.0	14	-5.6	-5.1	0.23
RTOG 0415	77.2	22.4	70.7	14	6.5	8.4	3.60
HYPRO	83.3	21.9	78.0	17.7	5.3	4.2	1.26
HYPO-RT-PC	78.9	9.5	78.0	9.7	0.9	-0.2	1.31
						Sum of squares	**14.3**

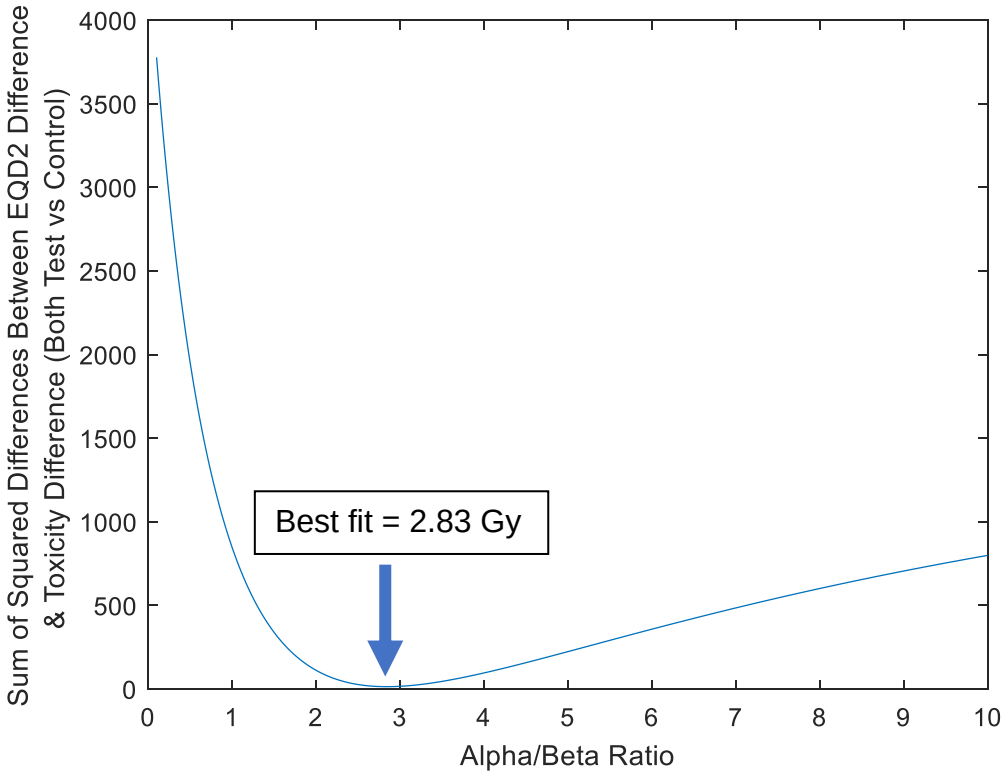


Figure 16. Fitting of α/β Ratio to Moderate Hypofractionation Trials

The fit values across the range of possible rectal α/β ratios to fit whole-trial level phase III hypofractionation data as outlined in **Table 14**. Showing the optimal fit (minima) as late rectal α/β ratio = 2.83 Gy. It can be seen that a value of 3 Gy is very close in terms of fit quality.

Plotting each trial arm EQD2 (with 2.8 Gy α/β ratio) versus trial arm G2+ toxicity shows a fairly good fit for dose response relationship in each trial (**Figure 17**), with similar dose-response lines.

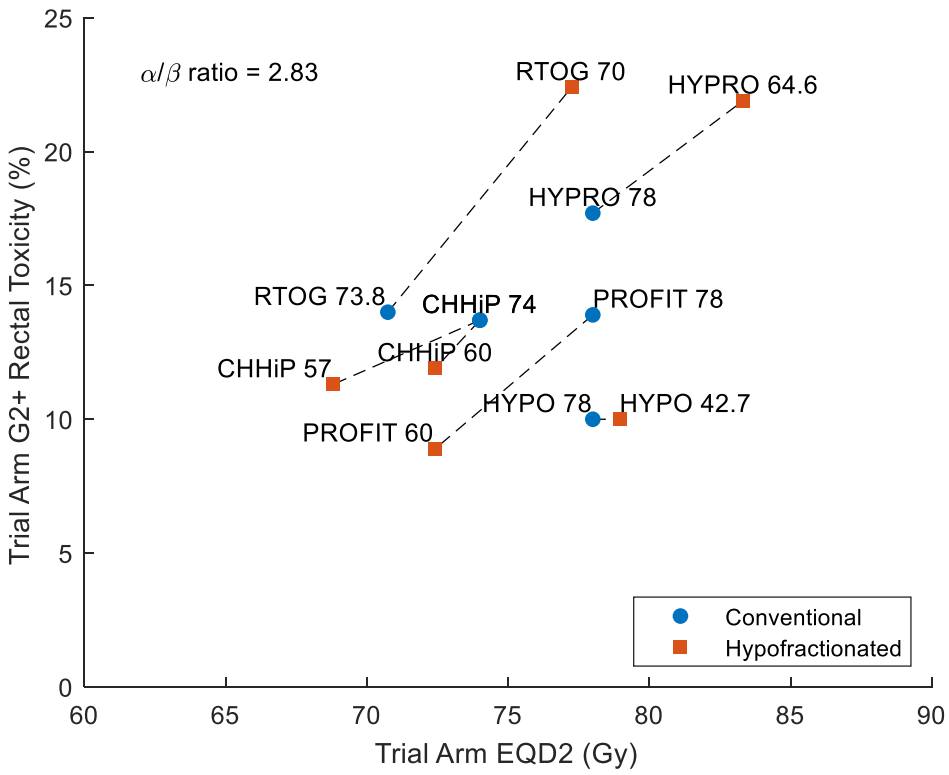


Figure 17. Fitted EQD2 vs Late Grade 2+ Rectal Toxicity for Moderate and Ultra Hypofractionation Phase III Trials

The EQD2 for each test and control arm is plotted against rectal toxicity, with the fitted late rectal α/β ratio = 2.83 Gy. Each control arm point (blue, circle) is connected to its relevant hypofractionation arm(s) (orange, square) by dashed line.

3.5 Discussion

3.5.1 Summary of Findings

In this study, I have used data from a large phase III trial of moderately hypofractionated radiotherapy (MHRT) for nmPCa. Through fitting an EQD2-corrected LKB model, estimates of the relative fraction size sensitivity (expressed as α/β ratio) for various clinician reported late rectal endpoints have been made. I have shown that these estimates do not vary markedly with inclusion of several possible dose modifying factors. To my knowledge, these are the first such individual rectal endpoint α/β ratio estimates in the literature.

3.5.2 Prior Studies in the Field

3.5.2.1 Rectal α/β ratio estimates

The prior work in this field has been discussed earlier (**Section 2.3.4.4**). The α/β ratio estimates here are generally lower than Brenner's whole-trial level analysis (5.4 Gy, 95% CI 3.9 – 6.9 Gy) and the individual patient LKB-EQD2 analyses by Tucker (4.8 Gy, **68%** CI 0.6 – 46 Gy) [88] and Zhu 7.17 Gy (95% CI 5.21 – 9.13 Gy) [89]. Having a wider range of fraction sizes should be beneficial to modelling. It is therefore interesting that the α/β ratio estimates here are more similar to Marzi *et al* (2.3 Gy, 95 % CI 1.1 – 5.6 Gy) [62]; the only other study where patients received moderately hypofractionated radiotherapy at ≥3 Gy per fraction.

3.5.2.2 The LKB-NoEQD2 Model in Context

Regarding the components of the traditional LKB model (*n*, *m*, *TD50*), it is reassuring that the LKB-NoEQD2 estimates for conventionally fractionated patients are similar to those previously reported for individual rectal endpoints [11,185–187]. Estimates from these cohorts for bleeding, stool frequency and proctitis are compared to this study in **Table 15**. The landmark Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) study meta-analysed LKB parameters from four of these studies, examining either G2+ rectal bleeding or G2+ late toxicity [154]. Comparing our G2+ rectal bleeding LKB-NoEQD2 values for 74 Gy patients versus these QUANTEC meta-analysis values, we see fairly similar findings: n = 0.13 (0.01-0.42) vs 0.09 (0.04–0.14) ; m = 0.21 (0.06-0.43) vs 0.13 (0.10–0.17); and TD50 = 74.0 (67.2-96.6) vs 76.9 (73.7–80.1) Gy.

Separately, it must be noted that the models for pain produced very wide confidence intervals (e.g. LKB-EQD2 α/β ratio estimate 3.6 Gy, 95% CI 0.01 – 840), suggestive of poor model fit for this endpoint. This is perhaps expected, given the relative subjectivity of pain and relative rarity as a radiotherapy related side effect.

Table 15. LKB-NoEQD2 Parameter Comparison with Literature Studies

Parameters n, m, TD50 for LKB model without EQD2 correction fitted on conventionally Fractionated 74Gy Patients. Comparing with other studies fitting similar endpoints. Peeters *et al* [185] incontinence data omitted as modelled only on anal wall OAR.

Endpoint	Study	Pts	n	95% CI *(68% CI)*	m	95% CI *(68% CI)*	TD50	95% CI *(68% CI)*
Bleeding G1+	This Study	644	0.26	0.08–0.70	0.38	0.17–0.76	51.7	44.6–60.8
	Gulliford *et al* [11]	361	0.14	0.09–0.16	0.26	0.18–0.48	59.2	57.8–61.9
Bleeding G2+	This Study	642	0.14	0.08–0.24	0.23	0.14–0.40	57.1	52.8–61.2
	Gulliford *et al*	361	0.12	0.10–0.16	0.14	0.12–0.16	68.2	64.9–69.3
	Peeters *et al* [185]	468	0.13	*(0.04–0.25)*	0.14	*(0.11–0.19)*	81.0	*(75–90)*
	Defraene *et al* [186]	512	0.18	–	0.15	–	79.0	–
	Rancati *et al* [187]	547	0.23	*(0.14–0.42)*	0.19	*(0.15–0.25)*	81.9	*(76.8–91.2)*
Frequency G1+	This Study	643	0.17	0.01–0.42	0.33	0.10–0.70	55.1	47.9–70.1
	Gulliford *et al*	344	0.30	0.16–0.6	0.60	0.41–>1	61.5	56.3–68.3
Frequency G2+	This Study	642	0.11	0.04–0.32	0.22	0.11–0.49	58.9	50.4–65.5
	Peeters *et al*	468	0.39	*(0.19–1.11)*	0.24	*(0.18–0.35)*	84.0	*(75–103)*
	Defraene *et al*	512	1.18	–	0.34	–	97.4	–
Proctitis G1+	This Study	691	0.10	0.01–0.16	0.24	0.09–0.47	59.5	55.9–70.6
	Gulliford *et al*	388	0.14	0.11–0.20	0.28	0.19–0.60	58.2	55.7–60.1
Proctitis G2+	This Study	691	0.06	0.01–0.09	0.15	0.08–0.29	63.7	60.5–70.7
	Gulliford *et al*	388	0.15	0.11–0.20	0.20	0.19–0.24	67.0	64.8–69.3

3.5.2.3 Results in the Context of the Phase III Hypofractionation Trials

In the context of these results, it is worth re-examining the α/β ratio assumptions (**Table 2**) and subsequent toxicity outcomes (**Table 1**) of the published phase III hypofractionation trials.

CHHiP assumed a late rectal α/β ratio = 3 Gy, isoeffective design, with the 60 Gy and 57 Gy arms reflecting uncertainty in the prostate α/β ratio (assumed α/β 2.5 Gy and 1.5 Gy respectively). Both 60 Gy and 57 Gy arms showed non-significantly reduced cumulative rectal grade 2+ toxicity by 5 years (11.9% & 11.3% vs 13.7% control arm), with the 60 Gy arm shown to be non-inferior for disease control [1].

PROFIT assumed late rectal α/β ratio = 3-5 Gy with isoeffective design (prostate α/β ratio 1 – 3 Gy), achieving non-inferior disease control with reduced late grade 2+ rectal toxicity in the test arm (8.9% vs 13.9%) [3]. The lower rectal toxicity seen for hypofractionated regimens in both CHHiP and PROFIT match anticipated therapeutic gains with rectal α/β exceeding that of the prostate.

RTOG 0415 assumed both tumour and late rectal α/β = 3 Gy, with the trial design escalating EQD2 to both [2]. The trial achieved non-inferior disease control with hypofractionation. Given the rectal dose escalation, the increased G2+ rectal toxicity in the hypofractionated arm (22.4% vs 14.0%) is not surprising.

The HYPRO trial adopted an isotoxic design, assuming the highest α/β ratio for late rectal toxicity (α/β = 4-6 Gy) [75]. Unfortunately, this study demonstrated increased late G2+ rectal toxicity (21.9% vs 17.7%), without superior disease control. It is worth noting that HYPRO is the only phase III MHRT study where the relative test vs control late rectal toxicity was worse than trial design anticipated. Based on the low α/β ratios seen here, this was likely due to the higher assumed rectal α/β ratio and therefore effective rectal dose delivered to the test arm.

Both large phase III randomised trials of prostate ultra-hypofractionation: PACE-B [188] and HYPO-RT-PC [95] have assumed a late rectal α/β = 3 Gy. The HYPO-RT-PC trial showed isoeffective cumulative grade 2 or worse late RTOG rectal toxicity for both arms: 42.7 Gy in 7 fractions (9.5%) and 78 Gy in 39 fractions (9.7%) [95].

Corrected for multiple testing, the LKB-EQD2 models with freely fitted α/β ratios did not significantly outperform the same model with fixed α/β = 3 Gy. I note that the upper bound of 95% CI for rectal bleeding G1+ was below 3 Gy and that the results were close to corrected significance. This is perhaps worth noting, given that the randomised ProtecT trial showed bloody stools to be the most common radiotherapy patient reported adverse event compared to radical prostatectomy, although the long term impact on bowel habits and bother was very small [21].

3.5.2.4 Meta-Analysis Fitting of Hypofractionation Trials

The pooled estimates seen in the individual patient level data LKB fitting (2.3-2.4 Gy) are lower than the typical 3 Gy used in clinical practice. However, it is reassuring that the whole-trial level meta-analysis suggests 2.83 Gy as best fit for EQD2 and rectal toxicity. Taken together, this would strongly suggest that a late rectal α/β ≤ 3 Gy should be used in any future clinical work.

3.5.3 Strengths of this Study

Strengths of this study are drawn from the nature of the inputted data. The CHHiP trial is the largest study of hypofractionated radiotherapy for PCa, with two thirds of patients' data used for this analysis. I have included only patients reporting zero baseline toxicity, in order to reduce possible pre-existent toxicity noise. Furthermore, data quality assurance has been undertaken: checking every rectal contour for protocol adherence and recalculating DVHs. This large, clean sample, combined with multiple dose-fractionation regimens, has permitted α/β ratio estimation with tight confidence intervals and good calibration for more frequently occurring endpoints. This is without the need to fix any of the parameters when modelling as has been done previously [62]. This study has also been aided by modern computing power facilitating usage of computationally intensive bootstrapping techniques. These have permitted nested model comparison using bootstrap-dependent estimates of test performance (632 estimate), reducing the potential influence of overfitting.

3.5.4 Limitations of this Study

Limitations must also be considered, starting with the modelling approach itself. The LKB model is a traditional parametric method for the fitting of radiotherapy data and more recently machine learning and artificial intelligence type modelling methodologies have been applied [189]. It does however, provide a model which permits fitting of data, with and without EQD2 correction, to estimate endpoint α/β ratios. Future toxicity modelling work with newer methodologies may benefit from these α/β ratio estimates, when using the LQ-model to rescale DVH data predictors from disparate dose-fractionation regimens.

Prostate inter-fraction motion (**recall 2.3.2.5**) means the use of CT planned doses in this study is a limitation, although necessary to utilise this large dataset. Further, I acknowledge that the endpoints modelled here are unlikely to recur in future trials, due to the amalgamation of multiple scales. This was a pragmatic choice based on the toxicity scales available, so there would be benefit to confirmatory studies with modern CRO scales (e.g. CTCAE) or PRO scales (e.g. EPIC [133]). Finally, despite the use of out-of-the-bag techniques, this is data from a single study and future validation on another hypofractionated prostate radiotherapy dataset would be desirable.

3.6 Conclusions

To my knowledge this study is the first to provide α/β ratio estimates for individual late rectal toxicity endpoints seen following moderately hypofractionated EBRT for nmPCa. For G1+ rectal bleeding, the α/β ratio 95% CI upper bound was lower than 3 Gy. For G1+ endpoints, the frequency-weighted pooled estimate was late rectal α/β ratio = 2.4 Gy. However, adjusting for multiple testing, no significant improvement from an LKB-EQD2 model with α/β = 3 Gy was demonstrated. The findings are reinforced by a value of 2.8 Gy being the best fit for rectal toxicity at a whole-trial level across the phase III hypofractionation trials. Future individual patient data level analysis on ultra hypofractionated trials is desirable, but at present I would suggest a late rectal α/β ratio of no more than 3 Gy be used when comparing dose-fractionation regimens.

Chapter 4. Clinician Reported Acute Toxicity in the PACE-B Trial

4.1 Publications and Proceedings Relating To Chapter

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

Douglas H Brand, Alison C Tree, Peter Ostler, Hans van der Voet, Andrew Loblaw, … , Emma Hall, Nicholas van As, on behalf of the PACE Trial Investigators. Lancet Oncology 2019 Nov;20(11):1531-1543.

DOI: 10.1016/S1470-2045(19)30569-8.

[Full Length Research Article]

PACE: Analysis of acute toxicity in PACE-B, an international phase III randomized controlled trial comparing stereotactic body radiotherapy (SBRT) to conventionally fractionated or moderately hypofractionated external beam radiotherapy (CFMHRT) for localized prostate cancer (LPCa).

Nicholas John Van As, **Douglas Brand**, Alison Tree, Peter James Ostler, William Chu, Andrew Loblaw, … , Emma Hall. Journal of Clinical Oncology 2019 37:7_suppl, 1-1. DOI: 10.1200/JCO.2019.37.7_suppl.1

[Oral Presentation at GU-ASCO 2019]

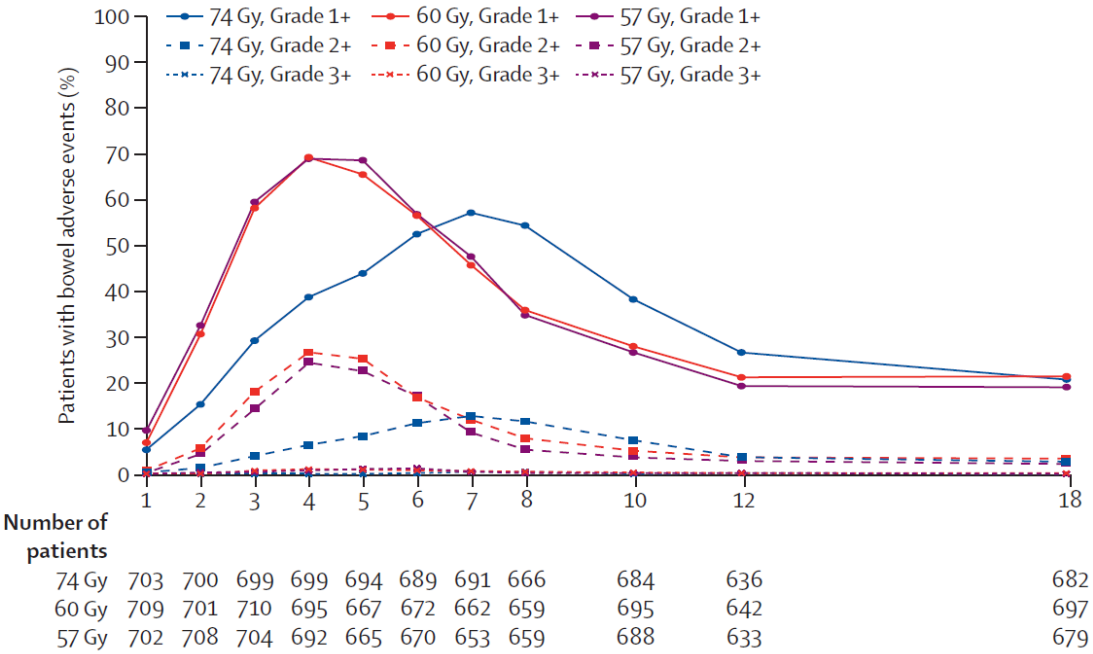
4.2 Background

Given the low apparent α/β ratio for PCa and excellent results from MHRT trials [1–3], ultrahypofractionation for nmPCa is of significant interest (**Section 2.3.5**). This has commonly been delivered over five fractions in a number of large prospective, early phase trials [94]. The advantages are several: possible gain in therapeutic ratio, reduction in number of attendances for patient ,and potential savings in total cost of healthcare delivery.

Trial designs for both phase III trials of ultrahypofractionation, PACE [188] and HYPO-RT-PC [95], have chosen non-inferiority designs for bPFS. Should the conclusion be non-inferiority of SBRT for disease control, then the uptake of this new technology will be determined by the balance of new toxicity versus monetary & time benefits.

Classical radiobiology would suggest that a reduction in overall treatment time will result in an increase in acute toxicity (**Recall 2.3.5.3.1**). In the prostate setting, the CHHiP trial demonstrated an earlier time course of acute toxicity with the two moderate hypofractionation delivered over 3.8/4 weeks (**Figure 18**), including a higher grade 2+ GI toxicity peak.

Panel A



Panel B

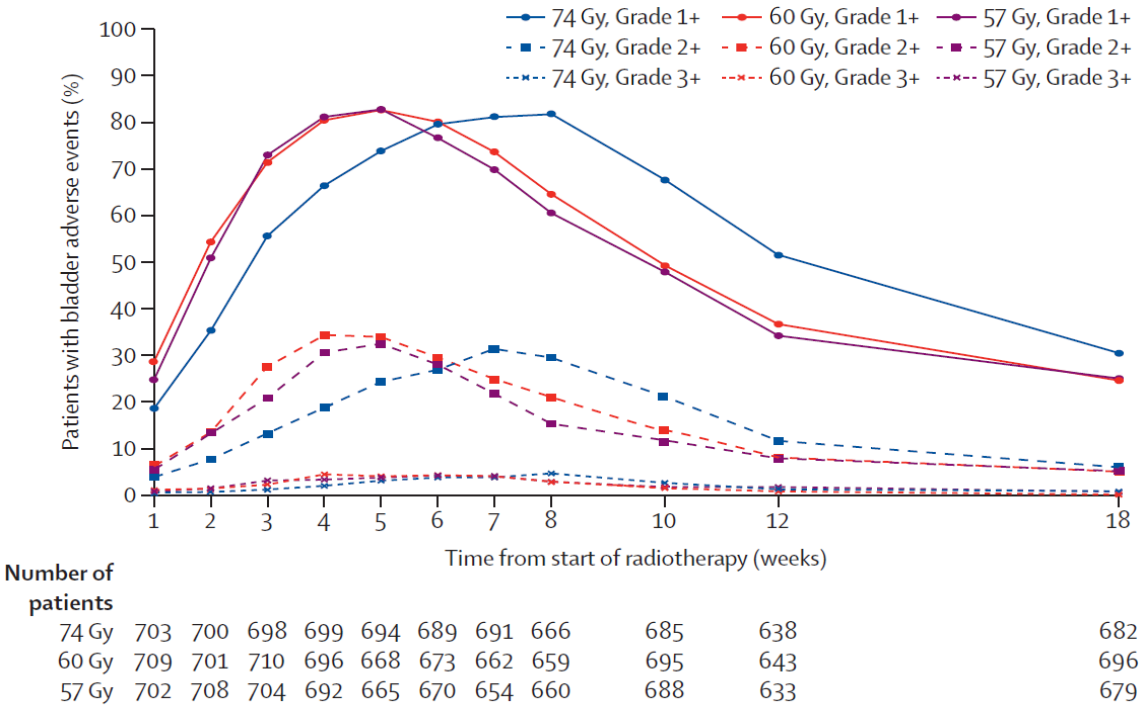


Figure 18. Acute RTOG side effect prevalence over time in the CHHiP trial.

Panel A: Gastrointestinal. Panel B: Genitourinary.

Reproduced, with modification, from Dearnaley et al [1], under CC-BY licence.

https://doi.org/10.1016/S1470-2045(16)30102-4

https://creativecommons.org/licenses/by/4.0/

Ultrahypofractionation allows even faster overall treatment time completion than MHRT. However, it is important that SBRT does not substantially worsen acute toxicity when MHRT already offers an efficacious and relatively convenient regimen for patients. This work will examine whether the effects seen with moderate hypofractionation in CHHiP may be more pronounced for an SBRT regimen; i.e. a faster time course and more pronounced peaks in proportions of patients experiencing toxicity. This will be examined through the CRO reported toxicity, which was collected at higher frequency in the PACE-B trial than PRO data.

A further consideration is whether the method of treatment delivery may modulate toxicity amongst patients receiving SBRT. Platforms permitting SBRT include CK and more conventional co-planar LINACs, typically utilising VMAT (recall **Section 2.3.2.6.2**). The CK platform requires real-time fiducial tracking, allowing intra-fractional adjustment to prostate motion, which reduces the required CTV to PTV margins. For patients receiving CK, these differences may reduce the dose delivered to normal tissues surrounding the prostate and therefore attenuate acute toxicity. There is not previous work in the literature to examine these platform effects from a toxicity perspective.

4.3 Hypotheses

1) Comparing conventionally fractionated or moderately hypofractionated radiotherapy (CFMHRT) and SBRT, clinician reported acute gastrointestinal toxicity may differ between modalities

2) Comparing CFMHRT and SBRT, clinician reported acute genitourinary toxicity may differ between modalities

3) Comparing CFMHRT and SBRT, clinician reported acute sexual toxicity may differ between modalities

4) For SBRT patients, clinician reported acute toxicity may differ between those receiving CyberKnife versus other treatment platforms

4.4 Methods

4.4.1 The PACE Study

PACE-B is a multicentre, international, phase III, open-label RCT. It aims to demonstrate non-inferiority of SBRT compared with CFMHRT for biochemical/clinical failure. CRO acute toxicity and PROs are pre-defined secondary endpoints, the analysis of which are reported here and in **Chapter 5**. The PACE trial was already designed and underway at the onset of this thesis work.

The PACE study (NCT01584258) comprises multiple cohorts (PACE-A, PACE-B, and PACE-C) with independent eligibility and randomisations. PACE-A randomises patients eligible for surgery between radical prostatectomy and surgery. PACE-C will be discussed later. PACE-B, recruited only patients suitable for radical radiotherapy, but not willing and/or not suitable for radical prostatectomy. The trial was approved by the London Chelsea Research Ethics Committee (REC) (ref 11/LO/1915) in the UK and the relevant institutional review boards in Ireland and Canada. The trial is sponsored by The Royal Marsden Hospital NHS Foundation Trust, funded by Accuray Incorporated (Sunnyvale, CA, USA) and endorsed by Cancer Research UK. The ICR-CTSU coordinated the trial.

4.4.2 Participants and Eligibility

For PACE-B, eligible patients were men aged ≥18 years, with WHO performance status 0-2 [190], life expectancy of ≥5 years and histologically

confirmed prostate adenocarcinoma. All patients had NCCN low or intermediate risk disease [18]. Low risk patients were: cT1c-T2a (TNM 6th edition [191]), N0/X, M0/X; Gleason score ≤ 6 ; PSA < 10ng/mL. Intermediate risk patients had ≥1 of: T2c; Gleason score 3+4 (Gleason 4+3 excluded); PSA 10-20 ng/mL. Distant staging was not mandated. A minimum of 10 biopsy cores, ≤18 months pre-randomisation, were required[27]. No PSA adjustment was made for 5-alpha reductase inhibitor (5-ARI) use at randomisation. Willingness and ability to participate with PRO measures was mandatory (reported in **Chapter 5**). Exclusion criteria were: disease of higher risk than permitted; prior ADT; prior active treatment for PCa; life expectancy < 5 years; bilateral hip prostheses; participation in another treatment protocol for PCa. Treating physicians could exclude patients for comorbidities making radiotherapy inadvisable (e.g. IBD, significant urinary symptoms). All participants provided voluntary, written, informed consent. ADT was not permitted.

4.4.3 Randomisation

Between 07/08/2012 and 04/01/2018, 874 men were randomised to PACE-B (441 CFMHRT; 433 SBRT) from 37 centres across three countries (UK, Ireland & Canada).

[27] Except for those progressing on active surveillance, whose last biopsy was suitable for PACE-B, and required treatment (e.g. PSA/MRI progression) –stratified as intermediate risk.

4.4.4 Treatment Procedures

Prostatic fiducial markers (3+) were strongly recommended. Bowel preparation (enemas) was suggested, plus moderate bladder filling. The RT planning CT scan was ≥7 days after fiducial placement. An RT planning MRI scan was strongly recommended, for fusion to the CT scan, improving anatomical definition. The CTV was either prostate only (low risk), or prostate + proximal 1cm SVs (intermediate risk). Recommended CFMHRT CTV to PTV expansion was 5-9mm isometric (posteriorly 3-7mm). Recommended SBRT CTV to PTV expansion was 4-5mm isometric (posteriorly 3-5mm). The dose constraints applied to OARs were amended during the trial. A history of the constraints, with numbers of patients randomised to each iteration, is in **Appendix 1** (**p383**). Final dose constraints used from 24th March 2016 (for 604/847 patients) are shown for CFMHRT (**Table 16**) and SBRT (**Table 17**).

Table 16. Final Dose Volume Constraints for CFMHRT Patients
Adapted from the PACE trial protocol.

Organ	Fractionation Regimen		Volume Constraints	
	78Gy/39 fractions	62Gy/20 fractions	Mandatory	Optimal
	Dose (Gy)	Dose (Gy)	% of organ	% of organ
Rectum	30	24	N/A	80%
	40	32	N/A	65%
	50	40	60%	50%
	60	48	50%	35%
	65	52	30%	N/A
	70	56	25%	15%
	75	60	5%	3%
Bladder	50	40	50%	N/A
	60	48	25%	N/A
	74	59	15%	5%

Table 17. Final Dose Volume Constraints for SBRT Patients

Adapted from the PACE trial protocol.

Note that some violations were considered minor variations:

Rectum V37 = 1-2cc; Bladder V37 10-20cc

Organ	Dose	Volume Constraints	
		Mandatory	Optimal
	(Gy)	% or cc	% or cc
Rectum	18.1	50%	N/A
	29	20%	N/A
	36	1 cc	N/A
Bladder	18.1	40%	N/A
	37	10 cc	5 cc

CFMHRT PTV dose was 78 Gy in 39 daily fractions or, after protocol amendment (24th March 2016), 62 Gy in 20 daily fractions. This change followed the CHHiP trial data supporting moderate hypofractionation [1], but with a higher dose (62 Gy versus 60 Gy), because PACE-B prohibited ADT. Data from PROFIT for 60 Gy in 20 fractions, without ADT, being equivalent to 78 Gy in 39 fractions, were not available at that time [2]. Post-amendment, centres had to choose either 78 Gy in 39 fractions or 62 Gy in 20 fractions as their control treatment for all subsequent patients. The SBRT PTV dose was 36·25 Gy in 5 fractions over 1-2 weeks (i.e. daily or alternate days, at centre discretion), with an additional secondary CTV dose target of 40 Gy. The CK treatment platform was initially mandatory for all SBRT, however a protocol amendment (24th October 2014) permitted SBRT on conventional LINACs. Detailed prescription objectives, along with minor variations permitted, are in the protocol (published previously [188]).

IGRT (preferably fiducial based) was mandated. For SBRT, intra-fractional motion monitoring was permitted; otherwise a repeat image was required for SBRT fraction delivery ≥3 minutes. Radiotherapy QA occurred for each centre to ensure consistency with trial protocol[28].

Participants were assessed alternate weeks during CFMHRT and on the final fraction for SBRT and weeks 2, 4, 8 and 12 after the end of treatment. Two CROs were collected: RTOG (GI and GU domains) at baseline and every visit; CTCAE at baseline and follow-up weeks 2, 4, 8 and 12, with additional end-of-treatment assessment for SBRT patients. Subsequent follow-up is ongoing.

4.4.5 Outcomes

This acute toxicity report is a pre-specified sub-analysis of the PACE-B trial. I prospectively wrote a statistical analysis plan (SAP) (**Appendix 2, p385**). This was revised with critical input from my supervisors and statisticians at the ICR-CTSU (Clare Griffin and Vicki Hinder). The co-primary sub-study endpoints were G2+ worst RTOG toxicity score, up to week-12 follow-up post-RT, for both GI and GU systems. Separately for GI and GU, the numerator was patients with recorded RTOG G2+ toxicity at any point after baseline, up to week-12 post-RT. The denominator was all patients with at least one RTOG score completed after baseline, up to week-12 post-RT.

[28] This included: pre-trial contouring and planning benchmark cases, central review of the first CFMHRT and SBRT patient, followed by principal investigator review of the first 3 cases for each clinician contouring at that centre.

Patients were missing if no such score was returned. This endpoint was pragmatically chosen, as only RTOG assessments were conducted for CFMHRT patients during radiotherapy (recall the peak in **Figure 18**). Secondary endpoints focussed on other CROs (CTCAE) and PROs (See **Chapter 5**). For each scale, the baseline, worst, worst above baseline and week-12 (i.e. residual) toxicity were of interest; with exact definitions detailed in the SAP (**Appendix 2, p400**).

4.4.6 Statistical analysis

Patients were analysed by treatment received (termed per-protocol analysis here), with those receiving one or more fractions of CFMHRT or SBRT included. Those patients receiving both CFMHRT and SBRT fractions were excluded unless the reason was toxicity related, where analysis was on first treatment fraction received. As the main trial was powered for non-inferiority of biochemical/failure, a retrospective power calculation was performed for this secondary analysis of acute toxicity. Acute RTOG G2+ toxicity proportions of 25% (GI) and 40% (GU) were assumed for CFMHRT, per CHHiP and PROFIT [1,3]. With two-sided $\alpha=0{\cdot}025$ for each endpoint, it was estimated that with 874 patients[29] recruited there would be 83% power to exclude a 10% increase in acute GI toxicity and 84·5% power to exclude an 11% increase in acute GU toxicity, for the SBRT arm (**Appendix 2, p394**).

[29] All randomised patients in this calculation due to prospective calculation and uncertainty in number that would be excluded.

The chi-square test was used to compare treatment groups for the co-primary endpoints. For secondary endpoints (full list in SAP, **Appendix 2, p400**), chi-square tests compared proportions, except for the use of Fisher's test when Chi-square assumptions were not met. Baseline toxicity scores were compared by Mann-Whitney test. To reduce the impact of multiple comparisons, a p-value of 0·001 was considered significant for secondary comparisons. The different durations of radiotherapy (CFMHRT 4 or 7.8 weeks; SBRT 1 or 2 weeks) led to differing time points of toxicity assessment. RTOG (assessed during radiotherapy) and CTCAE (assessed at end of radiotherapy for SBRT) graphical presentation is therefore both as four groups and just CFMHRT and SBRT (interpolation method detailed in **Appendix 3, p413**). Confidence intervals for the difference in proportions were calculated using normal approximation.

Exploratory examination of CK versus standard LINAC for SBRT patients was prospectively included in the trial protocol, when an amendment permitted standard LINACs (5th August 2014). The primary endpoint measures (worst RTOG G2+ GI and GU toxicity) were re-examined for SBRT patients only, split by CK and non-CK delivery, and interpreted at a significant p-value of 0·001. Since centre-level effects may influence this non-randomised analysis (e.g. variation in toxicity reporting), similar analysis was performed for CFMHRT patients; separated by whether their centre used CK or non-CK for SBRT treatments.

Analyses are based on a snapshot of data taken on 28/05/2019 and were conducted using STATA version 15.1 (StataCorp LLC, Texas, USA). The database was provided to me by ICR-CTSU. I undertook data QA and analyses using scripts wholly written and debugged by me personally. The Independent Data Monitoring Committee (IDMC) gave approval for release of acute toxicity results prior to release of primary endpoint (efficacy) results.

4.5 Results

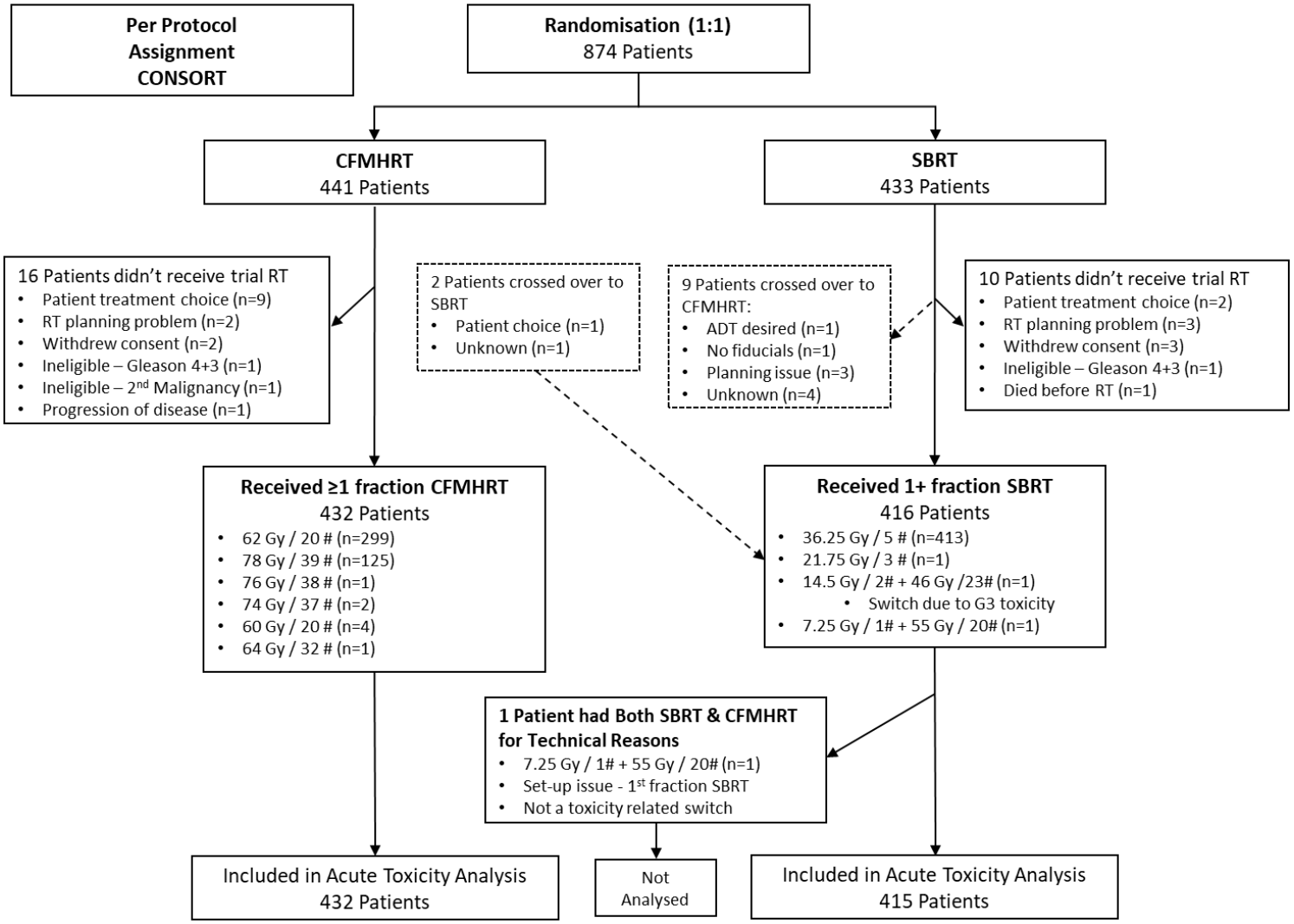
4.5.1 Patient Inclusion

The CONSORT diagram details per-protocol assignment (**Figure 19**), including ineligibility reasons and exact radiotherapy regimens delivered. Reasons for the eleven patients receiving non-protocol regimens, are completely detailed in **Appendix 4 (p415)**. Only one was toxicity related (an SBRT patient with G3 urinary toxicity). A single patient randomised to SBRT died due to myocardial infarction prior to receiving trial treatment and is excluded from this analysis; no other deaths were reported up to 12 weeks post radiotherapy. Median follow-up completed was 12 weeks (interquartile range (IQR) 12 – 12 weeks).

4.5.2 Baseline Characteristics

Baseline characteristics for each treatment group appeared balanced (**Table 18**). Four of 19 patients on 5-ARI at baseline had a PSA value of 10-20 ng/mL.

Figure 19. Per Protocol CONSORT



Patient flow through the trial, with deviations from expected treatments. Exact dose-fractionation regimens administered are shown. Two men received both SBRT and CFMHRT treatments: included is one patient who received two fractions of SBRT (14·5 Gy) then developed grade 3 toxicity (urosepsis) and switched to CFMHRT (further 46 Gy in 23 fractions). Excluded is one patient who received a single incomplete fraction of SBRT (<7·25 Gy, set-up issues) and switched to CFMHRT (further 55 Gy in 20 fractions).

Table 18. Baseline Characteristics by Per Protocol Treatment Arm

Novel abbreviations: PS = Performance Status

Baseline Characteristic	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	No.	%	No.	%	No.	%
Age (Years)						
Mean	69.5	N/A	69.3	N/A	69.4	N/A
Min	48	N/A	45	N/A	45	N/A
Max	86	N/A	84	N/A	86	N/A
Ethnic Origin						
Black	25	5.8%	35	8.4%	60	7.1%
East Asian	3	0.7%	4	1.0%	7	0.8%
Mixed Heritage	2	0.5%	2	0.5%	4	0.5%
Southern Asian	9	2.1%	19	4.6%	28	3.3%
White	386	89.4%	352	84.8%	738	87.1%
Other	7	1.6%	3	0.7%	10	1.2%
Family History of Prostate Cancer						
No	321	74.3%	300	72.3%	621	73.3%
Yes	85	19.7%	85	20.5%	170	20.1%
Unknown	26	6.0%	30	7.2%	56	6.6%
WHO Performance Status						
PS 0	382	88.4%	372	89.6%	754	89.0%
PS 1	48	11.1%	43	10.4%	91	10.7%
PS 2	2	0.5%	0	0.0%	2	0.2%
NCCN Risk Score						
Low	38	8.8%	30	7.2%	68	8.0%
Intermediate	394	91.2%	385	92.8%	779	92.0%
T-Stage						
T1c	78	18.1%	76	18.3%	154	18.2%
T2a	130	30.1%	105	25.3%	235	27.7%
T2b	57	13.2%	81	19.5%	138	16.3%
T2c	167	38.7%	153	36.9%	320	37.8%
Gleason Grade						
3+3	84	19.4%	61	14.7%	145	17.1%
3+4	348	80.6%	354	85.3%	702	82.9%
Pre-treatment PSA						
Mean	8.7	N/A	8.6	N/A	8.7	N/A
Median	8.0	N/A	8.0	N/A	8.0	N/A
Range	0.8 - 20	N/A	0.5 - 20	N/A	0.5 - 20	N/A
PSA Categories						
<10 ng/mL	299	69.2%	283	68.2%	582	68.7%
10 - 20 ng/mL	133	30.8%	132	31.8%	265	31.3%

Continued overleaf…

…Continuation of Table 18						
Pre-treatment Testosterone						
<1.7 nmol/L	0	0.0%	2	0.5%	2	0.2%
1.7+ nmol/L	391	90.5%	376	90.6%	767	90.6%
Unknown	41	9.5%	37	8.9%	78	9.2%
Active Surveillance Before Trial Enrolment						
Yes	160	37.0%	146	35.2%	306	36.1%
No	258	59.7%	256	61.7%	514	60.7%
Unknown	14	3.2%	13	3.1%	27	3.2%
Prostate Volume						
<40 mL	153	35.4%	160	38.6%	313	37.0%
40 - <80 mL	200	46.3%	170	41.0%	370	43.7%
80+ mL	16	3.7%	21	5.1%	37	4.4%
Unknown	63	14.6%	64	15.4%	127	15.0%
Alpha Blockers at Randomisation						
Yes	68	15.7%	67	16.1%	135	15.9%
No	361	83.6%	344	82.9%	705	83.2%
Unknown	3	0.7%	4	1.0%	7	0.8%
Aspirin at Randomisation						
Yes	74	17.1%	63	15.2%	137	16.2%
No	355	82.2%	347	83.6%	702	82.9%
Unknown	3	0.7%	5	1.2%	8	0.9%
Statin at Randomisation						
Yes	153	35.4%	126	30.4%	279	32.9%
No	275	63.7%	283	68.2%	558	65.9%
Unknown	4	0.9%	6	1.4%	10	1.2%
Anticholinergic for Bladder Symptoms at Randomisation						
Yes	16	3.7%	10	2.4%	26	3.1%
No	414	95.8%	400	96.4%	814	96.1%
Unknown	2	0.5%	5	1.2%	7	0.8%
5-Alpha Reductase Inhibitor at Randomisation						
Yes	9	2.1%	10	2.4%	19	2.2%
No	416	96.3%	387	93.3%	803	94.8%
Unknown	7	1.6%	18	4.3%	25	3.0%
Phosphodiesterase-5 Inhibitor at Randomisation						
Yes	12	2.8%	6	1.4%	18	2.1%
No	412	95.4%	392	94.5%	804	94.9%
Unknown	8	1.9%	17	4.1%	25	3.0%
Total	**432**	**100.0%**	**415**	**100.0%**	**847**	**100.0%**

4.5.3 Radiotherapy Delivery

Radiotherapy delivery techniques (planning, IGRT, margins) expectedly differed between arms (**Table 19**). Recorded supportive prescribing appears similar (**Table 20**). Despite fiducial recommendation for both arms, more SBRT patients received fiducial markers (303/415, 73·0%) than CFMHRT (245/427, 56·7%). Reported CTV to PTV margins differed from the protocol recommendations, with greater non-compliance amongst SBRT patients. Most non-compliance was towards the use of smaller than protocol recommended margins. CFMHRT, non-posterior margins: <5 mm (n=5, 1·2%); protocol 5-9 mm (n=406, 94·0%); 10 mm (n=5, 1·2%); Unknown (n=16, 3·7%). CFMHRT posterior margins: <3 mm (n=9, 2·1%); protocol 3-7 mm (n=407, 94·2%); Unknown (n=16, 3·7%). SBRT non-posterior margins: <4 mm (n=40, 9·6%); protocol 4-5 mm (n=366, 88·2%); >5 mm (n=6, 1·4%); unknown (n=3, 0·7%). SBRT posterior margins: <3 mm (n=53, 12·8%); 3-5 mm (n=357, 86·0%); >5 mm (n=2, 0·5%); Unknown (n=3, 0·7%).

Table 19. Treatment Delivery Techniques

Treatment Delivery Technique	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
Fiducial Markers Inserted?						
No	187	43.3%	112	27.0%	299	35.3%
Yes	245	56.7%	303	73.0%	548	64.7%
Number of Fiducial Markers						
0	187	43.3%	112	27.0%	299	35.3%
2	4	0.9%	11	2.7%	15	1.8%
3	184	42.6%	94	22.7%	278	32.8%
4	51	11.8%	189	45.5%	240	28.3%
5+	1	0.2%	9	2.2%	10	1.2%
Unknown	5	1.2%	0	0.0%	5	0.6%
Radiotherapy Delivery Method						
Step and Shoot IMRT	106	24.5%	3	0.7%	109	12.9%
VMAT	322	74.5%	242	58.3%	564	66.6%
Tomotherapy	4	0.9%	0	0.0%	4	0.5%
CyberKnife	0	0.0%	170	41.0%	170	20.1%
IGRT Method						
Planar Film - With Fiducials	94	21.8%	10	2.4%	104	12.3%
Planar Intra-fractional Tracking	2	0.5%	170	41.0%	172	20.3%
CBCT - No Fiducials	185	42.8%	112	27.0%	297	35.1%
CBCT - With Fiducials	124	28.7%	116	28.0%	240	28.3%
CBCT & Planar Film – No Fiducials	1	0.2%	0	0.0%	1	0.1%
CBCT & Planar Film – With Fiducials	23	5.3%	4	1.0%	27	3.2%
CBCT & Planar Intra-Fractional Tracking	3	0.7%	3	0.7%	6	0.7%
Overall Treatment Time						
1 week	0	0.0%	86	20.7%	86	10.2%
2 weeks	0	0.0%	305	73.5%	305	36.0%
3 weeks	0	0.0%	18	4.3%	18	2.1%
4 weeks	136	31.5%	2	0.5%	138	16.3%
5 weeks	162	37.5%	4	1.0%	166	19.6%
6 weeks	4	0.9%	0	0.0%	4	0.5%
7 weeks	2	0.5%	0	0.0%	2	0.2%
8 weeks	61	14.1%	0	0.0%	61	7.2%
9-10 weeks	67	15.5%	0	0.0%	67	7.9%
Totals	**432**	**100%**	**415**	**100%**	**847**	**100%**

Table 20. Supportive Medications Prescribed During Acute Toxicity Window

Supportive Medications Prescribed in Acute Toxicity Window	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
Alpha Blockers						
No	308	71.3%	299	72.0%	607	71.7%
Yes	50	11.6%	39	9.4%	89	10.5%
On at Randomisation	68	15.7%	67	16.1%	135	15.9%
Unknown	6	1.4%	10	2.4%	16	1.9%
Anticholinergics						
No	406	94.0%	392	94.5%	798	94.2%
Yes	7	1.6%	8	1.9%	15	1.8%
On at Randomisation	16	3.7%	10	2.4%	26	3.1%
Unknown	3	0.7%	5	1.2%	8	0.9%
Phosphodiesterase-5 Inhibitors						
No	402	93.1%	383	92.3%	785	92.7%
Yes	10	2.3%	8	1.9%	18	2.1%
On at Randomisation	12	2.8%	6	1.4%	18	2.1%
Unknown	8	1.9%	18	4.3%	26	3.1%
Totals	**432**	**100%**	**415**	**100%**	**847**	**100%**

4.5.4 Return Rates

RTOG and CTCAE form completion were >90% at all timepoints assessed (**Appendix 5, p416**). Of the non-completed forms, patient illness was the specified reason for 3 RTOG forms (2 CFMHRT, 1 SBRT) and 1 SBRT CTCAE form.

4.5.5 RTOG Acute Toxicity

Worst RTOG acute toxicity adverse events, by GI and GU organ systems, are shown in **Table 21**. Regarding the co-primary endpoints of interest for this sub-study**:** worst acute RTOG G2+ GI toxicity, compared between CFMHRT (53/432, 12·3%) and SBRT (43/415, 10·4%) did not differ significantly (difference −1.9%, 95% CI −6.2% to 2.4%, p=0·38). Worst acute RTOG G2+ GU toxicity, compared between CFMHRT (118/432, 27·3%) and SBRT (96/415, 23·1%) also did not differ significantly (difference −4.2%, 95% CI −10 to 1.7%, p=0·16).

Table 21. Worst Acute RTOG Toxicity by Organ System

RTOG Toxicity Grade by Organ	Per Protocol Treatment			
	CFMHRT		SBRT	
	n	Percentage	n	Percentage
Gastrointestinal				
Grade 0	115	26.6%	153	36.9%
Grade 1	264	61.1%	219	52.8%
Grade 2	49	11.3%	42	10.1%
Grade 3	4	0.9%	1	0.2%
Grade 4	0	0.0%	0	0.0%
Grade 5 (Death)	0	0.0%	0	0.0%
Genitourinary				
Grade 0	60	13.9%	83	20.0%
Grade 1	254	58.8%	236	56.9%
Grade 2	111	25.7%	86	20.7%
Grade 3	6	1.4%	8	1.9%
Grade 4	1	0.2%	2	0.5%
Grade 5 (Death)	0	0.0%	0	0.0%
Total	**432**	**100%**	**415**	**100%**

For RTOG secondary endpoints, baseline GI scores were similar and no significant differences were seen comparing CFMHRT and SBRT by any comparison for GI toxicity (**Table 22**), including worst RTOG GI G3+ toxicity (4/432, 0·9% vs 1/415, 0·2%, difference −0.7%, 95% CI −1.7 to 0.3%, p=0·37). For RTOG GU, baseline scores were similar and again no significant differences were seen by any comparison (**Table 23**), including worst RTOG GU G3+ toxicity (7/432, 1·6% vs 10/415, 2·4%, difference 0.8%, 95% CI −1.1 to 2.7%, p=0·47).

Table 22. Acute RTOG GI Toxicity: Baseline, Worst & Worst Above Baseline

RTOG Gastrointestinal (GI) Toxicity	Treatment						Statistical Comparisons
	CFMHRT			SBRT			
	No.	%	Grade X+ (%)	No.	%	Grade X+ (%)	
Baseline GI Grade							
Grade 0	377	93.8%	100%	365	93.6%	100%	p=0·90 Mann-Whitney comparing grade frequencies
Grade 1	23	5.7%	6.2%	22	5.6%	6.4%	
Grade 2	2	0.5%	0.5%	3	0.8%	0.8%	
Grade 3	0	0.0%	0.0%	0	0.0%	0.0%	
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing	30	N/A	N/A	25	N/A	N/A	
Worst GI Grade							**Comparisons of Grade X+ %**
Grade 0	115	26.6%	100%	153	36.9%	100%	
Grade 1	264	61.1%	73.4%	219	52.8%	63.1%	
Grade 2	49	11.3%	12.3%	42	10.1%	10.4%	−1.9% difference 95% CI −6.2 to 2.4% p=0.38 (Chi-square)
Grade 3	4	0.9%	0.9%	1	0.2%	0.2%	−0.7% difference 95% CI −1.7 to 0.3% p=0.37 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Worst GI Grade, Exceeding Baseline Grade							
No Baseline Data	30	N/A	N/A	25	N/A	N/A	
Baseline Not Exceeded	123	30.6%	100%	160	41.0%	100%	
Grade 1	230	57.2%	69.4%	195	50.0%	59.0%	
Grade 2	45	11.2%	12.2%	34	8.7%	9.0%	−3.2% difference 95% CI −7.5 to 1.1% p=0.14 (Chi-square)
Grade 3	4	1.0%	1.0%	1	0.3%	0.3%	−0.7% difference 95% CI −1.8 to 0.4% p=0.37 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Week 12 Post-RT GI Grade							
Grade 0	350	83.7%	100%	330	82.1%	100%	
Grade 1	66	15.8%	16.3%	65	16.2%	17.9%	
Grade 2	2	0.5%	0.5%	7	1.7%	1.7%	1.3% difference 95% CI −0.2 to 2.7% p=0.10 (Fishers')
Grade 3	0	0.0%	0.0%	0	0.0%	0.0%	N/A
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing	14	N/A	N/A	13	N/A	N/A	

Table 23. Acute RTOG GU Toxicity: Baseline, Worst & Worst Above Baseline

RTOG GU Toxicity	Per Protocol Treatment						Statistical Comparisons
	CFMHRT			SBRT			
	No.	%	Grade X+ %	No.	%	Grade X+ %	
Baseline GU Grade							
Grade 0	318	79.1%	100%	295	75.6%	100%	p=0·24 Mann-Whitney comparing grade frequencies
Grade 1	74	18.4%	20.9%	83	21.3%	24.4%	
Grade 2	10	2.5%	2.5%	10	2.6%	3.1%	
Grade 3	0	0.0%	0.0%	2	0.5%	0.5%	
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing	30	N/A	N/A	25	N/A	N/A	
Worst GU Grade							**Comparisons of Grade X+ %**
Grade 0	60	13.9%	100%	83	20.0%	100%	
Grade 1	254	58.8%	86.1%	236	56.9%	80.0%	
Grade 2	111	25.7%	27.3%	86	20.7%	23.1%	−4.2% difference 95% CI −10.0 to 1.7% p=0·16 (Chi-square)
Grade 3	6	1.4%	1.6%	8	1.9%	2.4%	0.8% difference 95% CI −1.1 to 2.7% p=0·47 (Fisher's)
Grade 4	1	0.2%	0.2%	2	0.5%	0.5%	
Worst GU Grade, Exceeding Baseline Grade							
No Baseline Data	30	N/A	N/A	25	N/A	N/A	
Baseline Not Exceeded	116	28.9%	100%	149	38.2%	100%	
Grade 1	186	46.3%	71.1%	162	41.5%	61.8%	
Grade 2	93	23.1%	24.9%	69	17.7%	20.3%	−4.6% difference 95% CI −10.4 to 1.2% p=0.12 (Chi-square)
Grade 3	6	1.5%	1.7%	8	2.1%	2.6%	0.8% difference 95% CI −1.2 to 2.8% p=0·47 (Fisher's)
Grade 4	1	0.2%	0.2%	2	0.5%	0.5%	
Week 12 Post-RT GU Grade							
Grade 0	291	69.6%	100%	278	69.2%	100%	
Grade 1	112	26.8%	30.4%	104	25.9%	30.8%	
Grade 2	14	3.3%	3.6%	20	5.0%	5.0%	1.4% difference 95% CI −1.4 to 4.2% p=0·33 (Chi-square)
Grade 3	1	0.2%	0.2%	0	0.0%	0.0%	p=1.0 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing	14	N/A	N/A	13	N/A	N/A	

RTOG acute toxicity is shown over time for GI (**Figure 20**) and GU toxicity (**Figure 21**). In both cases, it can be seen that SBRT has an earlier peak in toxicity, although the maximum proportions of patients experiencing each grade appear similar across all grades for GI and GU side effects.

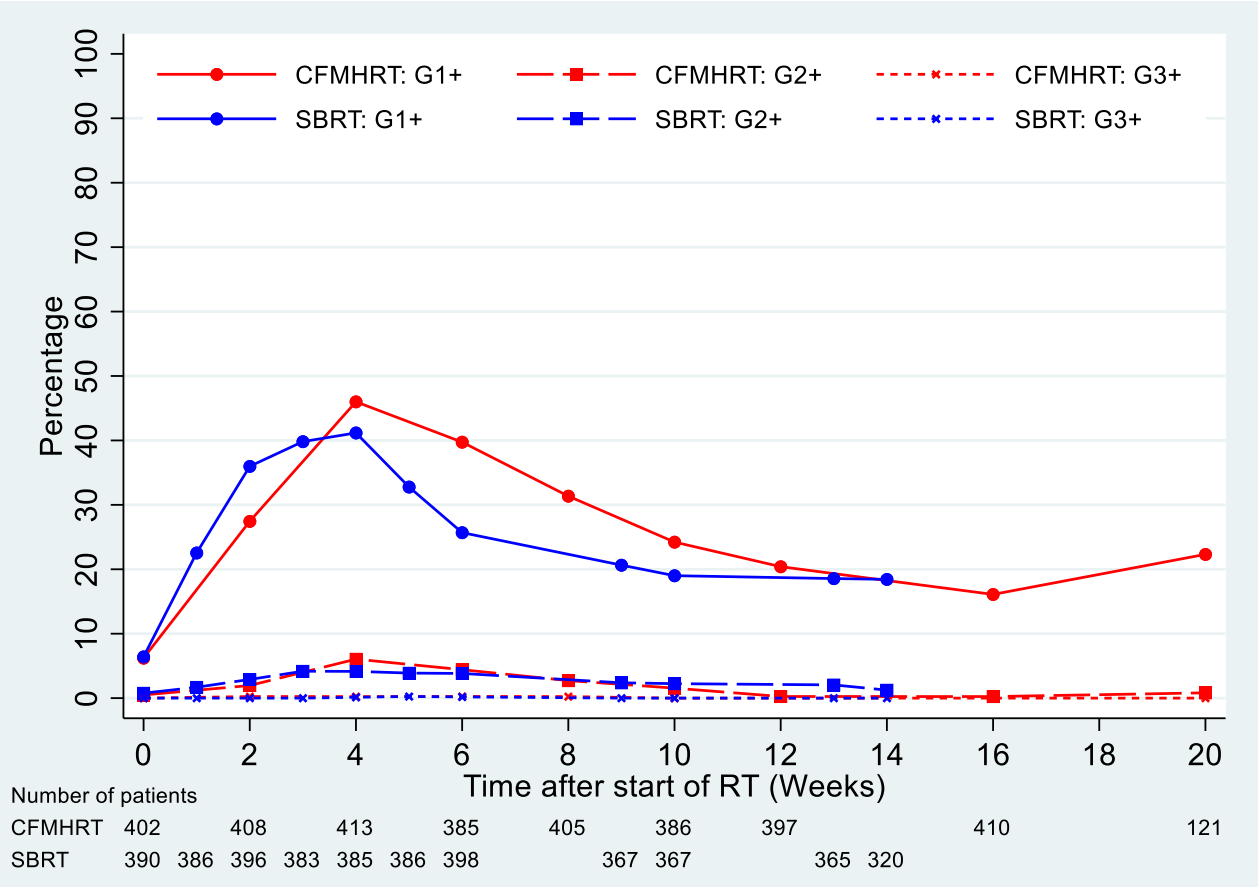


Figure 20. RTOG Acute Gastrointestinal Toxicity

Acute RTOG gastrointestinal Toxicity, by radiotherapy received. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of radiotherapy). Week 0 is the baseline toxicity score taken before start of radiotherapy.

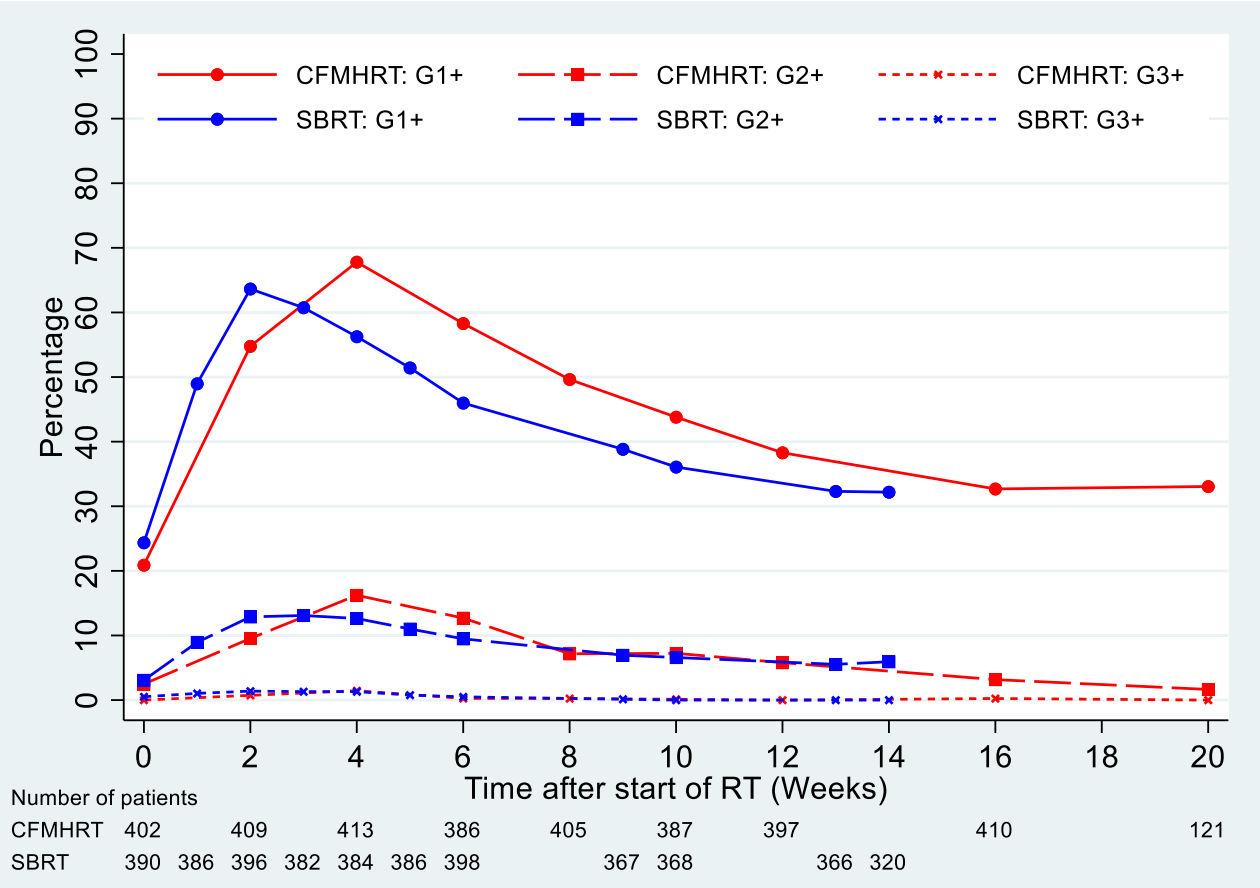


Figure 21. RTOG Acute Genitourinary Toxicity

Acute RTOG GU toxicity, by radiotherapy received. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of radiotherapy). Week 0 is the baseline toxicity score taken before start of radiotherapy.

Graphical representation of the four different durations of treatment separately (SBRT 1 week, SBRT 2 weeks, CFMHRT 4 weeks and CFMHRT 7·8 weeks) is shown in **Figure 22**. From 7.8 weeks down to 1 week, radiotherapy regimens delivered over shorter times appear to result in earlier peaking acute toxicity.

[Caption for figure overleaf]

Figure 22. Acute RTOG Toxicity By Treatment Duration

Acute RTOG toxicity, separated into four different overall treatment times permitted. For ease of display, SBRT patients receiving their treatment over more than the maximum recommended 2 weeks (n=24) are in the same line as the 2-week SBRT patients. Week 0 is the baseline toxicity score pre-RT. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of radiotherapy.
Novel Abbreviations: CFRT = Conventionally Fractionated RT

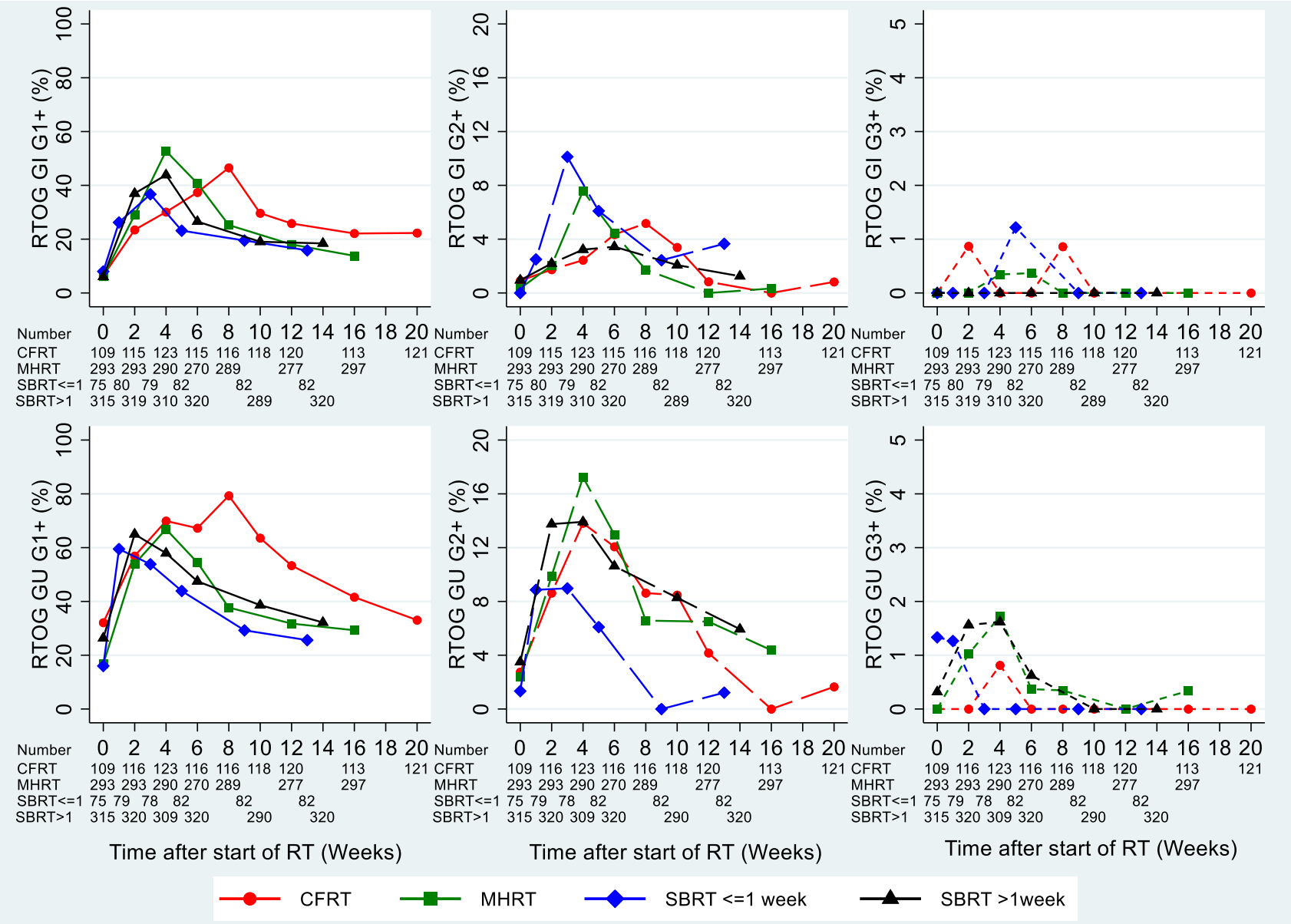


Figure 22: *Caption on preceding page*

4.5.6 CTCAE Acute Toxicity

CTCAE acute toxicity over time is demonstrated for composite GI (**Figure 23**) and GU (**Figure 24**) toxicity. Graphical representation of the four different durations of treatment separately (SBRT 1 week, SBRT 2 weeks, CFMHRT 4 weeks and CFMHRT 7·8 weeks) is shown in **Figure 25.** Similar to the results for RTOG, the shorter treatment duration regimens peak earlier, but other differences between CFMHRT and SBRT are not visually apparent.

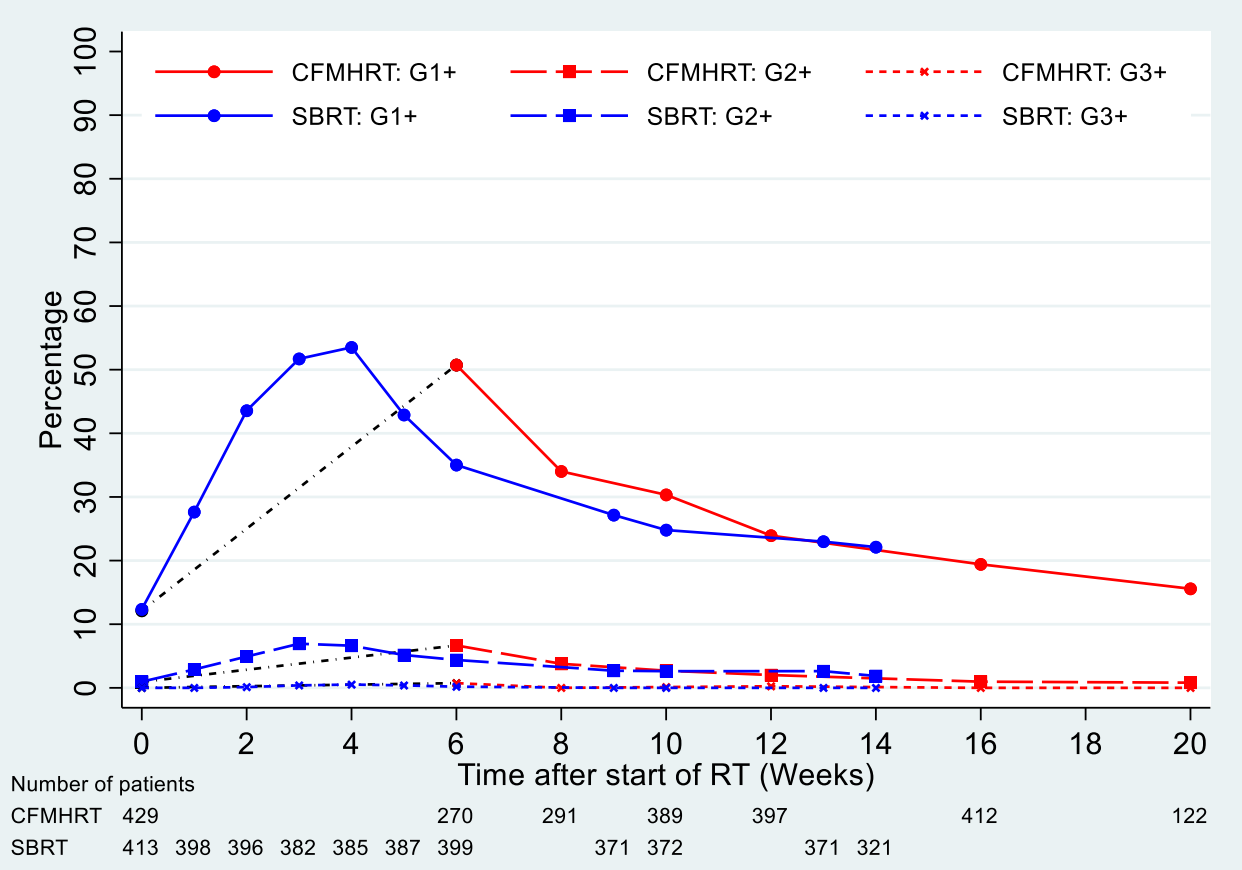


Figure 23. CTCAE Gastrointestinal Acute Toxicity

Acute composite CTCAE gastrointestinal toxicity, by treatment arm. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of radiotherapy). The initial points for CFMHRT are connected by grey dash-dot lines to emphasise that there were no CTCAE assessments during radiotherapy delivery. Week 0 is the baseline toxicity score taken before start of radiotherapy.

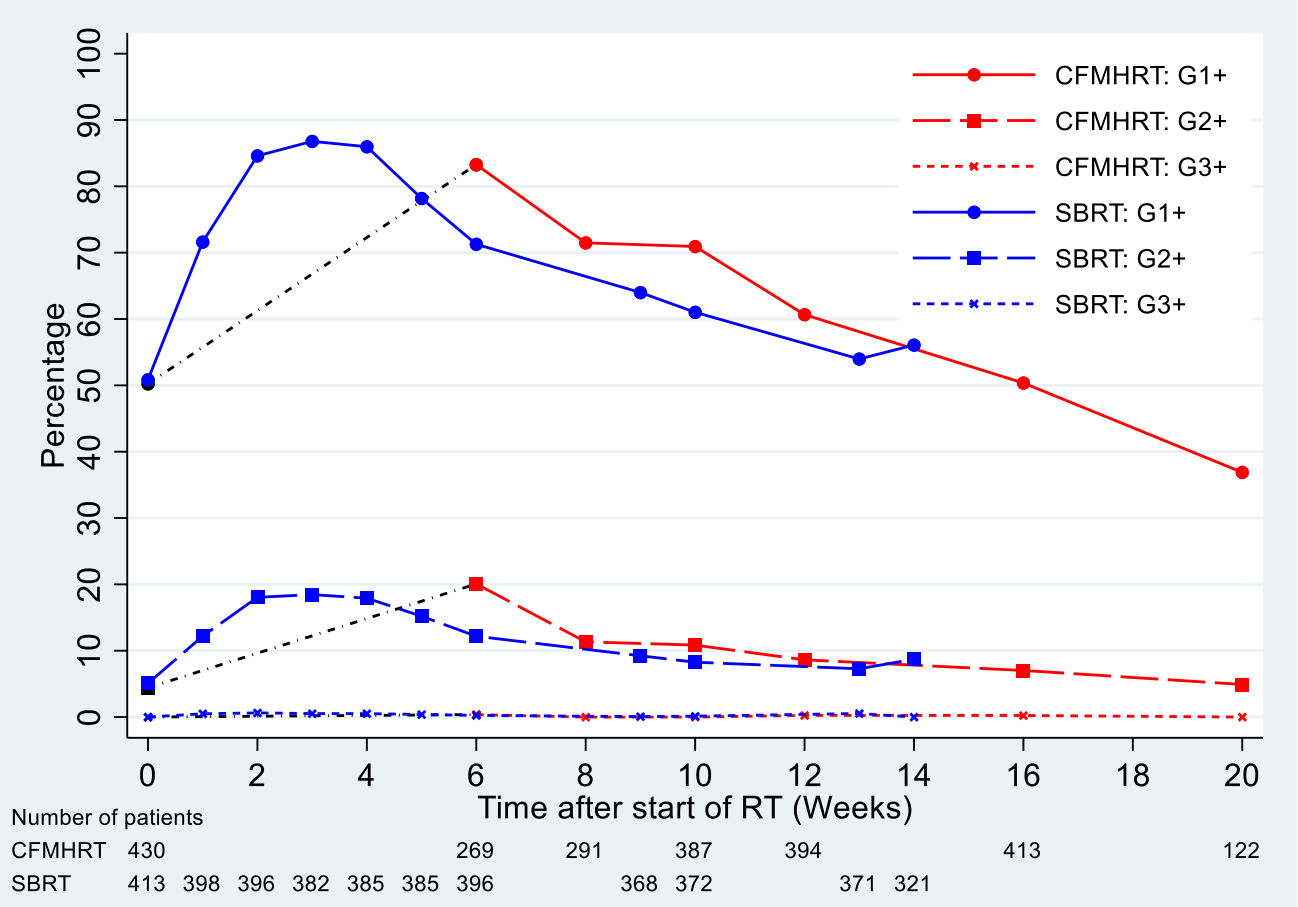


Figure 24. CTCAE Genitourinary Acute Toxicity

Acute composite CTCAE genitourinary toxicity, by treatment arm. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of radiotherapy). The initial points for CFMHRT are connected by grey dash-dot lines to emphasise that there were no CTCAE assessments during radiotherapy delivery. Week 0 is the baseline toxicity score taken before start of radiotherapy.

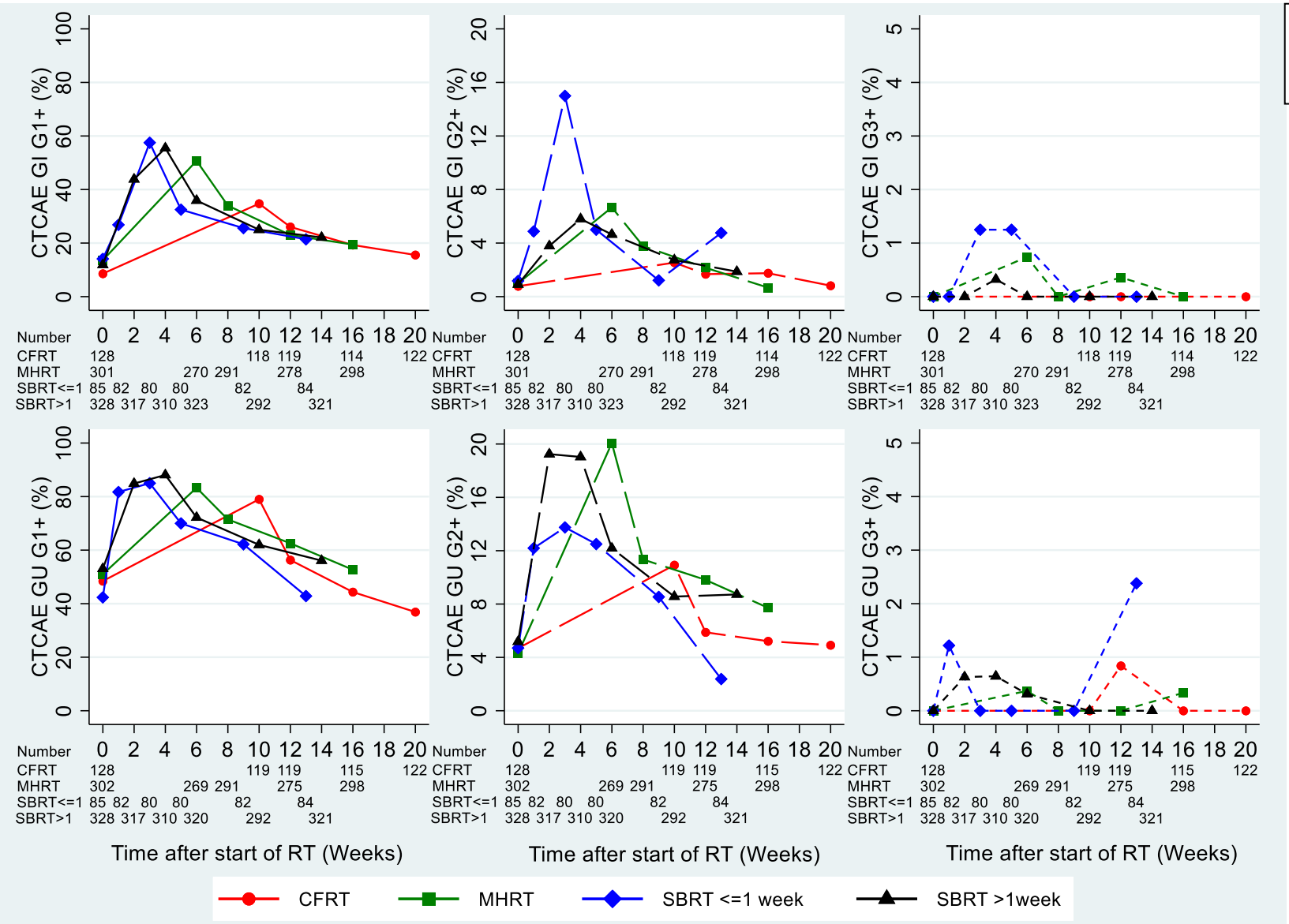


Figure 22: *Caption on next page*

[Caption for preceding page figure]

Figure 25. Acute CTCAE Toxicity By Treatment Duration

Acute CTCAE toxicity, separated into four different overall treatment times permitted. For ease of display, SBRT patients receiving their treatment over more than the maximum recommended 2 weeks (n=24) are displayed in the same line as the 2-week SBRT patients. Week 0 is the baseline toxicity score taken before start of radiotherapy. Numbers at risk for each arm are asynchronous because they are shown only at data collection timepoints (which are non-simultaneous relative to the start of RT).

Data for composite CTCAE GI toxicity, at baseline, worst, worst (exceeding baseline), and week-12 post-RT are summarised in **Table 24**, along with results of hypothesis testing performed. SBRT was statistically significantly worse for two of the CTCAE secondary endpoints analysed: CTCAE worst G2+ GI toxicity (36/430, 8·4% vs 65/415, 15·7%; difference 7.3%, 95% CI 2.9 to 11.7%, **p=0·0011**), corroborated by CTCAE worst G2+ GI toxicity exceeding baseline (34/427, 8·0% vs 63/413, 15·3%; difference 7.3%, 95% CI 3.0 to 11.6%, **p=0·00095**). It can be seen from **Figure 25** that this difference is likely driven by the 1-week SBRT regimen. Regarding most contributory individual endpoints, diarrhoea G2 and worst proctitis G2 occurred more frequently in the SBRT arm. There was no significant difference in worst CTCAE G2+ GI toxicity by week-12. No other significant differences in CTCAE GI secondary endpoints were seen comparing CFMHRT and SBRT, including worst CTCAE GI G3+ toxicity (3/430, 0·7% vs 3/415, 0·7%).

Table 24. CTCAE GI Composite Toxicity

CTCAE GI Composite Toxicity	Per Protocol Treatment						Statistical Comparisons
	CFMHRT			SBRT			
	No.	%	Grade X+ %	No.	%	Grade X+ %	
Baseline							
Grade 0	377	87.9%	100%	362	87.7%	100%	p=0·92 Mann-Whitney comparing grade frequencies
Grade 1	48	11.2%	12.1%	47	11.4%	12.3%	
Grade 2	4	0.9%	0.9%	4	1.0%	1.0%	
Grade 3	0	0.0%	0.0%	0	0.0%	0.0%	
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	3	N/A	N/A	2	N/A	N/A	
Worst							**Comparisons of Grade X+ %**
Grade 0	181	42.1%	100%	109	26.3%	100%	
Grade 1	213	49.5%	57.9%	241	58.1%	73.7%	
Grade 2	33	7.7%	8.4%	62	14.9%	15.7%	Difference 7.3% 95% CI 2.9 to 11.7% **p=0.0011 (Chi-square)**
Grade 3	3	0.7%	0.7%	3	0.7%	0.7%	Difference 0.03% 95% CI −1.1 to 1.2% p=1.0 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	2	N/A	N/A	0	N/A	N/A	
Worst, Exceeding Baseline							
Baseline Not Exceeded	212	49.6%	100%	145	35.1%	100%	
Grade 1	181	42.4%	50.4%	205	49.6%	64.9%	
Grade 2	31	7.3%	8.0%	60	14.5%	15.3%	Difference 7.3% 95% CI 3.0 to 11.6% **p=0.00095 (Chi-square)**
Grade 3	3	0.7%	0.7%	3	0.7%	0.7%	Difference 0.02% 95% CI −1.1 to 1.2% p=1.0 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing Data	5	N/A	N/A	2	N/A	N/A	
Week 12 Post-RT							
Grade 0	343	81.7%	100%	316	78.0%	100%	
Grade 1	74	17.6%	18.3%	79	19.5%	22.0%	
Grade 2	3	0.7%	0.7%	10	2.5%	2.5%	Difference 1.8% 95% CI 0.04 to 3.4% p=0.052 (Fisher's)
Grade 3	0	0.0%	0.0%	0	0.0%	0.0%	N/A
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	12	N/A	N/A	10	N/A	N/A	

Data for composite CTCAE GU toxicity, at baseline, worst, worst (exceeding baseline), and week 12 post RT are summarised in **Table 25**, along with results of hypothesis testing performed. No significant differences in CTCAE GU secondary endpoints were seen comparing CFMHRT and SBRT, including worst CTCAE GU G3+ toxicity (3/430, 0·7% vs 7/415, 1·7%).

Table 25. CTCAE GU Composite Toxicity

CTCAE GU Composite Toxicity	Per Protocol Treatment						Statistical Comparisons
	CFMHRT			SBRT			
	No.	%	Grade X+ %	No.	%	Grade X+ %	
Baseline							
Grade 0	214	49.8%	100%	203	49.2%	100%	p=0.79 Mann-Whitney comparing grade frequencies
Grade 1	197	45.8%	50.2%	189	45.8%	50.8%	
Grade 2	19	4.4%	4.4%	21	5.1%	5.1%	
Grade 3	0	0.0%	0.0%	0	0.0%	0.0%	
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	2	N/A	N/A	2	N/A	N/A	
Worst							**Comparisons of Grade X+ %**
Grade 0	48	11.2%	100%	15	3.6%	100%	
Grade 1	283	65.8%	88.8%	272	65.5%	96.4%	
Grade 2	96	22.3%	23.0%	121	29.2%	30.8%	Difference 7.8% 95% CI 1.9 to 13.8 p=0.010 (Chi-square)
Grade 3	3	0.7%	0.7%	7	1.7%	1.7%	Difference 1.0% 95% CI −0.5 to 2.5% p=0.22 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	2	N/A	N/A	0	N/A	N/A	
Worst, Exceeding Baseline							
Baseline Not Exceeded	198	46.3%	100%	164	39.7%	100%	
Grade 1	142	33.2%	53.7%	136	32.9%	60.3%	
Grade 2	85	19.9%	20.6%	106	25.7%	27.4%	Difference 6.8% 95% CI 1.0 to 12.6% p=0.021 (Chi-square)
Grade 3	3	0.7%	0.7%	7	1.7%	1.7%	Difference 1.0% 95% CI −0.5 to 2.5% p=0.22 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Missing Data	4	N/A	N/A	2	N/A	N/A	
Week 12 Post-RT							
Grade 0	218	51.9%	100%	189	46.7%	100%	
Grade 1	173	41.2%	48.1%	186	45.9%	53.3%	
Grade 2	28	6.7%	6.9%	28	6.9%	7.4%	Difference 0.5% 95% CI −3.0 to 4.0 p=0.78 (Chi-square)
Grade 3	1	0.2%	0.2%	2	0.5%	0.5%	Difference 0.3% 95% CI −0.6 to 1.1% p=0.62 (Fisher's)
Grade 4	0	0.0%	0.0%	0	0.0%	0.0%	
Not Graded	12	N/A	N/A	10	N/A	N/A	

4.5.7 Treatment Platform Comparison

4.5.7.1 SBRT Patients: Cyberknife vs Non-Cyberknife

For SBRT patients, RTOG GI G2+ worst (without reference to baseline) toxicity for non-CK (27/245, 11·0%) vs CK (16/170, 9·4%) delivery was not statistically different (difference −1.6%, 95% CI −7.5 to 4.3%, p=0.597); consistent with appearances over time in **Figure 26**.

Also for SBRT patients, RTOG G2+ worst GU (without reference to baseline) was statistically significantly worse for non-CK (75/245, 30·6%) than CK (21/170, 12·4%) delivery (difference −18.3%, 95% CI −10.7 to −25.9%, **p<0·001**); consistent with appearances over time in **Figure 27**.

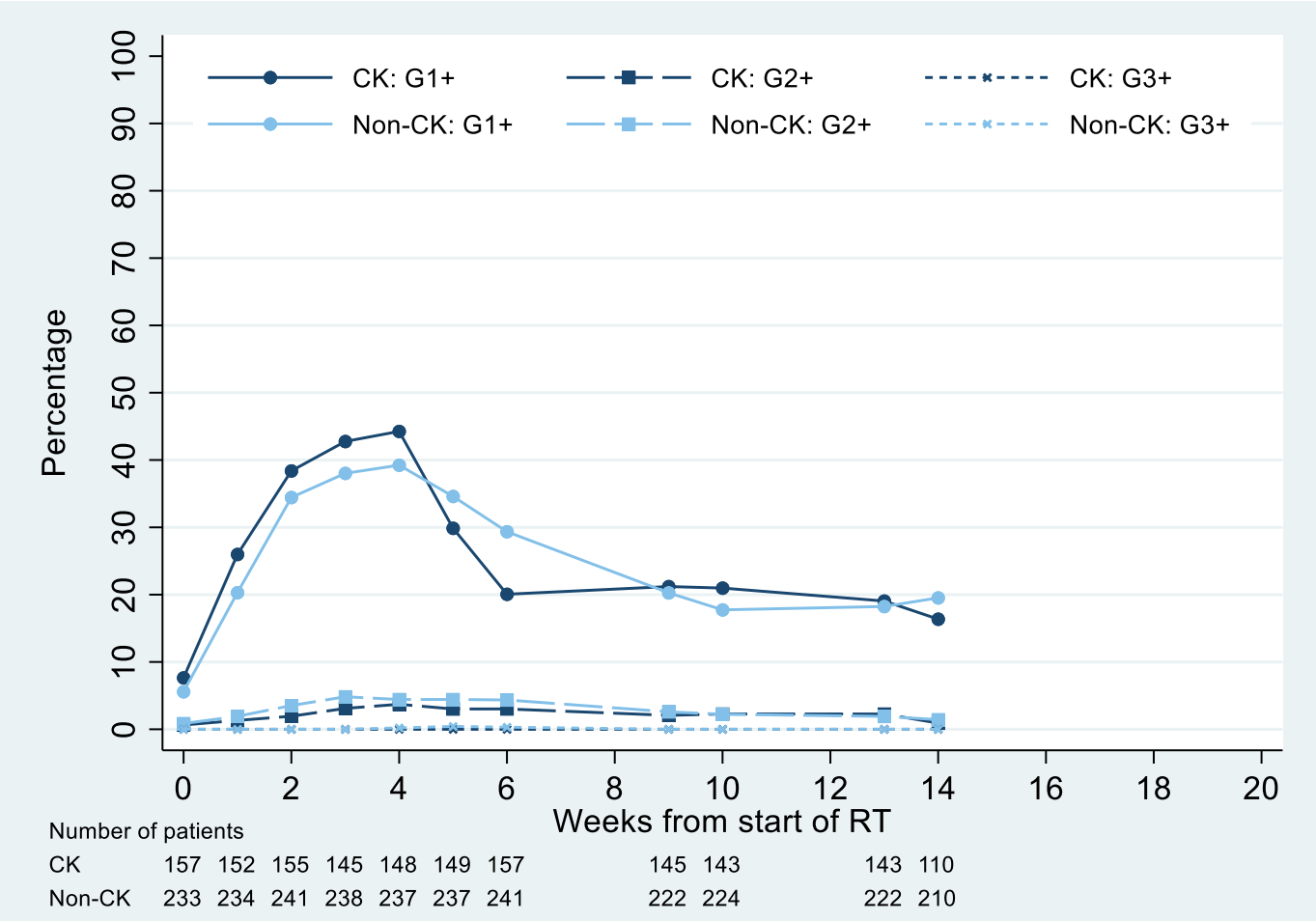


Figure 26. RTOG Gastrointestinal Acute Toxicity (SBRT Only) By Platform

RTOG acute gastrointestinal toxicity presented only for patients receiving SBRT, separated into those receiving CyberKnife and those receiving non-CyberKnife RT. X-axis scale matched to other RTOG graphs to facilitate comparison. Week 0 is the baseline toxicity score taken before start of radiotherapy.

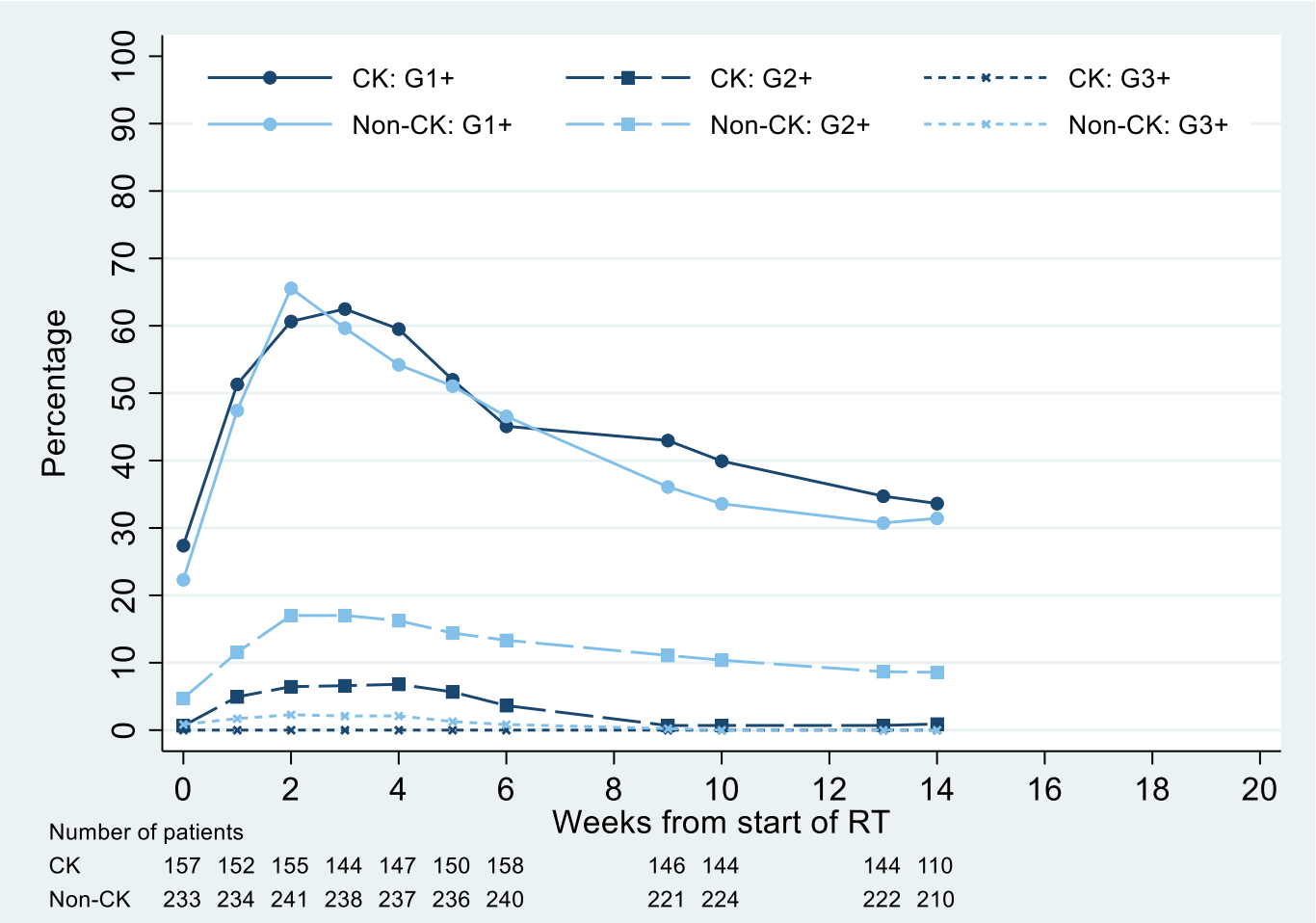


Figure 27. RTOG Genitourinary Acute Toxicity (SBRT Only) By Platform

RTOG acute genitourinary toxicity presented only for patients receiving SBRT, separated into those receiving CyberKnife and those receiving non-CyberKnife RT. X-axis scale matched to other RTOG graphs to facilitate comparison. Week 0 is the baseline toxicity score taken before start of radiotherapy.

4.5.7.2 CFMHRT Patients: CK-Centres vs Non-CK-Centres

Given the non-randomised nature of comparing non-CK vs CK, the CFMHRT toxicity in non-CK-centres vs CK-centres was examined. For CFMHRT patients, RTOG G2+ worst GI (without reference to baseline) in non-CK-using-centres (25/252, 9·9%) vs CK-centres (28/180, 15·6%) was not statistically different (difference 5.6%, 95% CI −0.8 to 12.1%, p=0·078); consistent with G2 & G3 appearances over time in **Figure 28**.

For CFMHRT patients, RTOG G2+ worst GU (without reference to baseline) in non-CK-using-centres (73/252, 29·0%) vs CK-using-centres (45/180, 25·0%) was also not statistically different (difference −4.0%, 95% CI −12.4 to 4.5%, p=0·361), (**Figure 29**).

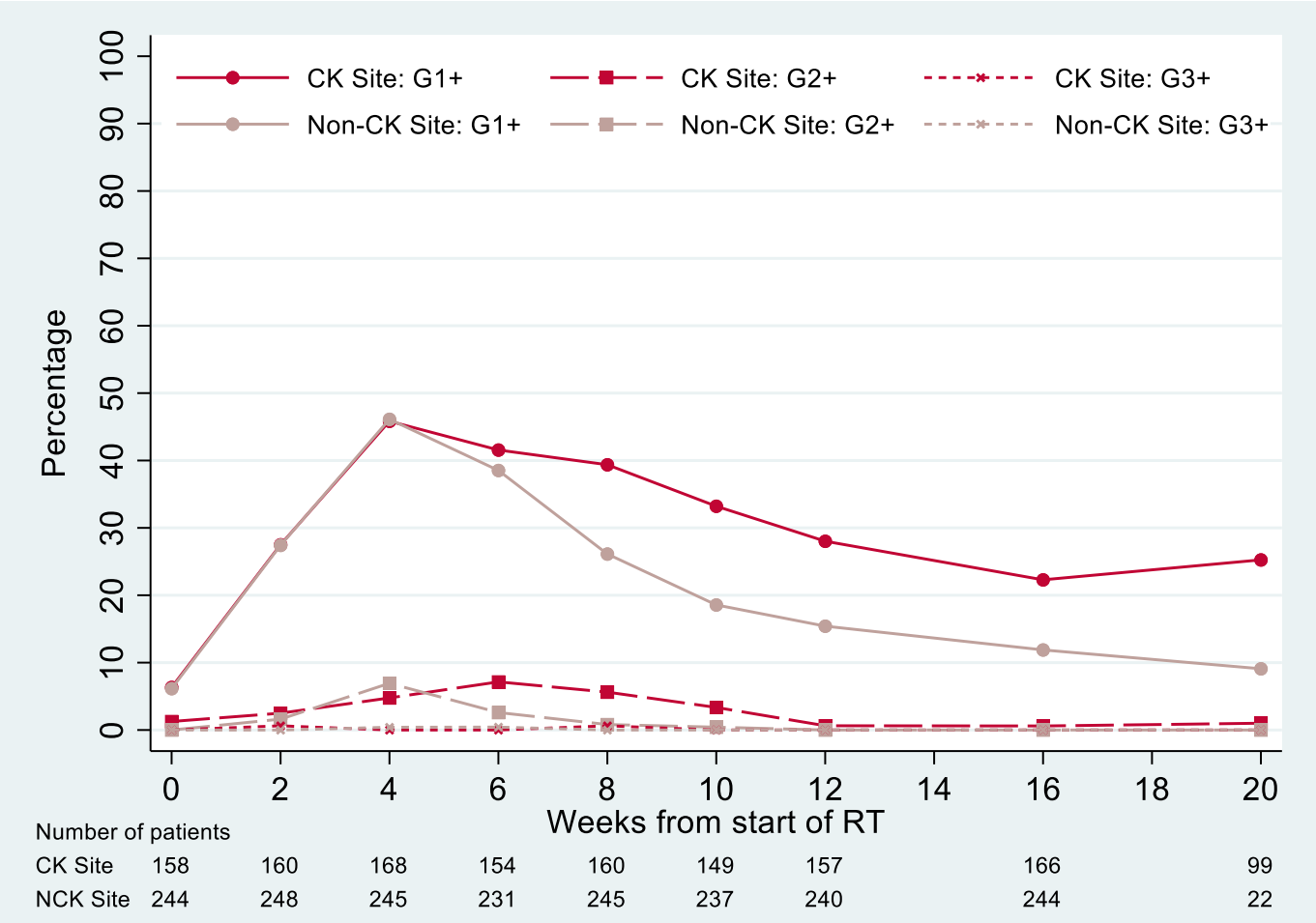


Figure 28. RTOG Gastrointestinal Acute Toxicity (For CFMHRT Patients Only) By Treating Centre CyberKnife Status

RTOG acute gastrointestinal toxicity presented only for patients receiving CFMHRT, separated into those receiving radiotherapy at a centre which performed their SBRT treatments on CyberKnife versus non-CyberKnife platforms. Week 0 is the baseline toxicity score taken before start of RT.

Figure 29. RTOG Genitourinary Acute Toxicity (CFMHRT Only) By Treating Centre CyberKnife Status

RTOG acute genitourinary toxicity presented only for patients receiving CFMHRT, separated into those receiving radiotherapy at a centre which performed their SBRT treatments on CyberKnife versus non-CyberKnife platforms. Week 0 is the baseline toxicity score taken before start of RT.

4.6 Discussion

4.6.1 Overview of the Results

This analysis of the PACE-B trial acute toxicity, occurring up to 12 weeks from radiotherapy delivery, does not suggest that patients suffer greater acute RTOG toxicity with SBRT than CFMHRT. Of all secondary clinician reported endpoints examined, only CTCAE worst G2+ GI composite toxicity (both with and without reference to baseline) showed significantly higher proportions experiencing toxicity with SBRT. Differences in CTCAE toxicity were resolved by week-12. Overall, these results do not suggest substantially higher acute toxicity with SBRT relative to CFMHRT.

4.6.2 Study Results in Context

Regarding comparison to relevant historical studies, it is notable that the control arm (CFMHRT) has lower acute toxicity than the preceding CHHiP trial [1], with control toxicity more comparable to the PROFIT trial [3] (**Table 26**). As the largest UK and Canadian studies from the MHRT trial era, these are the best available comparators. Regarding possible causes of this difference; whereas IGRT was mandatory in both PACE and PROFIT, it was only used in 30% of CHHiP participants. PACE also utilised smaller margins, plus has benefitted from use of highly conformal techniques such as VMAT. CHHiP used ADT for most patients, which was not permitted in PACE or PROFIT. However, the OTT 01-01 trial did not show a difference in acute toxicity between 432 patients randomised 1:1 to RT then either adjuvant ADT versus neo-adjuvant/concurrent ADT, suggesting this is unlikely to be

causative [192]. Both PROFIT and CHHiP assessed acute RTOG weekly during radiotherapy, versus two-weekly in PACE. Conceivably the higher RTOG G2+ cumulative percentage seen in CHHiP/PROFIT, versus PACE-B, may result from more frequent sampling due to recall selection bias.

Table 26. PACE-B RTOG Toxicity in Context of CHHiP & PROFIT

Trial and Fractionation	RTOG G2+ Acute Toxicity	
	GI (%)	GU (%)
PACE		
78 Gy in 39 fractions OR 62 Gy in 20 fractions	12.3	27.3
36.25 Gy in 5 fractions	10.4	23.1
CHHiP		
74 Gy / 37 fractions	25	46
60 Gy / 20 fractions	38	49
57 Gy / 19 fractions	38	46
PROFIT		
78 Gy / 39 fractions	10.5	27.4
60 Gy / 20 fractions	16.7	30.9

The most comparable phase III RCT is the Scandinavian HYPO-RT-PC trial, which randomised intermediate-high risk nmPCa patients (1:1) between 78 Gy in 39 fractions over 7·8 weeks and 42·7 Gy in 7 fractions over 2·5 weeks, without ADT [95]. Key differences were HYPO-RT-PC recruited 11% high risk patients 89% intermediate (versus 8% low, 92% intermediate in PACE-B), treated a CTV of prostate only, and mostly (80%) utilised 3DCRT. IGRT (fiducial markers or guidance catheter) plus planning MRI were used for all patients. The control arms differ between HYPO-RT-PC (all 78 Gy in 39 fractions) and PACE-B (70% receiving 62 Gy in 20 fractions). This difference is important, given the higher acute GI toxicity seen for moderate hypofractionation in the CHHiP trial [1].

HYPO-RT-PC made only a single end-of-treatment toxicity assessment during the acute toxicity window[30]. The study suggested statistically significantly higher RTOG GU and PRO acute toxicity with ultra-hypofractionation. Comparing RTOG toxicity for PACE against HYPO-RT-PC (estimates approximated from graphs in paper [95]), GI grade 2 (10·4% vs 7·5%), GI grade 3-4 (0·2% vs 1%), GU grade 2 (20·7% vs 22%), and GU grade 3-4 (2·4% vs 6%). These are comparable figures, although reported G3-4 toxicity for HYPO-RT-PC is higher than most reports of ultra-hypofractionation (Recall **Table 4**).

PACE-B outcomes appear broadly in line with results anticipated from earlier phase work with 5 fraction SBRT (Recall **Table 4**). In the largest example, a multicentre phase II study of 309 men [97] recorded cumulative acute toxicity of CTCAE GI G2+ 12% and CTCAE GU G2+ 26%, similar to 15·7% and 30·8% respectively for SBRT patients in PACE-B. This concordance with prior work is reassuring regarding the external validity of these results.

4.6.3 SBRT and Different Delivery Methods

SBRT is already standard of care in some global centres, and is an option for men with low/favourable-intermediate risk nmPCa in the NCCN guidelines [18]. HYPO-RT-PC has suggested similar oncological outcomes with ultra-hypofractionation [95]. This was attenuated by increased acute toxicity in the study, notably higher grade 3-4 toxicity than other reports of SBRT. This may

[30] Thus potentially under-reporting relative to PACE-B.

potentially be driven by the 3DCRT technique predominantly utilised in the HYPO-RT-PC study. Other earlier phase studies, most of which used the same 36·25 Gy dose as PACE (Recall **Table 4**), suggest good oncological outcomes and low late toxicity with SBRT, but the mature results of PACE-B are required before definite oncological outcome statements may be made.

The method of SBRT delivery, e.g. CK versus non-CK, may influence acute toxicity; a prespecified area of interest after the introduction of conventional LINAC SBRT. There are many reasons why there may be a systematic difference between CK and non-CK SBRT outcomes, including variations in dosimetry, image-guidance and treatment times (typically 45 minutes for CK and <5 minutes for conventional LINAC). Our post-hoc analysis of same primary endpoint RTOG metrics show similar G2+ GI toxicity, but less G2+ GU toxicity with CK. CFMHRT toxicity has been compared between those centres using CK versus non-CK, finding no statistically significant difference for either RTOG G2+ GI or GU. This has suggested the need for multivariate analysis of the data to address possible confounding factors. This hypothesis is pursued in **Chapter 7**.

It is also interesting to consider the impact of overall treatment time on toxicity amongst SBRT patients. Recalling the worse CTCAE GI G2+ toxicity with SBRT, it is interesting to note the difference in peak CTCAE G2+ GI toxicity at 2 weeks post-RT for the 1-week SBRT (15%) vs >1-week SBRT (6%). This difference fits with typical radiobiological assumptions regarding the aetiology of acute toxicity. As a non-randomised factor (centres chose

their time of delivery), it would be ideal to confirm any protective effect of overall treatment time in a multivariate model. Such a hypothesis will also be pursued in **Chapter 7**.

4.6.4 Potential Cost Savings

Cost savings from the adoption of SBRT could benefit universally insured healthcare systems, such as the UK. Switching from 37 fractions to 20 fractions following the results of the CHHiP study resulted in savings estimated at £28 million per year [193]. Cost savings might also be seen with eventual adoption of the PACE-B protocol into routine practice. However, it should be noted that only low to low-intermediate patients would be eligible, the reduction is 15 rather than 17 fractions and the complexity of treatment would increase if planning MRI and daily IGRT is adopted universally.

4.6.5 Strengths of This Study

Strengths of this data relate predominantly to trial design. This is a large phase III RCT, and represents the first phase III acute toxicity data on 5 fraction SBRT compared to standard fractionation. PACE-B reflects real world prostate radiotherapy practice, with multiple centres recruiting in the UK, Canada and Ireland. It incorporates modern planning practice, with no patients receiving 3DCRT. The control arm protocol amendment strengthened the trial by allowing most CFMHRT patients to receive moderate hypofractionation at 62 Gy in 20 fractions – close to the 60 Gy in 20 fractions regimen shown effective in CHHiP [1] and PROFIT [3]. The

PACE-B acute toxicity sampling frequency exceeded HYPO-RT-PC (assessed only at end of RT and then 6 months). In combination with high proportions of assessment form returned, this is a major strength given the dynamic nature of acute toxicity.

4.6.6 Limitations of this Study

The data here is acute toxicity only; while important for patients, it must also be balanced against late toxicity and efficacy, the data for which are awaited. These results also cannot necessarily be extrapolated to higher risk patients. Randomised data regarding toxicity after SBRT, with concurrent ADT, and a larger target volume will be acquired by the PACE-C trial.

The lack of treatment blinding is always a limitation for subjective endpoints, such as toxicity. Whilst blinding has been achieved in radiotherapy trials in the past [194,195], this is not feasible for most studies. This issue is discussed further in **Chapter 5**. There was greater fiducial usage for IGRT in SBRT patients compared to CFMHRT. Mandatory fiducials would have prevented some centres participating, slowing trial recruitment.

The multiple radiotherapy schedule durations meant that some undesirable interpolation was needed to present two arm graphs (RTOG & CTCAE). It also means that the concept of 12 weeks post-radiotherapy refers to a quite different period of time for someone receiving 1 week SBRT (i.e. 13 weeks), versus 7.8 week conventional fractionation (i.e. 19.8 weeks). Future trials

comparing treatments of different durations should consider a follow-up schedule fixed by radiotherapy start date rather than end date.

4.7 Conclusions

This study represents the first published prospective phase III acute toxicity results randomising patients between five fraction SBRT and either conventional or moderately hypofractionated radiotherapy. The lack of increased toxicity in the SBRT arm is reassuring given the higher acute toxicity suggested in the only previously published phase III ultra-hypofractionation trial (HYPO-RT-PC). This is particularly relevant, given the more abbreviated (5 fraction) investigational radiotherapy protocol utilised in PACE-B. Results on late toxicity and biochemical control from PACE-B will be reported in the next 3-4 years.

Chapter 5. Patient Reported Acute Toxicity in the PACE-B Trial

5.1 Publications and Proceedings Relating To Chapter

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

Douglas H Brand, Alison C Tree, Peter Ostler, Hans van der Voet, Andrew Loblaw, … , Emma Hall, Nicholas van As, on behalf of the PACE Trial Investigators. Lancet Oncology 2019 Nov;20(11):1531-1543.

DOI: 10.1016/S1470-2045(19)30569-8.

[Full length research article]

Patient Reported Acute Toxicity in PACE-B, an International Phase III Randomised Controlled Trial Comparing Stereotactic Body Radiotherapy to Conventionally Fractionated or Moderately Hypofractionated Radiotherapy (CFMHRT) for Localised Prostate Cancer

D.H. Brand, A. Tree, P. Ostler, H. van der Voet, … , E. Hall, N. van As.

International Journal of Radiation Oncology Biology & Physics. 2019.

Volume 105, Issue 1, Supplement, Pages S55–S56.

DOI: 10.1016/j.ijrobp.2019.06.490

[Oral presentation, ASTRO Annual Congress 2019]

5.2 Background

Acute toxicity has historically tended to be presented in the form of CRO measures. For the practice-changing moderate hypofractionation trials, CHHiP and PROFIT reported acute RTOG toxicity [1,3], while RTOG-0415 reported acute CTCAE toxicity [2]. However, there has been increasing recognition of the importance of PRO measures, since the patient may ultimately be best placed to judge the detriment to life caused by any side effect [196]. PROs are more relevant to shared decision making and health-related economic analyses (e.g. QALY assessment) [197] .

Furthermore, a number of groups have suggested the possibility that clinicians under-report toxicity compared to patients. Examples for PCa EBRT CRO vs PRO comparisons are limited to a single analysis of a small randomised trial (n=160), comparing two hypofractionated prostate radiotherapy schedules [10]. This demonstrated clinician under-reporting for late GI and GU side effects, following MHRT.

In the acute toxicity setting, CRO vs PRO differences have only been reported in non-prostate tumour sites. Flores *et al* examined acute toxicity amongst 199 patients undergoing rectal chemoradiation, showing increased reporting of diarrhoea and proctitis for PRO vs CRO measures [198]. In two major randomised trials of adjuvant breast RT, PRO skin toxicity was consistently higher than both CRO and photographic assessment [199,200]. From further afield, analysis of CRO vs PRO data from three trials of

chemotherapy (1 adjuvant breast, 2 first line lung) have also suggested CRO under-reporting of toxicities in general [201].

PROs have also gained importance in localised PCa treatment following results from the ProtecT study, the only phase III RCT comparing EBRT to other treatment options [21]. PRO data from this has been incorporated into the National Health Service (NHS) predict calculator for nmPCa, which helps patients to make decisions between radical modalities [202]. It seems prudent that current and future comparisons between treatment modalities should examine PRO metrics, to facilitate cross-comparison with the gold-standard ProtecT RCT data.

The HYPO-RT-PC study, previously described in **Chapter 4**, is the only published phase III trial comparing ultrahypofractionated to conventionally fractionated EBRT for nmPCa. To date, only two PRO questions, both from the PCSS scale, have been reported: urinary and bowel problems in general, on a scale 0-10 [95]. End of treatment PRO bowel problems were significantly worse in the ultrahypofractionated arm ($p<0.0001$)[31]. For GU effects, they suggested evidence of increased PRO urinary problems ($p=0.0066$). It should be noted that large numbers of comparisons were performed in the paper, without account for multiple testing.

[31] Results reported graphically in HYPO-RT-PC paper, so exact figures not available for discussion.

In summary, under-reporting of toxicity with CROs compared to PROs has been recognised. The seven fraction HYPO-RT-PC study suggests worse PRO acute toxicity with a seven-fraction prostate EBRT regimen. For this chapter, the aim is to investigate differences in acute toxicity outcomes for a range of PROs for between SBRT and CFMHRT in the PACE-B trial.

5.3 Hypotheses

1. Comparing CFMHRT and SBRT, patient reported gastrointestinal acute toxicity may differ between modalities.

2. Comparing CFMHRT and SBRT, patient reported genitourinary acute toxicity may differ between modalities.

3. Comparing CFMHRT and SBRT, patient reported sexual acute toxicity may differ between modalities.

5.4 Methods

5.4.1 The PACE Study

Details of the PACE trial (NCT01584258), participants, treatment procedures and primary endpoints were described in **Chapter 4**. Methods here therefore focus on aspects specific to the analysis of PROs, as described in the statistical analysis plan, written by me and reviewed by supervisory team and ICR-CTSU statisticians (Glare Griffin and Vicki Hinder) (**Appendix 2, p385 onwards**).

5.4.2 Patient Reported Outcome Measures

A similar ethos to the CROs was adopted for analysis of PRO acute toxicity. Namely that acute toxicity is characterised by onset, and then recovery, for most patients, which at the trial level is seen as a peaked acute toxicity graph following RT (recall **Figure 18**). It was considered important to capture information on both peak-toxicity and recovery phases of this dynamic response. Four PRO scales were collected as part of the PACE-B trial: EPIC-26; IPSS; IIEF-5 & Vaizey Incontinence Score.

5.4.2.1 EPIC-26 Assessments

EPIC-26 assesses five domains: urinary incontinence (UI), urinary obstructive (UO), bowel, sexual and hormonal. An additional question assesses overall urinary bother, which is not included in the two urinary domains. For acute toxicity, it was completed at baseline, then 4 weeks & 12 weeks post-radiotherapy.

Per recommended methodology, EPIC-26 scores were rescaled to a 0-100 point scale; higher scores representing better QoL [24]. Subdomains were scored, by averaging, if sufficient subdomain questions were completed[32] [24]. Both absolute subdomain scores and changes in subdomain scores from baseline were compared between arms. Additionally, each individual EPIC-26 question was examined graphically, between arms, although without formal statistical comparison to control multiplicity of testing.

A minimal clinically important difference (MCID) in EPIC-26 subdomain score was defined as: urinary incontinence (8 points) urinary obstructive (6 points), bowel (5 points), sexual (11 points), hormonal (5 points), based on data by Skolarus et al [25]. MCID values are calculated as ½ SD of baseline subdomain score, so such values were calculated for the PACE-B patients, to check appropriateness of the Skolarus scores. The proportions of patients experiencing a MCID deterioration at any point during the acute toxicity window were compared between arms.

5.4.2.2 IPSS Assessments

The IPSS focusses on prostatic obstruction, originally being validated to assess symptoms of benign prostatic hypertrophy [203]. It comprises seven questions assessing incomplete emptying, frequency, intermittency, urgency, weak stream, straining and nocturia. These are scored from 0 (not at all) to 5 (almost always), with a total score from 0 (best) to 35 (worst). A separate

[32] Urinary incontinence (4/4 questions), urinary obstructive (4/4 questions), bowel (5/6 questions), sexual (5/6 questions), hormonal (4/5 questions)

QoL question asks how patients would feel spending the rest of their life with such symptoms from 0 (delighted) to 6 (terrible). IPSS was collected at baseline then at 2-, 4-, 8- and 12-weeks post-RT.

Each question, the total IPSS score and IPSS bother question were plotted over time, both for absolute scores and changes from baseline. Total IPSS score and IPSS bother question were both compared between arms at baseline, worst, worst exceeding baseline and week-12 (recovery). IPSS was also analysed categorically with severity categories were assessed as none (0 points), mild (1-7 points), moderate (8-19 points), severe (20-35 points); based on standard scoring [26], with comparison of proportions between arms at baseline, worst and week-12 (recovery).

5.4.2.3 IIEF-5 Assessments

IIEF-5 is five question abridged questionnaire, validated for assessment of erectile function and sexual intercourse satisfaction [141]. Each question is scored from 1 to 5, with higher scores representing better sexual function. The total scale therefore runs from 5 (severe erectile dysfunction) to 25 (no erectile dysfunction). The scale was assessed only at baseline and week-12 post-RT. Scores were plotted over time and compared between arms at baseline and week-12 post-RT.

5.4.2.4 Vaizey Assessments

The Vaizey faecal incontinence scale is a seven-question tool which assesses a patient’s perspective on their faecal continence. Four questions

assess the frequency of incontinence (solid, liquid, gas), and lifestyle impact. A further three assess pad use, constipating medications and urgency. Higher scores are worse, with the total score ranging from zero (perfect continence) to 24 (totally incontinent). The Vaizey score was collected at baseline and weeks 4 & 12 post-RT. Vaizey total scores were plotted over time, both for absolute scores and changes from baseline. Vaizey total scores were compared between arms for baseline, worst, worst exceeding baseline and week-12 (recovery).

5.4.3 Statistical analysis

For this PRO analysis, the patients were analysed in the same per-protocol assignment as the CROs. I.e. those patients receiving ≥1 fractions of CFMHRT or SBRT radiotherapy included. Patients not receiving radiotherapy were excluded from this analysis. Patients missing a single timepoint remained eligible for analysis at other timepoints. Those with missing data at baseline were excluded from analyses with reference to baseline scores.

Graphical data was plotted with the 95% CI bars (calculated as ±1.96 x SD) at each data timepoint to facilitate comparison. Numbers-at-risk were included for each arm, at each timepoint. Red data is CFMHRT and blue data is SBRT throughout this chapter.

For summarised score data in tables, medians and interquartile ranges are used as measures of central tendency and data dispersion. Comparisons of scores in each arm were by Mann-Whitney tests. Comparisons of

proportions were by Chi-square, or by Fisher's exact test where the assumptions of Chi-square were not met. The 95% CIs for the difference in percentages of patients experiencing MCID reduction in EPIC-26 score were by normal approximation. To avoid inflation of type I error rate between CRO and PRO data, p-values in this PRO study were all interpreted at penalised significance level of $p<0.001$, to account for multiplicity of testing.

Analyses are based on a snapshot of data taken on 28/05/2019. Data was collected and provided to me by ICR-CTSU. Quality assurance of data and analyses were conducted using STATA version 15.1 (StataCorp LLC, Texas, USA). All code was entirely written and debugged by me personally.

5.5 Results

5.5.1 Return Rates

Return rates were generally good, for PRO instruments. Considering each timepoint collected, EPIC-26 rates ranged 84.5 – 93.5%; IPSS rates ranged 82.4 – 92.4%; IIEF-5 rates ranged 66.8 – 74.5%; Vaizey rates ranged 64.1 – 86.3%. Detailed return tables are provided in **Appendix 6, p417**.

5.5.2 Expanded Prostate Cancer Index Composite-26 Outcomes

Changes from baseline in EPIC-26 subdomain scores appear similar over time, comparing between CFMHRT and SBRT, with overlapping 95% CIs at each timepoint assessed (**Figure 30**). The typical acute toxicity response of worsening of function, followed by recovery is seen most clearly in the UO and bother domain, with bowel and sexual domains still (on average) reduced at the week-12 mark. The absence of UI symptoms is noted.

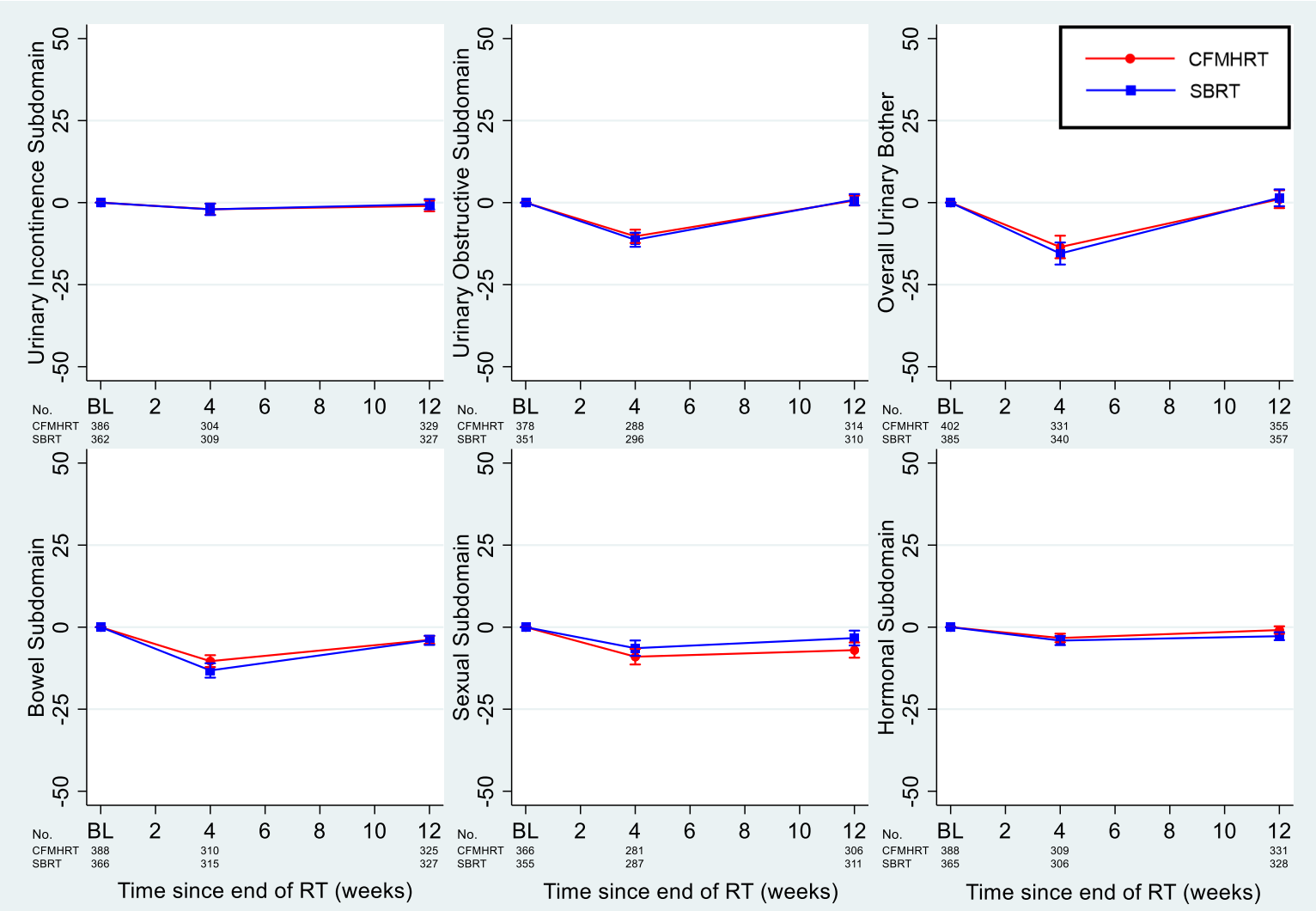


Figure 30. Changes from Baseline in EPIC-26 Subdomains

EPIC-26 subdomain score changes from baseline in the acute toxicity setting, by arm. Urinary bother is graphed separately, as it is not part of the UI or UO subdomains. Error bars show 95% CIs. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1-2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

Similar findings are seen when looking absolute scores for each EPIC-26 subdomain (i.e. without reference to baseline). CFMHRT and SBRT have visually similar plots in all subdomains, again with overlapping 95% CIs at each timepoint assessed. (**Figure 31**)**.** Absolute scores for CFMHRT vs SBRT are presented graphically over time for individual questions in the urine subdomains (**Figure 32**). Though little change in the UI questions (top row) is apparent, the UO acute toxicity (bottom row) show dysuria, daytime frequency and weak stream all worsen, although error bars for CFMHRT and SBRT overlap at all timepoints.

Individual questions for the bowel subdomain are presented in **Figure 33**, with all questions exhibiting some acute response, although, as might be anticipated, frequency and urgency show the most marked deteriorations. At worst, the error bars for “bowel frequency” and “overall bowel problem” do not overlap, with SBRT showing a greater deterioration.

Sexual subdomain individual questions are shown in **Figure 34**, with all questions showing small deteriorations with very similar appearances for CFMHRT and SBRT.

The individual hormone subdomain questions (**Figure 35**) expectedly do not show alteration in aspects specifically related to ADT (hot flushes and gynaecomastia), with fatigue being the only question showing a pronounced acute toxicity reaction.

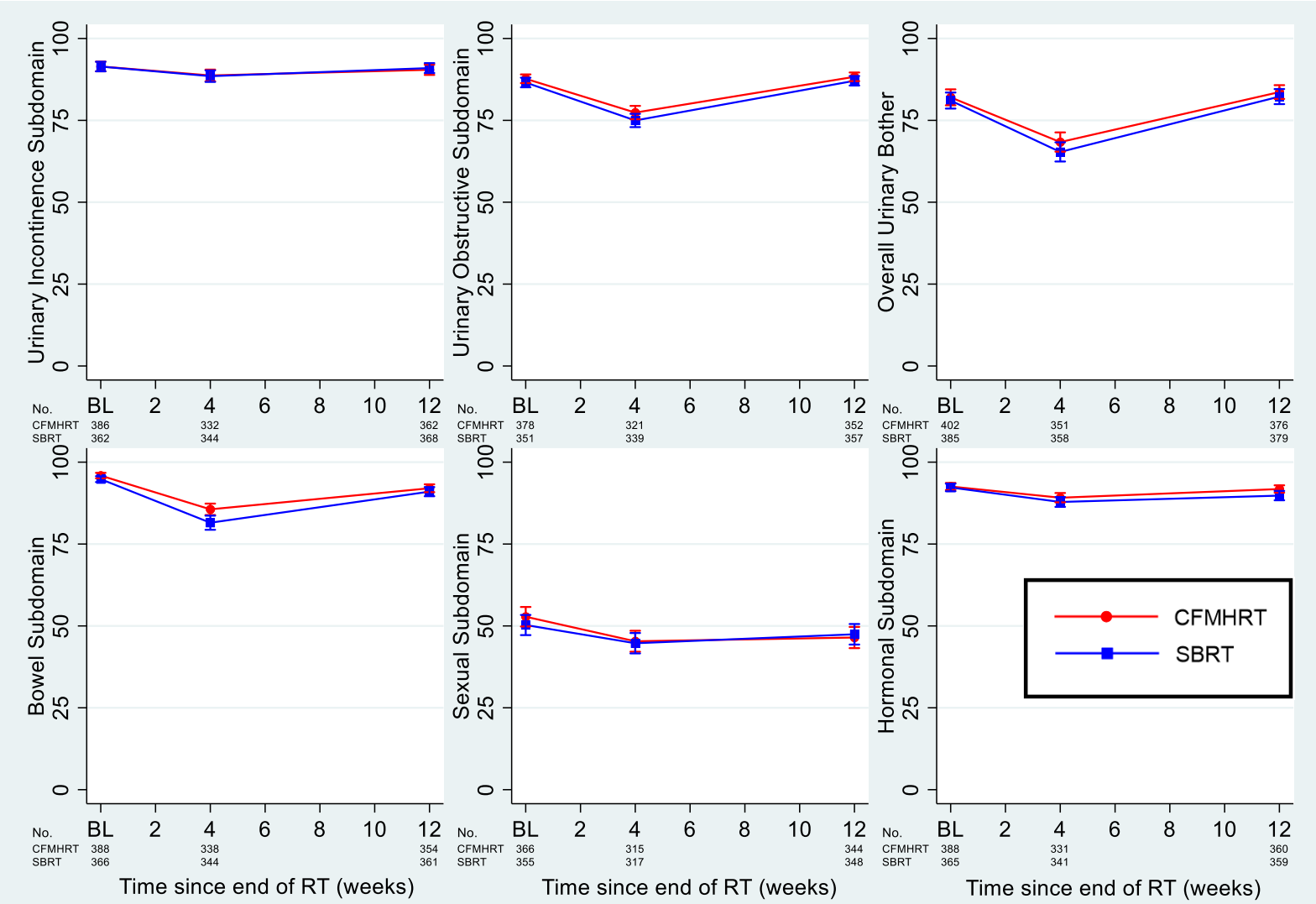


Figure 31. Mean average EPIC-26 subdomain scores

Urinary bother is graphed separately, as not form part of UI or UO subdomain scores. Error bars show 95% CIs. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

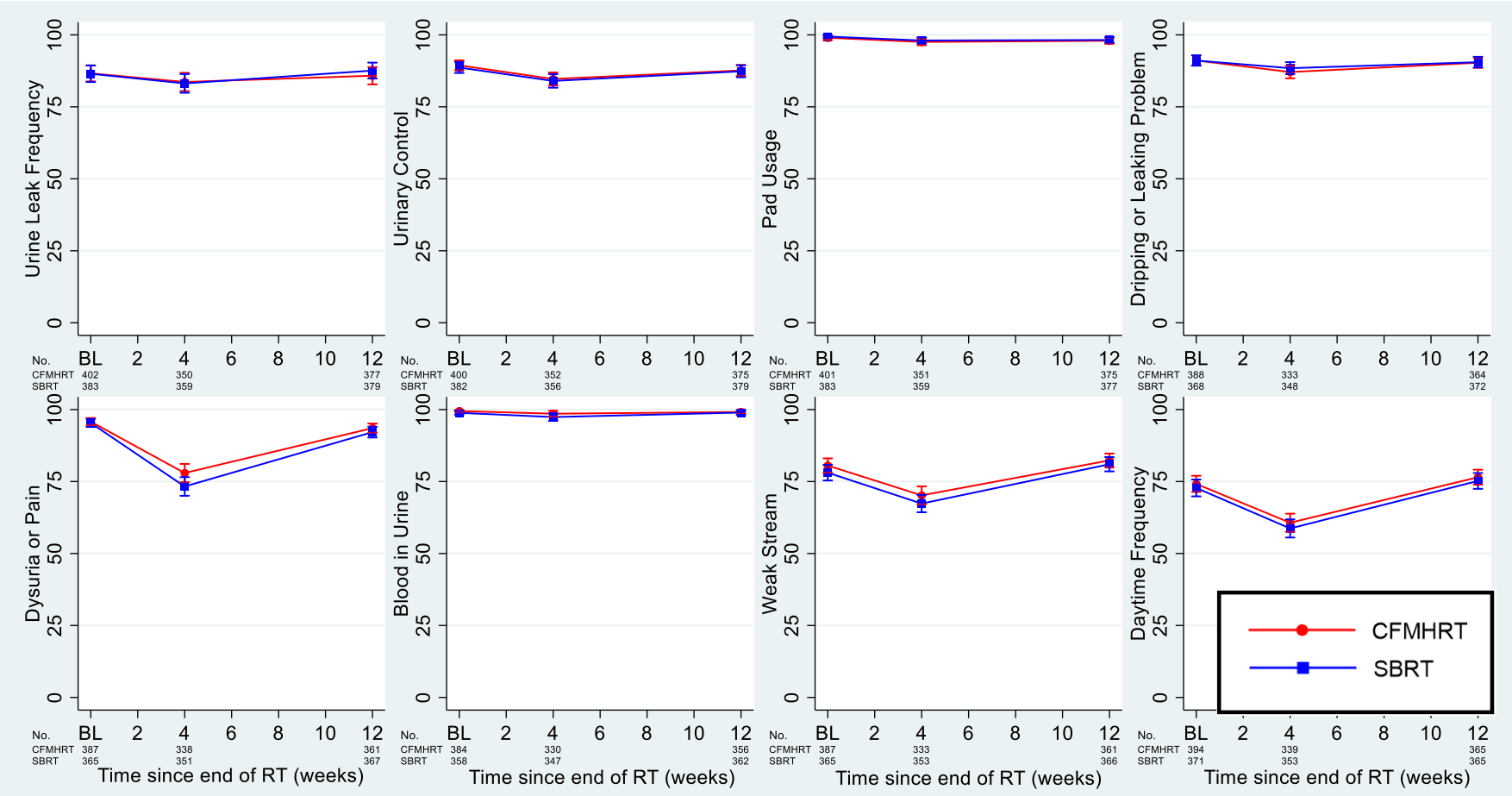


Figure 32. EPIC-26 Urine Questions - Absolute Scores

Top four boxes form the UI subdomain, bottom four form the UO subdomain. Error bars show 95% CIs for estimates of mean subdomain scores. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 = baseline toxicity score, pre-RT

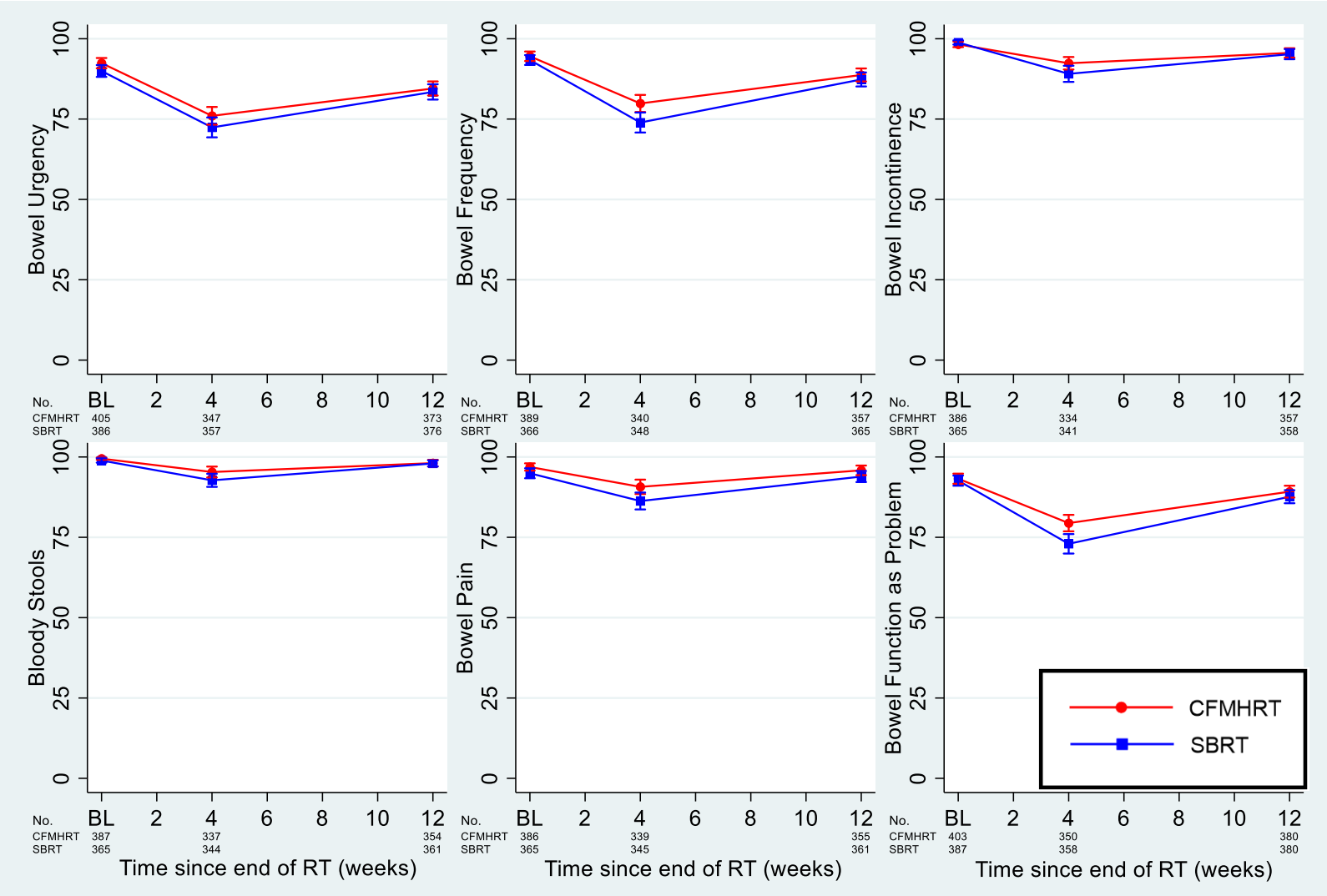


Figure 33. EPIC-26 Bowel Questions - Absolute Scores

Error bars show 95% CIs for estimates of mean subdomain scores. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

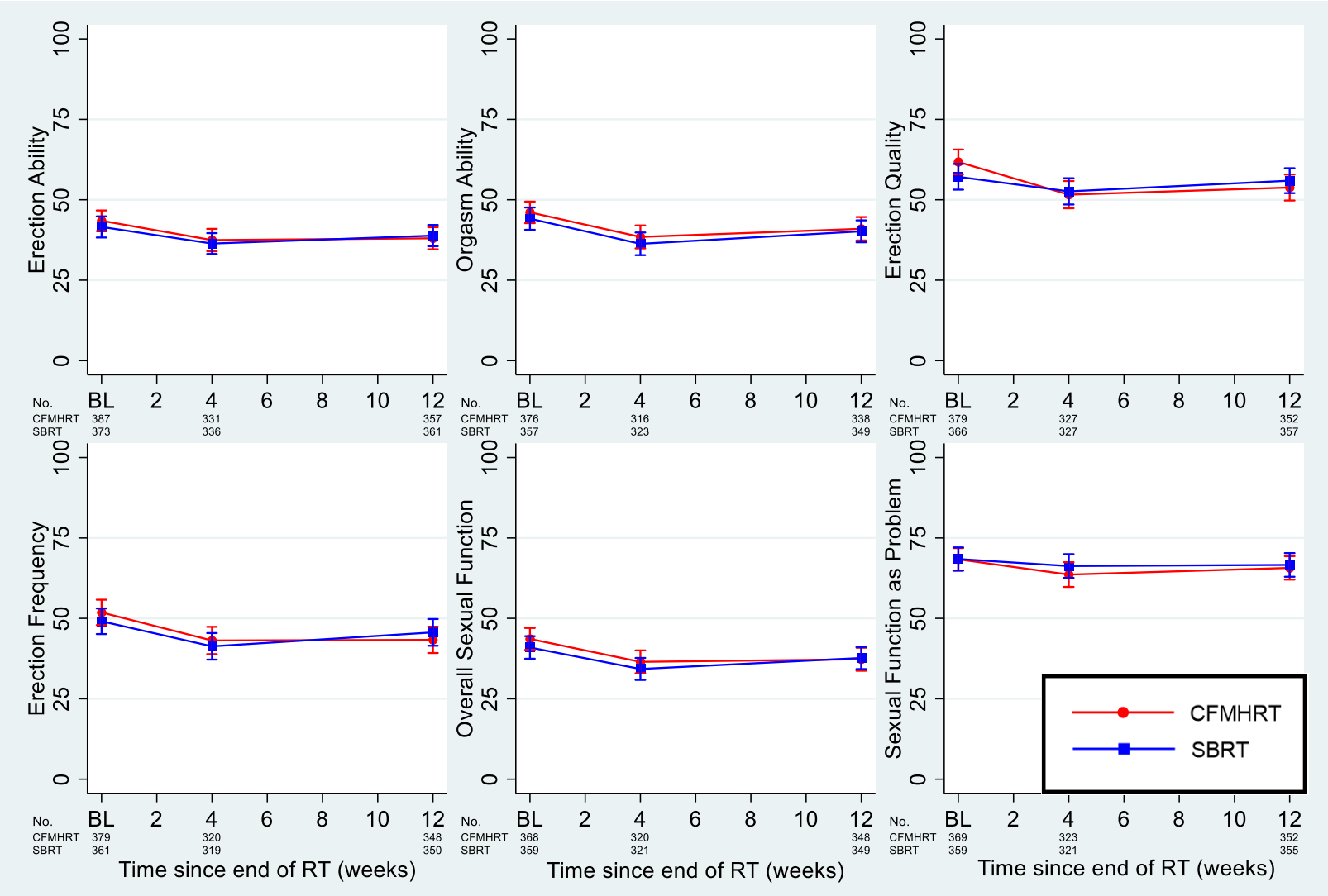


Figure 34. EPIC-26 Sexual Questions - Absolute Scores

Error bars = 95% CIs for estimates of mean subdomain scores. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

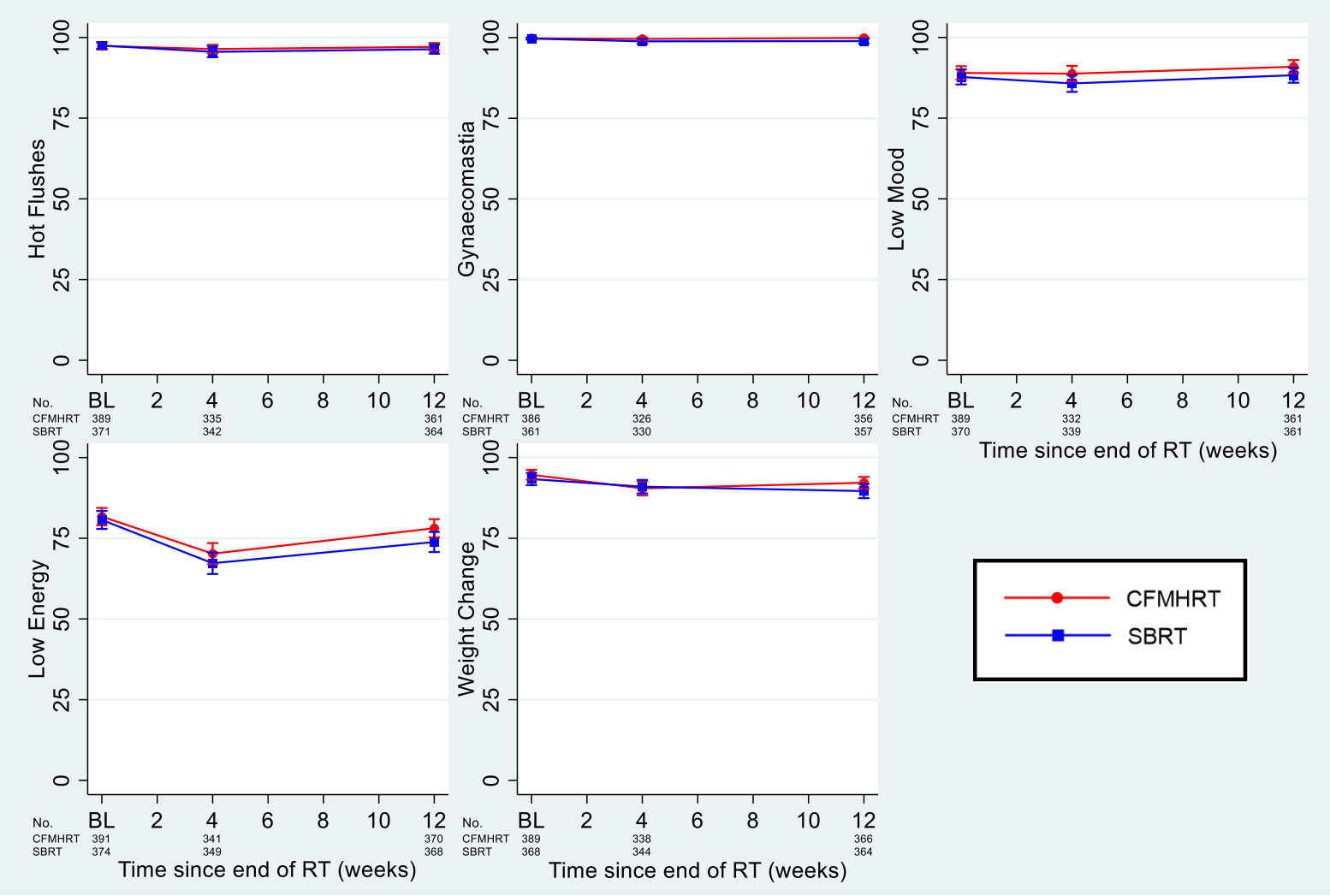


Figure 35. EPIC-26 Hormonal Questions - Absolute Scores

Error bars = 95% CIs. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

The median scores and interquartile ranges are presented for each EPIC-26 subdomain score and the urinary bother question at baseline, worst, worst minus baseline and week-12 post RT (**Table 27**). Accounting for multiple comparisons in interpretation of p-values, no statistically significant differences between CFMHRT and SBRT was seen at any timepoint (**Table 27, right hand column**).

Table 27. Comparison of Median Scores for EPIC-26 Subdomains

Due to multiple comparisons, p-value significance interpreted at p<0.001 level.

EPIC-26 Subdomain	Per Protocol Treatment						Mann-Whitney p-value
	CFMHRT			SBRT			
	n	Median	IQR	n	Median	IQR	
Urinary Incontinence							
Baseline	386	100	85.5 – 100	362	100	85.5 – 100	0.75
Worst	406	93.75	79.25 – 100	400	93.75	77.25 – 100	0.84
Worst Minus Baseline	368	0	- 8.375 – 0	355	0	- 8.25 – 0	0.91
12 weeks post-RT	362	100	85.5 – 100	368	100	85.5 – 100	0.72
Urinary Obstructive							
Baseline	378	87.5	81.25 – 100	351	87.5	81.25 – 100	0.33
Worst	399	81.25	68.75 – 93.75	399	81.25	62.5 – 87.5	0.053
Worst Minus Baseline	354	-6.25	-18.75 – 0	342	-6.25	-18.75 – 0	0.50
12 weeks post-RT	352	93.75	81.25 – 100	357	87.5	81.25 – 100	0.28
Urinary Bother							
Baseline	402	100	75 – 100	385	100	75 – 100	0.40
Worst	413	75	50 – 100	403	75	50 – 75	0.15
Worst Minus Baseline	390	0	-25 – 0	378	0	-25 – 0	0.32
12 weeks post-RT	376	100	75 – 100	379	100	75 – 100	0.65
Bowel							
Baseline	388	100	95.8 – 100	366	100	91.7 – 100	0.014
Worst	404	91.7	75 – 100	400	87.5	75 – 95.8	0.024
Worst Minus Baseline	369	-4.2	-16.7 – 0	359	-8.3	-20.8 – 0	0.081
12 weeks post-RT	354	95.8	87.5 – 100	361	95.8	87.5 – 100	0.61
Sexual							
Baseline	366	52.8	26.3 – 75	355	48.7	22.2 – 75	0.23
Worst	388	39.6	16.7 – 65.3	376	36.2	16.7 – 65.3	0.80
Worst Minus Baseline	342	-9.7	-25 – 0	333	-5.7	-21.1 – 1.3	0.081
12 weeks post-RT	344	44.5	18 – 72.1	348	44.5	18 – 74	0.63
Hormonal							
Baseline	388	97.5	90 – 100	365	95	90 – 100	0.74
Worst	403	93.75	80 – 100	391	90	80 – 100	0.019
Worst Minus Baseline	370	0	-10 – 0	350	-5	-12.5 – 0	0.020
12 weeks post-RT	360	95	85 – 100	359	95	85 – 100	0.11

The ½ SD (used to define MCIDs) for each baseline EPIC-26 subdomain score is presented in **Table 28**, with the Skolarus data [204] presented for comparison. It would appear that the values for MCIDs in the Skolarus paper are appropriate for application to the PACE-B dataset.

Table 28. EPIC-26 Minimal Clinically Important Differences Comparison

Showing the comparison of EPIC-26 subdomain MCIDs (1/2 SD of score for subdomain) between the Skolarus data and PACE-B baseline.

EPIC-26 Subdomain	**Skolarus Data** [204]	**PACE-B Baseline**
	½ Standard Deviation	**½ Standard Deviation**
Urinary Incontinence	6.4	7.3
Urinary Obstructive	7.0	6.8
Bowel	4.4	4.3
Sexual	13.8	14.7
Hormonal	5.7	5.6

No statistically significant difference, between arms, in the proportion of patients experiencing a clinically significant reduction from baseline occurred for any EPIC-26 subdomain score area, as assessed at any time (**Table 29**).

Table 29. EPIC-26 Score Reductions Exceeding Minimal Clinically Important Differences during Acute Toxicity Window

Data missing if either baseline and/or all follow-up data points missing.

EPIC-26 MCID Reduction at Any Timepoint	Per Protocol Treatment				Comparison and Chi-square Testing
	CFMHRT		SBRT		
	n	%	n	%	
Urinary Incontinence					
No	255	69.3%	255	71.8%	Difference −2.5% 95% CI −9.2 to 4.1% p=0.45
Yes	113	30.7%	100	28.2%	
Missing Data	64	N/A	60	N/A	
Urinary Obstructive					
No	137	38.7%	129	37.7%	Difference 1.0% 95% CI −6.2 to 8.2% p=0.79
Yes	217	61.3%	213	62.3%	
Missing Data	78	N/A	73	N/A	
Bowel					
No	189	51.2%	161	44.8%	Difference 6.4% 95% CI −0.9 to 13.6% p=0.085
Yes	180	48.8%	198	55.2%	
Missing Data	63	N/A	56	N/A	
Sexual					
No	174	50.9%	194	58.3%	Difference −7.4% 95% CI −14.9 to 0.1% p=0.054
Yes	168	49.1%	139	41.7%	
Missing Data	90	N/A	82	N/A	
Hormonal					
No	198	53.5%	162	46.3%	Difference 7.2% 95% CI −0.06 to 14.5% p=0.053
Yes	172	46.5%	188	53.7%	
Missing Data	61	N/A	65	N/A	

Neither were any significant differences seen when examining the proportions of CFMHRT vs SBRT patients exhibiting a MCID at week-12 only (**Table 30**).

Table 30. Patients with EPIC-26 Score Reductions Exceeding Minimal Clinically Important Differences at Week-12 Only

Data missing if either baseline and/or week 12 post-RT follow-up data points missing.

EPIC-26 MCID Reduction at Week 12 Only	Per Protocol Treatment				Comparison
	CFMHRT		SBRT		
	n	%	n	%	
Urinary Incontinence					
No	261	79.3%	275	84.1%	Difference −4.8% 95% CI −10.7 to 1.1% p=0.11
Yes	68	20.7%	52	15.9%	
Missing Data	103	N/A	88	N/A	
Urinary Obstructive					
No	206	65.6%	204	65.8%	Difference −0.2% 95% CI −7.7 to 7.2% p=0.96
Yes	108	34.4%	106	34.2%	
Missing Data	118	N/A	105	N/A	
Bowel					
No	233	71.7%	235	71.9%	Difference −0.2% 95% CI −7.1 to 6.7% p=0.96
Yes	92	28.3%	92	28.1%	
Missing Data	107	N/A	88	N/A	
Sexual					
No	187	61.1%	216	69.5%	Difference −8.3% 95% CI −15.8 to 0.9% p=0.029
Yes	119	38.9%	95	30.5%	
Missing Data	126	N/A	104	N/A	
Hormonal					
No	227	68.6%	199	60.7%	Difference 7.9% 95% CI 0.6 to 15.2% p=0.034
Yes	104	31.4%	129	39.3%	
Missing Data	101	N/A	87	N/A	

5.5.3 International Prostate Symptom Score Outcomes

IPSS sub-scores, total score and QoL over time appear similar between arms, both for change from baseline (**Figure 36**) and absolute scores (**Figure 37**). 95% CIs overlap at all timepoints examined.

[Caption for next page figure]

Figure 36. IPSS Scores : Change from Baseline

Changes from baseline IPSS scores, by time, for CFMHRT and SBRT. Patients included at any timepoint if both baseline and relevant timepoint score available. The IPSS total is formed by the sum of all subscores except for urinary QoL. Note that the time period between baseline scoring and week 2 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Error bars = 95% CIs. Week 0 is the baseline toxicity score taken before start of RT.

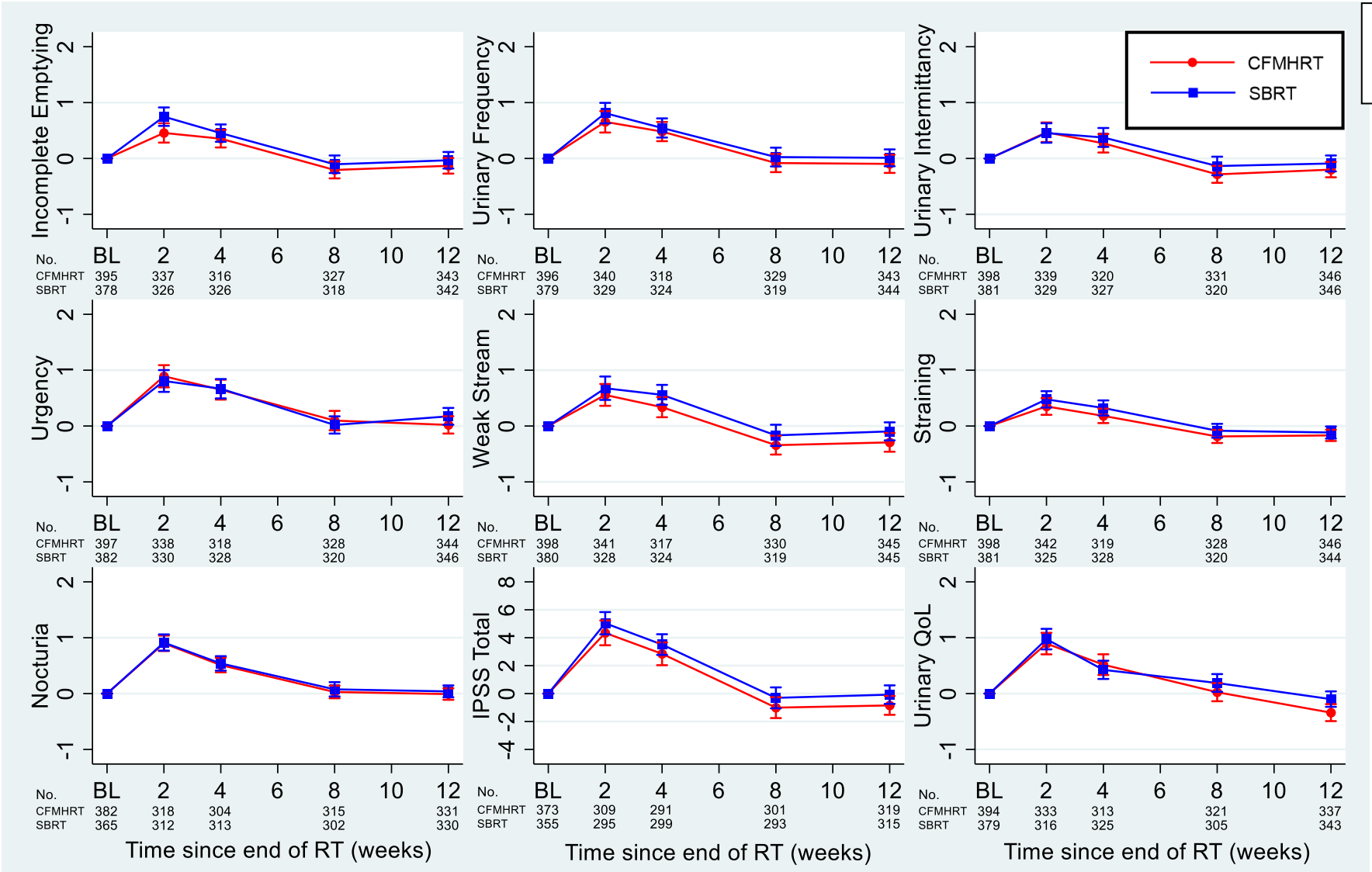
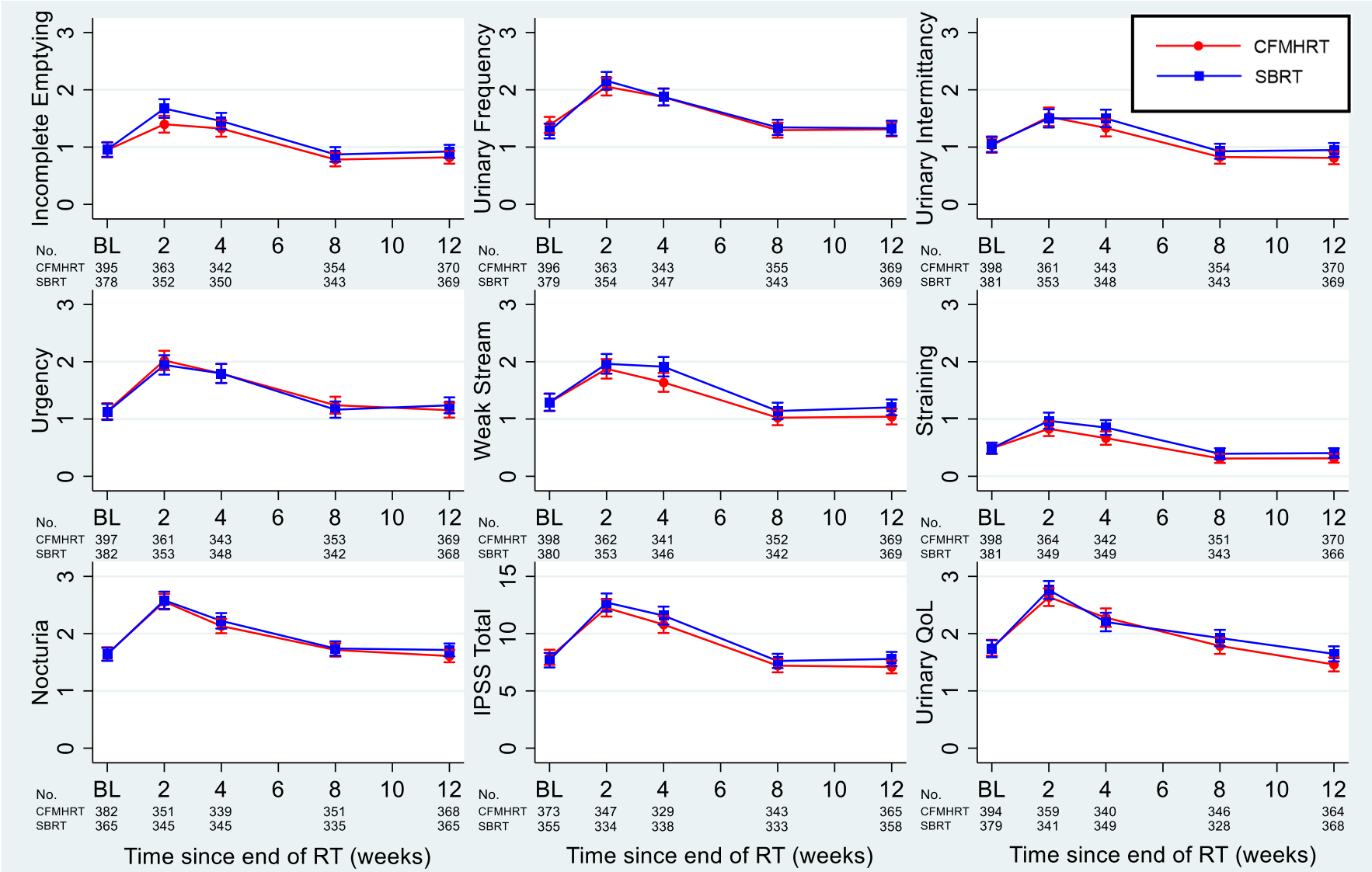


Figure 36: *Caption on prior page*



[Caption *for preceeding*]

[Caption for preceeding page figure]

Figure 37. IPSS Scores: Absolute Measures

Averages for IPSS subscores, total and quality of life score, by time, for CFMHRT and SBRT. The IPSS total is formed by the sum of all subscores except for urinary QoL. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Error bars = 95% CIs. Week 0 is the baseline toxicity score taken before start of RT.

No statistically significant differences, between treatment arms, were seen for median scores of: worst IPSS total, week-12 IPSS total, worst IPSS QoL, or week-12 IPSS QoL (**Table 31**).

Table 31. Comparison of Median IPSS Total Scores and Quality of Life Scores

IPSS Parameter	Per Protocol Treatment						Mann-Whitney p-value
	CFMHRT			SBRT			
	n	Median	IQR	n	Median	IQR	
IPSS Total Score							
Baseline	373	6	3 – 11	355	6	3 – 12	0.70
Worst	420	13	7 – 19	402	13	8 – 19	0.076
Worst Minus Baseline	365	5	1 – 10	348	6	2 – 10	0.035
12-week post RT	365	6	3 – 10	358	6.5	3 – 11	0.13
IPSS QoL Score							
Baseline	394	2	1 – 3	379	2	1 – 3	0.74
Worst	423	3	2 – 4	409	3	2 – 4	0.41
Worst Minus Baseline	387	1	0 – 2	376	1	0 – 2	0.20
12-week post RT	364	1	1 – 2	368	2	1 – 2	0.044

IPSS severity categories (none, mild, moderate, severe) over time appear similar between treatment arms (**Figure 38**) with no statistically significant differences in IPSS total score categories at baseline, worst and week-12 post-RT (**Table 32**).

Table 32. Comparison of IPSS Total Score Categories

IPSS Total Score Categories at Timepoint	Per Protocol Treatment				Chi-Square / Fishers
	CFMHRT		SBRT		
	n	%	n	%	p-value
Baseline					
None	20	5.4%	16	4.5%	0.82 (Chi-square)
Mild (1-7)	191	51.2%	193	54.4%	
Moderate (8-19)	139	37.3%	127	35.8%	
Severe (20-35)	23	6.2%	19	5.4%	
Worst					
None	1	0.2%	1	0.2%	0.15 (Fisher's)
Mild (1-7)	107	25.5%	80	19.9%	
Moderate (8-19)	232	55.2%	227	56.5%	
Severe (20-35)	80	19.0%	94	23.4%	
Week 12 Post-RT					
None	10	2.7%	12	3.4%	0.17 (Chi-square)
Mild (1-7)	212	58.1%	195	54.5%	
Moderate (8-19)	134	36.7%	131	36.6%	
Severe (20-35)	9	2.5%	20	5.6%	

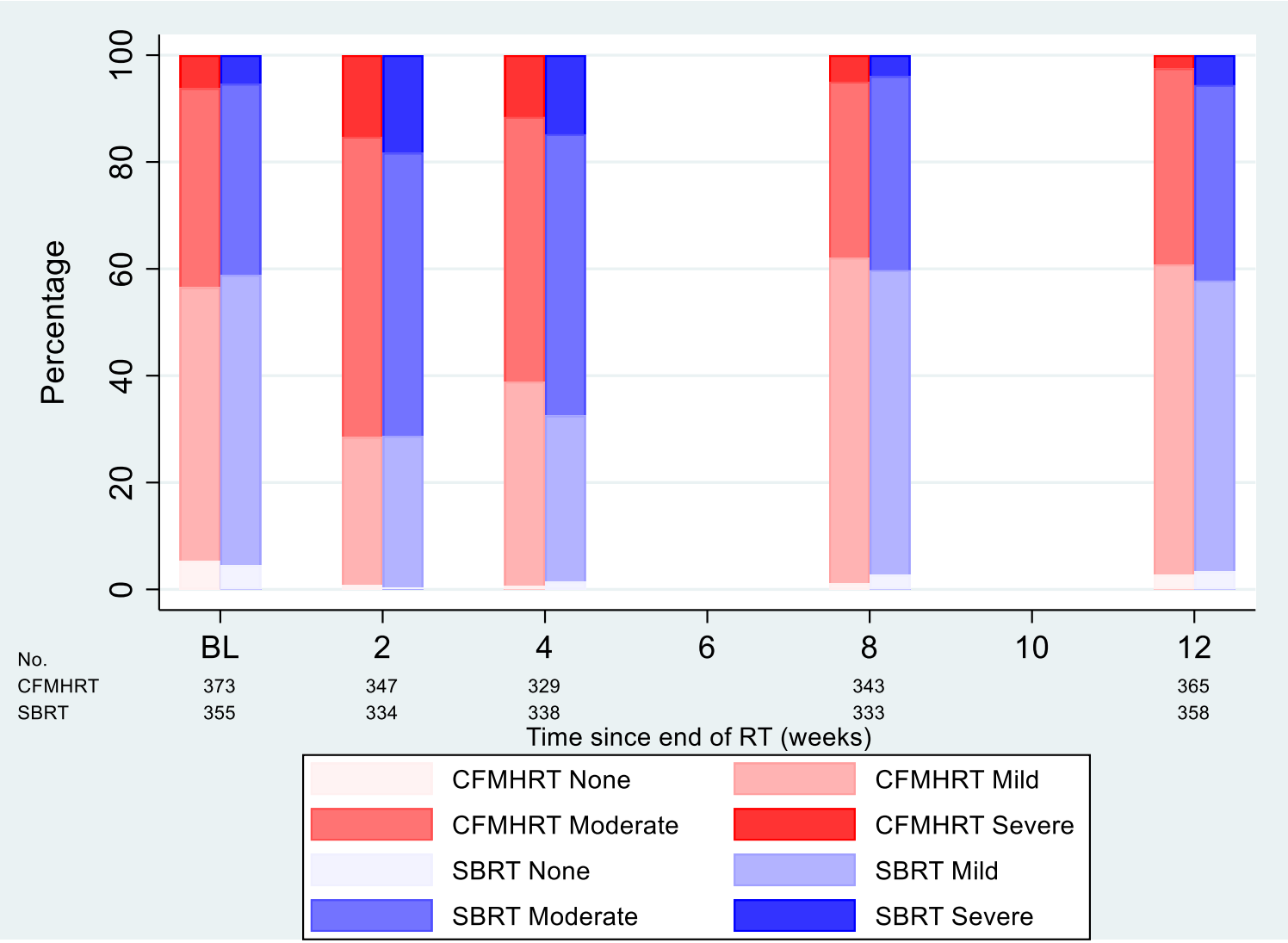


Figure 38. IPSS Severity Categories Over Time

Changes in IPSS severity categories after RT. Categories are defined by the IPSS total score; none (score 0), mild (score 1-7), moderate (score 8-19), severe (score 20-35). Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Week 0 is the baseline toxicity score taken before start of RT.

5.5.4 International Index of Erectile Function Outcomes

For IIEF-5, no statistically significant differences were seen between CFMHRT and SBRT at baseline, nor week-12 post-RT (**Table 33**).

Table 33. Comparison of Median IIEF-5 Scores

IIEF – 5 Scores	Per Protocol Treatment						Mann-Whitney p-value
	CFMHRT			SBRT			
	n	Median	IQR	n	Median	IQR	
Baseline	322	16	7 – 21	309	14	7 – 20	0.13
Week 12	280	12	5 – 20	286	12.5	5 – 20	0.86

IIEF-5 individual questions and total score are displayed graphically (**Figure 39**), showing a uniform minor deterioration in all questions between baseline and week-12 post-RT.

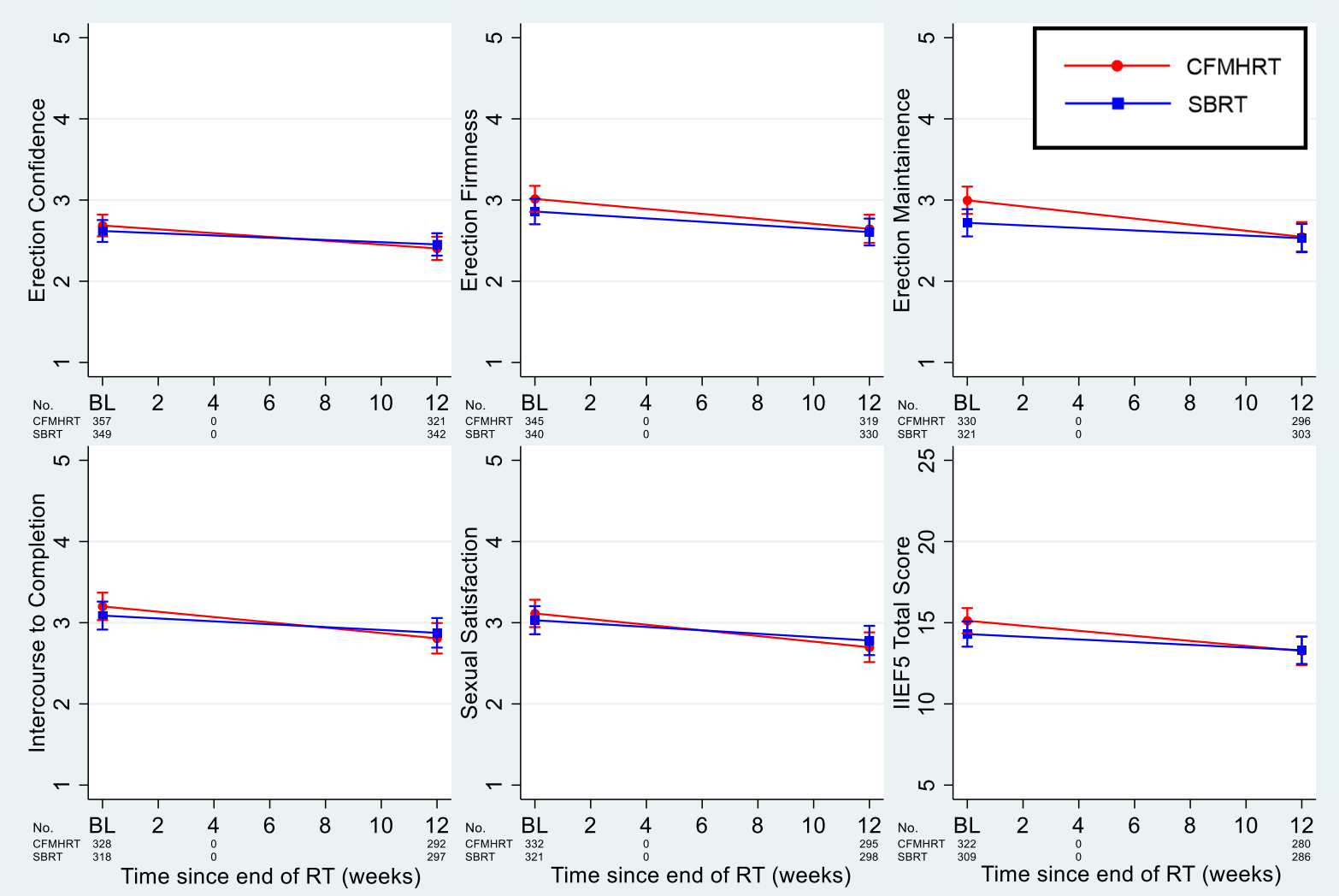


Figure 39. IIEF-5 Questions and Total Score

IIEF-5 individual questions and total score over time, for CFMHRT and SBRT. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Error bars = 95% CIs. Week 0 is the baseline toxicity score taken before start of RT. N.B. Higher score for IIEF-5 is better.

5.5.5 Vaizey Faecal Incontinence Scale Outcomes

No statistically significant differences between treatment arms were seen for Vaizey scores at baseline, worst, worst change from baseline, and week-12 post-RT (**Table 34**).

Table 34. Comparison of Median Vaizey Total Scores

Vaizey Score Timepoint	Per Protocol Treatment						Mann-Whitney
	CFMHRT			SBRT			
	n	Median	IQR	n	Median	IQR	p-value
Baseline	373	1	0 – 4	358	1	0 – 4	0.99
Worst	384	4	1 – 6	381	4	0 – 6	0.82
Worst Change cf. Baseline	214	2	0 – 4	223	1	0 – 4	0.84
Week 12 Post RT	349	2	0 – 4	352	2	0 – 4	0.75

Vaizey score changes from baseline appear similar for both treatment arms (**Figure 40**). Examining absolute scores, there is generally little reporting of incontinence and, perhaps surprisingly, no increase observed in rectal urgency at week 4 (**Figure 41**).

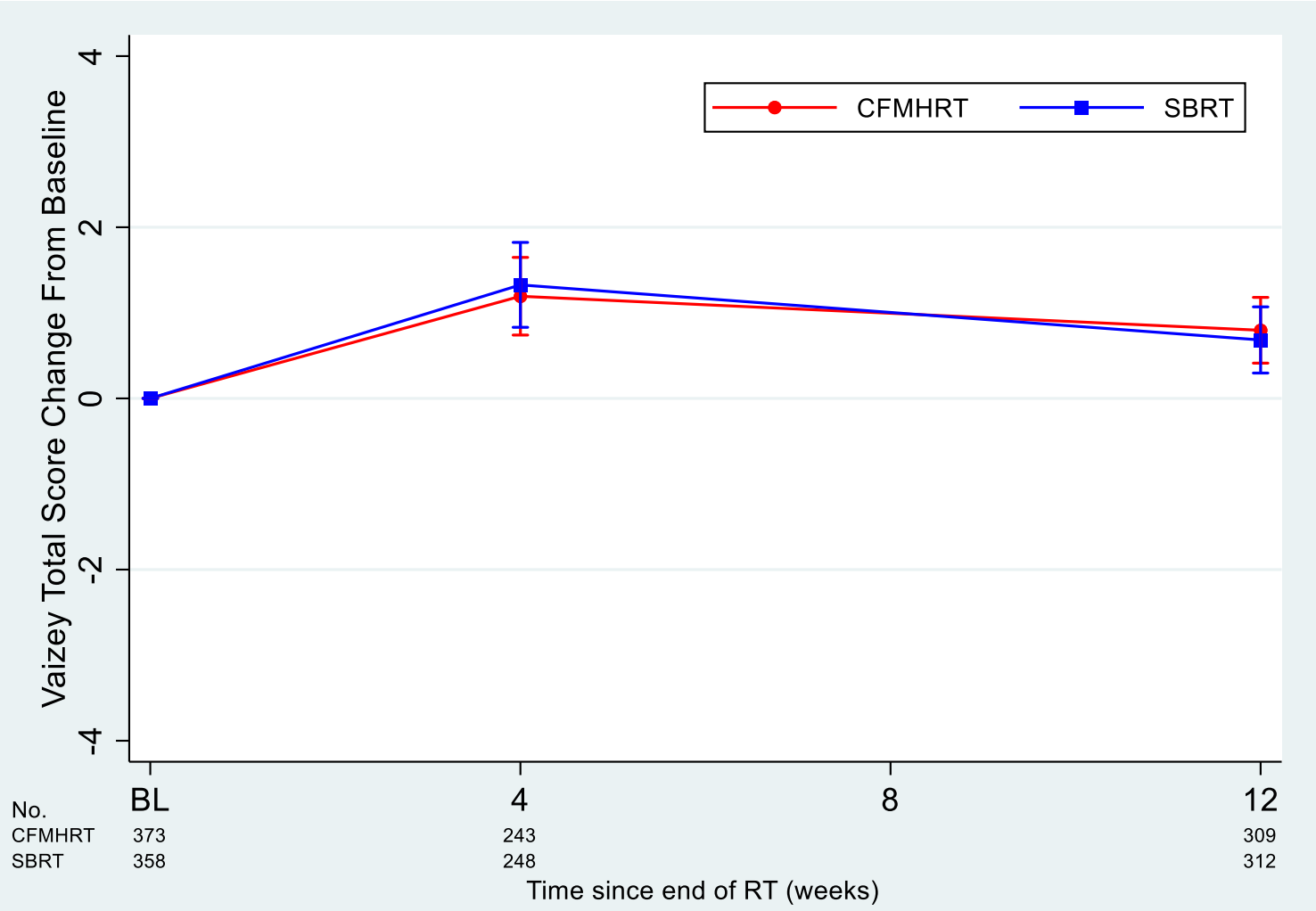


Figure 40. Change from Baseline Vaizey Total Scores

Changes from baseline Vaizey total scores, by time, for CFMHRT and SBRT. Patients included at any timepoint if both baseline and relevant timepoint score available. Note that the time period between baseline scoring and week 4 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). Error bars = 95% CIs. Week 0 is the baseline toxicity score taken before start of RT. N.B. Higher score for Vaizey is worse: 0 = perfect continence; maximum score = 24 = totally incontinent.

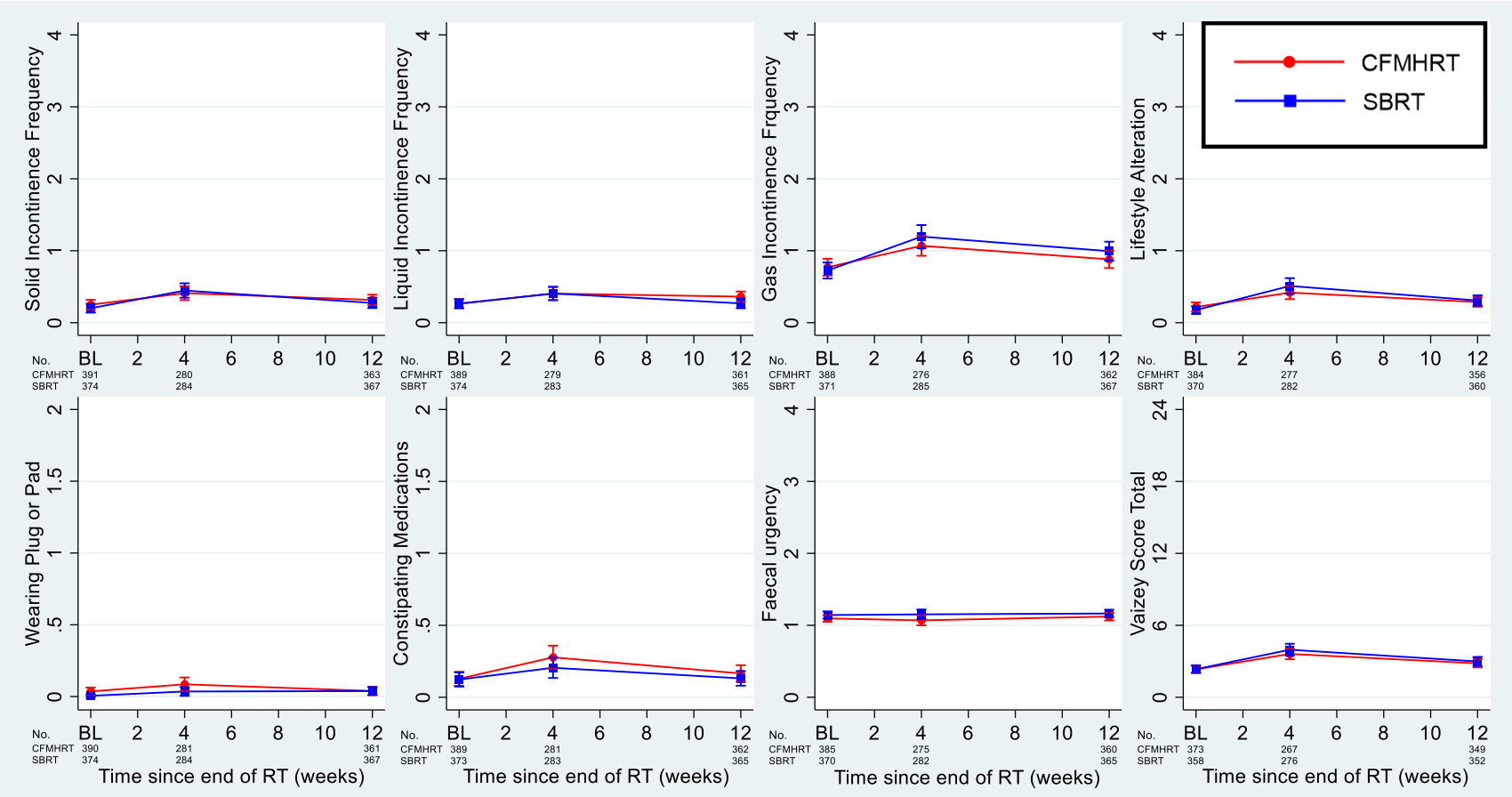


Figure 41. Vaizey Question Scores: Baseline to Week 12 Post-RT

Averages for Vaizey individual question scores and total, at baseline and week 12 post-RT, for CFMHRT and SBRT. Note that the time period between baseline scoring and week 12 post-RT follow-up is variable, since the total time of RT delivery varied (SBRT in 1 or 2 weeks; CFMHRT in 4 or 7.8 weeks). 95% CIs for each point mean estimate are displayed. Week 0 is the baseline toxicity score taken before start of RT. N.B. Higher score for Vaizey is worse: 0 = perfect continence; maximum score, 24 = totally incontinent.

5.6 Discussion

5.6.1 Overview of the Results

This analysis of the PRO acute toxicity data from PACE-B is in broad agreement with the main CRO analysis, with no statistically significant differences seen between CFMHRT and SBRT. This is reassuring, particularly in the context of detailed examination of every question from every scale, with none showing large differences between the two arms.

A few examined endpoints rest between the pre-specified penalised significance level ($p<0.001$) and typical $p=0.05$. Given the somewhat subjective nature of p-value penalisation, these will briefly be discussed, in case such signals might be hypothesis generating as points of interest in future SBRT studies, such as PACE-C.

CFMHRT had better EPIC-26 worst bowel score than SBRT (91.7 vs 87.5 $p=0.024$) although this is in context of better baseline scores. The difference was not seen on the worst, exceeding baseline, comparison. CFMHRT scored better than SBRT in worst (93.75 vs 90, $p=0.019$) and worst-minus-baseline (0 vs -5 , $p=0.02$) hormonal EPIC-26 scores, which can be seen to be driven by low energy in **Figure 35**. CFMHRT had lower (better) scores than SBRT for worst IPSS total-minus-baseline (5 vs 6, $p=0.035$) and IPSS quality of life week-12 (1 vs 2, $p=0.044$), however these are small and unlikely clinically meaningful differences.

5.6.2 Study Results in Context

The ProtecT trial PROs demonstrated that one of the biggest issues following surgery was urinary incontinence – worst immediately after surgery, with some recovery over time [21]. SBRT has the potential to deliver treatment in a total time similar to typical 2-hour radical prostatectomy (5 fractions x 20 minutes). Based on the data presented here, we can reassure patients that early onset incontinence remains an issue confined to surgical management of PCa. Similarly, the acute reduction in EPIC sexual function seen with EBRT (and surgery) in ProtecT appears to be less of an issue in PACE-B; likely the consequence of avoiding ADT in these patients. Further direct comparative data from the PACE-A randomisation will hopefully be available once accrual is completed.

As discussed in the introduction, the HYPO-RT-PC trial showed worse bowel and urinary problems at the end of treatment for ultrahypofractionated RT compared to conventional RT [95]. The absence of PRO differences in PACE-B compared to HYPO-RT-PC may be driven by a number of factors. Firstly, the first PRO assessment (IPSS) occurred at 2 weeks post-RT for PACE-B, compared to end-of-treatment for HYPO-RT-PC. Arguably some of the toxicity may have settled by 2 weeks, although it would be perfectly logical to argue the counterfactual; that not all toxicity will have been expressed on the final day of treatment. Larger margins for ultra hypofractionated patients in HYPO-RT-PC (7mm) compared to SBRT patients in PACE-B (4-5mm) may also be contributory; pooling of the OAR dosimetric data and relation to side effects would be of future interest.

The CHHiP trial collected PRO data at baseline and 10 weeks after the start of RT [143]. Reported graphically, (see *figure 3* in the QoL paper [143]) it showed highly overlapping confidence intervals at 10 weeks for the three dose regimens, across bowel function (UCLA-PCI), bowel summary (EPIC), urinary function (UCLA-PCI), urinary summary (EPIC), sexual function (UCLA-PCI) and sexual summary (EPIC). This supports the findings here, that hypofractionation and abbreviation of dose schedules does not necessarily lead to worse acute QoL metrics.

The PROFIT trial stated that PRO outcomes would be comprehensively reviewed in a separate paper [3], although to my knowledge this is yet to be published. RTOG-0415 did not collected QoL instruments until 6 months post-RT, with the paper not specifically analysing this timepoint [205]. Similarly, the HYPRO trial did not assess PRO instruments until 6 months post-RT, again likely missing the acute toxicity window [206].

5.6.3 PROs for Toxicity Assessment in Future Trials

For the assaying of trial toxicity, PROs offer a few potential advantages over CROs. Firstly, their collection can be done without requiring relatively expensive clinician time. This is particularly attractive for diseases such as PCa, where high incidence rates and large trials would suggest efforts are directed towards shortening clinician time taken up in patient follow-up assessment. Secondly, they could be administered electronically and

remotely via electronic PRO platforms, although such practice is not yet common in clinical trials.

5.6.4 Strengths of This Study

As per the CRO analysis, the key principle strength of this PRO analysis that it is the first in depth analysis of comparative SBRT acute toxicity from a large RCT. Unlike the CRO assessment schedule, each arm was assessed under exactly the same schedule, avoiding potential bias from additional assessment points. The assessments were made at a higher frequency during the relevant acute toxicity period than other trials; 4 PRO assessment timepoints ≤90 days post-RT in PACE-B vs 1 in HYPO-RT-PC. This permits closer examination of the possible dynamic treatment response seen with acute toxicity. Furthermore, the analysis has been made at a fine level of detail (individual question level), allowing confidence that there are not unexpected signals buried by conglomeration into PRO instrument total scores.

5.6.5 Limitations of this Study

Per **Chapter 4**, a key limitation is that this data reports acute toxicity only; implementation of this regimen should ideally await late toxicity and relapse data. As mentioned above, the PRO scales were collected (at the earliest) two weeks post-RT, so it is possible that the peak toxicity has been missed by these assessments. Examination of the RTOG toxicity by individual schedules (Recall **Figure 22**) suggests that toxicity is certainly dropping by 4 weeks, which was the first timepoint for EPIC-26 and Vaizey assessment.

Consideration in future trials might be made to optimise acute toxicity data recording: perhaps 1-week post-RT for all scales to assess peak toxicity, then at week-12 to assess recovery.

A further limitation is the limited comparability of scale information. PACE-B and HYPO-RT-PC have used differing PRO scales for bowel and bladder toxicity (recall **Table 6**), making direct trial comparison or meta-analysis difficult. Future studies might focus on coordinating the scales chosen by similar large phase III trials to aid future meta-analysis.

The lack of blinding in the PACE-B trial is a shortcoming; common to the vast majority of EBRT trials. It is well documented that patients experience a greater placebo effect through more intensive interventions; for example an intravenous placebo versus a tablet placebo [207]. Certain endpoints should be robust to the lack of blinding; most obviously OS. However, by both clinician and patient being aware of the therapy received, there exists opportunity for CROs or PROs to differentially report toxicity based on subjective opinions of the treatment that was delivered. Potentially more serious toxicity might be anticipated from the more “intensive” SBRT, although this would be difficult to definitively prove.

5.7 Conclusions

This is the most detailed analysis of randomised PRO acute toxicity data comparing an ultrahypofractionated regimen to standard EBRT treatments. It is reassuring that this PRO data supports the findings of the CRO analysis, with no significant excess in toxicity seen. Implementation of SBRT in UK practice would ideally await the primary outcome data from PACE-B. Should the HYPO-RT-PC regimen be permitted before then, this data will be reassuring for patients regarding the tolerability of ultra-hypofractionation.

Chapter 6. The Effect of Rectal Contour and Dosimetric Definitions on Toxicity Prediction

6.1 Conference Proceedings Related To Chapter

Dosimetric impact of central OAR review on rectal and bladder constraint attainment in PACE-B trial

Douglas Brand, Sarah Brueningk, Katie Fernandez, Sarah Gulliford, Emma Hall., Olivia Naismith, Alison Tree, Nicholas van As, On behalf PACE TMG. [Poster presentation at ESTRO 2020, Poster PH-0602] https://www.postersessiononline.eu/173580348_eu/congresos/ESTRO2020/aula/-PH_602_ESTRO2020.pdf

6.2 Background

6.2.1 Effect of Inter-observer Rectal Contouring Difference

A contemporary issue with RT planning is the difficulty in defining the contours of target and OAR tissues on the planning imaging[33]. Human demarcated contours are subject to personal interpretative biases, with inter-observer contouring variability noted across many tumour sites and OARs [208]. It has been shown that rectal inter-observer contouring differences can lead to differences of 10-20% in important relative-volume DVH parameters (e.g. V50Gy in 2Gy/fraction) [209]. Studies have applied inter-observer

[33] Commonly CT imaging, but PET-CT and MRI increasingly used in the planning process.

changes in dosimetry to pre-fitted rectal NTCP models; results have varied, showing either small [210] or clinically significant [211] differences in predicted toxicity. However, to my knowledge, no group has reported on the effect of contouring difference on direct toxicity prediction.

6.2.2 The Core Issue with Rectal Contouring Variation

The importance of obtaining a consistent rectal contour can be understood from a treatment planning perspective. Treatment planning systems optimise dose based upon weighted dosimetric objectives. Some objectives are tumour related (e.g. median dose), while others are OAR related; in the case of the rectum these would typically be dose-volume constraints on a relative volume cumulative DVH. For example, V60Gy<50% would instruct the optimiser to ensure that less than 50% of the rectum is receiving a dose of ≥60Gy. Dose-constraints are derived based on their ability to classify late rectal toxicity, from trials such as RT-01 [175].

Unfortunately, such constraints may be invalid if the rectal contour does not match those used to derive the dose-constraint. This can be demonstrated with a simple example (**Figure 42**). An inappropriately small rectal contour artificially increases V60Gy, which in a multi-objective treatment planning optimisation might result in underdosing to the tumour depending on choice of tumour vs OAR priorities. The converse can also occur, if too much sigmoid was included, then V60Gy would be artificially low, resulting in higher dose to the rectum than desired. It should be noted that absolute-

volume DVH metrics (i.e. volume receiving 60Gy in cubic centimetres), would be unaffected, making them an interesting alternative.

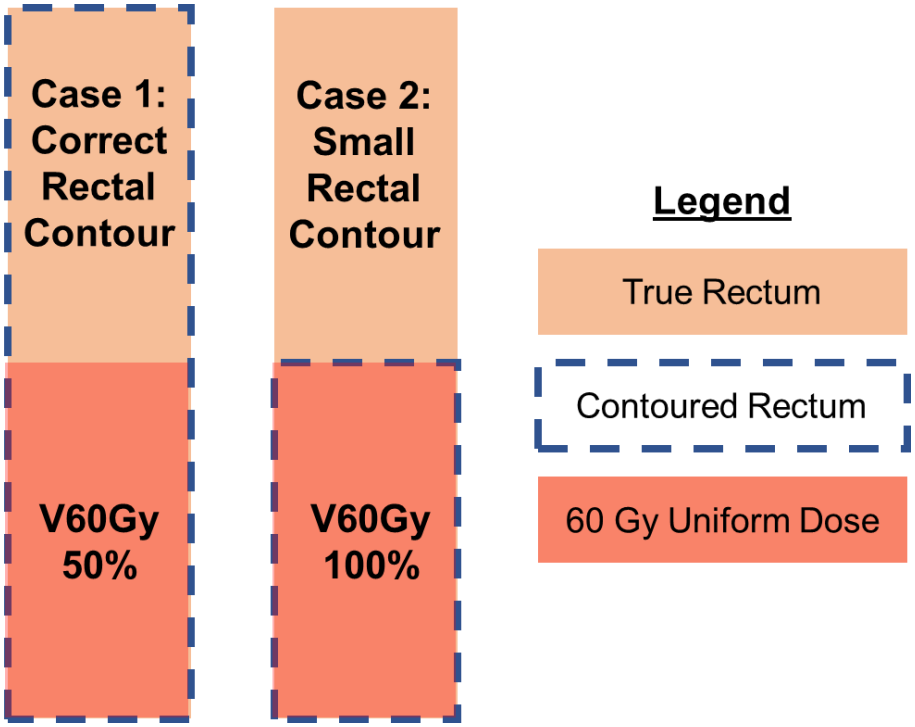


Figure 42. Explanatory Example of Incorrect Rectal Contouring Influencing Relative Dose Volume Constraint Attainment

The rectal volume is contoured correctly in Case 1, but has an incorrectly low superior border in Case 2. In this exaggerated example, it can be seen that the lower rectal volume of Case 2 results in an increased apparent relative V60Gy. If absolute volume V60Gy were used, then this would be the same in both cases.

6.2.3 Possible Remedies for Contouring Variation

6.2.3.1 Reducing Inter-Observer Variation

As discussed in the literature review (**2.5.2.3.1**), the best definition of the rectum, as a contoured structure, is a matter of debate. Major radiotherapy organisations, such as American Society for Radiation Oncology (ASTRO), still actively push for standardisation of normal tissue contouring [212]. This

appears desirable, since if all contouring was identical[34], then it would eliminate the effect shown in **Figure 42**.

6.2.3.1.1 Inter-centre Differences in Rectal Contouring Within Trials

Therefore, for dosimetric analyses of RT trial data, reducing inter-centre heterogeneity in contouring appears desirable. In **Chapter 3**, the rectums examined were all centrally checked for protocol accuracy, before the DVHs were recalculated for dosimetric work. This OAR contour central review, prior to dosimetric research, was also undertaken for MRC-RT01, another large UK EBRT study [175]. Such recontouring, in theory, reduces variability that might introduce noise into toxicity prediction models.

International practice regarding a central contour review process shows heterogeneity. The TROG-RADAR study has undertaken rectal dosimetric analyses on 754 patients, where two observers undertook central review of the rectal contour to ensure protocol compliance[35] [213]. However, for prior studies of the rectal α/β ratio, neither Marzi (single institution), nor Tucker (multi-institutional) recontoured the rectum [62,88]. In other hypofractionation trials, recontouring was not undertaken in dosimetric analysis of late rectal effects for RTOG-0415 [168], nor HYPRO [214]. No such dosimetric analysis of late rectal toxicity has yet been produced for the PROFIT trial [3].

[34] Both in the studies used to design constraints and in routine practice.
[35] Defined as “Outer rectal wall from the level of the ischial tuberosities until when the rectum turns horizontally to the sigmoid colon”

Given heterogeneity of UK/Australian and international practice, it ought to be considered whether such OAR recontouring continues to be worthwhile. Firstly, the time effort involved in such recontouring for CHHiP was very high. Assuming ≈15 minutes per patient[36], across 2400 patients, ≈600 hours was required. Secondly, dosimetric findings using recontoured OARs may have poorer external generalisability, since centres do not adhere to exact protocol definition in routine practice. Thirdly, this recontouring was initiated when 3D planning was still relatively new for clinicians: 1998-2001 was the recruitment window for the RT01 trial [215]. Further community experience may have reduced rectal contouring heterogeneity in more recent prostate EBRT trials.

6.2.3.2 Absolute Dose Volume Constraints

Rather than focussing on improving the superior and inferior rectal delineation, an alternative approach would be through use of absolute (cc) rather than relative (%) volumes for DVH analysis. This is attractive, since people can continue to outline the rectum as they see fit, with the most variable region (superior border) unlikely to influence the absolute DVH at significant levels (e.g. V30Gy) due to distance from the high dose region.

The concept of using rectal absolute dose constraints is not a new one, having been proposed since at least 2002 [216], although their absence from the seminal 2011 QUANTEC analysis may have hampered interest [154]. Multiple groups have attempted to fit both relative and absolute constraints

[36] Includes time for opening case, making new structure, contouring, DVH calculation and saving new DICOM.

for the rectum [216–221]. Improved performance of absolute volumes for rectal bleeding prediction was reported, although with low group size and event rates (n=112, 8% toxicity) [216]. Other larger studies (n=331 [219] & n=366 [221]) have found relative dose-volume metrics to be the better predictors. Finally it has been suggested that both relative and absolute constraints could be used, although this has not gained traction and there are obvious issues of multicollinearity [217]. The heterogeneity of data in this area suggests that a much larger study, with lower chance of type-II error, might be of benefit.

6.2.3.3 Rectal Truncation to Improve Consistency

Another option to reduce the effect of inter-observer differences in rectal contouring would be to simplify the rectal definition. Since most variability has been seen at the distal borders (mainly superior [222]), truncation of the rectum superiorly and inferiorly based upon the PTV position might reduce this variation. Such PTV-based truncation is attractive, since it would not involve any retraining for practitioners: contour as usual, then truncate algorithmically based on PTV. This would lessen the possibility of the relative volume effect in **Figure 42**, since the most variable regions would have been truncated. It is worth noting that current relative volume dose-constraints for the whole rectum [175] would need modifying, were truncation to be adopted.

Groups have investigated PTV based truncation for both acute [223] and late toxicity [12], without finding improved toxicity prediction. Indeed, the late toxicity study by Nitsche and colleagues [12] found no relationship between

dose and toxicity parameters, however, the sample size of 23 suggests a high possibility of type II statistical error. Reinvestigation of such an approach in a much larger cohort is therefore of interest.

6.2.4 Summary of Study Purpose

Using data from the CHHiP trial, this chapter aims to examine the effect of protocol-based central recontouring on contour morphology and dose volume metrics. Additionally, a recontoured selection of patients from PACE-B will be examined, to see if a more recent trial shows better protocol adherence for rectal contouring. For the CHHiP patients, where late rectal toxicity data is available, I will examine whether recontoured dose-volume metrics improved discrimination of subsequent late rectal toxicity. The aim will be to establish whether the considerable cost and effort of central rectal contour review (editing) is worthwhile for the development of predictive toxicity models. To my knowledge no group has looked at the influence of central contour review on actual toxicity outcomes.

Secondary interests are whether the use of absolute volumes or PTV-based rectal definitions improve toxicity prediction compared to standard-of-care whole rectum, relative volumes. The CHHiP trial cohort provides a uniquely large sample size, with prospectively collected toxicity data, making it ideal to examine these questions.

6.2.5 Hypotheses

1. Protocol-based recontouring (editing) of the rectum will alter rectal contour morphology.

2. Protocol-based recontouring (editing) of the rectum will alter dose-volume metrics.

3. There will be similar differences in contour and DVH parameters, between original and edited rectums, for PACE-B versus CHHiP.

4. In CHHiP patients, recontoured (edited) DVH metrics, expressed as relative (%) volumes, will be better than original for discriminating patients subsequently expressing late rectal toxicity.

5. In CHHiP patients and whole rectum, the use of absolute (cc) rather than relative (%) volumes DVH measures may improve toxicity prediction.

6. In CHHiP patients using relative-volumes, the use of PTV-based rectum truncation (e.g. PTV ± 2 cm) may improve toxicity prediction over the use of whole rectum.

6.3 Methods

6.3.1 Eligibility for CHHiP Trial Patients

The CHHiP study has been described (**Chapter 2 & Chapter 3**). Patients were eligible for inclusion in this sub-study if: i) a full protocol regimen was delivered (i.e. 74Gy/37Fr or 60Gy/20Fr or 57Gy/19Fr); ii) radiotherapy treatment plans adequate for the recontouring process (i.e. CT, dose, structures) were available. Patients were excluded if the original and edited rectal contour were not both available for analysis. Patients without toxicity data were excluded from toxicity analyses.

6.3.2 The CHHiP Trial Recontouring Method

To recall, CHHiP defined a solid structure rectum "*from the anus (usually at the level of the ischial tuberosities or 1cm below the lower margin of the PTV whichever is more inferior) to the recto-sigmoid junction"*. For PACE-B, the rectum was similarly defined as: "*a solid structure, including the lumen and rectal wall, extending from the anus to the rectosigmoid junction*."

The recontouring QA process for CHHiP rectums has already been described in detail in **Section 3.3.3**. For this chapter, I batch converted the DICOMs into Computational Environment for Radiotherapy Research (CERR) file format [224], a MATLAB-based radiotherapy planning system.

The recontouring QA process (**Section 3.3.3**) was commenced prior to the initiation of this doctoral project, for the purposes of other dosimetric

analyses [225]. Unfortunately, for 773 patients, the original rectum was not preserved as a structure within the recontoured DICOM. In eight cases, this was an irrevocable change to the original RT planning files received from the treating centre. For this thesis, the 765 patients with original RT planning DICOM files were identified. I converted these to CERR files. I then wrote a MATLAB script that searched these CERR files for the original rectum structure. These names were manually checked by me to ensure correct structure choice and the structures were scripted into the corresponding recontoured CERR files for each patient.

This created a disparity in the calculation of rectal DVHs for the original rectum between those where it was initially present (DVH calculated in VODCA) and those where it was added back in (DVH calculated in CERR). I therefore recalculated the edited and original rectal DVHs in CERR [224] for every patient to produce both relative and absolute cumulative DVHs.

6.3.3 Procedurally Generated Rectum Sizes

As discussed above, Nitsche *et al* noted many rectal definitions [12]. For this study, procedurally generated rectal sizes were created, which would be easily reproducible under normal contouring conditions. Both the original and edited rectum were copied and then truncated so that their inferior and superior slices were the same as the PTV (rectum PTV±0cm). This process was repeated with truncation at each X centimetre (1 to 5 inclusive) superior and inferior to the sup-inf extent of the PTV (rectum PTV±Xcm). For each of the new structures, the DVH was recalculated in CERR [224]. The choice of

truncations to take forward to toxicity analysis (0 and 2 cm) was made based on examination of morphology and dosimetric changes relative to whole rectum. This was performed on the planC files, generated by me using CERR, by scripts in MATLAB that I wrote and debugged in full.

6.3.4 PACE-B Trial Confirmatory Cohort

The PACE-B trial design has been described in **Chapter 4**. The confirmatory PACE-B trial cohort in this study was a non-random sample based upon central availability of the patient DICOMs at the time of the study. The sample size was determined by the limited time available to the observer.

All recontouring was performed by a single observer (KF), trained in the recontouring of rectums, overseen by a single clinician (myself). Prior to commencing the recontouring, a QA program of 10 cases was undertaken to ensure a second clinician (Alison Tree) approved of contouring by KF.

Again, the DVHs were recalculated in CERR for all patients, in order to achieve a standardised DVH calculation method across the study. Morphological and dosimetric analyses for the CHHiP patients were repeated for the PACE-B confirmatory cohort. PACE-B late toxicity data is not yet available.

6.3.5 Statistics

All statistics were performed by me personally.

6.3.5.1 Rectal Contour Morphology Comparisons

Morphology comparison was undertaken for edited versus original whole rectums, across a number of metrics. The cranio-caudal rectal length was calculated as Z-position of superior minus inferior slice). The difference between original vs edited rectal lengths were then calculated for the CHHiP and PACE-B trials, with inter-trial comparison by Mann-Whitney test, due to non-normality. This was repeated for each of the six different PTV ± X cm truncations.

Rectal volume was calculated as the sum of all volume bins in the absolute volume differential DVH for the whole rectum. Volume data was analysed by the same analysis steps as rectal length.

For each patient, the Dice similarity coefficient (DSC) was calculated for original vs edited whole rectum. This was repeated for all PTV ± X cm truncations. The DSC values were compared between the CHHiP trial and PACE-B by Mann-Whitney test, due to non-normality. For all contour morphology graphical representation was by Tukey box and whisker plots.

6.3.5.2 Rectal Dosimetric Comparisons

Cumulative DVHs, with both relative and absolute volumes, were calculated for every rectum (original, edited & all truncations). An EQD2 correction was applied, with an α/β = 3 Gy chosen (per results of **Chapter 3**). The dose-

levels of interest were chosen based on the 74 Gy in 37 fraction regimen of CHHiP, to provide a spread of data across the DVH. These were V30, V40, V50, V60, V70, V74 Gy. Dose levels for other fractionations were chosen to be EQD2 equivalent to these dose levels. The conversions, by dose-fractionation arm, for CHHiP and PACE-B are summarised in **Table 35**. Volumes (% for relative DVHs, cc for absolute DVHs) at each dose-level were extracted from each patient's rectal DVHs.

To give an example, the $EQD2_{\alpha/\beta=3Gy}$ for V30 Gy in the 74 Gy regimen is 22.9 Gy. Therefore, to calculate the 60 Gy dose level of interest, the total dose yielding $EQD2_{\alpha/\beta=3Gy}$ = 22.9 Gy over 20 fractions is found.

Table 35. EQD2 Corrected Rectal Dose-Levels of Interest with Corresponding Physical Doses by Regimen

The dose levels were chosen in the 74 Gy in 37 fraction regimen. These were then converted to EQD2. A dose level was then set in each other regimen to be equal in EQD2.

Dose Level From 74 Gy in 37 Fraction Regimen	EQD2-Corrected Dose Level α/β = 3 Gy	CHHiP			PACE		
		74 Gy	60 Gy	57 Gy	78 Gy	62 Gy	36.25 Gy
		Dose (Gy)	Dose (Gy)	Dose (Gy)	Dose (Gy)	Dose (Gy)	Dose (Gy)
V30	22.9	30	26.4	26.1	30.3	26.4	17.6
V40	32.6	40	34.5	34.1	40.4	34.5	22.0
V50	43.5	50	42.5	41.8	50.6	42.5	26.3
V60	55.5	60	50.3	49.5	60.8	50.3	30.5
V65	61.8	65	54.2	53.3	65.9	54.2	32.5
V70	68.5	70	58.0	57.0	71.0	58.0	34.6
V74	74.0	74	61.1	60.1	75.1	61.1	36.2

Broad graphical summarisation of the DVH data was presented for each trial arm, by mean averaging the volumes at each dose level. This was done for both relative (%) and absolute (cc) volumes.

The volumes at each DVH dose-level was plotted for original versus edited rectal contours, both for relative and absolute volumes. This was done for both CHHiP and PACE trials. At each dose-level, a comparison of the volumes for original vs edited was made by Wilcoxon Signed-Rank test. Plots were also made of the volume differences at each dose-level between edited and original contours, for both CHHiP and PACE trials. This was done for both relative and absolute volumes.

6.3.5.3 Toxicity Comparison

6.3.5.3.1 Endpoints

Seven different unified rectal toxicities were used: stool frequency G1+, stool frequency G2+, bleeding G1+, bleeding G2+, proctitis G1+, proctitis G2+, sphincter control G1+ and stricture/ulcer G1+. The amalgamation of these toxicity endpoints from RTOG, LENTSOM and RMH scales were described in **Section 3.3.4**. Pain was not included, due to very low apparent dose-response relationship in Chapter 3. Toxicity endpoint frequencies were described in **Table 9**.

6.3.5.3.2 Original versus Edited Contours: Individual Dose Bins

Since most dose constraints at present are based on individual dose bins, an initial analysis is presented by individual dose bin. For each toxicity endpoint

and dose bin, the area under curve (AUC) for (e.g.) original V30 and edited V30 was calculated, with statistical comparison of the AUC values by the DeLong method [226]. The 95% confidence intervals for the AUC were computed by 2000 bootstrap resamples (using 2.5^{th} – 97.5^{th} centile method), stratified for toxicity status. To examine whether editing induced changes in the optimal receiver operating characteristic (ROC) curve cut-point, the Youden statistic (maximum of sensivity+specificity-1) was calculated. In some cases, this followed a bimodal bootstrap distribution, so the mean Youden statistics of the bootstrap data is presented. Again 95% CI bootstrap interval for the Youden statistic was calculated.

6.3.5.3.3 Original versus Edited Contours: Logistic Model

The combined toxicity prediction of the dose data was estimated by logistic model. Separately for original and edited contours, all seven relative volume dose bins were fitted simultaneously to a logistic regression model against each toxicity endpoint. Of course, it would be possible to undertake variable selection or feature reduction (e.g. principle components), however the aim here is not to establish a final model for practice, but to compare the predictive abilities of the original and edited contours. Some overfitting may occur, but this will be balanced between the two methods. AUCs of the whole logistic model were compared between edited and original contours by DeLong method. This process was repeated for 2000 bootstraps (stratified by toxicity) in order to provide AUC 95% confidence intervals and estimates of test performance. Estimates of test performance for sensitivity, specificity,

positive predictive value (PPV) and negative predictive value (NPV) were generated by the 632 method [178], seen previously in Chapter 3. This was preferred over 632+, due to quicker calculation and very low chance of near perfect prediction. Estimates are termed for $sensitivity_{632}$, $specificity_{632}$, NPV_{632} and PPV_{632}. Youden cut-points were selected, but are not reported in tabular form, since units of such a figure derived from logistic regression are not of relevance to this study.

6.3.5.3.4 Relative Volumes versus Absolute Volumes

Logistic models were fitted for each toxicity endpoint with the corresponding seven dose bin values for both relative and absolute DVH curves. This was bootstrapped and analysed in the same manner as described above for the original vs edited logistic model analysis. Comparison between the AUC for relative volumes and absolute volumes was by DeLong method.

6.3.5.3.5 Truncated Rectum Analyses

The same approach was taken for truncated analyses, except having two comparisons: whole rectum versus respectively PTV±0cm and PTV±2cm.

6.3.5.4 Significance Levels

Due to multiple testing, corrections were applied for the interpretation of significance levels for p-values. The most important significance tests are the eight logistic model comparisons for each of the four key hypotheses (original vs edited; absolute vs relative volumes; whole rectum vs PTV±0cm; whole rectum ± 2 cm). These are interpreted as significant at the 0.001 level,

by Bonferroni correction, further penalised slightly to allow some significance for exploratory tests. All other endpoints are considered exploratory in nature (morphology, dosimetry, individual dose bin analyses) and will be interpreted at the 0.0001 significance level.

6.3.6 Software

Recontouring was undertaken in VODCA (version 5.4.1, MSS Medical Software Solutions GmbH, Hagendorn, Switzerland). Conversion into CERR files used MATLAB (multiple versions up to 2020a, Mathworks, MA, USA) and the CERR suite (GitHub commits up to 10/06/20) [224]. Processing of CERR files to copy structures, truncate structures and add DVHs used custom made MATLAB scripts with use of core CERR commands. Quality assurance of the new CERR DVHs versus the VODCA DVHs was done by automated MATLAB extraction of both DVHs to Excel (version 2019, Microsoft, WA, USA) where any outliers were identified for manual examination in CERR study viewer; corrections being made where necessary. DSC metrics were calculated using CERR core code. All DVH, DSC, size and DVH information for structures was extracted from CERR into tabular format using custom made MATLAB scripts. This data was imported into Stata (version 16-MP 4-core, Statacorp, USA) and R (version 4.0.2) for bootstrapping, logistic model fitting and analysis of ROC curves. Graphs were produced in MATLAB and Stata using custom made scripts.

6.4 Results

6.4.1 Patient Population

From the CHHiP trial, 2350/3216 of randomised patients were included in one or more parts of this analysis. **Figure 43** [*overleaf*] is a CONSORT-style flowchart indicating reasons for exclusion of all randomised patients. These patients are similar to the trial as a whole (**Table 36**).

Table 36.Treatment Arm and Disease Characteristic of CHHiP Patients

Characteristic	This Study		Whole CHHiP Trial	
	No.	%	No.	%
Age	**69 years**	**44-85 (range)**	**69 years**	**44-85 (range)**
Arm (Intent to Treat)				
57 Gy / 19 Fractions	802	34%	1077	33%
60 Gy / 20 Fractions	791	34%	1074	33%
74 Gy / 37 Fractions	757	32%	1065	33%
Regimen Received				
57 Gy / 19 Fractions	799	34%	N/A	N/A
60 Gy / 20 Fractions	789	34%	N/A	N/A
74 Gy / 37 Fractions	762	32%	N/A	N/A
NCCN Risk Group				
Low risk	325	14%	484	15%
Intermediate risk	1,750	75%	2347	73%
High risk	275	12%	385	12%
Gleason score				
≤6	792	34%	1122	35%
7	1,488	63%	1995	62%
8	70	3%	99	3%
Clinical T Stage				
T1	887	38%	1170	36%
T2	1,275	54%	1766	55%
T3	186	8%	277	9%
Missing	2	<1%	3	<1%
Pre-Treatment PSA				
<10 ng/mL	1,141	49%	1567	49%
10-20 ng/mL	1,072	46%	1415	44%
≥20 ng/mL	137	6%	208	6%
Missing	0	0%	26	<1%
Total	**2350**	**100%**	**3216**	**100%**

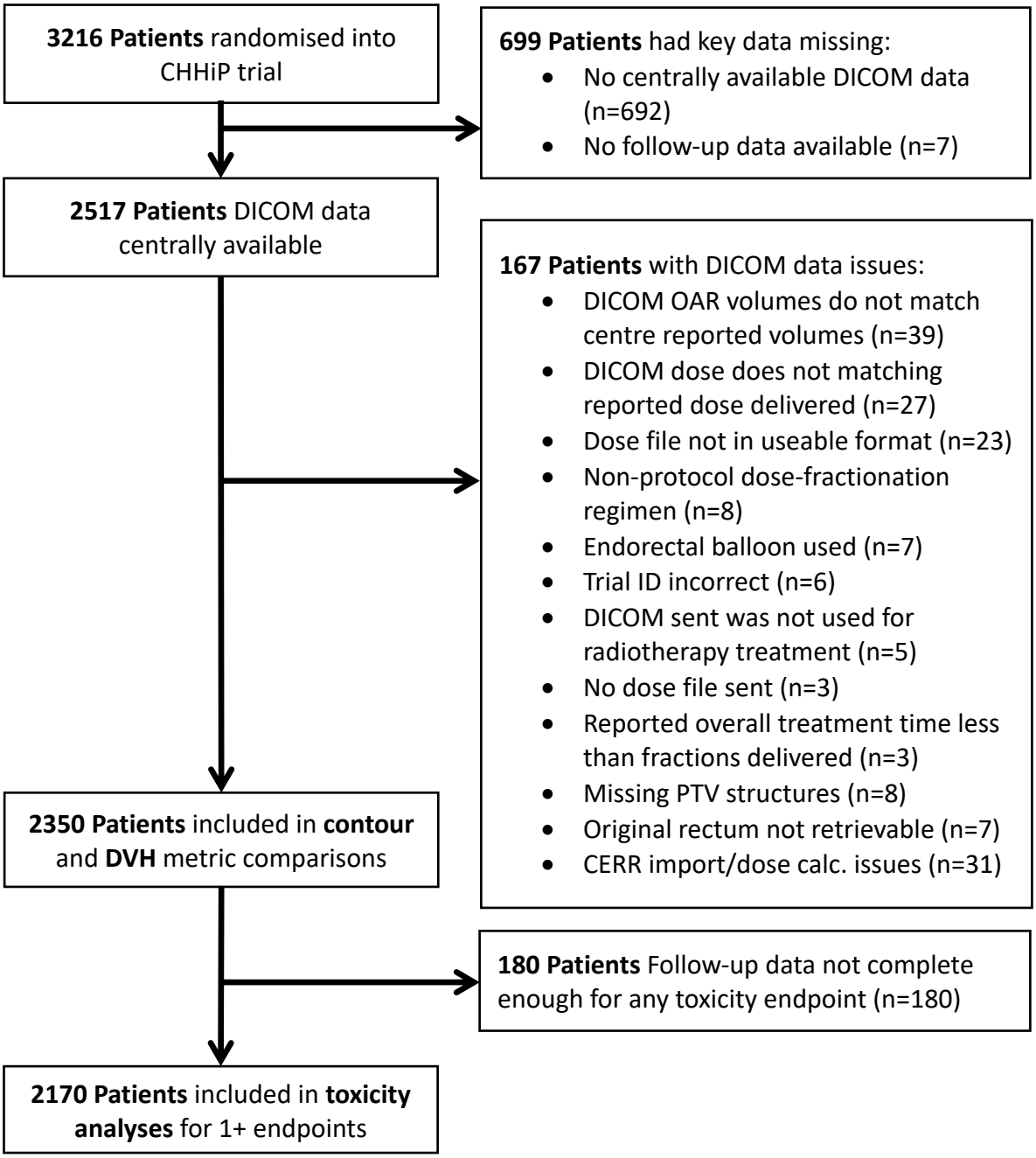


Figure 43. CONSORT-Style Flow Diagram for CHHiP Trial Patients

Reasons for exclusion from this substudy for any patient randomised into the CHHiP

From the PACE-B trial, 304/874 randomised patients were included, with a similar CONSORT-style diagram in **Figure 44**. This represents 36% (304/848) of patients who received radiotherapy within the PACE-B trial. No patient is included in toxicity analyses, since late toxicity is not yet reported.

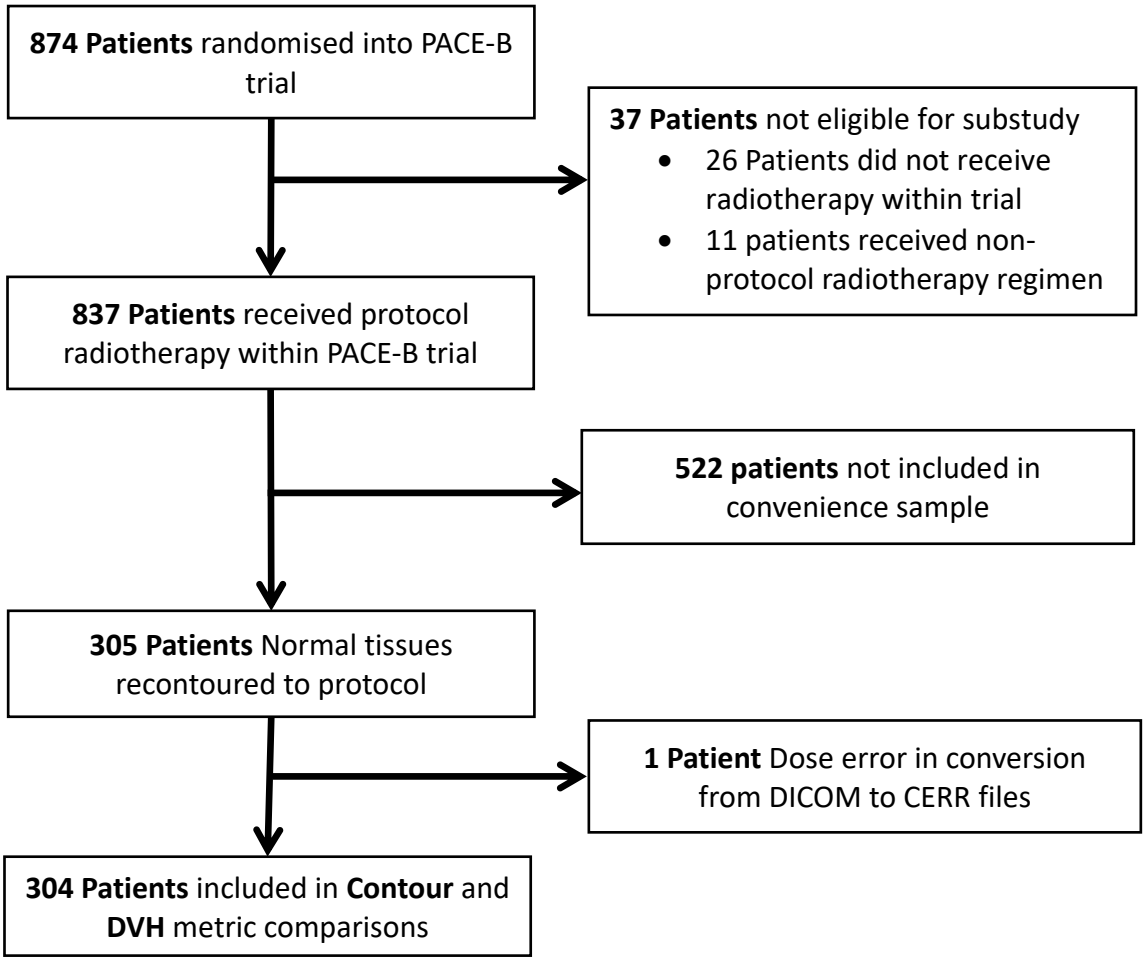


Figure 44. CONSORT-Style Flow Diagram for PACE-B Trial Patients
Showing reasons for exclusion from this substudy for any patient who was randomised into the PACE-B Trial.

The treatment and disease characteristic for the PACE-B patients included in this substudy are compared to those who received protocol radiotherapy (thus being eligible for the convenience sample) in **Table 37**. The substudy has more 62 Gy in 20 fraction patients, at the expense of 78 Gy in 39 fractions. Within SBRT patients, CK delivery is strongly underrepresented.

Table 37. Treatment Arm and Disease Characteristic of PACE-B Patients

Treatment and disease characteristics compared between this study and the CHHiP trial as a whole.

Characteristic	**This Substudy**		**PACE-B Trial Protocol Regimen Patients**	
	No.	**%**	**No.**	**%**
Age	**69 years**	**44-85 (range)**	**69 years**	**44-85 (range)**
Regimen Received				
78 Gy in 39 Fractions	14	5%	125	15%
62 Gy in 20 Fractions	142	47%	299	36%
36.25 Gy in 5 Fractions	148	49%	413	49%
NCCN Risk Group				
Low risk	20	7%	68	8%
Intermediate risk	284	93%	769	92%
Gleason score				
3+3	39	13%	145	17%
3+4	265	87%	692	83%
Clinical T Stage				
T1c	87	29%	154	18%
T2a	82	27%	233	28%
T2b	41	14%	137	16%
T2c	94	31%	313	37%
Pre-Treatment PSA				
<10 ng/mL	213	70%	577	69%
10-20 ng/mL	91	30%	260	31%
Delivery Technique				
Step and Shoot IMRT	19	6%	105	13%
VMAT	277	91%	559	67%
Tomotherapy	0	0%	4	1%
Cyberknife	8	3%	169	20%
Total	**304**	**100%**	**837**	**100%**

The cause of this imbalance of 62 Gy patients in the convenience sample for this substudy is likely related to trial recruitment period. Since efforts to retrieve DICOMs were not commenced until near completion of the trial, centres appear to have returned the more recent DICOMs first, which have been included in the convenience sample (**Figure 45**). This would disproportionately favour 62 Gy over 78 Gy patients, since protocol amendment only permitted 62 Gy from 24/03/2016.

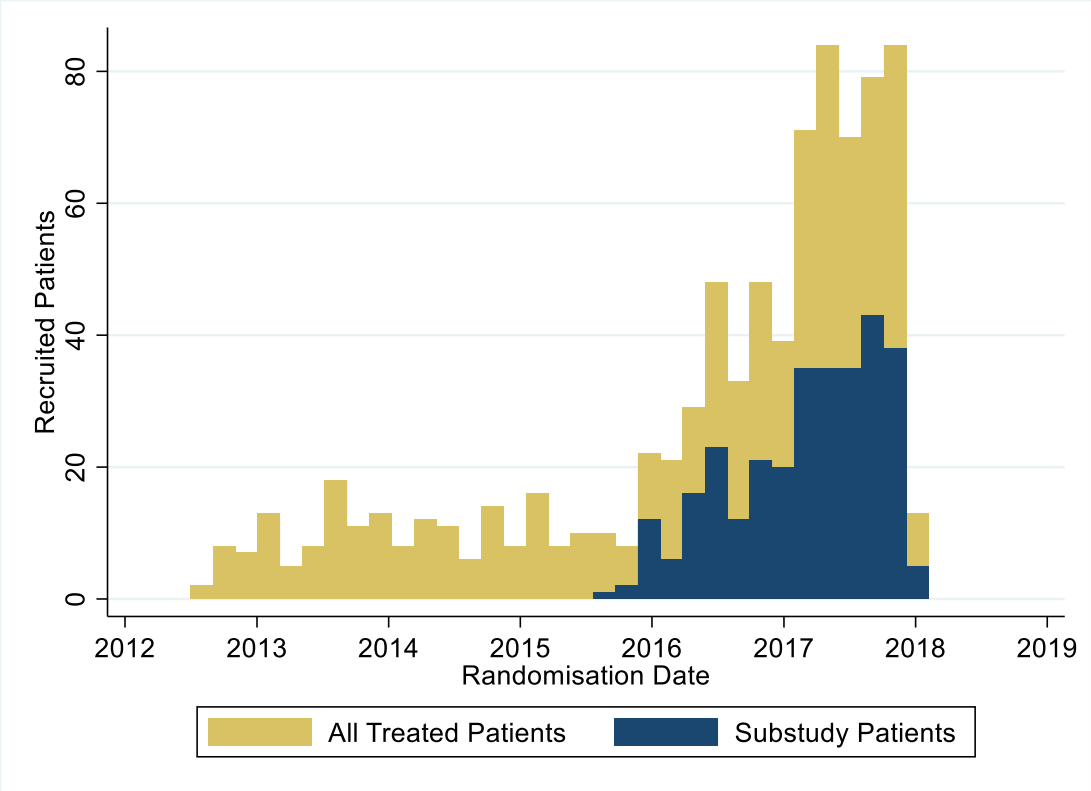


Figure 45. PACE-B Substudy Patients By Recruitment Time

A plot comparing the recruitment times of all PACE-B patients receiving protocol radiotherapy (yellow) to those included in this substudy (blue). It is clear that the convenience sample has favoured more recent patients.

The cause of the relative lack of CK patients is partly explained by the inclusion of more recent patients, since all SBRT was CK until amendment 24/10/2014. Secondly, it is partly explained by some centres being better at returning DICOMs than others (**Figure 46**). Without identifying centres, CK treatment facilities were slow at returning DICOMs.

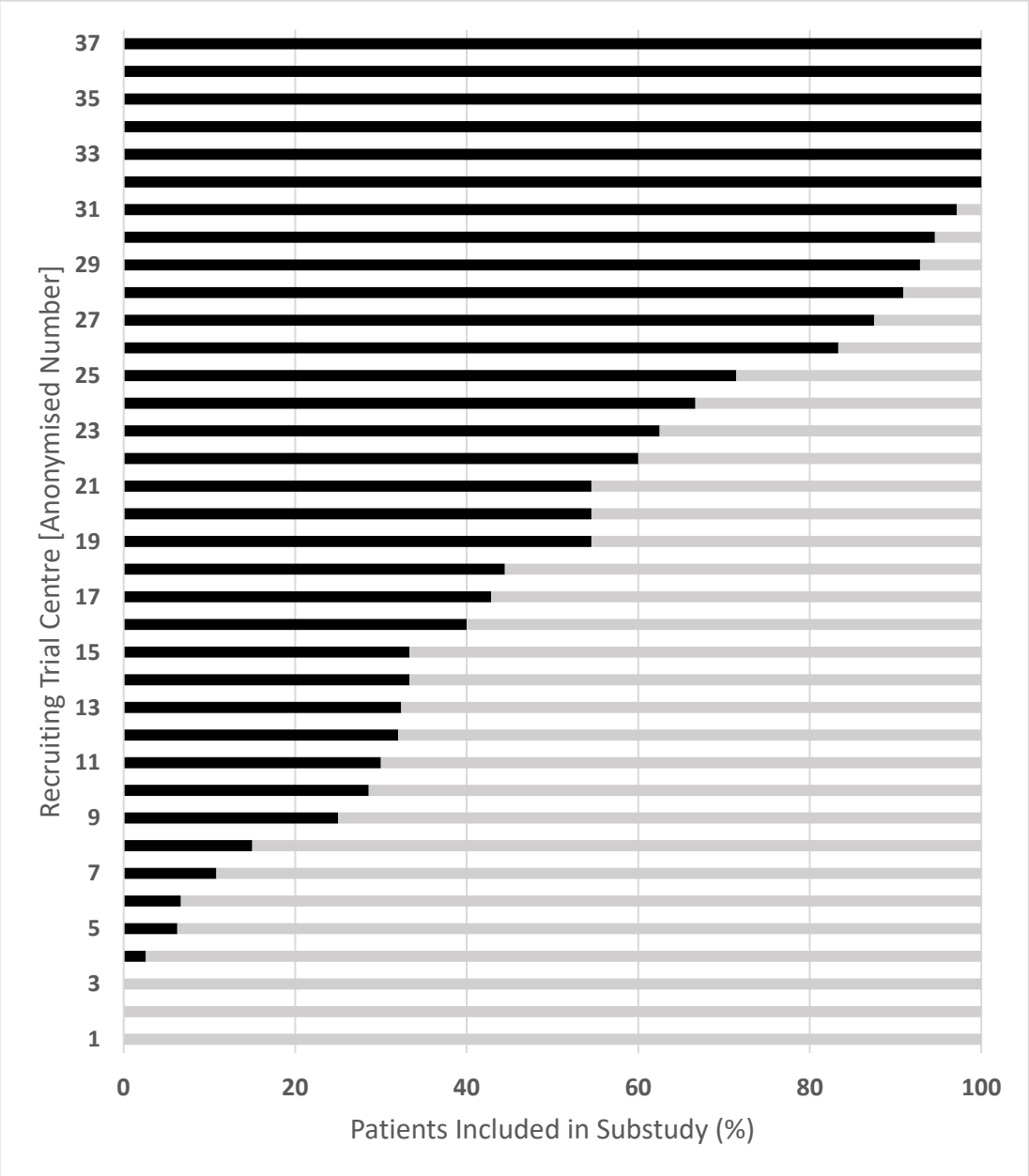


Figure 46. Percentage of Patients in Convenience Sample by Treating Centre
By centre, percentage of patients, who received a protocol regimen, that have been included in this substudy. Centres anonymised, so recruitment numbers not given.

6.4.2 Contour Metrics

6.4.2.1 Length Analyses

Rectal length analyses are shown in **Table 38**. It can be seen that the median whole rectum length contoured was similar for both trial groups (~9-10 cm), across original and edited. The edited vs original whole rectum lengths were longer for the CHHiP group (median 0.5 cm), but shorter for the PACE-B group (median 0.2 cm); a significant difference between trial groups. By graphical comparison (**Figure 47**), the generally less positive edited vs original differences in CHHiP versus PACE-B are apparent.

In both trial groups, the difference between edited and original rectal lengths decreased with the proximity of rectal truncation to the PTV from an inflexion point at around PTV±2cm or less. Edited minus original length differences for CHHiP versus PACE-B were significantly larger for all truncations examined.

Table 38. Rectal Lengths for Original and Edited Contours, By Truncation

Looking at how edited rectum length differed to original rectum length for the CHHiP and PACE-B trial groups. Mann-Whitney Tests compare the original:edited rectum differences between the two trial groups.

Novel abbreviations: Edit. = Edited; Med. = median; Orig. = Original.

Rectum Definition	CHHiP Trial (n=2350)						PACE-B Trial (n=304)						Mann-Whitney Test of Differences: CHHiP v PACE-B
	Original		Edited		Difference (Edit - Orig)		Original		Edited		Difference (Edit - Orig)		
	Med. (cm)	IQR (cm)	Med. (cm)	IQR (cm)	Med. (cm)	IQR (cm)	Med. (cm)	IQR (cm)	Med. (cm)	IQR (cm)	Med. (cm)	IQR (cm)	
Whole Rectum	9.3	8.3 - 10.5	10	9 - 11	0.5	0 - 1.5	9.9	8.9 - 10.8	9.6	8.6 - 10.6	-0.2	-1 - 0.6	**<0.0001**
PTV ± 5 cm	9.3	8.3 - 10.5	10	9 - 11	0.5	0 - 1.5	9.9	8.9 - 10.8	9.6	8.6 - 10.6	-0.2	-1 - 0.6	**<0.0001**
PTV ± 4 cm	9.3	8.3 - 10.5	10	9 - 10.8	0.5	0 - 1.5	9.9	8.9 - 10.8	9.6	8.6 - 10.6	-0.2	-1 - 0.6	**<0.0001**
PTV ± 3 cm	9.3	8.3 - 10.5	10	9 - 10.8	0.5	0 - 1.5	9.8	8.9 - 10.5	9.5	8.6 - 10.4	-0.2	-0.9 - 0.5	**<0.0001**
PTV ± 2 cm	9	8.1 - 10	9.5	8.8 - 10.3	0.3	0 – 1.0	9.1	8.4 - 9.8	9	8.2 - 9.6	0	-0.8 - 0.3	**<0.0001**
PTV ± 1 cm	8.3	7.5 - 9	8.5	7.8 - 9	0	0 - 0.5	7.8	7.2 - 8.3	7.7	7.1 - 8.2	0	-0.3 - 0	**<0.0001**
PTV ± 0 cm	6.9	6 - 7.5	6.9	6 - 7.5	0	0 - 0	6.0	5.5 - 6.6	6	5.4 - 6.6	0	0 - 0	**<0.0001**

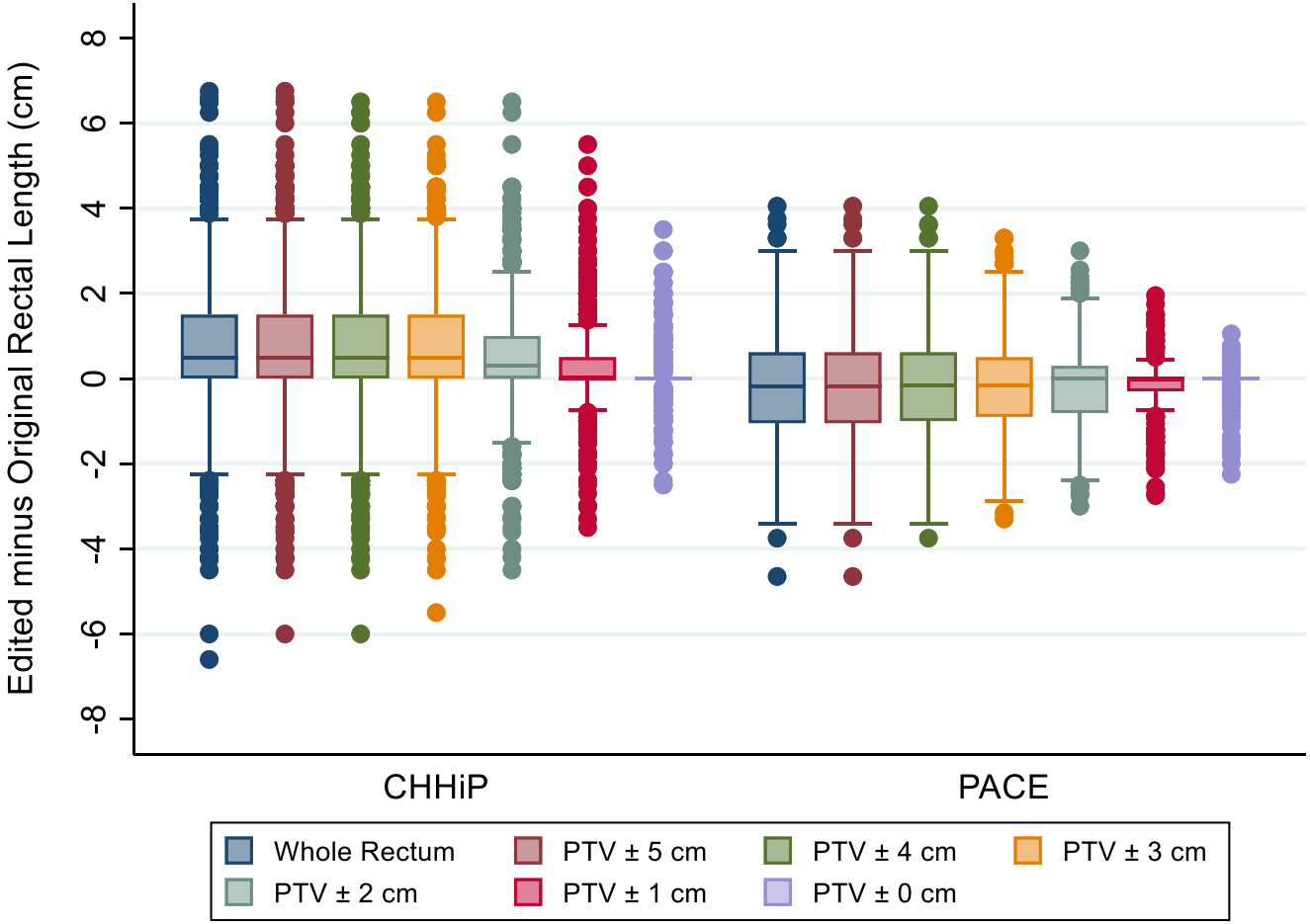


Figure 47. Rectal Length Differences (Edit. Minus Orig.): CHHiP vs PACE-B

Boxplots of edited minus original rectal lengths, by rectal truncation and trial: CHHiP (n=2350) and PACE-B (n=304).

6.4.2.2 Volume Analyses

Rectal volume data for both CHHiP and PACE-B is presented in **Table 39**. It is interesting to note that despite similar rectal lengths across the two trials, the rectal volumes for original and edited contours are lower for PACE-B. It should be noted that CHHiP protocol recommended the rectum "*ideally be empty of both faeces and flatus*", with enemas "*permissible*". PACE-B recommended the routine use of enemas. The whole rectal edited volumes were smaller than original volumes, for both CHHiP (median -0.7cm^3) and PACE-B (median -5 cm^3).

Differences in volume unsurprisingly reduced as truncation occurred closer to the PTV. For whole rectum and all truncation levels, the edited minus original differences were significantly more negative for PACE-B than CHHiP.

Examining boxplots of edited minus original volumes (**Figure 48**), fewer large changes in volume occurred for PACE-B patients, with none where the edited contour was much larger than original (e.g. +50cm^3).

Table 39. Rectal Volumes for Original and Edited Contours, By Truncation

Rectum Definition	CHHiP Trial (n=2350)						PACE-B Trial (n=304)						Mann-Whitney Test of Differences: CHHiP v PACE-B
	Original		Edited		Difference (Edit - Orig)		Original		Edited		Difference (Edit - Orig)		
	Med. (cm^3)	IQR (cm^3)	Med. (cm^3)	IQR (cm^3)	Med. (cm^3)	IQR (cm^3)	Med. (cm^3)	IQR (cm^3)	Med. (cm^3)	IQR (cm^3)	Med. (cm^3)	IQR (cm^3)	
Whole Rectum	66	52.6 - 86.3	64.3	51 - 84	-0.7	-7.9 - 2.7	59.7	49.6 - 71.1	52.8	44.4 - 63.2	-5	-11.5 - -0.9	**<0.0001**
PTV ± 5 cm	66	52.6 - 86.2	64.3	51 - 84	-0.7	-7.8 - 2.7	59.7	49.6 - 71	52.8	44.4 - 63.2	-5	-11.5 - -0.9	**<0.0001**
PTV ± 4 cm	66	52.6 - 86.2	64.2	51 - 83.9	-0.7	-7.8 - 2.7	59.5	49.5 - 70.9	52.8	44.4 - 63.1	-5	-11.4 - -0.8	**<0.0001**
PTV ± 3 cm	65.9	52.5 - 85.7	63.8	50.9 - 83.4	-0.6	-7.5 - 2.6	58.3	49 - 70.6	52.3	44 - 62.1	-5	-10.9 - -0.7	**<0.0001**
PTV ± 2 cm	64.8	51.5 - 84.3	62.8	50.1 - 81.4	-0.4	-6.8 - 2.1	55.4	46.9 - 67.6	50.4	42.5 - 59.8	-4.5	-9.1 - -0.9	**<0.0001**
PTV ± 1 cm	61.3	48.8 - 78.8	58.9	47 - 77	-0.4	-6.2 – 1.0	49	41.2 - 59.3	45.5	38.1 - 53.9	-3.1	-6.7 - -0.6	**<0.0001**
PTV ± 0 cm	53.1	41.9 - 69	50.9	40.4 - 66.6	-0.2	-4.8 - 0.5	39.8	33 - 49.1	37.8	30.7 - 45.5	-1.6	-3.5 - -0.3	**<0.0001**

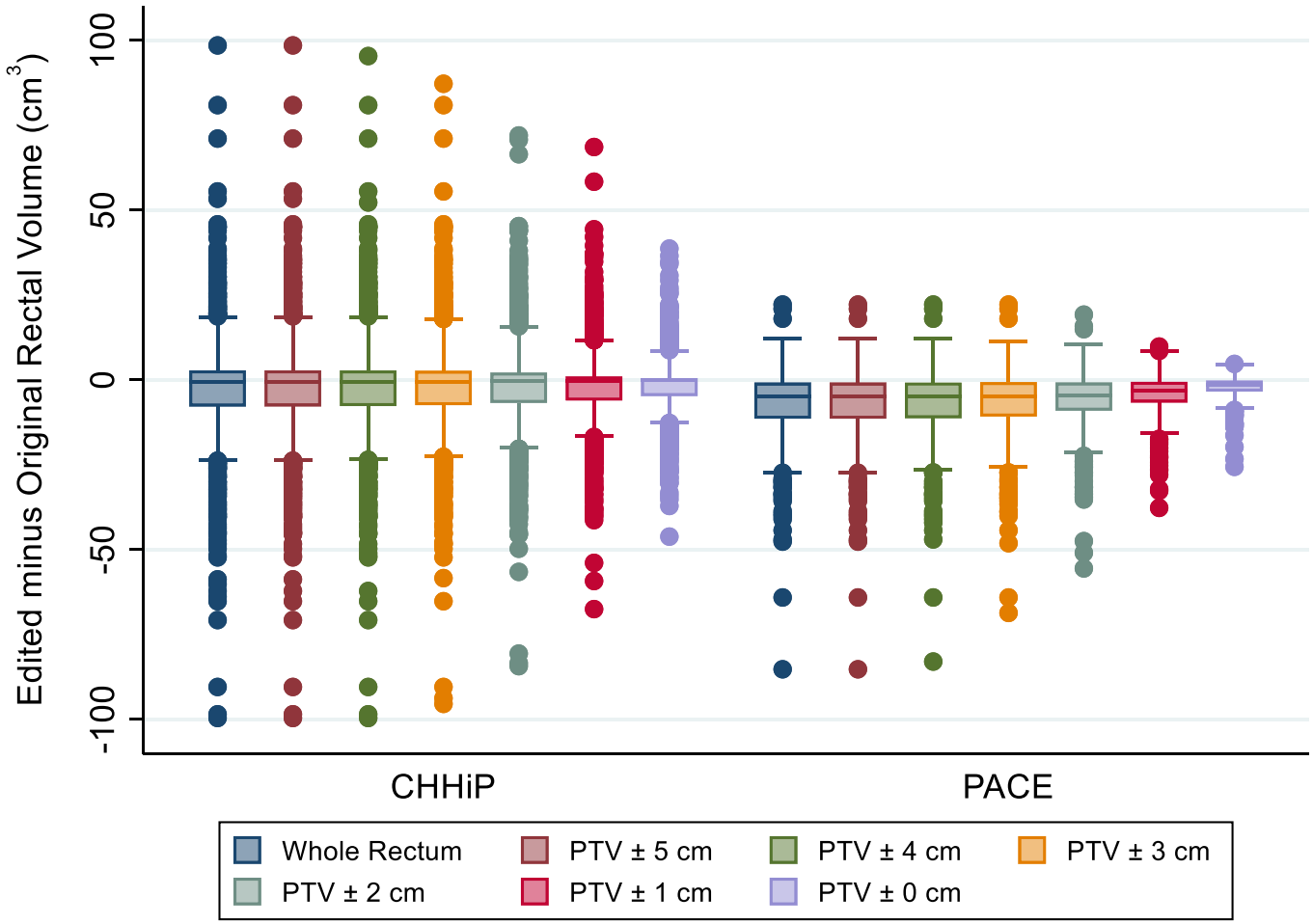


Figure 48. Rectal Volume Differences (Edit minus Orig): CHHiP vs PACE-B

Boxplots of rectal volume differences (edited minus original), by rectal truncation and by trial: CHHiP (n=2350) and PACE-B (n=304).

6.4.2.3 Dice Similarity Coefficient Analyses

Boxplots of the DSC values are compared for CHHiP versus PACE-B in **Figure 49**. High DSC values for the whole rectum are seen in both CHHiP (median 0.91) and PACE-B (median 0.90). It is apparent that fewer outlier low DSC values (below lower whisker) are seen in the PACE-B trial. Regarding truncated contours, median DSC values increase as truncation is performed nearer to the PTV. Mann-Whitney testing showed significantly higher DSC values for CHHiP versus PACE-B when examining whole rectum and truncations at PTV ± 5 cm & PTV ± 4 cm.

It is important to remember for all comparisons in this contouring section that the Mann-Whitney test is rank based and a statistically significant result does not relate to the size of numeric difference, nor indicate a clinically meaningful difference.

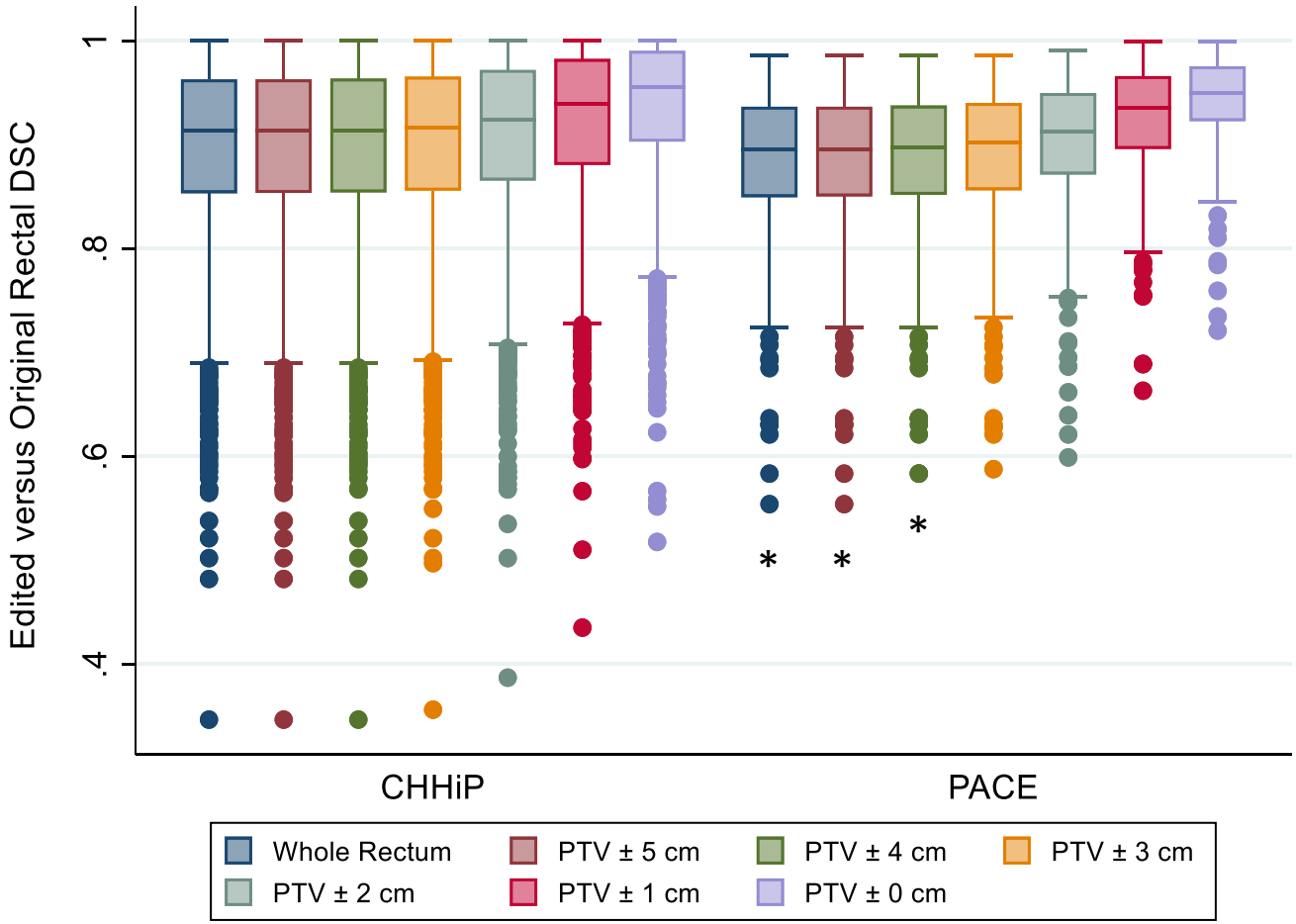


Figure 49. DSC Values for Original:Edited Rectal Contours: CHHiP vs PACE-B

Tukey boxplots of DSC values for original compared to edited rectal contours, by rectal truncation and by trial.

* = PACE-B significantly different to CHHiP, at given truncation, by Mann-Whitney test at a <0.0001 adjusted significance level.

6.4.3 Dosimetric Comparisons

6.4.3.1 Dose Volume Histograms by Trial

As an overview of EQD2-corrected rectal dosimetry in the CHHiP and PACE-B trial groups, the cumulative original rectal DVH dose-levels of interest are presented in **Figure 50** for relative volumes and **Figure 51** for absolute volumes.

It is readily apparent that the CHHiP arms are similar in morphology, with small EQD2-driven difference; not surprising given the planning methodology was identical.

For PACE-B, we again see the systematic difference between 78 and 62 Gy regimens, noting the 78 Gy to be the “hottest” regimen, with the highest volume at the maximum dose-level. It is notable that the SBRT arm has much lower EQD2-corrected dose-level volumes across most of the DVH, which may be interesting when late toxicity is reported in due course.

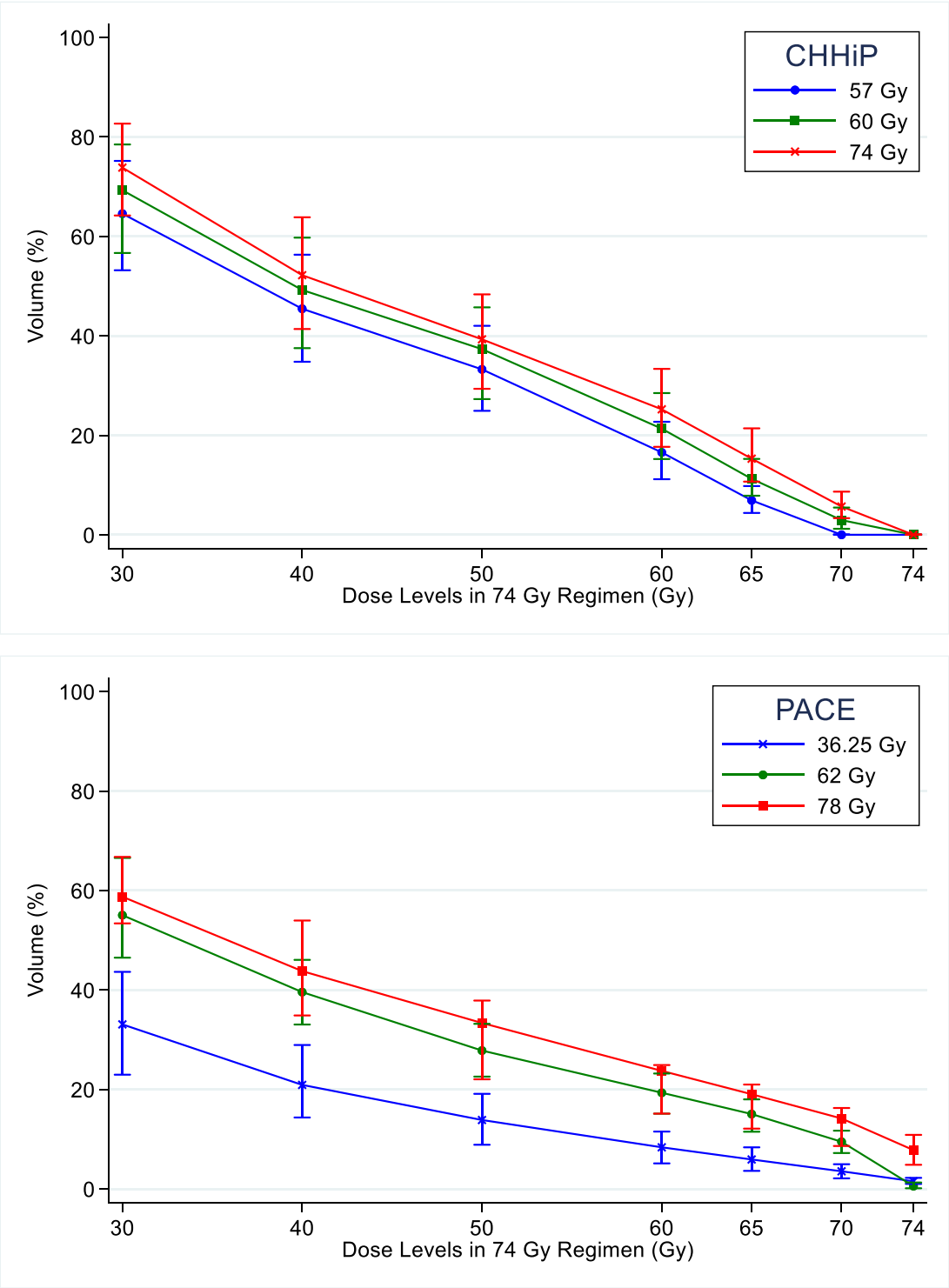


Figure 50. Original Rectal Relative Dose-Volume Bin Volumes, By Trial Arms for CHHiP and PACE-B

Medians for CHHiP (top panel) and PACE-B (bottom panel), by trial arm, for relative dose bins, EQD2 corrected to 74Gy arm constraints per methods. IQRs for each arm are shown.

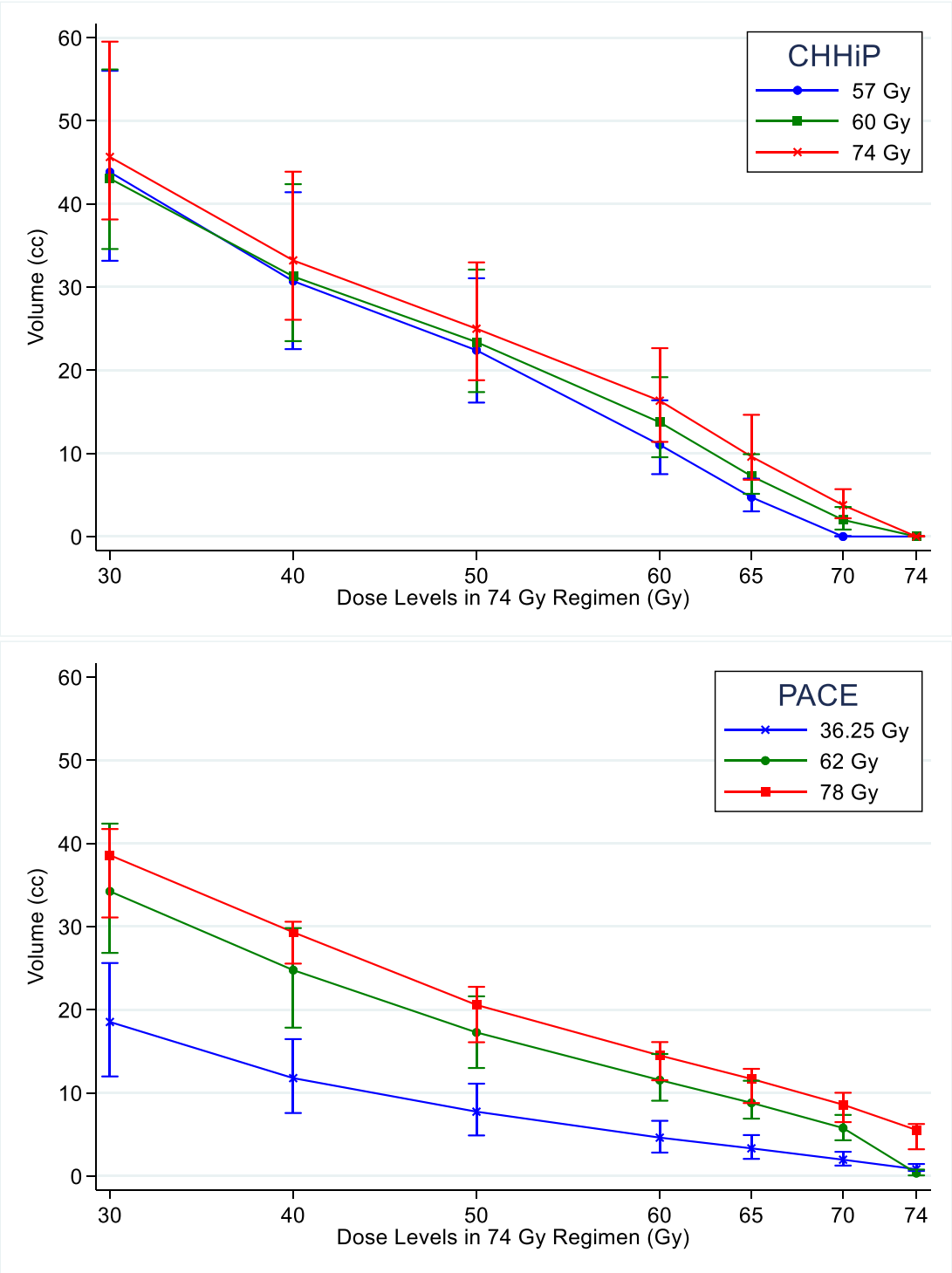


Figure 51. Original Rectal Absolute Dose-Volume Bin Volumes, By Trial Arms for CHHiP and PACE-B

Medians for CHHiP (top panel) and PACE-B (bottom panel), by trial arm, for absolute dose bins, EQD2 corrected to 74Gy arm constraints per methods. IQRs for each arm are shown.

6.4.3.2 Original vs Edited Contours: Relative Volumes at the Dose-Volume Constraint Levels of Interest

For the CHHiP trial, the median relative volumes at each dose constraint of interest are summarised in **Figure 52**. Comparing whole rectum original and edited contours, it can be seen that the median and IQR are very similar at each dose constraint of interest. For whole rectum, signed rank comparison of original vs edited bin volumes was significant at several dose bins examined (V30, V60, V65, V70). Plot differences between edited and original volumes did not substantively alter with truncation, although only V65 and V70 remained significantly different at all truncations examined.

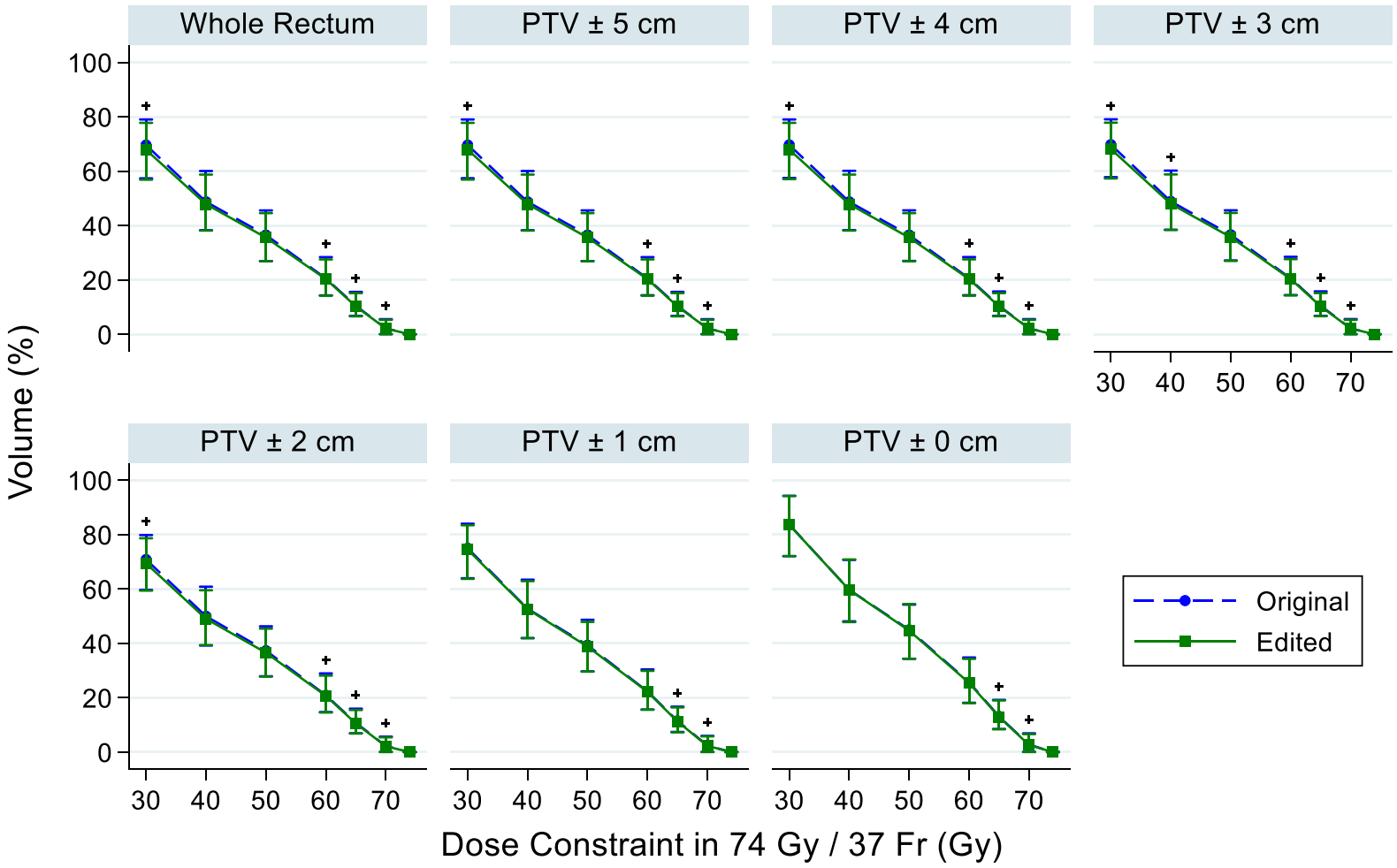


Figure 52. CHHiP Relative DVH Volumes: Original vs Edited, By Truncation

Median and IQRs for examined rectal dose bins, EQD2 corrected to 74Gy arm constraints per methods.
+ = edited versus original dose bin volume significantly different by Wilcoxon Signed rank (p<0.0001).

Similar relative volume plots are produced for the PACE-B trial in **Figure 53**. For the whole rectum, we again see the median volumes and IQRs are similar for original and edited rectal contours at each dose bin. Signed rank comparison of whole rectum original vs edited bin volumes was significant at all bins examined. Truncation of the rectum at 0cm abrogated these significant differences at higher dose bins, although original and edited plot morphology continue to appear similar.

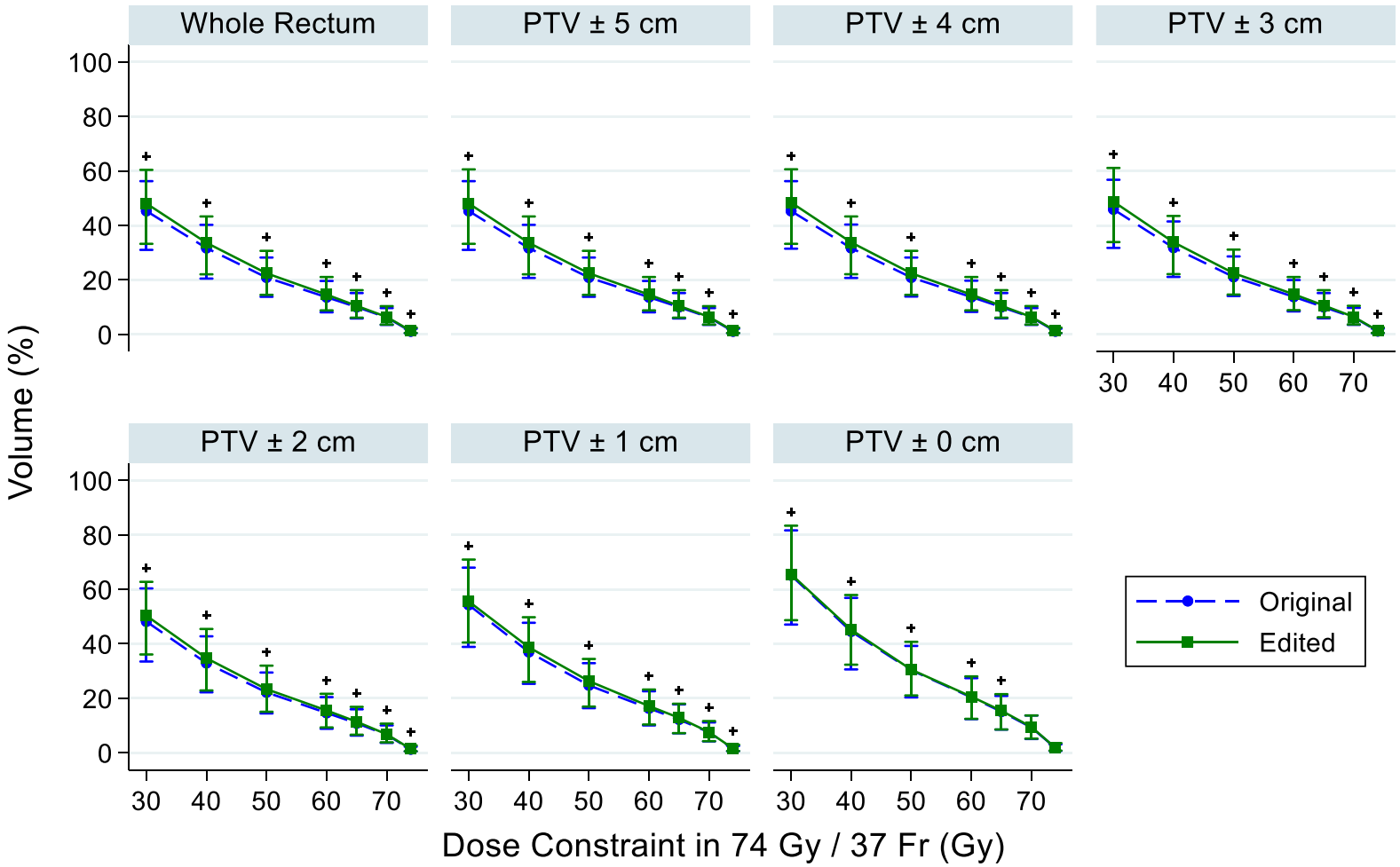


Figure 53. PACE-B Relative DVH Volumes: Original vs Edited, By Truncation

Median and IQRs for examined rectal dose bins, EQD2 corrected to CHHiP 74Gy arm constraints per methods.
+ = edited versus original dose bin volume significantly different by Wilcoxon Signed rank (p<0.0001).

6.4.3.3 Original Vs Edited Contours: Absolute Volumes at the Dose-Volume Constraint Levels of Interest

For the CHHiP trial, the median absolute volumes at each dose constraint of interest are summarised in **Figure 54**. For whole rectum, as was seen with relative volumes (**Figure 52**), there are small differences between edited and original medians and IQRs, which reduce at higher dose bins. Signed-rank comparison of original vs edited absolute volumes was significant at all dose-volume bins except 74 Gy.

These findings were similar across all rectal truncations. It should be appreciated that truncation makes very little difference to the absolute DVH volumes, implying that little rectum receives ≥30 Gy, outside of the PTV±0cm truncation.

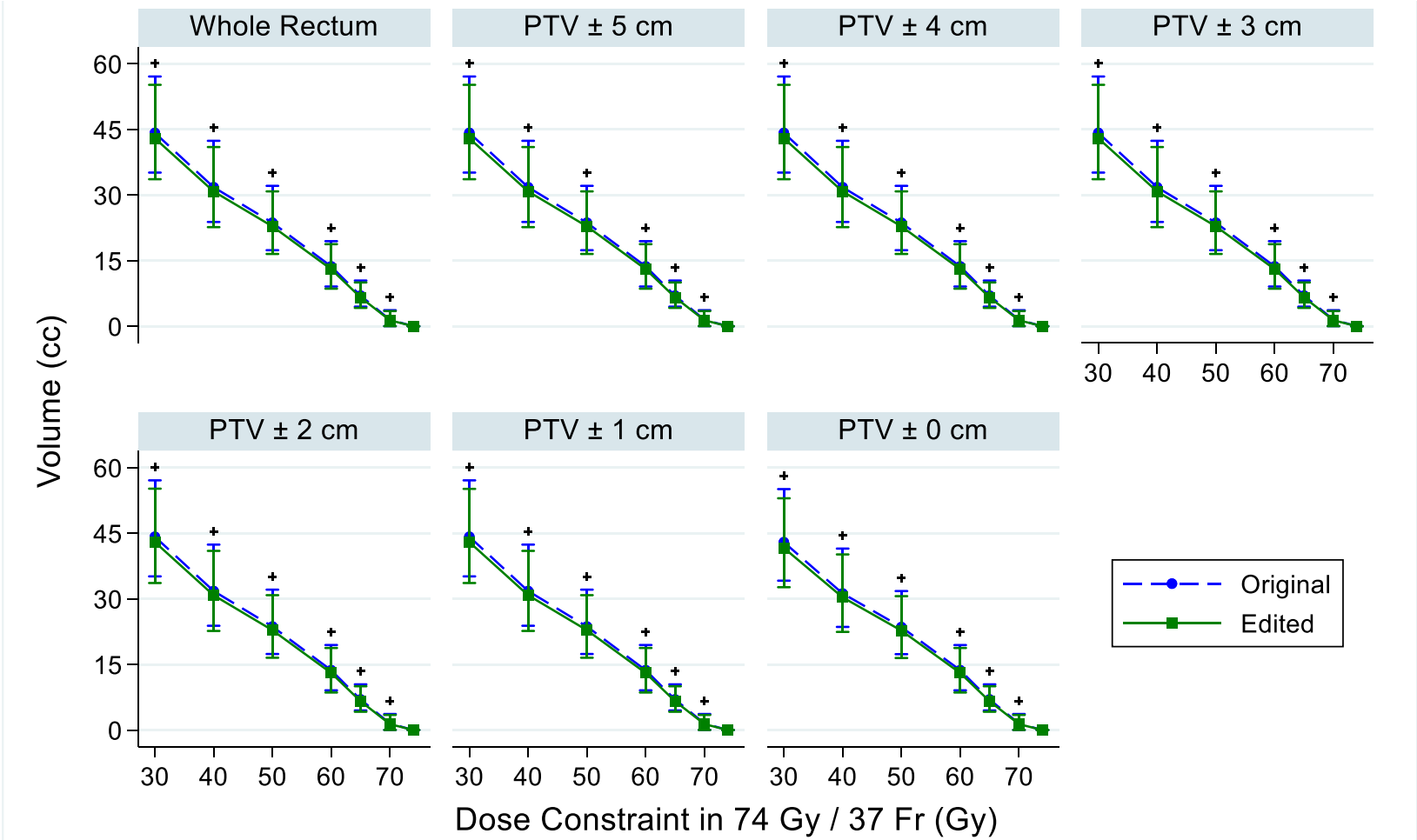


Figure 54. CHHiP Absolute DVH Volumes: Original vs Edited, By Truncation

Median and IQRs for examined rectal absolute dose bins, EQD2 corrected to CHHiP 74Gy arm constraints per methods.
+ = edited versus original dose bin volume significantly different by Wilcoxon Signed rank (p<0.0001).

A similar absolute dose-volume plot is produced for the PACE-B trial in **Figure 56** [*overleaf*]. Again, for whole rectum, the differences between edited and original contour medians and IQRs are small. These are significant for V30Gy – V60Gy, across all truncations, except V60Gy in PTV±0cm.

Again, the extremely similar plots across all truncations implies that little rectum receives ≥30 Gy, outside of PTV±0cm. This, perhaps surprising, finding is explained by a fairly typical example in **Figure 55**.

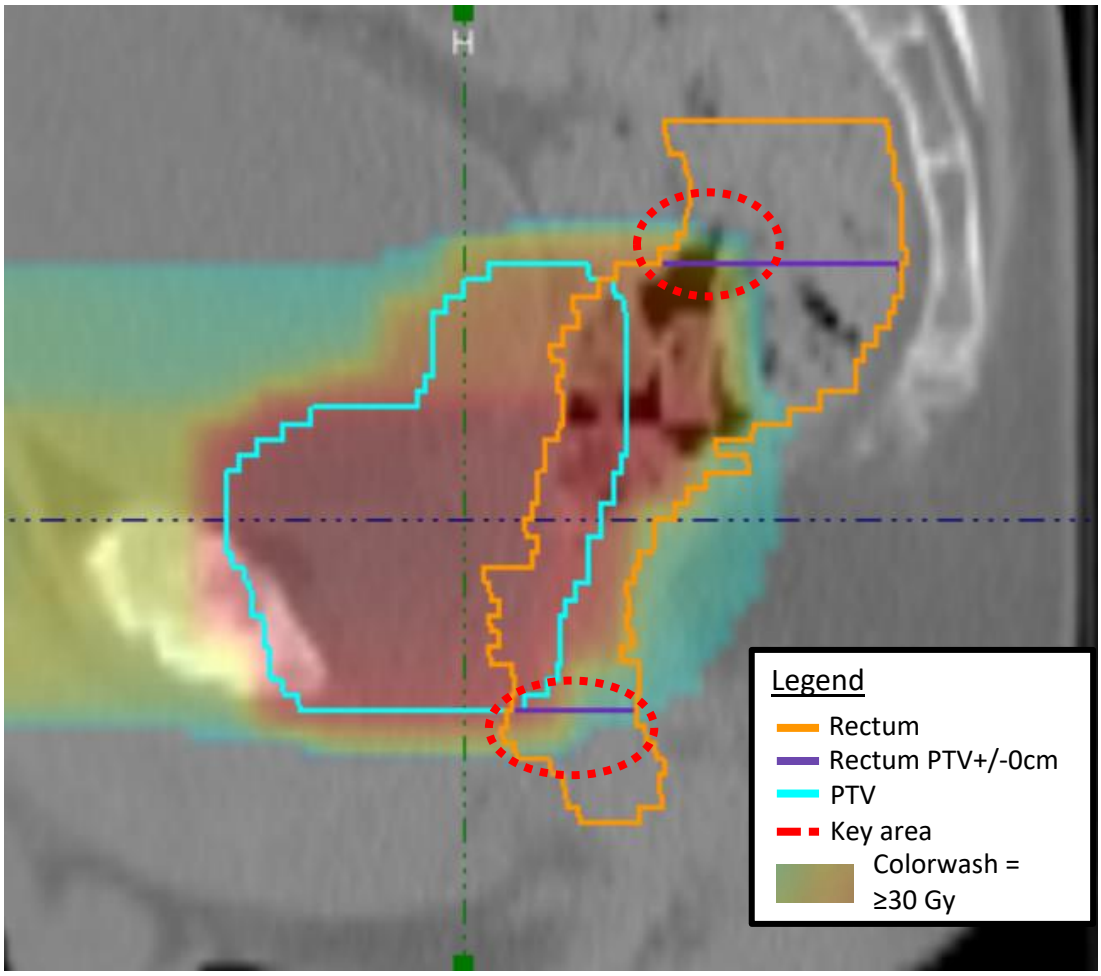


Figure 55. Example Showing Higher Doses Confined Largely to PTV+/-0cm

An example of a mid-rectal sagittal plane. In orange is the whole rectum, while the purple solid line demarks the PTV+/- 0cm truncation (PTV maximal extent in pale blue). The coloured dose is ≥30 Gy. The small amount of dose ≥30 Gy in the orange whole rectum, but outside the purple PTV+/-0cm is apparent (~3cm^3 in this patient) – red dotted highlight. It can be seen to be a small proportion of the ≥30 Gy dose to the orange rectum as a whole.

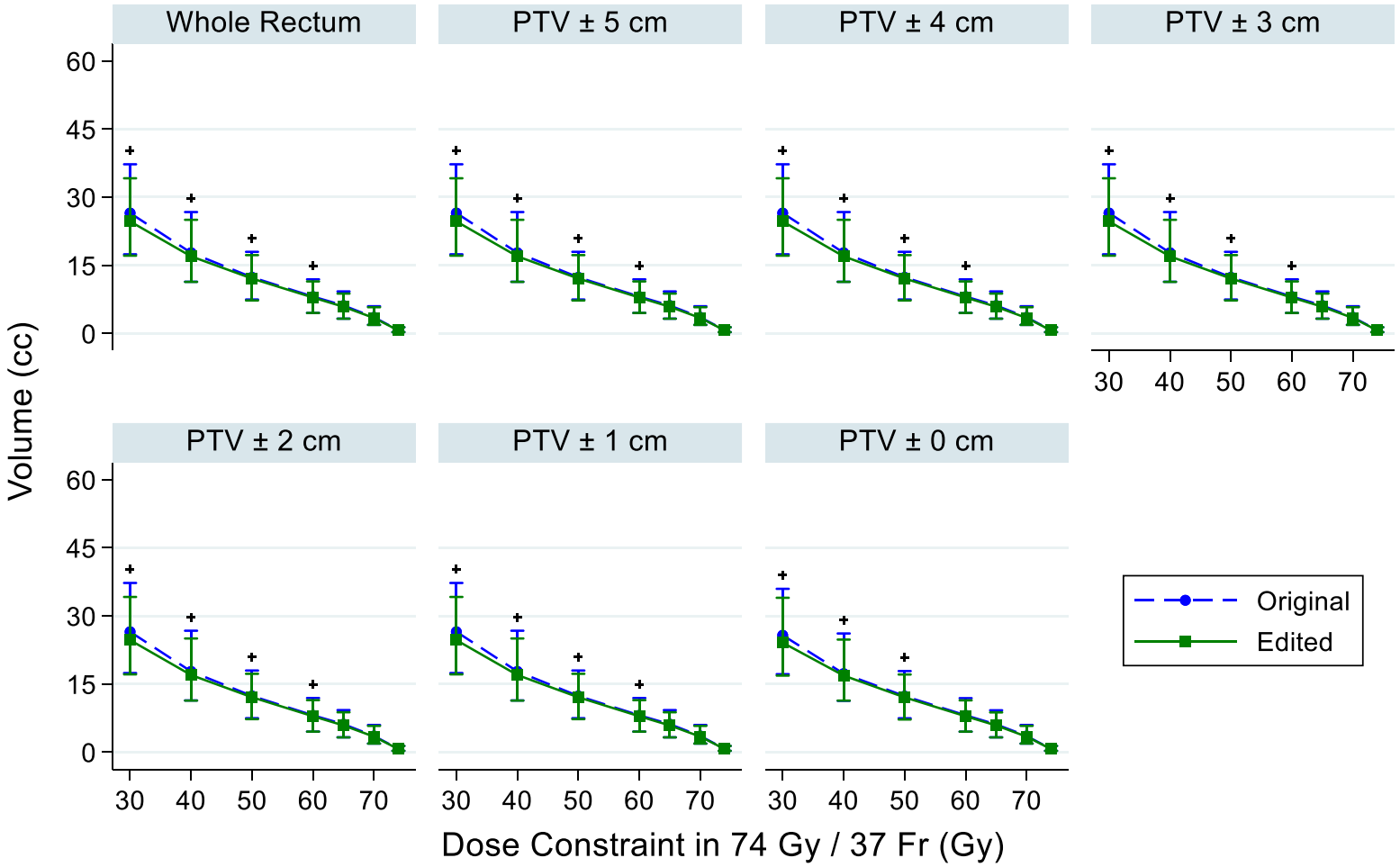


Figure 56. PACE-B Absolute DVH Volumes: Original vs Edited, By Truncation

Median and IQRs for examined rectal absolute dose bins, EQD2 corrected to CHHiP 74Gy arm constraints per methods.
+ = edited versus original dose bin volume significantly different by Wilcoxon Signed rank (p<0.0001).

6.4.4 Toxicity Metrics

From this point the analyses are limited to the CHHiP trial, since PACE-B late toxicity is yet to report.

6.4.4.1 Whole Rectum Relative Volumes: Original vs Edited Contour

6.4.4.1.1 Individual Dose Bins

The whole rectum relative volume individual dose-bin analyses for original vs edited contours are presented in **Table 40**. The key information in this table is that the p-values comparing original and edited AUC values are resoundingly negative. None approach adjusted significance, nor even an unadjusted 0.05 threshold. At an individual dose bin level, there is no evidence of benefit to toxicity prediction from protocol-based re-contouring.

In terms of the absolute magnitude of AUC values, the AUC for any individual dose bin ranges between 0.49-0.64, similar to prior analyses of CHHiP in prediction of rectal toxicity [225,227]. Most endpoint dose bins exhibit an AUC lower 95% confidence interval above 0.5 (point of no effect), i.e. overall, most are weak true predictors for toxicity.

Table 40. Original Versus Edited Contours: Individual Dose Constraint Analysis of Relative DVH Volumes

Comparing original vs edited individual relative volume DVH dose bin levels. Values for original and edited are very similar across every endpoint scale and dose-bin. Note every original versus edited AUC comparison (right column) is non-significant.

Endpoint	DVH Level (Gy)	Original				Edited				Original vs edited AUC p-value
		AUC	AUC 95% CI	Youden Index (%)	Youden Index 95% CI	AUC	AUC 95% CI	Youden Index (%)	Youden Index 95% CI	
Frequency G1+ n=1986	30	**0.52**	0.50–0.55	61.0	51.2–68.9	**0.52**	0.50–0.55	63.7	54.7–68.8	0.9914
	40	**0.56**	0.53–0.58	46.6	44.9–49.8	**0.56**	0.53–0.58	46.2	41.2–50.7	0.9986
	50	**0.57**	0.54–0.59	35.4	32.1–38.5	**0.57**	0.54–0.59	34.5	29.6–39.7	0.7762
	60	**0.56**	0.54–0.59	23.4	18.9–28.3	**0.57**	0.54–0.59	23.5	19.5–26.5	0.7581
	65	**0.54**	0.52–0.57	13.5	6.0–19.3	**0.54**	0.52–0.57	12.2	5.7–19.5	0.9990
	70	**0.53**	0.50–0.55	3.5	0.1–10.3	**0.53**	0.50–0.56	3.1	0.0–8.1	0.4762
	74	**0.51**	0.50–0.53	0.2	0.0–0.7	**0.52**	0.50–0.53	0.1	0.0–0.3	0.2948
Frequency G2+ n=1982	30	**0.57**	0.53–0.61	68.9	54.5–75.2	**0.56**	0.52–0.60	66.1	62.4–72.4	0.4133
	40	**0.59**	0.56–0.63	48.4	45.7–53.0	**0.59**	0.55–0.62	46.5	42.8–55.2	0.4442
	50	**0.59**	0.55–0.63	39.6	36.9–41.7	**0.58**	0.55–0.62	37.0	29.7–44.1	0.3134
	60	**0.58**	0.54–0.62	27.8	27.2–29.1	**0.57**	0.54–0.61	26.7	25.6–28.8	0.3669
	65	**0.58**	0.54–0.61	15.4	7.2–21.8	**0.57**	0.54–0.61	13.2	6.3–20.0	0.7897
	70	**0.56**	0.52–0.60	4.2	0.6–8.6	**0.56**	0.52–0.60	4.0	0.5–8.1	0.7583
	74	**0.51**	0.49–0.54	0.4	0.0–0.7	**0.52**	0.50–0.54	0.3	0.0–0.5	0.2655
Bleeding G1+ n=1969	30	**0.53**	0.50–0.56	67.5	45.5–90.3	**0.54**	0.51–0.57	67.1	48.7–78.2	0.3109
	40	**0.57**	0.55–0.60	48.6	39.6–54.8	**0.58**	0.55–0.60	48.2	40.0–55.6	0.3187
	50	**0.59**	0.56–0.61	38.3	30.2–41.3	**0.59**	0.57–0.62	38.0	31.4–42.0	0.4855
	60	**0.59**	0.56–0.62	22.1	17.5–28.5	**0.60**	0.57–0.62	22.3	18.3–26.9	0.3589
	65	**0.58**	0.56–0.61	10.2	7.7–14.5	**0.59**	0.56–0.61	9.1	7.4–14.4	0.2763
	70	**0.56**	0.53–0.59	1.7	0.0–3.9	**0.56**	0.53–0.59	2.0	0.0–4.1	0.4346
	74	**0.52**	0.50–0.54	0.1	0.0–0.8	**0.52**	0.50–0.54	0.1	0.0–0.6	0.7560
Bleeding G2+ n=1967	30	**0.54**	0.50–0.58	78.3	54.4–91.2	**0.53**	0.50–0.57	76.2	58.8–88.7	0.6093
	40	**0.57**	0.54–0.61	47.6	36.9–63.3	**0.57**	0.53–0.60	46.7	39.4–65.7	0.3036
	50	**0.59**	0.55–0.62	38.6	28.7–41.4	**0.57**	0.54–0.61	34.7	30.1–41.7	0.1382
	60	**0.59**	0.55–0.62	23.5	18.8–29.3	**0.58**	0.55–0.62	25.9	18.5–26.9	0.6042
	65	**0.59**	0.55–0.63	14.2	7.8–18.6	**0.59**	0.55–0.63	12.5	7.4–17.8	0.7208
	70	**0.56**	0.52–0.60	2.4	0.0–7.8	**0.56**	0.53–0.60	2.9	0.0–7.0	0.3611
	74	**0.53**	0.51–0.56	0.1	0.0–0.3	**0.53**	0.51–0.56	0.1	0.0–0.4	0.6036
Proctitis G1+ n=2105	30	**0.49**	0.47–0.52	65.3	42.8–92.2	**0.49**	0.46–0.51	62.8	39.7–93.3	0.6695
	40	**0.53**	0.51–0.56	44.3	36.5–55.3	**0.53**	0.50–0.55	43.6	36.8–52.3	0.5624
	50	**0.55**	0.53–0.58	35.8	27.3–39.1	**0.55**	0.52–0.57	33.0	25.9–36.1	0.5080
	60	**0.57**	0.54–0.59	22.1	17.2–27.4	**0.57**	0.54–0.59	20.6	17.2–26.6	0.7875
	65	**0.56**	0.54–0.59	12.5	7.4–19.1	**0.56**	0.54–0.59	11.6	6.0–18.8	0.6072
	70	**0.55**	0.53–0.58	1.7	0.0–8.0	**0.55**	0.53–0.58	1.1	0.0–7.0	0.7174
	74	**0.52**	0.50–0.53	0.2	0.0–0.4	**0.52**	0.50–0.53	0.1	0.0–0.6	0.7290

Continued overleaf …

Table 40 cont…		Original				Edited				Original vs edited AUC p-value
Endpoint	**DVH Level (Gy)**	**AUC**	**AUC 95% CI**	**Youden Index (%)**	**Youden Index 95% CI**	**AUC**	**AUC 95% CI**	**Youden Index (%)**	**Youden Index 95% CI**	
Proctitis G2+ n=2104	30	**0.50**	0.46–0.54	78.2	53.6–92.0	**0.51**	0.47–0.55	71.0	53.7–89.9	0.2988
	40	**0.53**	0.49–0.56	43.8	40.9–51.4	**0.53**	0.50–0.57	44.7	40.7–51.8	0.5030
	50	**0.54**	0.51–0.58	36.0	32.4–38.3	**0.55**	0.51–0.59	33.9	26.9–38.9	0.6912
	60	**0.56**	0.52–0.60	23.3	17.5–30.4	**0.56**	0.53–0.60	21.2	12.7–29.2	0.4572
	65	**0.57**	0.53–0.61	13.6	6.5–21.8	**0.57**	0.53–0.61	10.9	6.2–18.8	0.3324
	70	**0.56**	0.52–0.60	2.4	0.1–8.1	**0.57**	0.53–0.60	2.6	0.1–8.1	0.4770
	74	**0.52**	0.50–0.55	0.3	0.0–0.6	**0.52**	0.50–0.55	0.2	0.0–0.7	0.9060
Sphincter Control G1+ n=2154	30	**0.52**	0.48–0.55	62.2	51.5–85.1	**0.52**	0.48–0.56	64.5	53.8–69.5	0.6551
	40	**0.56**	0.52–0.60	44.8	36.3–54.4	**0.56**	0.53–0.60	46.5	40.1–49.5	0.8182
	50	**0.59**	0.55–0.63	36.3	29.6–41.9	**0.59**	0.55–0.63	35.6	29.7–42.6	0.7701
	60	**0.60**	0.56–0.63	24.3	18.0–29.1	**0.60**	0.57–0.64	25.5	18.0–29.4	0.4952
	65	**0.57**	0.53–0.61	13.7	6.8–22.2	**0.57**	0.53–0.61	12.6	6.1–21.1	0.8244
	70	**0.55**	0.51–0.59	1.9	0.8–3.6	**0.55**	0.51–0.59	1.6	0.2–6.7	0.6770
	74	**0.52**	0.49–0.54	0.1	0.0–1.5	**0.52**	0.49–0.54	0.1	0.0–1.0	0.7660
Stricture Or Ulcer G1+ n=2161	30	**0.58**	0.51–0.65	78.7	57.9–91.5	**0.59**	0.52–0.65	67.5	54.7–83.2	0.6129
	40	**0.63**	0.57–0.69	51.5	36.5–66.1	**0.64**	0.58–0.69	47.7	41.8–55.6	0.4683
	50	**0.63**	0.57–0.69	36.5	30.9–49.3	**0.63**	0.57–0.69	35.4	34.3–40.1	0.8432
	60	**0.62**	0.55–0.69	25.2	19.1–31.2	**0.62**	0.56–0.68	25.7	17.9–31.5	0.9296
	65	**0.61**	0.55–0.68	16.3	8.9–22.3	**0.62**	0.55–0.68	14.2	8.3–22.1	0.8846
	70	**0.62**	0.55–0.68	3.6	0.8–6.9	**0.61**	0.55–0.68	3.3	0.8–6.2	0.4161
	74	**0.55**	0.50–0.60	0.1	0.0–0.3	**0.55**	0.50–0.60	0.1	0.0–0.4	0.9231

6.4.4.1.2 Edited vs Original Rectum: Logistic Model

Although original vs edited individual dose bin toxicity prediction differences are not seen, it is possible that the combined information from all dose levels might permit discernment of such a difference.

Metrics of the logistic models fitted to all dose constraint bins for each endpoint are presented in **Table 41**. The absolute values of AUC range are expectedly somewhat better than individual bins, ranging 0.57 to 0.65. Low 95% CI for all model AUCs is above 0.5, implying each model is a true weak predictor of toxicity. This is corroborated by $sensitivity_{632}$, $specificity_{632}$, PPV_{632} and NPV_{632}, which are generally modest. Differences seen between original and edited models are generally reciprocal trade-offs of sensitivity for specificity and vice-versa (e.g. proctitis G1+).

It thus seems clear that editing, by central protocol-based review, of rectal contours does not improve toxicity prediction. I will proceed to examine if any changes to the analysis of original rectal contours (absolute volumes or rectal truncation) might improve toxicity prediction. From this point of the results, only the original rectal contours are considered, since these are the contours that will occur in typical radiotherapy practice.

Table 41. Original vs Edited Rectal Volumes for the Prediction of Rectal Toxicity

Separately for original and edited rectal contours, logistic models were fitted to whole rectum relative DVH (%) V30, V40, V50, V60, V65, V70 and V74 data for each endpoint. The predictive ability (AUC) of these is compared between original and edited contours, with no difference by DeLong comparison.

Novel Abbreviations: Sens = sensitivity; Spec = specificity.

Endpoint	n	Original						Edited						AUC: Original vs edited p-value
		AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	
Frequency G1+	1986	**0.57**	0.54–0.60	0.46	0.66	0.45	0.67	**0.57**	0.54–0.60	0.47	0.64	0.44	0.66	0.9179
Frequency G2+	1982	**0.60**	0.56–0.64	0.61	0.56	0.18	0.90	**0.60**	0.56–0.63	0.67	0.47	0.17	0.90	0.5431
Bleeding G1+	1969	**0.60**	0.57–0.63	0.42	0.73	0.44	0.72	**0.60**	0.57–0.62	0.43	0.71	0.43	0.72	0.8507
Bleeding G2+	1967	**0.60**	0.57–0.64	0.48	0.66	0.19	0.88	**0.59**	0.55–0.63	0.46	0.67	0.19	0.88	0.2050
Proctitis G1+	2105	**0.58**	0.56–0.61	0.55	0.55	0.40	0.70	**0.58**	0.55–0.60	0.62	0.49	0.39	0.71	0.2884
Proctitis G2+	2104	**0.57**	0.54–0.61	0.66	0.43	0.12	0.92	**0.58**	0.54–0.61	0.64	0.46	0.12	0.92	0.9438
Sphincter Control G1+	2154	**0.61**	0.57–0.65	0.67	0.48	0.14	0.92	**0.61**	0.57–0.65	0.56	0.60	0.15	0.92	0.5949
Stricture/Ulcer G1+	2161	**0.65**	0.59–0.71	0.65	0.57	0.05	0.98	**0.66**	0.60–0.71	0.77	0.46	0.05	0.98	0.9758

6.4.4.2 Absolute versus Relative DVH Volumes

Original whole rectum logistic models fitted to relative versus absolute DVH volumes are compared in **Table 42** [*overleaf*], note the relative volume logistic model data is identical to **Table 41**, but reproduced for easier comparison.

The relative volume models all showed similar or marginally better AUCs, although such differences were not significant. There is certainly no suggestion that absolute volume DVH metrics improve rectal toxicity prediction over relative volumes.

Table 42. Relative vs Absolute Volumes for the Prediction of Rectal Toxicity

For original whole rectum contours, logistic models were fitted, separately, to relative and absolute DVH (%) V30, V40, V50, V60, V65, V70 and V74 data for each endpoint. The predictive ability (AUC) of these is compared between relative and absolute volume models, with no differences by DeLong testing.

Endpoint	n	Relative Volumes						Absolute Volumes						AUC: Relative vs Abs p-value
		AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	
Frequency G1+	1986	**0.57**	0.54–0.60	0.46	0.66	0.45	0.67	**0.56**	0.54–0.59	0.54	0.57	0.43	0.67	0.6305
Frequency G2+	1982	**0.60**	0.56–0.64	0.61	0.56	0.18	0.90	**0.57**	0.53–0.61	0.34	0.74	0.18	0.88	0.0675
Bleeding G1+	1969	**0.60**	0.57–0.63	0.42	0.73	0.44	0.72	**0.59**	0.56–0.62	0.56	0.57	0.39	0.73	0.1299
Bleeding G2+	1967	**0.60**	0.57–0.64	0.48	0.66	0.19	0.88	**0.59**	0.55–0.62	0.49	0.62	0.18	0.88	0.1719
Proctitis G1+	2105	**0.58**	0.56–0.61	0.55	0.55	0.40	0.70	**0.57**	0.55–0.60	0.60	0.51	0.40	0.70	0.2808
Proctitis G2+	2104	**0.57**	0.54–0.61	0.66	0.43	0.12	0.92	**0.57**	0.53–0.61	0.67	0.43	0.13	0.92	0.9126
Sphincter Control G1+	2154	**0.61**	0.57–0.65	0.67	0.48	0.14	0.92	**0.60**	0.57–0.64	0.41	0.71	0.16	0.91	0.6504
Stricture/Ulcer G1+	2161	**0.65**	0.59–0.71	0.65	0.57	0.05	0.98	**0.63**	0.57–0.69	0.61	0.58	0.05	0.98	0.2653

6.4.4.3 Whole Rectum versus Truncated Rectum

PTV based truncations of the rectum (relative volumes) will now be examined, with PTV ± 2 cm and PTV ± 0 cm having been chosen for toxicity analysis based upon general visual inspection of contour metric inflexion points in earlier figures: **Figure 47, Figure 48 & Figure 49**.

Table 43 [*overleaf*] compares toxicity prediction for whole rectum versus truncated rectum (PTV ± 2 cm) by logistic model fitted to all seven relative-volume dose-bins. Note again that the whole rectum, relative volumes is the same results as original rectum in **Table 41**, reproduced to facilitate comparison.

The AUC results are identical to two decimal places, unsurprisingly resulting in no significant differences seen by AUC comparison. Given bleeding G1+ would be approaching unadjusted significance, it is worth noting that the DeLong test is a non-parametric AUC comparison, so p-values are determined by the relative predictions compared between models, rather than absolute prediction magnitudes. Hence even were a significant test result to be observed, consideration would need to be made of whether any absolute difference was clinically meaningful (which these are not).

Table 43. Whole Rectum versus PTV+/-2cm Truncation for Toxicity Prediction

Separately for whole rectal and truncated rectum in PTV ± 2 cm, logistic models were fitted to relative DVH (%) V30, V40, V50, V60, V65, V70 and V74 data for each endpoint. The predictive ability (AUC) of these is compared between whole rectum and PTV ± 2 cm contours.

Endpoint	n	Whole Rectum						Truncated Rectum: PTV+/-2cm						AUC: Whole Rectum vs PTV+/-2cm p-value
		AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	
Frequency G1+	1986	**0.57**	0.54–0.60	0.46	0.66	0.45	0.67	**0.57**	0.55–0.60	0.45	0.66	0.45	0.66	0.2361
Frequency G2+	1982	**0.60**	0.56–0.64	0.61	0.56	0.18	0.90	**0.60**	0.56–0.64	0.61	0.56	0.18	0.90	0.9898
Bleeding G1+	1969	**0.60**	0.57–0.63	0.42	0.73	0.44	0.72	**0.60**	0.58–0.63	0.49	0.66	0.41	0.73	0.0512
Bleeding G2+	1967	**0.60**	0.57–0.64	0.48	0.66	0.19	0.88	**0.60**	0.57–0.64	0.45	0.69	0.20	0.88	0.9854
Proctitis G1+	2105	**0.58**	0.56–0.61	0.55	0.55	0.40	0.70	**0.58**	0.56–0.61	0.54	0.56	0.41	0.70	0.7567
Proctitis G2+	2104	**0.57**	0.54–0.61	0.66	0.43	0.12	0.92	**0.57**	0.54–0.61	0.62	0.47	0.12	0.91	0.9585
Sphincter Control G1+	2154	**0.61**	0.57–0.65	0.67	0.48	0.14	0.92	**0.61**	0.57–0.65	0.60	0.55	0.14	0.92	0.1463
Stricture/Ulcer G1+	2161	**0.65**	0.59–0.71	0.65	0.57	0.05	0.98	**0.65**	0.59–0.71	0.68	0.54	0.05	0.98	0.5129

The whole rectum (relative volumes) is compared to the smaller truncation of PTV ± 0 cm, in **Table 44** . Note again that the whole rectum, relative volumes is the same results as original rectum in **Table 41**, reproduced to facilitate comparison. Again, no significant differences are seen, with very similar AUC estimates.

Table 44. Whole Rectum versus PTV+/-0cm Truncation for Toxicity Prediction

Separately for whole rectal and truncated rectum in PTV ± 0 cm, logistic models were fitted to relative DVH (%) V30, V40, V50, V60, V65, V70 and V74 data for each endpoint. The predictive ability (AUC) of these is compared between whole rectum and PTV ± 0 cm contours.

Endpoint	n	Whole Rectum						Truncated Rectum: PTV+/-0cm						AUC: Whole Rectum vs PTV+/-0cm p-value
		AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	AUC	AUC 95% CI	Sens 632	Spec 632	PPV 632	NPV 632	
Frequency G1+	1986	**0.57**	0.54–0.60	0.46	0.66	0.45	0.67	**0.57**	0.55–0.60	0.54	0.57	0.43	0.67	0.3831
Frequency G2+	1982	**0.60**	0.56–0.64	0.61	0.56	0.18	0.90	**0.59**	0.55–0.63	0.56	0.58	0.18	0.90	0.3569
Bleeding G1+	1969	**0.60**	0.57–0.63	0.42	0.73	0.44	0.72	**0.60**	0.58–0.63	0.49	0.67	0.42	0.73	0.4947
Bleeding G2+	1967	**0.60**	0.57–0.64	0.48	0.66	0.19	0.88	**0.60**	0.57–0.64	0.51	0.63	0.19	0.89	0.6724
Proctitis G1+	2105	**0.58**	0.56–0.61	0.55	0.55	0.40	0.70	**0.58**	0.56–0.61	0.53	0.57	0.40	0.69	0.6827
Proctitis G2+	2104	**0.57**	0.54–0.61	0.66	0.43	0.12	0.92	**0.57**	0.53–0.61	0.46	0.62	0.13	0.91	0.3075
Sphincter Control G1+	2154	**0.61**	0.57–0.65	0.67	0.48	0.14	0.92	**0.61**	0.58–0.65	0.56	0.60	0.15	0.92	0.2668
Stricture/Ulcer G1+	2161	**0.65**	0.59–0.71	0.65	0.57	0.05	0.98	**0.66**	0.60–0.71	0.70	0.53	0.05	0.98	0.9836

6.5 Discussion

6.5.1 Summary of Results

With data from the CHHiP trial, this chapter has addressed a number of questions surrounding the rectal contour for prostate EBRT. Firstly, whether the central review of rectal contours from trials is useful to aid subsequent toxicity modelling. Although significant differences were seen in rectal morphology and dosimetry, the magnitudes of effect were small. Thus, no significant difference in toxicity prediction was seen through the use of edited versus original contours. Given the substantial time cost of central rectal contour review, this would suggest the process should be omitted prior to dosimetric analysis of large prostate EBRT trials.

It is worth noting that prediction of toxicity with dosimetry alone was generally weak; as was also observed in the big-RT study examining combining dosimetry, clinical and genomic factors for a smaller subset (n=721) of the CHHiP patients (AUC 0.52 for rectal bleeding with dosimetry alone vs 0.71 with combined information [227]). Our values are slightly lower than AUCs in another reported dosimetric analysis from CHHiP (e.g. n=538, AUC 0.60 – 0.65 for rectal bleeding) [225]. This may be due to the case:control design of that study, which enriched for patients expressing toxicity. It is worth considering that progressively decreasing toxicity seen in EBRT for nmPCa (e.g. as seen from RT01 to CHHiP [1,215]) will likely hinder development of externally generalisable toxicity models.

To help consider external generalisability of these first results, the morphological and dosimetric analyses have been repeated on PACE-B. For edited vs original contours, length volume and DSC were more consistent for PACE-B patients than CHHiP. This may be driven by the increasing familiarity of treating centres with per-protocol rectal contouring over the decade 2002 (start of CHHiP) to 2012 (start of PACE-B).

For edited minus original whole rectum relative volume dosimetry, PACE-B had larger differences vs CHHiP (**Figure 53 vs Figure 52**), but these were still very small in magnitude. I feel that the morphological and dosimetric differences between CHHiP and PACE-B are insufficient to suggest revision of the conclusions drawn from the CHHiP trial.

Secondly, the possible use of absolute volume DVHs rather than relative volume DVHs has been considered for original contours. While in theory this might improve consistency, no significant difference in toxicity prediction was seen between the two approaches. The direction of effect was towards relative volumes being non-significantly better. This is perhaps driven by the tighter spread of relative volumes seen at each dose level, compared to absolute volumes (compare IQR for relative and absolute original volumes of **Figure 52 & Figure 54**). Since relative dose-volume constraints are in more common usage [175,225], no change to current practice is suggested by this study.

Thirdly, the possibility of using truncated versions of the original rectum have been investigated. No difference in toxicity prediction was seen between original rectum and PTV ± 2cm, nor PTV ± 0cm, although again acknowledging generally low AUCs. This could be interpreted as a benefit, given the reduced volume to contour, however I believe this would be incorrect. Maintaining a consistent definition of the rectum is beneficial to the implementation of multicentre trials and to the external generalisability of dosimetric analysis from such trials. For patients receiving non-co-planar radiotherapy (e.g. CK delivery), contouring the whole rectum is critical to prevent possible excess dose in parts of the rectum at distance from the PTV. This dose-dumping in non-contoured rectum may also occur with co-planar delivery, if using a PTV±0cm definition, as can be seen in the sagittal patient example earlier, (**Figure 55**). It is desirable that a rectal definition is chosen that would be suitable for all delivery methods. Therefore, without any evidence of an improvement with rectal truncation, I would suggest that standard practice remains the whole rectum.

6.5.2 Results in Context

6.5.2.1 Observer Effects on Rectal Morphology and Dosimetry

QA work for the RT-01 trial had 13 participating centre observers contour three cases, with large inter-observer differences (up to 7cm) noted for the superior border of the rectum [222]. This observation prompted a clearer definition of the required rectal contour superior border (the rectosigmoid junction), which was carried over into the CHHiP trial. The CHHiP results here show that differences of up to 7cm in length are seen between the

edited and original rectal lengths in this study, although in general differences were smaller (IQR ~ -2 – 4 cm). It is pleasing to see that the magnitude of such difference is even smaller in the PACE-B data, where differences were < 5 cm in magnitude.

A similar QA process, for the AIROPROS01-02 trial, had 18 observers contour the rectum on four prostate cases, with a strict definition of the inferior and superior rectal limits [209]. This definition perhaps drove the lower inter-observer variability seen c.f. the RT-01 QA study, with no observer >1cm from the global mean for the cranial or caudal border. Despite tighter contouring consistency, they noted interobserver rectal cumulative DVH differences of 15-20% (over the range V40Gy – V65Gy). The edited vs original whole rectum relative volume DVH differences were generally much smaller in this study (modulus median < 3%)

6.5.2.2 Observer Effects on Rectal NTCP

Pre-fitted NTCP models have been used to interpret inter-observer difference in rectal DVH dosimetry. Fiorino *et al* reported such a process, with ten patients' rectums contoured by three observers [210]. Dosimetry was fed into a pre-fitted Lyman Kutcher-Burman NTCP model, assuming a dose of 75.6 Gy . As was seen with the AIROPROS01-02 study, interobserver differences in contouring resulted in some large DVH parameter changes, e.g. up to 10-12% in the V50-V65 range. However, the SD in $NTCP_{75.6}$ probability was just 0.7%.

Roach *et al* reported two NTCP-modelled cohorts: firstly, three observers of 35 patients; second, 10 observers of five patients [211]. Unlike Fiorino, they used a Simultaneous Truth And Performance Level Estimation (STAPLE) meta-contour as the comparative metric [228], rather than maximum inter-observer differences. This permitted scrutiny of the SD in inter-observer rectal NTCP estimates: cohort one (SD 1.2%) and cohort two (SD 2.5%). This implies that the least extreme 95% of observer variability will cover a 10% range of NTCP difference. Such magnitudes of effect approach a clinically significant range of rectal toxicity probability. Again, note that the NTCP models were pre-fitted.

Note that neither of these studies incorporated the actual toxicity data for each patient, instead inferring it from dosimetry, via pre-fitted NTCP models. This chapter extends beyond such work through the fitting of actual toxicity data, allowing direct demonstration that central review of contours, to reduce inter-observer effects, makes no difference to dosimetric prediction of subsequent rectal toxicity.

6.5.2.3 Importance of Non-Contouring differences

It must be remembered that inter-centre differences beyond the rectal contouring process may influence final rectal DVH parameters. Rasch *et al* compared 22 prostate EBRT (78 Gy / 39 fractions) patients, contoured by

two centres within the Dutch CKVO 96-10 trial [229]. Rectal definition[37] was the same, but each centre had differing methods of PTV expansion. Rectal absolute volume DVH V74Gy differed by ~1cm^3 depending on the rectal contour, but by ~5cm^3 depending on the method of PTV expansion. This underscores the importance of controlling other sources of variance in multicentre clinical trial rectal dosimetry, which may be larger than interobserver differences.

6.5.2.4 Absolute Volume Dosimetry

In 2002, Kupelian *et al* reported on 128 patients receiving EBRT (70-78 Gy in 2-2.5 Gy fractions), fitting multivariate models for late rectal bleeding (8% rate) that included both relative and absolute rectal volume receiving prescription dose [216]. Absolute volume was an independent predictor of toxicity, while relative volume was not. This is a small study, with a low event rate, and the cut-off they suggest (15cm^3) would apply to none of our CHHiP patients (not presented above, but absolute V74Gy max = 11.3cc).

Koper *et al* reported on 199 patients receiving 66 Gy (2 Gy per fraction) of EBRT, with a 33% rate of any rectal bleeding at 3 years [218]. They fit Kaplan-Meier models for rectal bleeding to both rectal wall and solid rectum, trying different “cut-off” dose values to create two groups, then comparing the log-rank p-values found, between different cut-off values. Issues around

[37] “*From the most caudal slice where the tuber ischiadicum was still visible to the most cranial slice, where the rectum was still adjacent to the sacrum, or at the most caudal level of the sacroiliacal joints.*”

multiple testing are not discussed. They remark that for the solid rectum models, the relative volume outperformed the absolute volumes (and both were weaker than rectal wall) although the figures are not stated. With incomplete information reported for solid rectum, it is difficult to draw comparison with this study.

Vargas *et al* reported in 2005 on 331 patients treated with EBRT (63-79.2 Gy, 1.8 Gy per fraction) in a prospective dose-escalation protocol, with a 10.3% rate of late rectal toxicity G2+ [219]. Unfortunately, median follow-up was only 1.6 years. They examined the rectum as a solid and as a wall, though findings were similar for both. By linear regression, they reported closer association of relative, rather than absolute volume dosimetric parameters. However, both for relative and absolute volumes, dosimetric parameters (examined univariately) had significant associations with chronic toxicity. The benefit of relative volumes was concluded from lower p-values in the separately fitted models. As a prospective and larger dataset, this is probably the most robust prior study, although follow-up time (1.6-year median) is short for a late toxicity analysis.

More recently in 2018, Kotabe *et al* reported a small retrospective study of 82 patients receiving 76 Gy (2 Gy per fraction) EBRT as primary radiotherapy [220]. A late rectal bleeding rate of 3.2% was observed at 4-years, meaning a very small amount of data to be drawn from. Unusually, despite the study examining late toxicity, they report EQD2 adjusting the dosimetry by an α/β ratio of 10 Gy, which would appear more appropriate for acute toxicity. They

fitted relative and absolute rectal DVH parameters sequentially, identifying only relative V60Gy as significant for rectal bleeding. However, this study is hampered by the very small size, low event rate and unusual EQD2 dose adjustment.

Paleny *et al* (2019) retrospectively reported on 285 patients[38] receiving various forms of EBRT as salvage, adjuvant and primary radiotherapy for PCa (60-78 Gy in 2 Gy fractions) [221]. Only 3% of patients experienced grade 2+ late radiation proctitis. For late radiation proctitis, multiple relative dose-volume parameters were significantly related by univariate logistic regression to G1+ late radiation proctitis, but no absolute volume parameter was. This study is hampered by retrospective nature, heterogenous patient group and low event rate of relevant toxicity (i.e. G2 and above).

Overall, the data on this subject is highly heterogenous, with multiple retrospective and small studies, often with low event rates. Results for and against absolute volumes have been seen. The data in this chapter provide a sample size larger than every preceding study combined, prospectively collected, with good follow-up duration (the 5-year follow-up dataset) and reasonable toxicity event rates (Recall **Table 9**). For the solid rectum there is no evidence for the benefit of absolute volumes over relative volumes, the current standard-of-care.

[38] Other analyses including pelvic RT patients are reported, but are not relevant to this discussion.

6.5.2.5 PTV-based Rectal Truncation

Nitsche *et al* have retrospectively studied the effect of several different rectal contour definitions [12]. A single observer contoured 13 different rectal definitions on 23 prostate EBRT patients, choosing the three definitions resulting in most disparate rectal DVH distributions: the RTOG definition (as above); rectum in superior/inferior range of PTV ± 1cm; rectum in superior/inferior range of PTV ± 0cm. These definitions were then contoured on two cohorts: 97 primary EBRT and 66 salvage RT patients. They were unable to show relationship between various DVH parameters and worst acute/late rectal inflammation G1+ (CTCAE). Although innovative, this study suffers from small sample size and non-prospective toxicity collection.

In a similar study, but for acute toxicity, Onal *et al* applied 4 rectal definitions to 94 patients: i) Rectum in sup/inf range of prostate; ii) PTV ± 1cm; iii) 11cm superior to anal verge; iv); Anal verge to sigmoid flexure [223]. Significant differences between the 4 methods were seen in mean dose (maximum range 49.6 – 57.5 Gy) and all relative volume DVH dose bins V30Gy – V70 Gy (in 10 Gy intervals). The ordering of dose bin values was consistent: method 1 > 2 > 3 > 4. This is not surprising given the use of relative volume and the decreasing low dose rectum likely present in methods 2 and 1. For all definitions, both mean dose and V70Gy (%) were significantly higher for G2+ versus G0–1 acute rectal toxicity patients.

The data in this chapter is far larger (n~2000 for all endpoints), with demonstrable relationships between dosimetry and toxicities (seen as AUCs

95% CIs not encompassing 0.5 of no effect). The size of this study means we can be confident that PTV-truncation based definitions for solid rectum do not result in better rectal toxicity prediction. However, it is worth noting that in all cases, model AUCs were weak predictors of toxicity (all <0.7), supporting previous observations that dosimetry alone is no longer strongly related to toxicity prediction[39] [227].

6.5.3 Strengths of the Study

To my knowledge, a sample size >2000 patients has never been utilised to study the effect of rectal contouring on toxicity prediction. This is a strength, since it generates tight 95% CIs for key parameters such as prediction model AUCs, and diminishes the chance of type II error, which might occur in typical small contouring studies. Additionally, the data is prospectively collected, with standardised toxicity scales and good follow-up duration (5-year dataset). This is an improvement over many studies where retrospective data is used, meaning toxicity data has been collected in an ad-hoc manner in clinic.

The work on the value of central contour review is, to my knowledge, the first such study reported. It is reassuring to know that local centres are now delineating the rectum with sufficient accuracy as to be indistinguishable for late rectal toxicity prediction. The result can be implemented immediately by allowing omission of central review the remaining PACE-B contours, an

[39] Due to dose constraints being applied in modern EBRT trials.

immediate saving of several hundred hours. Given the relatively large number of prostate EBRT trials that are undertaken, its findings might also be relevant to analyses for thousands more patients (e.g. PIVOTAL-BOOST [ISRCTN80146950], PACE-C in the UK).

6.5.4 Limitations of the Study

The use of one central reviewer per patient contour is a limitation. There will always be marginal cases when interpreting the rectal border, so ideally one would have multiple observers to obtain a gold standard STAPLE contour for each rectum. However, the use of a >2000 patient cohort makes such an approach impractical, meaning this study instead relies upon the sample size reducing the impact of any such marginal case.

Per the unified approach of this thesis, I have limited this study to the examination of solid rectum rather than rectal wall. It is possible that the putative benefit of absolute volumes (c.f. relative) may be better seen by examination of rectal wall. This would potentially negate the noise from rectal contents (e.g. faeces) in the high dose field. It would be possible to procedurally generate rectal wall structures for the CHHiP patients; however, this represents work beyond the scope of this thesis.

The findings of this study are applicable to human-led contouring. This represents a limitation given, in my opinion, the very strong likelihood of auto-contouring emerging into routine clinical practice within the next decade [230]. It is probable that a limited number of such algorithms might enter UK

practice, making it more straightforward to compare and correct for systematic inter-algorithmic differences in future dosimetric analyses.

It must be acknowledged that the sampling method for the PACE-B patients was a convenience sample, owing to the availability of contouring time. It would appear the PACE-B analysis is biased towards more recent patients. This may help external generalisability, since any general improvement in contouring during the time period of the trial will be represented. I have also shown that contribution by centre was imbalanced, with CK centres underrepresented. It is therefore conceivable, although perhaps unlikely given their high recruitment numbers, that contouring practice in those centres would be different.

As for all work in this thesis, the doses examined are the dosimetry on a single planning CT scan, as oppose to accumulated organ dose. This limitation has been discussed previously in **Section 3.5.4** and similar thoughts apply here.

6.6 Conclusions

Using data from the CHHiP trial, this study has demonstrated that central rectal contour review confers no significant improvement for the prediction of rectal toxicity. This will be put into immediate effect in **Chapter 7**, where original rectum dosimetry for the entire PACE-B cohort will be utilised. This avoids hundreds of hours that would be required to centrally review the 500+ patients not represented in this study. The results would also be applicable to

other large UK national radiotherapy trials such as PIVOTAL-BOOST and PACE-C, once dosimetric information for those trials is collated.

This study has then demonstrated no change in toxicity prediction from the use of absolute vs relative volume DVHs. Given the widespread acceptance of relative volume DVH dose constraints, the continued usage of these seems appropriate. Future work might consider confirming that this hypothesis holds in a rectal wall OAR setting.

Finally, PTV-based truncation of the rectum (at PTV ± 0 cm & PTV ± 2 cm) has also failed to improve toxicity prediction. Whole rectal constraints will always be essential for non-co-planar planning, therefore in the interests of uniformity, it seems reasonable that whole rectum remains the status quo for all prostate EBRT patients.

To conclude, whole rectum relative DVHs appear to be suitable to remain as the status quo for toxicity modelling. However, it is clear that dosimetry data alone results in weak predictors. Efforts to combine dosimetry from whole rectum relative DVHs with other predictor modalities (e.g. clinical factor, genomic, baseline symptom data) should be considered for future work.

Chapter 7. Risk Factors for Acute Toxicity After Hypofractionated EBRT

7.1 Introduction

In **Chapter 4 & Chapter 5**, acute toxicity from the PACE-B trial was analysed. Such information is useful to inform patients of their average risk of toxicity and to optimise clinical practice. However, the development of models for toxicity is desirable, in order to allow risk assessment on a more individualised patient basis.

To control for potential confounding factors, multivariate predictive modelling is preferable (recall **2.5.1**), especially when the predictor of interest was not a randomisation or stratification factor. Multivariate models for acute toxicity following prostate radiotherapy/brachytherapy have been reported, however there is limited consensus on key predictors. GI toxicity risk factors reported have been inconsistent between studies, including: age [231], baseline symptoms [206,232], ADT usage [206], prostate volume [206], RT technique [233] and dosimetry [232]. Statin use has been reported as GI protective [231]. Significant reported GU toxicity risk factors include age [206,232], prior trans-urethral resection of prostate (TURP) [232], baseline symptoms [206,232,234], prostate volume [206], ADT usage [234], brachytherapy techniques [234] and bladder dosimetry [232]. The single references for many of the above predictors (e.g. ADT) indicates reproducibility is an issue between studies.

The PACE-B trial provides a unique opportunity to study acute toxicity risk factors in the ultrahypofractionation setting: a large prospectively collected dataset, with granular detail on potential predictors, including dosimetry. This chapter sets out to establish multivariate predictive models for acute GI and GU endpoints (CRO and PRO), investigating both previously reported and novel risk factors. It aims to provide estimates of covariates for important risk factors, in a format useful for inference in clinical practice. Additionally, such models' performances will be assessed on unseen data, via cross-validation, to reduce the possibility of overfitting, an approach not undertaken previously [206,231,233,234].

7.1.1 Hypotheses

1) Risk factors for acute GI and GU side effects (both CRO and PRO) following hypofractionated EBRT can be established with multivariate modelling.

7.2 Methods

7.2.1 Inclusion Criteria for PACE-B Patients

PACE-B has been described in **Chapter 4 & Chapter 5**. Patients were excluded from this substudy for: missing DICOM data, receiving a non-protocol regimen, or requiring re-planning during treatment. In each model, those with missing data for the model endpoint (toxicity status) were excluded. For predictor selection stages, patients were excluded for missing

data in any of the candidate model predictors. For the complete case final models, those missing data in any final model predictor were excluded.

Two sets of models were made. Set one for all patients, termed Whole-Trial models. Set two for SBRT patients only, termed SBRT-Only models. It is of specific interest to find variables predicting SBRT acute toxicity, since this may be the predominant nmPCa treatment modality in future, should PACE-B confirm the HYPO-RT-PC efficacy results [95].

7.2.2 Endpoints of Interest

The chosen endpoints, along with associated model format (logistic/linear), are shown in **Table 45**. The definitions for each endpoint are as described in **Chapter 4 & Chapter 5**, with subdomain scores in EPIC-26 again calculated from constituent questions by recommended methodology [134]. All models were fitted with a constant term. Individual numerical predictors were plotted against the numerical endpoints to consider inclusion of squared terms in the model, however no relationship appeared to warrant this. There are eight endpoints, with each having both Whole-Trial and SBRT-Only models, for a total of 16 models generated.

Table 45. Endpoints and Associated Model Type

Endpoints used in modelling. The CRO scores (RTOG and CTCAE) were scored positive for toxicity G2+, and EPIC-26 subdomain scores are the worst recorded score for that subdomain; in both cases at any follow-up after start of radiotherapy, up to 12 weeks. Novel Abbreviations: MVar = multivariate

Endpoint	Type	Model
GI Endpoints		
Acute RTOG GI G2+	Binary	MVar Logistic
Acute CTCAE GI G2+	Binary	MVar Logistic
Worst Acute Epic-26 Bowel Domain	Numeric	MVar Linear
GU Endpoints		
Acute RTOG GU G2+	Binary	MVar Logistic
Acute CTCAE GU G2+	Binary	MVar Logistic
Worst Acute IPSS Score	Numeric	MVar Linear
Worst Acute EPIC-26 UO Domain	Numeric	MVar Linear
Worst Acute EPIC-26 UI Domain	Numeric	MVar Linear

7.2.3 Predictors

7.2.3.1 Predictor Sources

Predictors, being defined as variables for inclusion in the model, came from two prospectively collected sources: i) the PACE-B trial database, managed by ICR-CTSU; ii) the DICOM files for each patient. I personally retrieved several hundred DICOMs[40] from the treating centres, with Olivia Naismith, physicist, retrieving the rest. The DICOM files were opened to ensure that the structures and dose aligned correctly with CT, before recalculating the DVHs, based on the original OARs (not recontoured, per the results in **Chapter 6**). This was done by myself, Katie Fernandez and Joanna Parker. This ensured that a single DVH calculation algorithm (VODCA) had produced all DVH metrics. These were then processed in MATLAB/Stata to extract predictors, using scripts written and debugged entirely myself.

[40] I unfortunately did not record the precise number of DICOMs that I retrieved.

7.2.3.2 Radiotherapy Regimen Predictors

The radiotherapy regimens were one-hot encoded to handle multiple regimens (**Figure 57**). As discussed in **Section 4.6.3**, CK treatment was pre-specified as hypothesis of interest, at the time of protocol amendment to permit VMAT treatments for SBRT patients in the trial. Only Whole-Trial models had SBRT and moderate hypofractionation as predictors.

SBRT, Cyberknife SBRT = 1 Cyberknife = 1 Mod. Hypofractionation = 0	62 Gy in 20 Fr SBRT = 0 Cyberknife = 0 Mod. Hypofractionation = 1
SBRT, No Cyberknife SBRT = 1 Cyberknife = 0 Mod. Hypofractionation = 0	78 Gy in 39 Fr SBRT = 0 Cyberknife = 0 Mod. Hypofractionation = 0

Figure 57. Explanation of Treatment Regimen Encoding
The three predictors SBRT, CK and Moderate Hypofractionation allow description of the 4 key delivery methodologies of interest.

7.2.3.3 Baseline Symptom Scales

Ordinal baseline symptom scales were changed to binary, due to low numbers of high-grade events. RTOG & CTCAE GU G2+ at baseline were both used as predictors. RTOG & CTCAE GI G2+ baseline scores were also considered, however rates were too low for bootstrapping, so baseline G1+ was used for GI models. Baseline IPSS and baseline EPIC-26 UO were strongly correlated. Therefore, baseline IPSS was used in GU models for RTOG, CTCAE and IPSS endpoints. Baseline EPIC-UO was used for EPIC-26 UI and EPIC-26 UO models.

7.2.3.4 Baseline Medications

Baseline statin usage was included for the GI model due to the reported protective effect by Palumbo *et al* [231]. For the GU models, baseline anti-cholinergic and alpha blocker usage was combined, due to the relatively low frequencies of use as individual categories. ADT was not given in PACE-B.

7.2.3.5 Dosimetry Predictors and Treatment Time

The dosimetry predictors were derived from the dose constraints for the SBRT arm (**Table 46**). Note that V50(%) & V80(%) refers to the percentage of organ receiving 50% or 80% of prescription dose, while V100(cc) refers to the volume in cc receiving 100% prescription dose. Strong correlation was noted between rectal V50(%) and V80(%) (Pearson r 0.81, $p<0.0001$), however it was decided to keep both due to these being the dose constraints used in the trial. Therefore, the use of any dosimetric predictor was also examined, since correlation may result in one or other being selected.

Table 46. Derivation of Dosimetric Predictors

How the dosimetric predictors were calculated for each arm. Note that the original top dose constraints in SBRT arm were at V36 Gy for rectum and V37 Gy for bladder. For simplicity it was decided to use V100% (36.2 Gy) for both.

Predictor	Data Type	36.25 Gy in 5 Fractions	62 Gy in 20 Fractions	78 Gy in 39 Fractions
GI				
Rectum V50(%)	Relative volume (%)	18.1 Gy	31.0 Gy	39.0 Gy
Rectum V80(%)	Relative volume (%)	29.0 Gy	49.6 Gy	62.4 Gy
Rectum V100(cc)	Absolute volume (cc)	36.2 Gy	62.0 Gy	78.0 Gy
GU				
Bladder V50(%)	Relative volume (%)	18.1 Gy	31.0 Gy	39.0 Gy
Bladder V100(cc)	Absolute volume (cc)	36.2 Gy	62.0 Gy	78.0 Gy

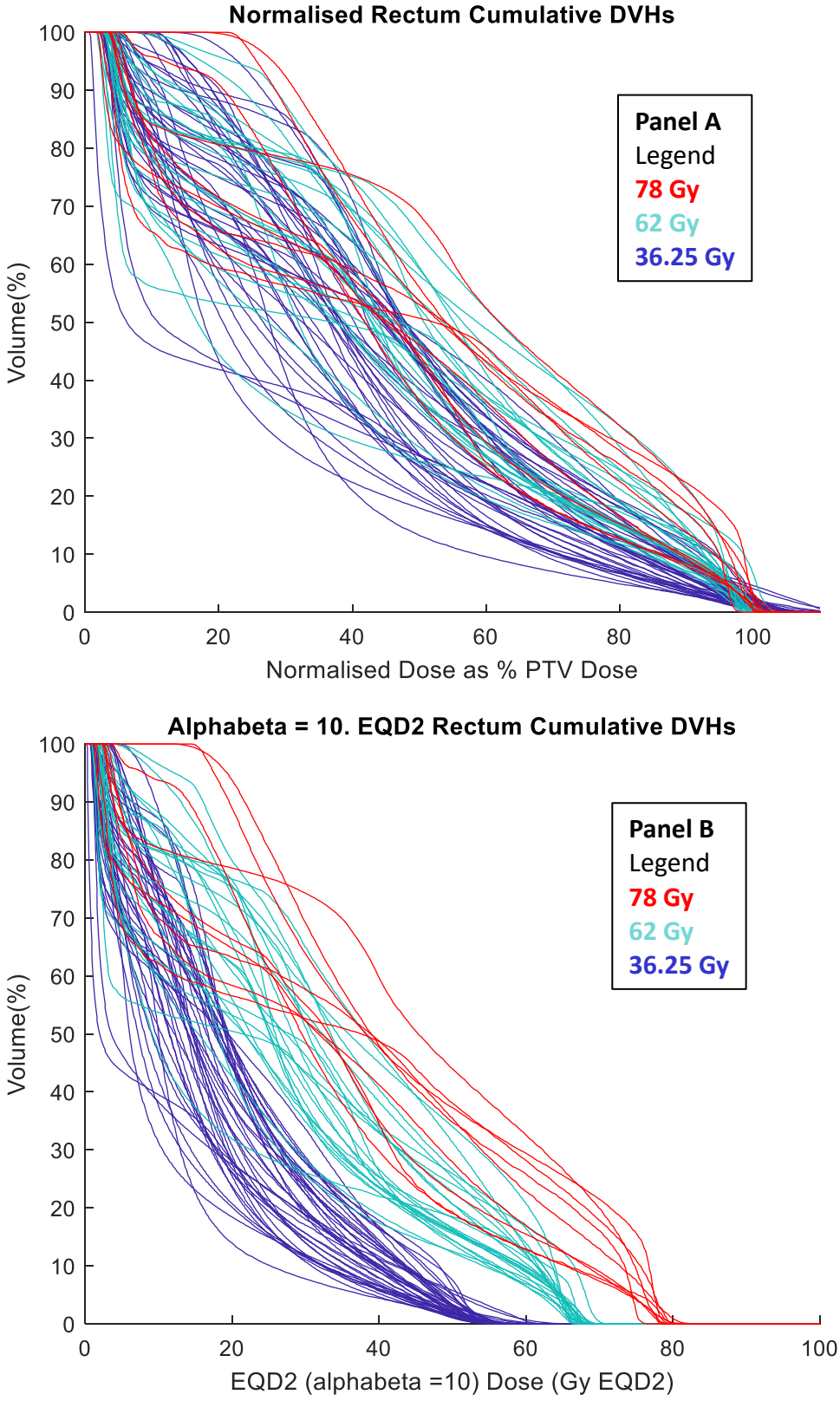
EQD2-adjusted doses were considered; however, this approach was not undertaken for two reasons. Firstly, uncertainty in the appropriate α/β ratio to apply to any acute EQD2 correction. While commonly a general acute α/β ratio is assumed to be 10 Gy [235], there are, to my knowledge, no human data to support this for bowel toxicity. Secondly, application of EQD2 with an α/β of 10 Gy will result in substantially non-overlapping DVHs between SBRT and CFMHRT patients (**Figure 58**). Lower effective doses with SBRT is at odds with the data seen in **Chapter 4 & Chapter 5**, where GI toxicity did not differ strongly between the two arms (with SBRT worse for CTCAE G2+ GI effects).

Adding overall treatment time into the model would potentially ameliorate this mismatch between trial level toxicity and EQD2 adjusted dose. While appropriate for SBRT-Only models, overall treatment time cannot reasonably be used as a predictor in the Whole-Trial models, as it would simply act as a surrogate for treatment arm. For SBRT-Only models, it was coded as ≥1 week (yes/no).

On initial modelling, V100(cc) did not model well in the Whole-Trial models, with biologically implausible negative coefficients (i.e. more dose → less toxicity). This likely represents V100(cc) acting as a surrogate for treatment arm, since planning objectives (Chapter 4) differed markedly between SBRT and CFMHRT. Therefore, V100(cc) was omitted from Whole-Trial models.

Figure 58. Comparison of Percentage and EQD2-Adjusted Rectal Dosimetry

Panel A: Percentage-adjusted rectal DVH dosimetry. Showing the overlap between arms, taken forward to the modelling process. Panel B showing EQD2 (α/β = 10 Gy) adjusted rectal dosimetry with much lower SBRT arm doses (in green).Both have random 10% of patients displayed for clarity



7.2.3.6 Other Predictors

The margin sizes were available as two predictors (in mm): posterior margin (margin posterior) & all other margins (margin non-posterior). These were strongly correlated ($r=0.7$; $p<0.0001$), so given anatomical consideration, margin posterior was selected for GI models and margin non-posterior for GU models.

PTV volume was considered, but correlated strongly with prostate volume ($r=0.7$; $p<0.0001$). Although different treatment volumes were used for low (CTV = Prostate) and intermediate risk (CTV = Prostate+1cm SV) patients, only ~10% of patients were low risk, so prostate volume was chosen as a more universally recognisable predictor, that is available pre-treatment. History of prior TURP was not available in the PACE trial database.

7.2.3.7 Final Predictor Selection

The complete candidate predictor set is shown in **Table 47**, with detail on which predictors were used specifically in GI or GU and Whole-Trial or SBRT-Only models.

Table 47. Predictors Used in Modelling Process

Predictors	Model Type	Source	Type	Notes on Measurement
GI & GU Models				
SBRT	Whole Trial	Database / DICOM	Binary	1 = 36.25 Gy in 5 fractions
Mod. Hypofractionation	Whole Trial	Database / DICOM	Binary	1 = 62 Gy in 20 fractions
Cyberknife	Both	Database / DICOM	Binary	1 = Treated on Cyberknife
Fiducials	Both	Database / DICOM	Binary	1 = Fiducials used
RT Time >1 week	SBRT only	Database	Binary	1 = Over 1 week from fraction 1 to final fraction
WHO PS 1+	Both	Database	Binary	1 = Baseline WHO Performance Status ≥1
Age	Both	Database	Numeric	Age at randomisation in years
Prostate Volume	Both	Database / DICOM	Numeric	Pre-treatment prostate volume in cc
RT Centre Caseload	Both	Database	Numeric	Number of PACE-B patients treated at that centre
GI Models Only				
Baseline RTOG GI G1+	Both	Database	Binary	1 = Grade 1 or more RTOG GI score at baseline
Baseline CTCAE GI G1+	Both	Database	Binary	1 = Grade 1 or more CTCAE GI score at baseline
Baseline EPIC-26 Bowel	Both	Database	Numeric	0-100 per EPIC-26 domain scoring (Bowel)
Statin Use	Both	Database	Binary	1 = Statin usage at baseline
Margin Posterior	Both	Database / DICOM	Numeric	Posterior margin in mm
Rectum V50(%)	Both	DICOM	Numeric	Percentage of rectum ≥50% prescription dose
Rectum V80(%)	Both	DICOM	Numeric	Percentage of rectum ≥80% prescription dose
Rectum V100(cc)	SBRT Only	DICOM	Numeric	Amount of rectum (cc) ≥100% prescription dose
GU Models Only				
RTOG GU Baseline G2+	Both	Database	Binary	1 = Grade 1 or more RTOG GU score at baseline
CTCAE GU Baseline G2+	Both	Database	Binary	1 = Grade 1 or more CTCAE GU score at baseline
Baseline IPSS Score	Both	Database	Numeric	0-35 per IPSS scoring
Baseline EPIC-26 UI	Both	Database	Numeric	0-100 per EPIC-26 domain scoring (Urinary Obstructive)
Baseline EPIC-26 UO	Both	Database	Numeric	0-100 per EPIC-26 domain scoring (Urinary Incontinence)
GU meds at baseline	Both	Database	Binary	1 = Anti-cholinergic and/or alpha blocker at baseline
Margin Non-Posterior	Both	Database / DICOM	Numeric	Non-Posterior margins in mm
PTV $Dmax_{EQD10}$	Whole Trial	DICOM	Numeric	PTV Max Dose adjusted by EQD2 (α/β = 10 Gy)
PTV Dmax	SBRT Only	DICOM	Numeric	PTV Max Dose (Gy)
Bladder V50%(%)	Both	DICOM	Numeric	Percentage of bladder ≥50% prescription dose
Bladder V100%(cc)	SBRT Only	DICOM	Numeric	Amount of bladder (cc) ≥100% prescription dose

7.2.4 Model Choice and Variable Selection Technique

The prior studies in this field have utilised simple logistic regression model selection approaches, or have not provided sufficient detail to be certain on the methods applied. It is known that stepwise selection is sensitive to changes in inputted data [160], so consideration was made to improve the variable selection methodology.

Briefly, I therefore decided to adopt a hybrid modelling approach, using bootstrapping to investigate backward-selection model stability, as has been recommended [160]. This is then supplemented with cross-validation (CV) of candidate models (some generated from combinations of common bootstrap predictor frequencies that might have been omitted by backward step-wise selection), in order to provide estimates of performance on unseen data; acting to counter the possibility of overfitting [147]. A CV method was used rather than a holdout set due to a relatively small dataset compared to potential predictors. The process is now described in detail.

7.2.4.1 Three Stage Variable Selection

7.2.4.2 Stage One: Bootstrapped Backward Selection

The first stage was backward selection, as has previously been used in other studies in the field [232,234]. This was bootstrapped, allowing assessment of the sensitivity of predictor selection to data variability [160]. Firstly, for each model, 2000 bootstraps (with replacement) of the model dataset were

created[41]. Bootstrap stratification can be used, without biasing model performance estimates [236], where a there is a desire to preserve class balance[42]. Some specific bootstrap strata of interest were used: endpoint status and SBRT status for the Whole Trial models; endpoint status and Cyberknife status for the SBRT Only models.

For each bootstrap, a backward selection model (i.e. eliminating variables one-by-one from the model with all predictors) was fitted, with variables eliminated when variable $p \geq 0.157$, as has been recommended [160]. For each predictor, the frequency of bootstrap model inclusion was calculated. The most frequently selected 100 unique models[43] were taken forward to stage three for CV model performance assessment.

7.2.4.3 Stage Two: Generating Other Candidate Models

The bootstrap inclusion frequencies were then used to generate further candidate models for stage three performance assessment. Predictors were separated into three categories:

i) Always included predictors (>90% bootstrap inclusion frequency)

ii) Sometimes included predictors (>20%, ≤90% bootstrap inclusion frequency).

iii) Never included predictors (≤20% bootstrap inclusion frequency)

[41] i.e. with patients removed as outlined at start of methods.
[42] Although not every predictor can be stratified as individual strata become too small.
[43] 100 models chosen as a number that could be processed with reasonable speed.

Candidate models were then generated, including the always-included predictors and each unique combination of the sometimes-included predictors. The exact choice of these cut-offs was pragmatic, to keep the maximum number of candidate models per endpoint to a reasonable number for processing. These candidate models also went forward to stage three.

7.2.4.4 Stage Three: Cross-validated Estimation of Candidate Model Performance

The third stage was to assess the performance of each candidate model from stages one and two. This was done using 5-fold CV with an 80:20% data split, to allow testing of model performance on unseen data [237]. For binary endpoints this performance was assessed by AUC (trapezoidal rule); for numeric endpoints by root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{j=1}^{n} (y_j - \hat{y}_j)^2} \tag{13}$$

Where: n = number of patients (iterated by *j*); y = endpoint value (e.g. IPSS score)

The overall candidate model performance was calculated as the average across the five CV folds of the AUC or RMSE. Since only limited stratification was feasible, random assortment of the CV folds may result in some more extreme examples in one CV fold [238]. Therefore, to select an overall best model for each endpoint, the top performing 100 unique models after one 5-fold CV were re-assessed, using 100x 5-fold CV. 100x CV was chosen based on visual assessment of estimate stability, averaging over repeated

CV estimation. Bootstrapping (x1000) was used to provide percentile 95% CIs for estimates, reported for the 1st, 20th and 100th rank models[44].

7.2.4.5 Final Model Selection

The top performing model was then fitted again using every patient with complete data for the endpoint and each selected predictor (complete case). Since more patients could be included than during variable selection (i.e. those missing data for predictors not included in final model), final model prediction accuracy was rechecked with another round of 100x 5-fold CV.

7.2.5 Modelling Diagram

A flow diagram of the complete modelling process is shown in **Figure 59.**

(Regarding figure on next page)

Figure 59. Modelling Flow Diagram

Showing the modelling process follow for each model, to produce the final complete case model estimates and missing data sensitivity analyses.

[44] It is very computationally expensive to nest cross-validation and bootstrapping, so these three were chosen to provide overview of spread of estimates and 95% CIs.

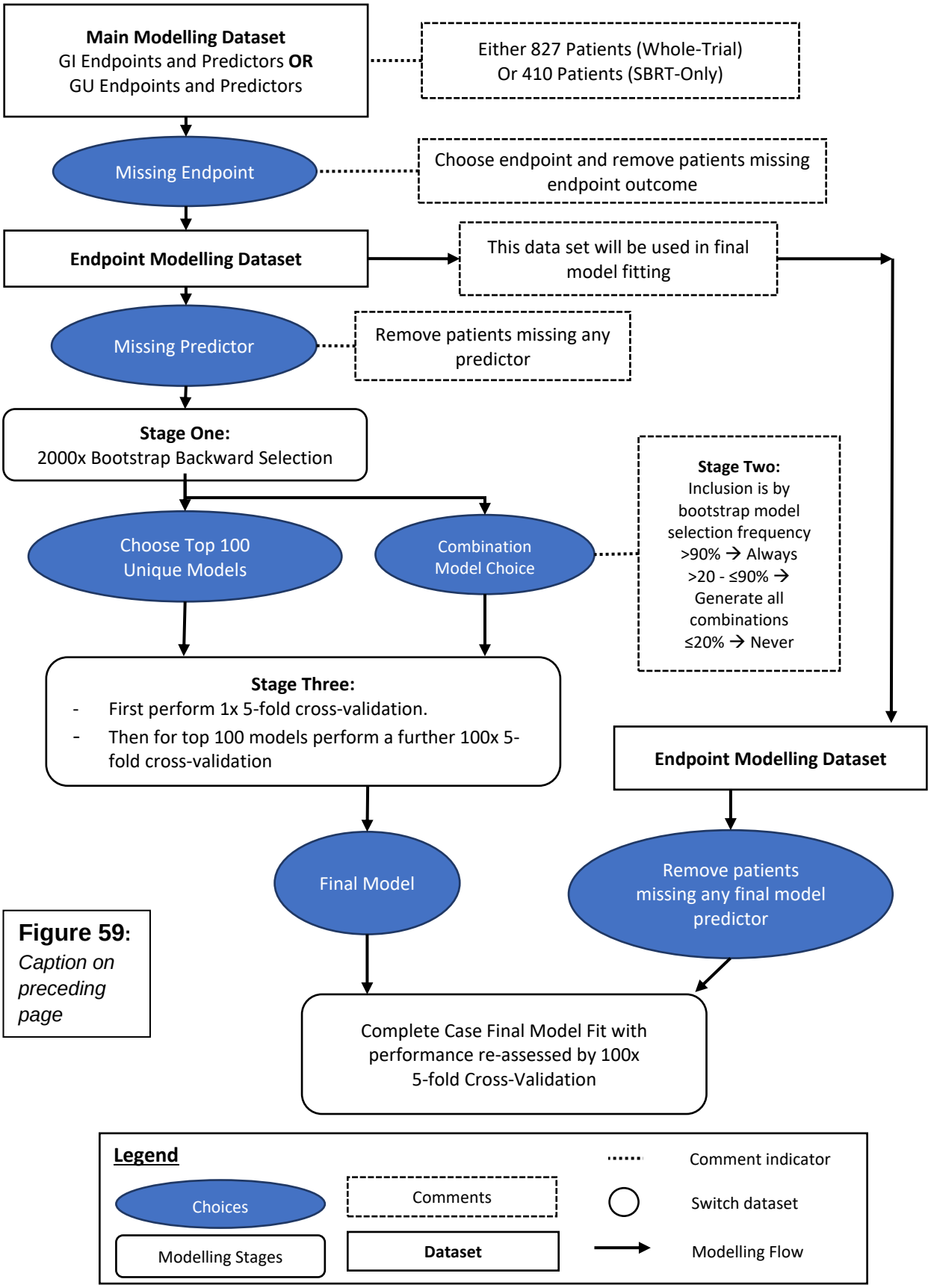


Figure 59: *Caption on preceding page*

7.2.6 Software

VODCA (version 5.4.1, MSS GmbH, Switzerland) was used to load DICOM images, check alignments and save DVHs. Additionally, VODCA was used to inspect plans to find certain data when missing from the main trial database: prostate volume, fiducials, margin sizes. MATLAB (2020a, Mathworks) was used to read DVHs and extract predictors for use in modelling. This was done by converting data to CERR format [224] and then interrogating the CERR files via custom scripts. Stata (Version 16, Statacorp) was used for the handling of predictors and for all modelling work, via custom .do files.

7.3 Results

7.3.1 Included Patients

A total of 874 men were randomised into the PACE-B study. Of these, 827 men received treatment with a protocol radiotherapy regimen, using a single DICOM plan, making them eligible for one or more models in this study. All exclusions are detailed in **Figure 60.**

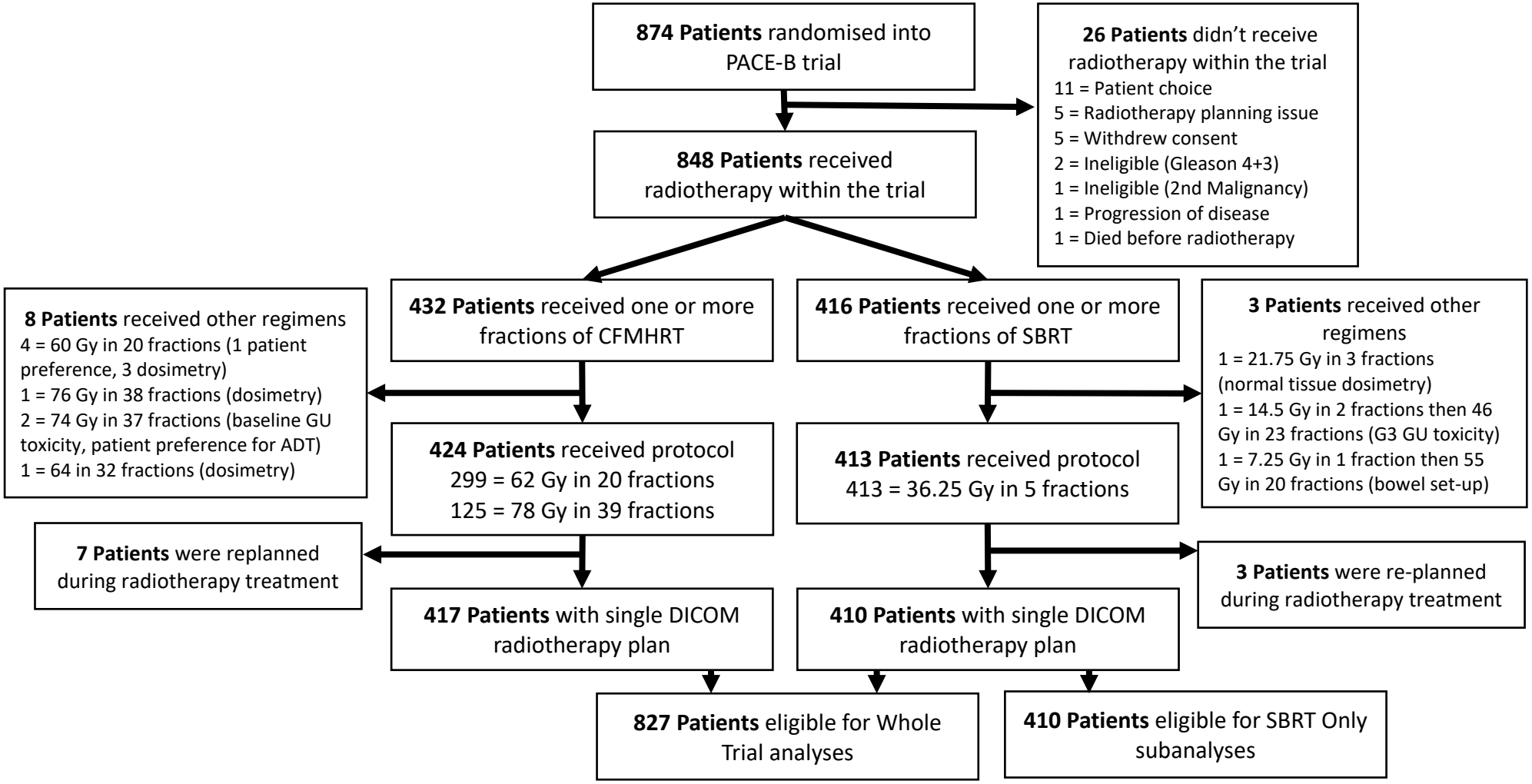


Figure 60. CONSORT Style Patient Eligibility Flowchart

7.3.2 Description of Endpoints Modelled

The endpoint data used for modelling, shown separately for the Whole-Trial and SBRT-Only models, is summarised in **Table 48**. It can be seen that all endpoints have relatively low rates of missing data (0 - 5.6%).

Table 48. Summary of Endpoint Data Used in Modelling

For CRO endpoints (RTOG, CTCAE) this is G2+ at any follow-up from RT to 12 weeks. For EPIC-26, it is worst subdomain score during the same time window.

Endpoint	Whole-Trial Models (n=827)			SBRT-Only Models (n=410)		
	Rate / Median	IQR	Missing n (%)	Rate / Median	IQR	Missing n (%)
GI						
RTOG GI G2+	11.6%	N/A	0 (0%)	10.5%	N/A	0 (0%)
CTCAE GU G2+	11.7%	N/A	1 (0.1%)	15.6%	N/A	0 (0%)
EPIC Bowel	87.5	75 - 100	41 (5%)	87.5	75 - 95.8	15 (3.7%)
GU						
RTOG GU G2+	25.3%	N/A	0 (0%)	22.9%	N/A	0 (0%)
CTCAE GU G2+	26.9%	N/A	1 (0.1%)	30.7%	N/A	0 (0%)
IPSS	13	8 - 19	22 (2.7%)	13	8 - 19	13 (3.2%)
EPIC-26 UI	93.75	79.25 - 100	38 (4.6%)	93.75	79.25 - 100	15 (3.7%)
EPIC-26 UO	81.25	62.5 - 87.5	46 (5.6%)	81.25	62.5 - 87.5	16 (3.9%)

7.3.3 Predictor Data Summary and Missingness

Predictor data, i.e. the patient and treatment related factors to be used for the GI and GU models, are summarised in **Table 49**. It can be seen that missing data was only present for the baseline symptoms scores, along with statin usage (GI only) and GU medications at baseline (GU only). Missingness rates were general low, ranging 0 – 15%.

Table 49. Summary of Predictor Data and Missingness

Frequency of occurrence reported for binary predictors (i.e. % yes).

Median and IQR reported for numerical predictors.

Predictors	**Whole-Trial Models (n=827)**			**SBRT-Only Models (n=410)**		
GI & GU Models	**Frequency / Median**	**IQR**	**Missing n (%)**	**Frequency / Median**	**IQR**	**Missing n (%)**
SBRT	49.6%	N/A	0 (0%)	100%	N/A	0 (0%)
Mod. Hypofractionation	35.8%	N/A	0 (0%)	0%	N/A	0 (0%)
Cyberknife	20.1%	N/A	0 (0%)	40.5%	N/A	0 (0%)
Fiducials	64.4%	N/A	0 (0%)	72.7%	N/A	0 (0%)
RT Time >1 week	N/A	N/A	N/A	79.5%	N/A	0 (0%)
WHO PS 1+	10.8%	N/A	0 (0%)	10.2%	N/A	0 (0%)
Age	69.7	65.5 - 73.9	0 (0%)	69.7	65.3 - 73.9	0 (0%)
Prostate Volume	42	32 - 57	0 (0%)	40	31 - 56	0 (0%)
RT Centre Caseload	77	25 - 112	0 (0%)	77	25 - 112	0 (0%)
GI Models Only						
Baseline RTOG GI G1+	6.3%	N/A	54 (6.5%)	6.5%	N/A	25 (6.1%)
Baseline CTCAE GI G1+	12%	N/A	5 (0.6%)	12.3%	N/A	2 (0.5%)
Baseline EPIC-26 Bowel	100	95.8 - 100	92 (11.1%)	100	91.7 - 100	49 (12%)
Statin Use	42.8%	N/A	7 (0.8%)	40.5%	N/A	5 (1.2%)
Margin Posterior	4	3 - 5	0 (0%)	3	3 - 4	0 (0%)
Rectum V50(%)	41.7	30.3 - 48.7	0 (0%)	33.5	25.1 - 43	0 (0%)
Rectum V80(%)	14.8	10.6 - 19.3	0 (0%)	11.1	8 - 14.7	0 (0%)
Rectum V100(cc)	0.5	0.1 - 0.9	0 (0%)	0.8	0.5 - 1.2	0 (0%)

Continued Overleaf

Table 49 continued…	**Whole-Trial Models (n=827)**			**SBRT-Only Models (n=410)**		
GU Models Only	**Frequency / Median**	**IQR**	**Missing n (%)**	**Frequency / Median**	**IQR**	**Missing n (%)**
RTOG GU Baseline G2+	22.3%	N/A	54 (6.5%)	24.2%	N/A	25 (6.1%)
CTCAE GU Baseline G2+	50.2%	N/A	4 (0.5%)	50.2%	N/A	2 (0.5%)
Baseline IPSS Score	6	3 - 11	114 (13.8%)	6	3 - 12	59 (14.4%)
Baseline EPIC-26 UI	100	85.5 - 100	96 (11.6%)	100	85.5 - 100	52 (12.7%)
Baseline EPIC-26 UO	87.5	81.3 - 100	115 (13.9%)	87.5	81.3 - 100	63 (15.4%)
GU meds at baseline	17.5%	N/A	9 (1.1%)	17.6%	N/A	6 (1.5%)
Margin Non-Posterior	5	5 - 7	0 (0%)	5	4 - 5	0 (0%)
PTV $Dmax_{EQD10}$	71.7	69.2 - 74.5	0 (0%)	69.4	67.1 - 73.1	0 (0%)
PTV Dmax	63	44.2 - 65.2	0 (0%)	44.2	43.2 - 45.8	0 (0%)
Bladder V50%(%)	24.7	16.4 - 34.8	0 (0%)	22.3	14.5 - 33.2	0 (0%)
Bladder V100%(cc)	6.8	3.7 - 10.2	0 (0%)	7.8	5.4 - 10.5	0 (0%)

7.3.4 Bootstrap Selection Frequencies of Predictors

The predictor selection frequencies of the bootstrapped backward elimination models are shown for GI Whole-Trial models in **Table 50**. Frequencies of these selections guided whether predictors would be always included in the combinatorial models (>90% frequency), or never be included (≤20% frequency). SBRT was frequently selected (>90%) for CTCAE and EPIC-26 Bowel, but not RTOG. Similar tables are shown for GI SBRT-Only in **Table 51**

GU Whole-Trial is shown in **Table 52**, with GU SBRT-Only in **Table 53**. All PRO measures (including GI EPIC-26 bowel measures) selected the corresponding baseline score at high frequency (>90%), but this was not true for CRO measures. Amongst GU SBRT-Only models, CK was selected at high frequency for all CRO endpoints (>90%).

Table 50. GI Whole-Trial Models: Predictor Bootstrap Selection Frequencies

Showing frequency of selection of each predictor amongst 2000 backward elimination bootstrap model fits.

Bold = >90% - always included in combinatorial models. Greyed out = ≤20% - never included in combinatorial models.

Red = average covariate deleterious; Green = average covariate protective

RTOG GI G2+		**CTCAE GI G2+**		**EPIC-26 Bowel**	
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	**Predictor**	**Selection Freq.**
Rectum V80(%)	77.4%	**SBRT**	**99.6%**	**Baseline EPIC-26 Bowel**	**100.0%**
Baseline RTOG GI G1+	47.0%	Cyberknife	78.9%	**SBRT**	**100.0%**
Margin Posterior	46.8%	Mod. Hypofractionation	58.2%	Age	81.1%
Cyberknife	42.3%	Statin Use	54.7%	Rectum V80(%)	71.8%
WHO PS 1+	35.1%	Rectum V50(%)	49.4%	Fiducials	71.1%
Rectum V50(%)	31.6%	Margin Posterior	48.8%	Baseline RTOG GI G1+	62.5%
Baseline EPIC-26 Bowel	30.9%	Baseline CTCAE GI G1+	47.7%	Mod. Hypofractionation	41.8%
Age	30.8%	RT Centre Caseload	37.7%	Rectum V50(%)	39.1%
Baseline CTCAE GI G1+	23.7%	Rectum V80(%)	36.7%	Statin Use	24.5%
RT Centre Caseload	22.6%	Baseline EPIC-26 Bowel	35.8%	WHO PS 1+	23.3%
Mod. Hypofractionation	21.3%	WHO PS 1+	35.6%	RT Centre Caseload	19.9%
SBRT	19.2%	Age	29.1%	Baseline CTCAE GI G1+	16.6%
Fiducials	19.0%	Fiducials	23.2%	Cyberknife	15.3%
Prostate Volume	18.8%	Baseline RTOG GI G1+	22.0%	Prostate Volume	14.6%
Statin Use	18.5%	Prostate Volume	13.9%	Margin Posterior	13.2%

Table 51. GI SBRT Only Models: Predictor Bootstrap Selection Frequencies

Showing frequency of selection of each predictor amongst 2000 backward elimination bootstrap model fits.

Bold = >90% so always included in combinatorial models. Greyed out = ≤20% - never included in combinatorial models.

Red = average covariate deleterious; Green = average covariate protective

RTOG GI G2+		**CTCAE GI G2+**		**EPIC-26 Bowel**	
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	**Predictor**	**Selection Freq.**
RT Time >1 week	78.8%	RT Time >1 week	79.8%	**Baseline EPIC-26 Bowel**	**100.0%**
WHO PS 1+	68.7%	WHO PS 1+	62.0%	Age	86.3%
RT Centre Caseload	63.5%	Cyberknife	61.5%	RT Time >1 week	81.2%
Statin Use	62.1%	Statin Use	56.2%	Rectum V80(%)	57.2%
Rectum V100(cc)	53.4%	Rectum V50(%)	51.7%	Fiducials	49.5%
Rectum V80(%)	46.7%	Rectum V80(%)	36.0%	Prostate Volume	46.8%
Rectum V50(%)	40.8%	Age	29.0%	Rectum V50(%)	42.8%
Fiducials	36.3%	RT Centre Caseload	24.5%	WHO PS 1+	29.8%
Age	31.8%	Margin Posterior	23.8%	RT Centre Caseload	25.6%
Margin Posterior	28.7%	Baseline CTCAE GI G1+	22.7%	Margin Posterior	25.1%
Cyberknife	25.8%	Baseline EPIC-26 Bowel	22.5%	Statin Use	19.0%
Baseline CTCAE GI G1+	24.8%	Fiducials	22.2%	Cyberknife	18.9%
Baseline EPIC-26 Bowel	22.2%	Prostate Volume	19.0%	Rectum V100(cc)	17.1%
Prostate Volume	18.8%	Rectum V100(cc)	18.3%	Baseline CTCAE GI G1+	11.9%
Baseline RTOG GI G1+	4.5%	Baseline RTOG GI G1+	18.3%	Baseline RTOG GI G1+	4.6%

Table 52. GU Whole-Trial Models: Predictor Bootstrap Selection Frequencies

Showing frequency of selection of each predictor amongst 2000 backward elimination bootstrap model fits.

Bold = >90% - always included in combinatorial models. Greyed out = ≤20% - never included in combinatorial models.

Red = average covariate deleterious; Green = average covariate protective

RTOG GU G2+		**CTCAE GU G2+**		
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	
Baseline IPSS Score	**92.5%**	**SBRT**	**99.4%**	
Cyberknife	**92.4%**	**Margin Non-Posterior**	**94.9%**	
Prostate Volume	87.9%	**Cyberknife**	**92.1%**	
RTOG GU Baseline G2+	82.1%	Baseline IPSS Score	86.7%	
Margin Non-Posterior	65.7%	Prostate Volume	80.0%	
Mod. Hypofractionation	50.3%	Mod. Hypofractionation	66.3%	
SBRT	45.4%	Age	53.3%	
Bladder V50%(%)	44.0%	Bladder V50%(%)	52.2%	
Baseline EPIC-26 UI	43.9%	RT Centre Caseload	47.7%	
PTV $Dmax_{EQD10}$	37.5%	WHO PS 1+	45.1%	
Age	28.1%	RTOG GU Baseline G2+	37.4%	
RT Centre Caseload	24.2%	Baseline EPIC-26 UI	30.6%	
GU meds at baseline	21.4%	PTV $Dmax_{EQD10}$	28.6%	
Fiducials	19.2%	Fiducials	20.2%	
WHO PS 1+	18.5%	CTCAE GU Baseline G2+	19.1%	
CTCAE GU Baseline G2+	15.0%	GU meds at baseline	14.7%	

Table continued overleaf for PRO endpoints…

…Continuation of Table 52…

IPSS		EPIC-26 UI		EPIC-26 UO	
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	**Predictor**	**Selection Freq.**
Baseline IPSS Score	**100.0%**	**Baseline EPIC-26 UI**	**100.0%**	**Baseline EPIC-26 UO**	**100.0%**
SBRT	78.6%	**WHO PS 1+**	**99.9%**	CTCAE GU Baseline G2+	83.9%
RTOG GU Baseline G2+	75.7%	**Fiducials**	**92.0%**	GU meds at baseline	76.1%
Bladder V50%(%)	66.5%	GU meds at baseline	50.1%	SBRT	75.5%
WHO PS 1+	56.0%	Bladder V50%(%)	42.0%	Age	60.0%
Age	43.3%	Age	29.9%	Margin Non-Posterior	56.4%
Mod. Hypofractionation	42.1%	SBRT	28.3%	Mod. Hypofractionation	51.4%
RT Centre Caseload	41.7%	Margin Non-Posterior	27.5%	Fiducials	47.7%
GU meds at baseline	31.4%	Cyberknife	27.5%	PTV $Dmax_{EQD10}$	38.0%
PTV $Dmax_{EQD10}$	30.1%	PTV $Dmax_{EQD10}$	20.0%	Bladder V50%(%)	37.7%
Fiducials	29.4%	Mod. Hypofractionation	19.8%	Cyberknife	36.8%
Prostate Volume	28.7%	CTCAE GU Baseline G2+	19.7%	Baseline EPIC-26 UI	34.0%
Cyberknife	26.6%	Baseline EPIC-26 UO	16.8%	WHO PS 1+	32.4%
Margin Non-Posterior	23.4%	RT Centre Caseload	16.8%	RT Centre Caseload	27.3%
CTCAE GU Baseline G2+	18.6%	RTOG GU Baseline G2+	11.4%	RTOG GU Baseline G2+	20.2%
Baseline EPIC-26 UI	10.7%	Prostate Volume	10.0%	Prostate Volume	14.4%

Table 53. GU SBRT-Only Models: Predictor Bootstrap Selection Frequencies

Showing frequency of selection of each predictor amongst 2000 backward elimination bootstrap model fits.

Bold = >90% - always included in combinatorial models. Greyed out = ≤20% - never included in combinatorial models.

Red = average covariate deleterious; Green = average covariate protective

RTOG GU G2+		**CTCAE GU G2+**		
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	
Cyberknife	**96.5%**	**Cyberknife**	**99.3%**	
Fiducials	87.6%	Bladder V50%(%)	83.2%	
Prostate Volume	86.8%	Fiducials	79.0%	
RTOG GU Baseline G2+	73.2%	Prostate Volume	74.8%	
Baseline IPSS Score	72.0%	RT Time >1 week	71.8%	
Bladder V50%(%)	58.8%	RT Centre Caseload	71.7%	
Margin Non-Posterior	56.5%	Baseline IPSS Score	64.0%	
Bladder V100(cc)	44.1%	Margin Non-Posterior	50.2%	
RT Centre Caseload	39.2%	RTOG GU Baseline G2+	48.5%	
CTCAE GU Baseline G2+	32.5%	Baseline EPIC-26 UI	46.5%	
WHO PS 1+	26.7%	CTCAE GU Baseline G2+	26.9%	
RT Time >1 week	22.7%	PTV Dmax	26.0%	
Baseline EPIC-26 UI	18.7%	Age	23.7%	
GU meds at baseline	17.9%	Bladder V100(cc)	21.2%	
Age	17.8%	WHO PS 1+	19.1%	
PTV Dmax	16.8%	GU meds at baseline	17.3%	

Table continued overleaf for PRO endpoints…

…Continuation of Table 53…

IPSS		EPIC-26 UI		EPIC-26 UO	
Predictor	**Selection Freq.**	**Predictor**	**Selection Freq.**	**Predictor**	**Selection Freq.**
Baseline IPSS Score	**100.0%**	**Baseline EPIC-26 UI**	**100.0%**	**Baseline EPIC-26 UO**	**100.0%**
RT Time >1 week	76.7%	**WHO PS 1+**	**92.2%**	RT Time >1 week	80.0%
Bladder V100(cc)	39.8%	Fiducials	72.9%	GU meds at baseline	71.6%
CTCAE GU Baseline G2+	37.1%	Bladder V100(cc)	58.1%	Margin Non-Posterior	68.8%
RT Centre Caseload	33.1%	GU meds at baseline	56.8%	Bladder V100(cc)	67.0%
Cyberknife	25.6%	Bladder V50(%)	38.4%	Age	64.2%
RTOG GU Baseline G2+	24.5%	Cyberknife	32.3%	CTCAE GU Baseline G2+	56.1%
PTV Dmax	21.0%	CTCAE GU Baseline G2+	31.7%	RTOG GU Baseline G2+	35.0%
WHO PS 1+	16.6%	RTOG GU Baseline G2+	27.9%	Fiducials	34.1%
Margin Non-Posterior	14.8%	Prostate Volume	24.3%	RT Centre Caseload	30.9%
Prostate Volume	14.6%	RT Centre Caseload	17.6%	PTV Dmax	28.0%
Bladder V50%(%)	14.4%	Baseline EPIC-26 UO	15.0%	WHO PS 1+	24.7%
Fiducials	13.7%	Age	14.3%	Prostate Volume	22.1%
GU meds at baseline	12.1%	RT Time >1 week	13.6%	Baseline EPIC-26 UI	20.7%
Age	10.6%	PTV Dmax	13.6%	Cyberknife	20.3%
Baseline EPIC-26 UI	6.7%	Margin Non-Posterior	7.3%	Bladder V50%(%)	16.0%

7.3.5 Variable Selection Process Summary

7.3.5.1 GI Model Selection

A summary of the top-20 model information for Whole-Trial and SBRT-Only GI models is shown in **Table 54**. Selection of variables following 100x 5-fold CV is contrasted to those following initial backward stepwise selection (in brackets).

The importance of dosimetry is clear, with different choices in each model likely due to collinearity inherent to DVH derived dosimetry (outlined above in **7.2.3.5**). CRO and PRO predictors tend to be chosen for their respective models, save SBRT-Only CTCAE G2+ GI, where neither is commonly selected. There is consistent selection of overall treatment time ≥1 week in the SBRT-Only models

Predictor selection frequencies are generally similar between the bootstrapped backward selection and cross-validation selection. Larger differences (>10/20) were seen in the Whole-Trial RTOG GI G2+ model, with Cyberknife in more top-20 models following cross-validation (20/20) than after bootstrapped backward selection (7/20). In the SBRT-Only RTOG GI G2+ model, Baseline EPIC-26 Bowel was selected in 17/20 top cross-validated models, but only 1/20 top bootstrap backward-selected models.

Table 54. GI Models: Summary of Predictors in Top-20 Models

The frequency of predictor selection in the top-20 ranked models following 100x 5-fold cross-validation are shown. In brackets are the corresponding frequency of selection amongst the top 20 unique models selected by initial bootstrapped backward stepwise selection.

White on Dark = selected in 20/20 models; **Dark highlight** = selected in 15-19/20. **Light highlight** = selected in 10–14/20, Faded text = 0/20

Predictor	Whole Trial			SBRT Only		
	RTOG GI G2+	CTCAE GI G2+	EPIC-26 Bowel	RTOG GI G2+	CTCAE GI G2+	EPIC-26 Bowel
SBRT	0 (0)	**20 (20)**	**20 (20)**	N/A	N/A	N/A
Mod Hypofractionation	2 (0)	13 (17)	7 (5)	N/A	N/A	N/A
Cyberknife	**20 (7)**	14 (20)	1 (0)	5 (3)	**20 (18)**	1 (3)
Fiducials	0 (1)	0 (0)	**19 (19)**	3 (5)	2 (1)	14 (11)
RT Time >1 week	N/A	N/A	N/A	**20 (20)**	**20 (20)**	**19 (19)**
WHO PS 1+	2 (3)	0 (4)	1 (2)	**19 (16)**	**17 (14)**	1 (4)
Age	6 (4)	0 (1)	**19 (18)**	2 (2)	0 (5)	**20 (19)**
Prostate Volume	1 (0)	0 (0)	1 (2)	0 (0)	0 (0)	6 (10)
RT Centre Caseload	0 (2)	1 (7)	3 (1)	**19 (16)**	1 (0)	2 (4)
Statin Use	1 (1)	**20 (11)**	3 (2)	12 (17)	**19 (14)**	0 (1)
Margin Posterior	**17 (8)**	7 (12)	1 (0)	0 (2)	1 (2)	7 (5)
Any CRO Predictor	**18 (14)**	**18 (12)**	6 (12)	14 (3)	7 (4)	1 (0)
Baseline RTOG GI G1+	**15 (11)**	0 (1)	6 (12)	0 (0)	5 (1)	0 (0)
Baseline CTCAE GI G1+	5 (3)	**18 (12)**	0 (2)	14 (3)	3 (3)	1 (0)
Any PRO Predictor	2 (3)	4 (3)	**20 (20)**	**17 (1)**	3 (2)	**20 (20)**
Baseline EPIC-26 Bowel	2 (3)	4 (3)	**20 (20)**	**17 (1)**	3 (2)	**20 (20)**
Any Rectal Dosimetry	**20 (20)**	**17 (19)**	**20 (20)**	**18 (16)**	**20 (19)**	**20 (20)**
Rectum V50(%)	3 (4)	11 (12)	4 (5)	10 (5)	10 (15)	7 (9)
Rectum V80(%)	**20 (20)**	6 (7)	**20 (16)**	8 (11)	11 (4)	14 (11)
Rectum V100(cc)	N/A	N/A	N/A	10 (11)	1 (0)	3 (1)

The performance metrics (AUC/RMSE) and confidence intervals for the 1st, 20th and 100th ranked models are shown in **Table 55**, in order to give a feel for the spread of model performances. It is readily apparent that although performance estimates decrease from rank 1 → rank 100, the confidence intervals for these estimates are highly overlapping. This implies uncertainty in the status of the true "best" model.

For CRO models, although the AUC metrics are fairly low, the lower bound of the AUC 95% CI does not cross the 0.5 (no effect) for top-20 ranked models.

Table 55. Performance Metrics for 1st, 20th and 100th Ranked GI Models
Recall a larger AUC is better for CRO models, while a smaller RMSE is better for PRO models.

Endpoint	Model	Perf. Metric	Rank 1 Perf. Metric	Rank 1 95% CI	Rank 20 Perf. Metric	Rank 20 95% CI	Rank 100 Perf. Metric	Rank 100 95% CI
RTOG GI G2+	Whole Trial	AUC	**0.58**	0.51 - 0.65	**0.56**	0.5 - 0.66	**0.54**	0.49 - 0.66
CTCAE GI G2+	Whole Trial	AUC	**0.63**	0.59 - 0.69	**0.62**	0.58 - 0.69	**0.60**	0.57 - 0.68
EPIC-26 Bowel	Whole Trial	RMSE	**17.5**	16.2 - 18.7	**17.6**	16.2 - 18.7	**17.6**	16.3 - 18.7
RTOG GI G2+	SBRT Only	AUC	**0.63**	0.55 - 0.77	**0.6**	0.53 - 0.74	**0.56**	0.46 - 0.74
CTCAE GI G2+	SBRT Only	AUC	**0.64**	0.56 - 0.74	**0.63**	0.56 - 0.74	**0.59**	0.53 - 0.74
EPIC-26 Bowel	SBRT Only	RMSE	**19**	16.8 - 20.8	**19.1**	16.9 - 20.8	**19.3**	17.1 - 20.8

7.3.5.2 GU Model Selection

A summary of the top-20 model information for Whole-Trial and SBRT-Only GU models is shown in **Table 56**. The crucial importance of baseline PRO predictors is very clear, being fitted as one predictor in 100% of top-20 models for every GU endpoint.

Compared to GI models, the role of dosimetry is less clear, although it appears important in SBRT-Only models other than EPIC-26 UI, which is quite different to the other endpoints. The very strong selection of CK in 100% of top-20 CRO models is apparent, as well as the scant selection of this in PRO models.

Overall treatment time >1 week is strongly selected in several SBRT-Only models (except RTOG and EPIC-26 UI), which complements the findings seen in GI models.

[Caption for table overleaf]

Table 56 GU Models: Summary of Predictors in Top-20 Models

The frequency of predictor selection in the top-20 ranked models following 100x 5-fold cross-validation are shown. In brackets are the corresponding frequency of selection amongst the top 20 unique models selected by initial bootstrapped backward stepwise selection. Note that PTV Dmax is EQD2 adjusted (α/β = 10 Gy) for Whole-Trial models.

White on Dark = selected in 20/20 models; **Dark highlight** = selected in 15-19/20. **Light highlight** = selected in 10–14/20, Faded text = 0/20

Table 56. GU Models: Summary of Predictors in Top-20 Models

[*Caption on preceding page*]

Predictor	Whole-Trial					SBRT-Only				
	RTOG GU G2+	CTCAE GU G2+	IPSS	EPIC-26 UI	EPIC-26 UO	RTOG GU G2+	CTCAE GU G2+	IPSS	EPIC-26 UI	EPIC-26 UO
SBRT	1 (8)	**20 (20)**	14 (20)	2 (3)	**19 (20)**	N/A	N/A	N/A	N/A	N/A
Mod Hypofractionation	**15 (10)**	**16 (16)**	10 (7)	1 (0)	13 (12)	N/A	N/A	N/A	N/A	N/A
Cyberknife	**20 (20)**	**20 (20)**	0 (2)	3 (2)	0 (1)	**20 (20)**	**20 (20)**	5 (6)	1 (6)	0 (0)
Fiducials	0 (0)	3 (0)	4 (1)	**20 (20)**	7 (8)	**20 (20)**	**19 (19)**	3 (0)	**18 (13)**	0 (7)
RT Time >1 week	N/A	N/A	N/A	N/A	N/A	1 (1)	**20 (19)**	11 (17)	2 (0)	**19 (17)**
WHO PS 1+	0 (0)	5 (8)	12 (16)	**20 (20)**	4 (4)	6 (3)	1 (0)	0 (1)	**20 (20)**	0 (1)
Age	0 (1)	**16 (11)**	0 (4)	3 (3)	**17 (15)**	0 (0)	3 (1)	0 (0)	1 (0)	14 (13)
Prostate Volume	14 (19)	**20 (20)**	3 (1)	0 (0)	0 (0)	**20 (20)**	**20 (17)**	0 (0)	2 (4)	4 (1)
RT Centre Caseload	1 (4)	4 (10)	4 (13)	0 (1)	1 (1)	3 (5)	14 (16)	1 (6)	2 (1)	0 (1)
GU meds at baseline	5 (2)	0 (0)	7 (2)	8 (11)	**19 (19)**	0 (0)	1 (1)	0 (0)	8 (12)	**19 (18)**
Margin Non-Posterior	7 (16)	**20 (20)**	0 (4)	0 (2)	**17 (12)**	11 (13)	3 (7)	0 (0)	0 (0)	**19 (17)**
PTV Dmax *	0 (6)	1 (3)	7 (1)	0 (0)	2 (3)	1 (0)	5 (2)	0 (3)	1 (0)	3 (1)
Any CRO Predictor	**20 (20)**	1 (9)	**20 (20)**	4 (5)	**20 (20)**	**19 (14)**	3 (9)	11 (10)	3 (3)	12 (13)
Baseline RTOG GU G2+	**20 (20)**	1 (9)	**20 (20)**	2 (1)	0 (0)	**19 (14)**	1 (9)	6 (4)	1 (3)	5 (8)
Baseline CTCAE GU G2+	0 (0)	0 (0)	0 (1)	2 (4)	**20 (20)**	0 (3)	2 (0)	5 (6)	2 (3)	11 (12)
Any PRO Predictor	**20 (20)**	**20 (20)**	**20 (20)**	**20 (20)**	**20 (20)**	**19 (19)**	**20 (18)**	**20 (20)**	**20 (20)**	**20 (20)**
Baseline IPSS Score	**20 (20)**	**20 (20)**	**20 (20)**	N/A	N/A	**19 (19)**	**20 (14)**	**20 (20)**	N/A	N/A
Baseline EPIC-26 UI	7 (10)	2 (1)	0 (0)	**20 (20)**	1 (2)	0 (0)	3 (6)	0 (0)	**20 (20)**	2 (1)
Baseline EPIC-26 UO	N/A	N/A	N/A	3 (2)	**20 (20)**	N/A	N/A	N/A	1 (0)	**20 (20)**
Any Bladder Dosimetry	7 (5)	9 (10)	**20 (15)**	7 (8)	2 (3)	**16 (14)**	**20 (20)**	10 (10)	3 (13)	**18 (15)**
Bladder V50(%)	7 (5)	9 (10)	**20 (15)**	7 (8)	2 (3)	14 (13)	**20 (20)**	2 (1)	2 (7)	0 (0)
Bladder V100(cc)	N/A	N/A	N/A	N/A	N/A	9 (10)	3 (0)	8 (9)	1 (12)	**18 (15)**

The performance metrics (AUC/RMSE) and confidence intervals for the 1st, 20th and 100th ranked GU models are shown in **Table 57**, in order to give a feel for the spread of model performances. AUC metrics are low-intermediate, although generally better than GI models. The lower bound of CRO model AUC 95% CIs do not cross the 0.5 (no effect), even for the 100th ranked models. As with the GI models, it is important to note that the 95% CI for the 1st ranked model (CRO & PRO) always encompasses the 100th rank model point performance estimate. This implies many models with statistically similar performance.

Table 57. Performance Metrics for 1st, 20th and 100th Ranked GU Models

Endpoint	Model	Perf. Metric	Rank 1 Perf. Metric	Rank 1 95% CI	Rank 20 Perf. Metric	Rank 20 95% CI	Rank 100 Perf. Metric	Rank 100 95% CI
RTOG GU G2+	Whole Trial	AUC	**0.67**	0.62 - 0.73	**0.66**	0.61 - 0.72	**0.65**	0.61 - 0.71
CTCAE GU G2+	Whole Trial	AUC	**0.67**	0.63 - 0.72	**0.67**	0.63 - 0.73	**0.66**	0.62 - 0.72
IPSS	Whole Trial	RMSE	**6.6**	6.2 - 7	**6.6**	6.1 - 6.9	**6.6**	6.2 - 7
EPIC-26 UI	Whole Trial	RMSE	**14.3**	13.1 - 15.5	**14.3**	12.9 - 15.4	**14.4**	13.1 - 15.5
EPIC-26 UO	Whole Trial	RMSE	**16.8**	15.6 - 17.7	**16.8**	15.6 - 17.8	**16.9**	15.7 - 17.9
RTOG GU G2+	SBRT Only	AUC	**0.73**	0.68 - 0.8	**0.72**	0.68 - 0.8	**0.71**	0.67 - 0.78
CTCAE GU G2+	SBRT Only	AUC	**0.70**	0.64 - 0.77	**0.69**	0.64 - 0.77	**0.67**	0.63 - 0.76
IPSS	SBRT Only	RMSE	**6.3**	5.7 - 6.9	**6.4**	5.7 - 6.9	**6.4**	5.8 - 6.9
EPIC-26 UI	SBRT Only	RMSE	**14.8**	12.7 - 16.5	**14.8**	12.6 - 16.5	**14.9**	12.8 - 16.6
EPIC-26 UO	SBRT Only	RMSE	**17.4**	15.5 - 18.6	**17.4**	15.7 - 18.7	**17.6**	15.7 - 18.9

7.3.6 Final Model Fits

7.3.6.1 Whole-Trial RTOG GI G2+ Final Model

The final complete case Whole-Trial logistic model fit for RTOG GI G2+ is shown in **Table 58.** Key information on the model is as follows: n = 773, events = 86, events per predictor = 21.5, overall model chi-square p-value 0.017, training AUC 0.61 (95% CI 0.54 – 0.67), test AUC (5-fold CV x 100) = 0.59 (95% CI 0.53 – 0.66). Note that the confidence intervals in complete case final fits differ from those earlier. This is due to inclusion of patients with complete case data for the final model predictors, who may have been excluded earlier due to missing data in a candidate predictor.

Table 58. Whole-Trial RTOG GI G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Cyberknife	0.568	0.258	1.251	0.1602
Baseline RTOG GI G1+	2.265	1.079	4.755	**0.0307**
Margin Posterior	0.9	0.738	1.098	0.3007
Rectum V80(%)	1.044	1.008	1.082	**0.0169**
Constant	0.102	0.038	0.272	**<0.0001**

Relevant to all models that will be presented, it should be remembered that 95% CIs for the coefficients shown may cross no-effect, yet inclusion of the predictor still improves the prediction of the model in cross-validation. Initial discussion will be limited to those predictors with statistically significant coefficients. Consideration will be made later towards the direction of effect of those non-significant predictors, although clearly a larger sample would be needed to constrain the (often) broad intervals shown.

A relatively simple model has been specified for Whole-Trial RTOG GI G2+, with significant detrimental effects from baseline RTOG G1+ and Rectum V80(%). Recall that the constant is not a predictor, but a component of the logistic model – can be thought of as representing risk in the absence of any predictor[45].

7.3.6.1.1 Whole-Trial RTOG GI G2+ Sensitivity Analysis

Although not at statistical significance, the surprising direction of risk effect from posterior margin is noted (more margin reduces risk). Sensitivity analysis was therefore performed (same n=773), excluding non-posterior margin from the model, to see the effect on other covariates (**Table 59**).

Table 59. Whole-Trial RTOG GI G2+ Sensitivity Analysis

The final model fitted with the exclusion of posterior margin as a predictor.

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Cyberknife	0.715	0.370	1.382	0.318
Baseline RTOG GI G1+	2.250	1.073	4.716	**0.032**
Margin Posterior	Excluded	Excluded	Excluded	Excluded
Rectum V80(%)	1.037	1.003	1.072	**0.031**
Constant	0.068	0.038	0.272	**<0.0001**

While some minor changes are seen to coefficients, there is no change to the direction of effect, nor significance for the remaining predictors.

[45] Though the constant value is somewhat theoretical if including continuous variables that are unlikely to ever equal zero.

7.3.6.2 Whole-Trial CTCAE GI G2+ Final Model

The top Whole-Trial CTCAE GI G2+ model from 100x CV was selected as the final model. The final complete case Whole-Trial logistic model fit for CTCAE GI G2+ is shown in **Table 60.** Key information on the model is as follows: n = 814, events = 95, events per predictor = 19, overall model chi-square p-value = 0.0011, training AUC 0.65 (95%CI 0.59 - 0.70), test AUC (5-fold CV x 100) = 0.62 (95% CI 0.58 – 0.68).

Table 60. Whole-Trial CTCAE GI G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
SBRT	2.981	1.634	5.44	**0.0004**
Cyberknife	0.625	0.344	1.134	0.1221
Baseline CTCAE GI G1+	2.007	1.143	3.523	**0.0152**
Statin Use	0.853	0.547	1.329	0.4819
Rectum V50(%)	1.014	0.994	1.035	0.1832
Constant	0.044	0.015	0.127	**<0.0001**

Here, SBRT and baseline CTCAE G1+ are significant risk factors for toxicity. SBRT with Cyberknife carries an overall 1.863x odds ratio over CFMHRT patients within the model.

7.3.6.3 Whole-Trial EPIC-26 Bowel Final Model

The top Whole-Trial EPIC-26 Bowel model from 100x CV was selected as the final model. The final complete case Whole-Trial multivariable linear model fit for EPIC-26 Bowel is shown in **Table 61.** Key information on the model is as follows: n = 711, overall model F-test p-value < 0.0001, training RMSE 17.1 (95% CI 15.9 – 18.3), test RMSE (5-fold CV x 100) = 17.3 (95% CI 16.0 – 18.3) .

Given the identical unit scales, the coefficient for EPIC-26 Bowel baseline score (0.686) is very high, indicating that 2/3 of the baseline score goes into the prediction of EPIC-26 acute toxicity scoring. Strong negative effects from SBRT and Rectum V80(%) are seen. Fiducials are protective. Age is protective, a finding addressed in discussion.

Table 61. Whole-Trial EPIC-26 Bowel Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
SBRT	-7.605	-10.81	-4.401	**<0.0001**
Fiducials	3.392	0.695	6.089	**0.0138**
Age	0.204	0.001	0.407	**0.049**
Baseline EPIC-26 Bowel	0.686	0.532	0.839	**<0.0001**
Rectum V80(%)	-0.471	-0.71	-0.233	**0.0001**
Constant	12.496	-8.289	33.282	0.2383

7.3.6.4 SBRT-Only RTOG GI G2+ Final Model

The top SBRT-Only RTOG GI G2+ model from 100x CV was selected as the final model. The final complete case Whole-Trial logistic model fit for SBRT-Only RTOG GI G2+ is shown in **Table 62.** Key information on the model is as follows: n = 335, events = 39, events per predictor = 4.86, overall model chi-square p-value = 0.024, training AUC 0.70 (95% CI 0.61 – 0.77), test AUC (5-fold CV x 100) = 0.62 (95% CI 0.55 – 0.76).

This is a relatively complicated model with a fairly low events per predictor. A strong protective effect is seen from overall treatment time over 1 week. Increased centre caseload (i.e. treating more patients) in the PACE-B trial is also protective. We see a strong adverse effect from Rectum V80(%).

Table 62. SBRT-Only RTOG GI G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
RT Time >1 week	0.391	0.175	0.875	**0.0222**
WHO PS 1+	2.619	0.941	7.286	0.0652
RT Centre Caseload	0.992	0.985	0.998	**0.0166**
Baseline CTCAE GI G1+	0.475	0.149	1.514	0.2082
Baseline EPIC-26 Bowel	0.975	0.941	1.009	0.152
Statin Use	0.575	0.274	1.207	0.1434
Rectum V80(%)	1.103	1.008	1.206	**0.032**
Rectum V100(cc)	0.539	0.238	1.22	0.1379
Constant	3.17	0.08	125.619	0.5388

7.3.6.4.1 SBRT-Only RTOG GI G2+ Sensitivity Analysis

The final model has been fitted with a protective direction of effect for Baseline CTCAE GI G1+ & Rectum V100(cc) volume, which are clinically unlikely directions of effect. Sensitivity analyses (same n=355) were performed removing both Rectum V100(cc) and CTCAE GI G1+ (**Table 63**)[46]. The direction of effect is similar for all coefficients, although Rectum V80(%) is no longer significant.

Table 63. SBRT-Only RTOG GI G2+ Sensitivity Analysis

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
RT Time >1 week	0.435	0.198	0.956	**0.038**
WHO PS 1+	2.131	0.790	5.754	0.135
RT Centre Caseload	0.993	0.986	0.999	**0.034**
Baseline CTCAE GI G1+	N/A	N/A	N/A	N/A
Baseline EPIC-26 Bowel	0.984	0.951	1.018	0.358
Statin Use	0.636	0.306	1.322	0.225
Rectum V80(%)	1.067	0.983	1.158	0.123
Rectum V100(cc)	N/A	N/A	N/A	N/A
Constant	0.870	0.026	29.449	0.938

[46] Individually removing each of Rectum V100(cc) and CTCAE GI G1+ was checked, however removal of either did not resolve the unusual coefficient direction of the other. Fit tables omitted for space reasons.

7.3.6.5 SBRT-Only CTCAE GI G2+ Final Model

The top SBRT-Only CTCAE GI G2+ model from 100x CV was selected as the final model. The final complete case SBRT-Only logistic model fit for CTCAE GI G2+ is shown in **Table 64.** Key information on the model is as follows: n = 405, events = 63, events per predictor = 12.6, overall model chi-square p-value = 0.0128, training AUC = 0.65 (95% CI 0.58 – 0.73), test AUC (5-fold CV x 100) = 0.62 (95% CI 0.54 – 0.72).

The only predictor to reach significance is an adverse effect for performance status ≥1. The protective effect for overall treatment time >1 week is close to significance.

Table 64. SBRT-Only CTCAE GI G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Cyberknife	0.6	0.329	1.096	0.0965
RT Time >1 week	0.531	0.278	1.015	0.0556
WHO PS 1+	2.418	1.108	5.276	**0.0266**
Statin Use	0.611	0.339	1.099	0.0998
Rectum V80(%)	1.061	0.993	1.134	0.0812
Constant	0.195	0.068	0.564	**0.0025**

7.3.6.6 SBRT-Only EPIC-26 Bowel Final Model

The top SBRT-Only EPIC-26 Bowel model from 100x CV was selected as the final model. The final complete case SBRT-Only multivariable linear model fit for EPIC-26 Bowel is shown in **Table 65.** Key information on the model is as follows: n = 354, overall model F-test p-value <0.0001, training RMSE 18.5 (95%CI 16.6 – 20.0) , test RMSE (5-fold CV x 100) = 18.8 (16.7 – 20.4).

Given identical unit-scale to the endpoint, baseline EPIC-26 Bowel is a very strong predictor (0.703 coefficient). Once again, as for the RTOG GI SBRT-Only models, we see a protective effect (positive coefficient) for overall treatment time >1 week. A protective effect for fiducial markers is seen here, but was not seen in the CRO models. Rectum V80(%) is significantly detrimental.

Table 65. SBRT-Only EPIC-Bowel Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
Fiducials	4.63	0.225	9.036	**0.0394**
RT Time >1 week	5.115	0.198	10.033	**0.0415**
Age	0.299	-0.007	0.605	0.0557
Baseline EPIC-26 Bowel	0.703	0.478	0.929	**<0.0001**
Rectum V80(%)	-0.754	-1.215	-0.294	**0.0014**
Constant	-5.082	-35.438	25.274	0.7422

7.3.6.7 Whole-Trial RTOG GU G2+ Final Model

The top Whole-Trial RTOG GU G2+ model from 100x CV was selected as the final model. The final complete case Whole-Trial logistic model fit for RTOG GU G2+ is shown in **Table 66.** Key information on the model is as follows: n = 637, events = 156, events per predictor = 22.3, overall model chi-square p-value <0.0001, training AUC 0.69 (95% CI 0.64 – 0.73), test AUC (5-fold CV x 100) = 0.67 (95%CI 0.63 – 0.73).

Here we see a very strong protective effect from Cyberknife. For both RTOG and IPSS baseline, higher scores predict for toxicity. Prostate volume is seen as a significant detrimental factor, although the coefficient is not particularly large in clinical magnitude; a 10cc increase would equate to a 1.14 odds ratio; a 40cc increase being only a 1.68 odds ratio.

Table 66. Whole-Trial RTOG GU G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Mod Hypofractionation	0.636	0.381	1.06	0.0827
Cyberknife	0.246	0.12	0.506	**0.0001**
Prostate Volume	1.013	1.004	1.022	**0.0058**
Baseline RTOG GU G2+	1.746	1.126	2.708	**0.0127**
Baseline IPSS Score	1.049	1.016	1.082	**0.0029**
Baseline EPIC-26 UI	1.01	0.995	1.024	0.1912
Margin Non-Posterior	1.111	0.958	1.289	0.1645
Constant	0.033	0.006	0.19	**0.0001**

7.3.6.7.1 Whole-Trial RTOG GU G2+ Sensitivity Analysis

In the above model, although not at significance, the EPIC-26 UI baseline score is fitted as a weak detrimental factor; implying better baseline function would predict for increased likelihood of an RTOG GU G2+ event at any time. Given the potential for overfitting, a sensitivity analysis was performed (same n=637) to check effect on coefficients from exclusion of Baseline EPIC-26 UI (**Table 67**). There are no substantive changes to coefficients, nor significance.

Table 67. Whole-Trial RTOG GU G2+ Sensitivity Analysis

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Mod Hypofractionation	0.649	0.390	1.082	0.097
Cyberknife	0.248	0.121	0.510	**<0.0001**
Prostate Volume	1.013	1.004	1.022	**0.006**
Baseline RTOG GU G2+	1.721	1.112	2.665	**0.015**
Baseline IPSS Score	1.040	1.011	1.070	**0.007**
Baseline EPIC-26 UI	N/A	N/A	N/A	N/A
Margin Non-Posterior	1.106	0.954	1.283	0.183
Constant	0.088	0.034	0.229	**<0.0001**

7.3.6.8 Whole-Trial CTCAE GU G2+ Final Model

The top Whole-Trial CTCAE GU G2+ model from 100x CV was selected as the final model. The final complete case Whole-Trial logistic model fit for CTCAE GU G2+ is shown in **Table 68.** Key information on the model is as follows: n = 712, events = 193, events per predictor = 27.6, overall model chi-square p-value <0.0001, training AUC 0.68 (95% CI 0.64 – 0.73), test AUC (5-fold CV x 100) = 0.67 (95% CI 0.63 – 0.72).

Here we see a strong adverse odds ratio for SBRT (5.3). Cyberknife's protective odds ratio implies that relative to 78 Gy in 39 fraction patients, for those having SBRT the overall odds ratio would be 2.4. Prostate volume is selected, with again a significant but modest adverse effect. Here the odds ratio for a 10cc increase would be 1.13, while for a 40cc increase it would be 1.61. Likewise, as for RTOG, non-posterior margin size is selected by the model, independently to the different regimen predictors. Here a 3mm margin increase[47] will double your odds of CTCAE GU G2+ acute toxicity.

Table 68. Whole-Trial CTCAE GU G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
SBRT	5.273	2.622	10.606	**<0.0001**
Mod Hypofractionation	1.354	0.726	2.523	0.3409
Cyberknife	0.457	0.272	0.767	**0.0031**
Age	1.021	0.993	1.05	0.1462
Prostate Volume	1.012	1.004	1.021	**0.0055**
Baseline IPSS Score	1.05	1.022	1.078	**0.0004**
Margin Non-Posterior	1.301	1.099	1.542	**0.0023**
Constant	0.003	0.0003	0.034	**<0.0001**

[47] 1.3 x 1.3 x 1.3 ≈ 2

7.3.6.9 Whole-Trial IPSS Final Model

The top Whole-Trial IPSS model from 100x CV was selected as the final model. The final complete case Whole-Trial multivariable linear model fit for IPSS (worst acute score) is shown in **Table 69.** Key model information: n = 663, overall model F-test p-value <0.0001, training RMSE 6.45 (95% CI 6.04 – 6.80), test RMSE (5-fold CV x 100) = 6.5 (95% CI 6.1 – 6.9).

Both SBRT and moderate hypofractionation are detrimental relative to 2 Gy per fraction (n.b. higher IPSS score is worse). Unlike the CRO endpoints, a detrimental dosimetric factor (bladder V50(%)) is included. Baseline RTOG and IPSS are both significant independent detrimental factors.

Table 69. Whole-Trial IPSS Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
SBRT	2.467	0.894	4.041	**0.0022**
Mod Hypofractionation	1.671	0.061	3.281	**0.0419**
WHO PS 1+	1.293	-0.335	2.92	0.1193
Baseline RTOG GU G2+	1.707	0.497	2.917	**0.0057**
Baseline IPSS Score	0.534	0.455	0.613	**<0.0001**
Bladder V50(%)	0.038	0.001	0.075	**0.042**
Constant	6.169	4.249	8.088	**<0.0001**

7.3.6.10 Whole-Trial EPIC-26 Urinary Incontinence Final Model

The top Whole-Trial EPIC-26 UI model from 100x CV was selected as the final model. The final complete case Whole-Trial multivariable linear model fit for EPIC-26 UI is shown in **Table 70.** Key information on the model is as follows: n = 700, overall model F-test p-value <0.0001, training RMSE 14.3 (95% CI 13.0 – 15.3) , test RMSE (5-fold CV x 100) = 14.4 (13.1 – 15.5).

A simple model, there is a protective effect of fiducials on urinary incontinence. Baseline EPIC-UI status and performance status are unsurprising detrimental factors.

Table 70. Whole-Trial EPIC-26 UI Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
Fiducials	2.97	0.75	5.19	**0.0088**
WHO PS 1+	-8.934	-12.468	-5.4	**<0.0001**
Baseline EPIC-26 UI	0.593	0.519	0.666	**<0.0001**
Constant	31.791	24.807	38.775	**<0.0001**

7.3.6.11 Whole-Trial EPIC-26 Urinary Obstructive Final Model

The top Whole-Trial EPIC-26 UO model from 100x CV was selected as the final model. The final complete case Whole-Trial multivariable linear model fit for EPIC-26 UO is shown in **Table 71.** Key model information: n = 672, overall model F-test p-value <0.0001, training RMSE 16.6 (95% CI 15.4 – 17.5), test RMSE (5-fold CV x 100) = 16.8 (95% CI 15.6 – 17.7).

As for the similar scale IPSS, we see SBRT as an important detrimental factor. This is another model where CRO and PRO baseline predictors are fitted as significant independent predictors of worse QoL. Perhaps expectedly, GU medications at baseline is a significantly detrimental predictor.

Table 71. Whole-Trial EPIC-26 UO Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
SBRT	-6.493	-10.933	-2.052	**0.0042**
Mod Hypofractionation	-3.478	-7.714	0.757	0.1073
Age	0.145	-0.059	0.348	0.1632
Baseline CTCAE GU G2+	-3.49	-6.239	-0.742	**0.0129**
Baseline EPIC-26 UO	0.442	0.339	0.545	**<0.0001**
GU meds at baseline	-4.655	-8.029	-1.281	**0.0069**
Margin Non-Posterior	-0.906	-2.093	0.281	0.1346
Constant	41.017	22.846	59.187	**<0.0001**

7.3.6.12 SBRT-Only RTOG GU G2+ Final Model

The top SBRT-Only RTOG GI G2+ model from 100x CV was selected as the final model. The final complete case SBRT-Only logistic model fit for RTOG GU G2+ is shown in **Table 72.** Key information on the model is as follows: n = 334, events = 75, events per predictor = 12.5, overall model chi-square p-value <0.0001, training AUC = 0.75 (0.69 – 0.81), test AUC (5-fold CV x 100) = 0.73 (0.69 – 0.79).

We see a very strong protective effect from the use of Cyberknife. Prostate volume is a significant adverse risk factor. The coefficient is larger than Whole-Trial CRO models (i.e. >1.02), however a 10cc increase would still only equate to a 1.23x odds ratio; a 40 cc increase being a 2.3x odds ratio. Baseline RTOG GU G2+ is significantly detrimental.

Table 72. SBRT-Only RTOG GU G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Cyberknife	0.184	0.086	0.392	**<0.0001**
Fiducials	1.764	0.937	3.32	0.0785
Prostate Volume	1.021	1.007	1.035	**0.0034**
Baseline RTOG GU G2+	2.211	1.177	4.151	**0.0136**
Baseline IPSS Score	1.039	0.995	1.085	0.0853
Margin Non-Posterior	1.436	0.866	2.383	0.1609
Constant	0.013	0.001	0.187	**0.0014**

7.3.6.13 SBRT-Only CTCAE GU G2+ Final Model

The top SBRT-Only CTCAE GI G2+ model from 100x CV was selected as the final model. The final complete case SBRT-Only logistic model fit for CTCAE GU G2+ is shown in **Table 73.** Key information on the model is as follows: n = 351, events = 110, events per predictor = 15.7, overall model chi-square p-value <0.0001, training AUC 0.72 (95% CI 0.66 – 0.78), test AUC (5-fold CV x 100) = 0.69 (95% CI 0.64 – 0.77).

As for the SBRT-Only RTOG GI G2+ model, Cyberknife is a very strong protective factor. Similarly, prostate volume is detrimental, although less so. Similar to SBRT-Only GI models, we see overall treatment time >1 week as a protective factor. Bladder V50%(%) is detrimental. Interestingly IPSS as baseline has been selected over baseline CTCAE.

Table 73. SBRT-Only CTCAE GU G2+ Complete Case Final Model Fit

Predictor	Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Cyberknife	0.162	0.07	0.376	**<0.0001**
Fiducials	1.561	0.86	2.836	0.1433
RT Time >1 week	0.424	0.227	0.791	**0.007**
Prostate Volume	1.016	1.004	1.029	**0.0118**
RT Centre Caseload	1.006	0.999	1.012	0.0822
Baseline IPSS Score	1.044	1.004	1.086	**0.031**
Bladder V50%(%)	1.027	1.007	1.048	**0.0082**
Constant	0.137	0.046	0.405	**0.0003**

7.3.6.14 SBRT-Only IPSS Final Model

The top SBRT-Only IPSS model from 100x CV was selected as the final model. The final complete case SBRT-Only multivariable linear model fit for IPSS is shown in **Table 74.** Key information on the model is as follows: n = 344, overall model F-test p-value <0.0001, training RMSE 6.2 (95% CI 5.7 – 6.7) , test RMSE (5-fold CV x 100) = 6.3 (95% CI 5.7 – 6.8).

A simple model has been selected. Here the only significant predictor is baseline IPSS. Once again overall treatment time is close to significance.

Table 74. SBRT-Only IPSS Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
RT Time >1 week	-1.571	-3.251	0.109	0.0668
Baseline IPSS Score	0.656	0.548	0.765	**<0.0001**
Bladder V100(cc)	0.169	-0.012	0.35	0.0666
Constant	8.995	6.853	11.137	**<0.0001**

7.3.6.15 SBRT-Only EPIC-26 Urinary Incontinence Final Model

The top SBRT-Only EPIC-26 UI model from 100x CV was selected as the final model. The final complete case SBRT-Only multivariable linear model fit for EPIC-26 UI is shown in **Table 75.** Key information on the model is as follows: n = 351, overall model F-test p-value <0.0001, training RMSE 14.5 (95% CI 12.6 – 16.2), test RMSE (5-fold CV x 100) = 14.7 (95% CI 12.9 – 16.4).

An identical predictor set has been selected as in the Whole-Trial EPIC-26 UI model. The protective effect from fiducials is higher here (4.03 vs 2.97).

Table 75. SBRT-Only EPIC-26 UI Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
Fiducials	4.029	0.614	7.444	**0.0209**
WHO PS 1+	-9.687	-15.003	-4.37	**0.0004**
Baseline EPIC-26 UI	0.601	0.492	0.711	**<0.0001**
Constant	29.952	19.437	40.467	**<0.0001**

7.3.6.16 SBRT-Only EPIC-26 Urinary Obstructive Final Model

The top SBRT-Only EPIC-26 UO model from 100x CV was selected as the final model. The final complete case SBRT-Only multivariable linear model fit for EPIC-26 UO is shown in **Table 76.** Key information on the model is as follows: n = 333, overall model F-test p-value <0.0001, training RMSE 17.0 (95% CI 15.2 – 18.4), test RMSE (5-fold CV x 100) = 17.4 (95% CI 15.6 – 18.8).

Overall treatment time >1 week is again a protective factor in this SBRT-Only model. Baseline EPIC-26 UO and GU medications are detrimental, as they were in the Whole Trial model. Here non-posterior margin is a significant detrimental factor.

Table 76. SBRT-Only EPIC-26 UO Complete Case Final Model Fit

Predictor	Coefficient	Lower 95% CI	Upper 95% CI	p-value
RT Time >1 week	5.183	0.392	9.975	**0.0341**
Age	0.228	-0.077	0.532	0.1426
Baseline EPIC-26 UO	0.51	0.372	0.647	**<0.0001**
GU meds at baseline	-7.048	-12.142	-1.954	**0.0068**
Margin Non-Posterior	-3.07	-6.091	-0.05	**0.0464**
Bladder V100(cc)	-0.394	-0.896	0.109	0.1242
Constant	30.723	1.397	60.048	**0.0401**

7.3.7 Final Model Summaries

The selected predictors for every GI model, coloured by direction of effect, are summarised in **Table 77**. The selected predictors for every GU model are similarly summarised in **Table 78**.

Table 77. GI Models: Summary of Selected Predictors

Summary of the selected models for each Whole-Trial and SBRT-Only model. The purpose is comparison of the direction of effect for risk factors, so odds ratios (RTOG and CTCAE) and linear coefficients (EPIC-26) are presented together. Recall that negative coefficients are detrimental for EPIC-26 (higher score is better).

Green highlight = protective; Red highlight = detrimental; **Bold** = significant predictor coefficient/odds ratio in final model

* = Exclusion of predictor has been checked in sensitivity analysis – direction of effect for other model predictor coefficients unchanged.

** = Loses significance on exclusion of predictors in model with unexpected direction of effect (denoted *)

Endpoint	Predictor Odds Ratios (CRO models) and RMSEs (PRO models)														
	SBRT	CK	Fidx	RT Time >1 week	WHO PS 1+	Age	RT Centre Caseload	Baseline RTOG GI G1+	Baseline CTCAE GI G1+	Baseline EPIC-26 Bowel	Statin Use	Margin Posterior	Rectum V50 (%)	Rectum V80 (%)	Rectum V100 (cc)
Whole-Trial															
RTOG GI G2+		0.568		N/A				**2.265**				0.9*		**1.044**	
CTCAE GI G2+	**2.981**	0.625		N/A					**2.007**		0.853		1.014		
EPIC-26 Bowel	**-7.605**		**3.392**	N/A		**0.204**				**0.686**				**-0.471**	
SBRT-Only															
RTOG GI G2+	N/A			**0.391**	2.619		**0.992**		0.475*	0.975	0.575			**1.103****	0.539*
CTCAE GI G2+	N/A	0.6		0.531	**2.418**						0.611			1.061	
EPIC-26 Bowel	N/A		**4.63**	**5.115**		0.299				**0.703**				**-0.754**	

Table 78. GU Models: Summary of Selected Predictors

Summary of the selected models for each Whole-Trial and SBRT-Only model. The purpose is comparison of the direction of effect for risk factors, so odds ratios (RTOG and CTCAE) and linear coefficients (IPSS, EPIC-26) are presented together. Recall that negative coefficients are detrimental for EPIC-26 (higher score is better), while positive coefficients are detrimental for IPSS (lower score is better). Note base = baseline.

Green highlight = protective; Red highlight = detrimental; **Bold** = significant predictor coefficient/odds ratio in final model

* = Exclusion of predictor has been checked in sensitivity analysis – direction of effect for other model predictor coefficients unchanged.

GU Endpoint	Predictor Odds Ratios (CRO models) and RMSEs (PRO models)																	
	SBRT	Moderate Hypofrac-tionation	CK	Fidx	RT Time >1 week	WHO PS 1+	Age	Prostate Volume	RT Centre Case-load	Base RTOG GU G2+	Base CTCAE GU G2+	Base IPSS Score	Base EPIC-26 UI	Base EPIC-26 UO	GU meds at baseline	Margin Non-Post.	Bladder V50% (%)	Bladder V100% (cc)
Whole-Trial																		
RTOG G2+		0.636	**0.246**					**1.013**		**1.746**		**1.049**	1.01*			1.111		
CTCAE G2+	**5.273**	1.354	**0.457**				1.021	**1.012**				**1.05**				**1.301**		
IPSS	**2.467**	**1.671**				1.293				**1.707**		**0.534**					**0.038**	
EPIC-26 UI				**2.97**		**-8.934**							**0.593**					
EPIC-26 UO	**-6.493**	-3.478					0.145				**-3.49**			**0.442**	**-4.655**	-0.906		
SBRT-Only																		
RTOG G2+			**0.184**	1.764				**1.021**		**2.211**		1.039				1.436		
CTCAE G2+			**0.162**	1.561	**0.424**			**1.016**	1.006			**1.044**					**1.027**	
IPSS					-1.571							**0.656**						0.169
EPIC-26 UI				**4.029**		**-9.687**							**0.601**					
EPIC-26 UO					**5.183**		0.228							**0.51**	**-7.048**	**-3.07**		-0.394

7.4 Discussion

7.4.1 Model Performance

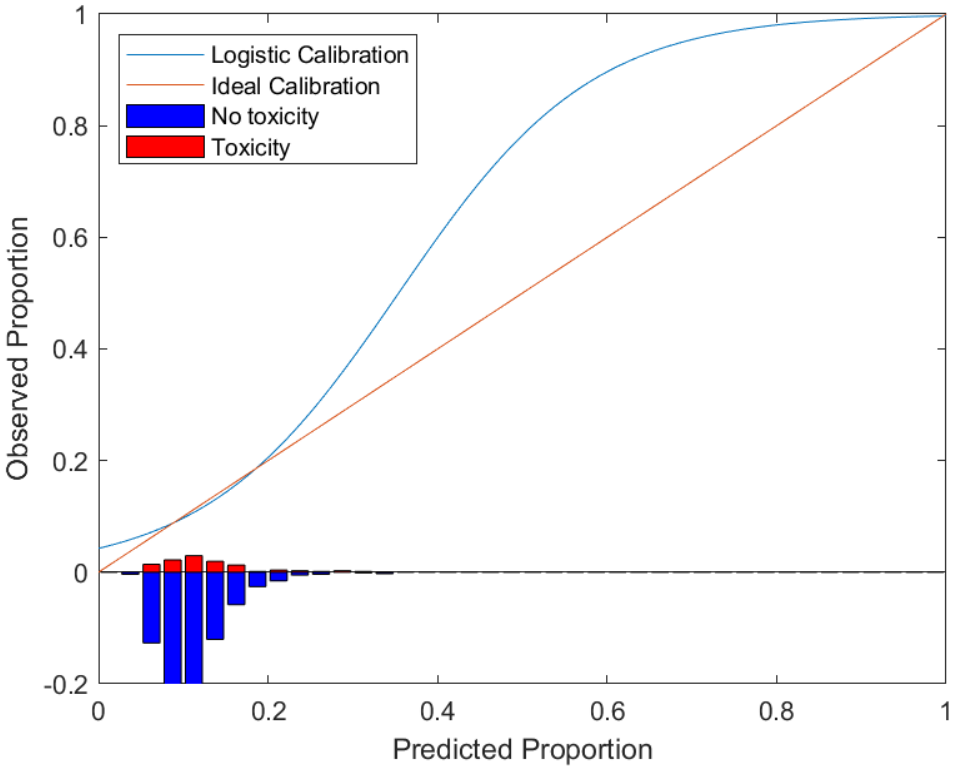
This chapter has provided multivariate prediction models of acute toxicity, including the provision of clinically interpretable risk factor effects. These have been produced using CV methodology to assess performance on data outside the fitting set. CRO model prediction performance is generally weak to moderate, as seen by AUCs ranging 0.59 – 0.73 (**Table 79**). The similarity in test AUC estimates between RTOG and CTCAE models in each domain are interesting, suggesting possible ceilings to the prediction utility of these dosimetric and clinical predictors.

Table 79. Summary of CRO Model Performance

CRO Model	Whole-Trial		SBRT-Only	
	Test AUC	95% CI	Test AUC	95% CI
GI Models				
RTOG GI G2+	0.59	0.53 – 0.66	0.62	0.55 – 0.76
CTCAE GI G2+	0.62	0.58 – 0.68	0.62	0.54 – 0.72
GU Models				
RTOG GU G2+	0.67	0.63 – 0.73	0.73	0.69 – 0.79
CTCAE GU G2+	0.67	0.63 – 0.72	0.69	0.64 – 0.77

Examples of calibration for the best and worst models are shown in **Figure 61**. The range of calibration quality is readily apparent, with the worst (RTOG GI G2+ Whole-Trial) lying far from ideal calibration.

Panel A: RTOG GI G2+ Whole-Trial Calibration (Worst, Test AUC = 0.59)



Panel B: RTOG GU G2+ SBRT-Only Calibration (Best, Test AUC = 0.73)

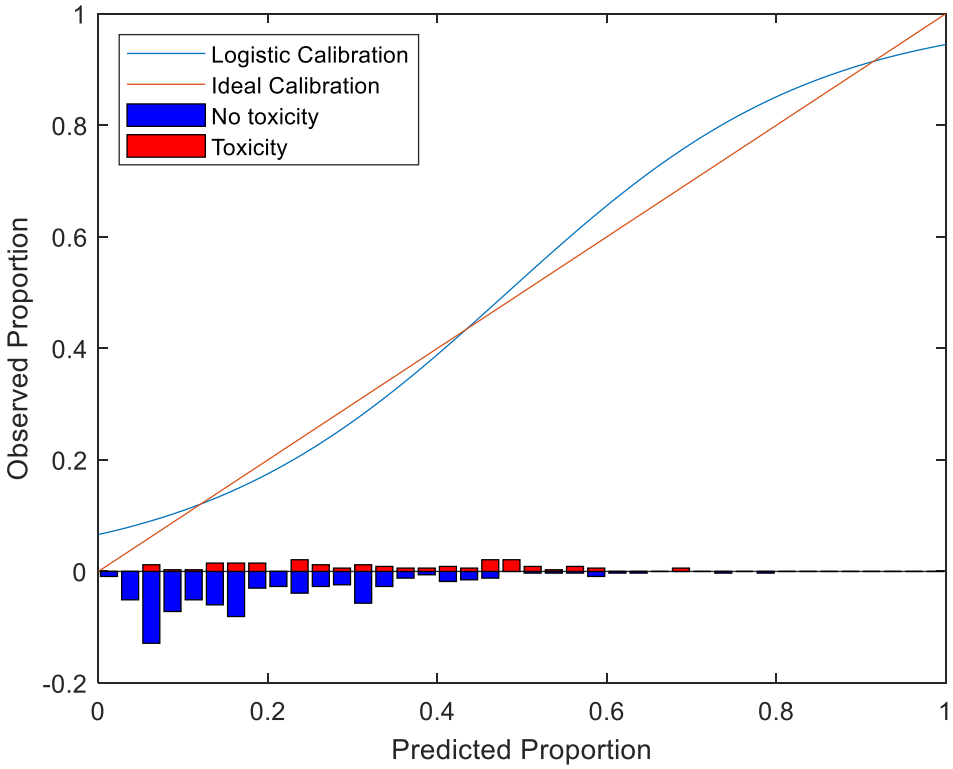


Figure 61. Worst and Best CRO Calibration Curves

Showing calibration for the worst and best AUCs seen across CRO models.

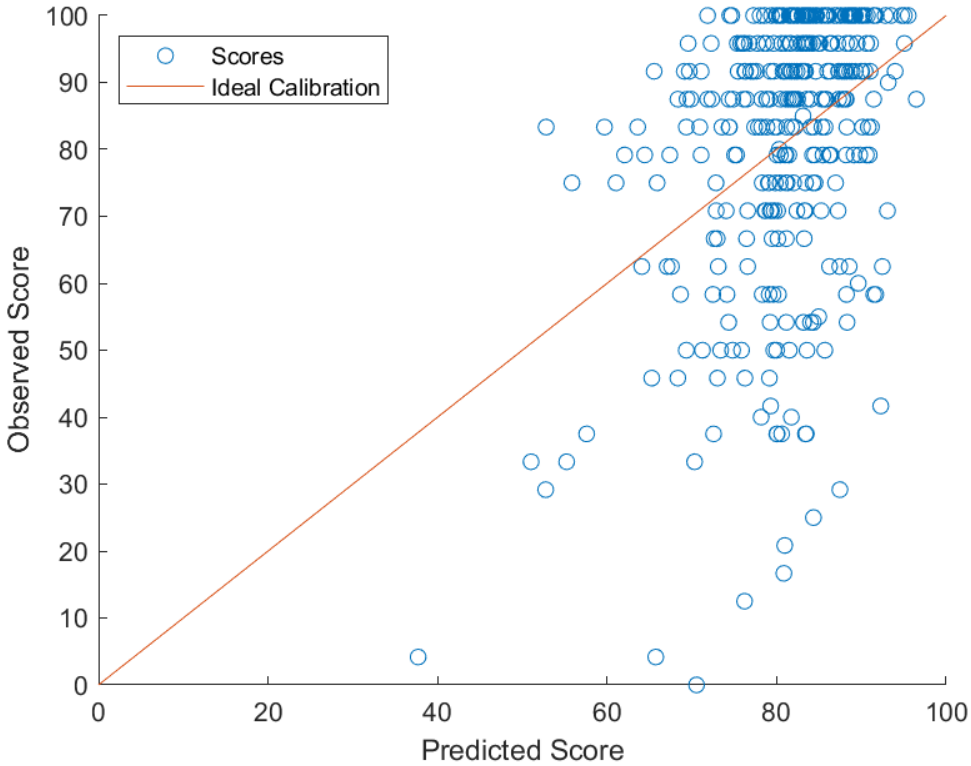
PRO model prediction accuracy is presented in (**Table 80**), including correction of the RMSE values by the MCID for each scale. The GU models are generally consistent with average prediction error around 2-2.5x MCID. Accuracy is insufficient to predict those with an MCID, although may be more useful for those with larger differences (potentially representing more relevant QoL change). The EPIC-26 bowel prediction is much less accurate, with 4-4.5x MCID.

Table 80. Summary of PRO Model Performance & Correction for MCID
Summary of the RMSE data for each PRO model in top half of table. In the bottom half, these figures are corrected by dividing by the minimal clinically important difference, calculated as 0.5 x SD of baseline score for each PACE-B scale: MCIDs: EPIC-26 bowel = 4.3; IPSS = 3.2; EPIC-26 UI = 7.3; EPIC-26 UO = 6.8.

PRO Model	**Whole-Trial**		**SBRT-Only**	
	Test RMSE	**95% CI**	**Test RMSE**	**95% CI**
GI Models				
EPIC-26 Bowel	17.3	16.0 – 18.3	18.8	16.7 – 20.4
GU Models				
IPSS	6.5	6.1 – 6.9	6.3	5.7 – 6.8
EPIC-26 UI	14.4	13.1 – 15.5	14.7	12.9 – 16.4
EPIC-26 UO	16.8	15.6 – 17.7	17.4	15.6 – 18.8
	MCID Corrected Test RMSE	**MCID Corrected 95% CI**	**MCID Corrected Test RMSE**	**MCID Corrected 95% CI**
GI Models				
EPIC-26 Bowel	4.0	3.7 – 4.3	4.4	3.9 – 4.7
GU Models				
IPSS	2.0	1.9 – 2.2	2.0	1.8 – 2.1
EPIC-26 UI	2.0	1.8 – 2.1	2.0	1.8 – 2.2
EPIC-26 UO	2.5	2.3 – 2.6	2.6	2.3 – 2.8

Examples of calibration for the best and worst models are shown in **Figure 62**. It is clear the model struggles to predict those with the worst observed scores in the SBRT-Only EPIC-26 Bowel model (Panel A).

Panel A: EPIC-26 Bowel SBRT-Only Calibration (Worst, Corrected RMSE = 4.4)



Panel B: IPSS SBRT-Only Calibration (Best, Corrected RMSE = 2.0)

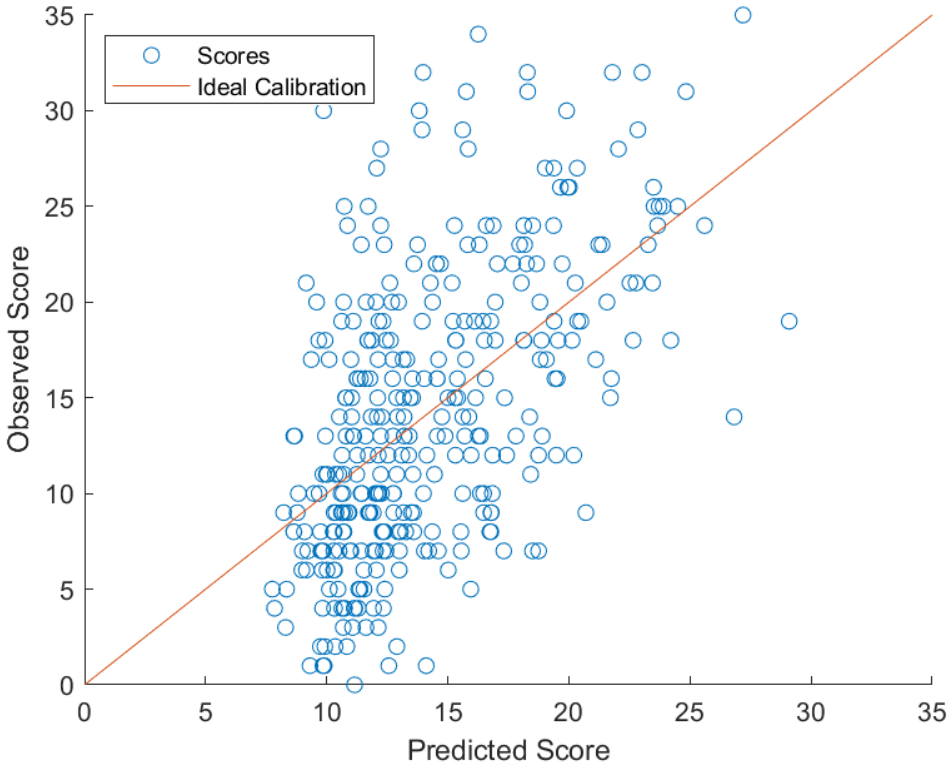


Figure 62. Worst and Best PRO Calibration Curves

Showing calibration for the worst and best minimal clinically important difference corrected RMSEs seen across PRO models.

7.4.2 Interpretation of Risk Factors in Context

7.4.2.1 SBRT & Moderate Hypofractionation

Amongst Whole-Trial models, SBRT was significantly detrimental for CTCAE GI (OR 3.0) and GU (OR 5.2), along with EPIC-26 Bowel (-7.6 points), IPSS (+2.5 points), and EPIC-26 UO (-6.5 points). For the PRO endpoints, the differences are around the magnitude of an MCID in each of these scales, however it is important to remember that the convenience of SBRT would likely outweigh such a difference in the acute setting, due to the transient nature of symptoms. However, SBRT was not selected for the RTOG models. As discussed in **Chapter 4**, this may be driven by RTOG having unbalanced assessments; assessing CFMHRT patients 2 weekly during treatment, while only at end of treatment for SBRT patients. It may be argued that RTOG is thus a fairer reflection of the peak toxicity a CFMHRT patient might expect to experience at any point up to 12 weeks post-RT. Alternatively, one might argue that RTOG will over-report CFMHRT toxicity due to repeated questioning and over-detection of borderline toxicity.

The effect for moderate hypofractionation (c.f. conventional) is only significantly worse in the IPSS model and only marginally (+1.7 points). Overall, this data does not alter conclusions one might draw from the very detailed analysis in the CHHiP trial [1], where we can see the earlier and more prevalent acute toxicity with moderate hypofractionation compared to conventionally fractionated treatment but no eventual impact on late toxicity.

7.4.2.2 Cyberknife & Fiducials

CK usage was selected for three GI CRO models, with a protective odds ratio, but 95% CIs for these estimates crossed no effect. All GU CRO models fitted CK as a significant protective effect. The odds ratios seen for these CRO GU endpoints are strong (0.16-0.46), so it is perhaps surprising that CK is not selected as a predictor in the GU PRO models. Some possible explanations could be considered. Firstly, that there is a difference which is best detected <2 weeks post-RT (i.e. before the first measurement of PRO); i.e. acute toxicity is settling by 2 weeks. Secondly, the lack of treatment blinding might be contributory to clinician assessment; however, this was a large multi-centre study and the effect size is consistently strong. Thirdly a centre scoring effect might be considered (i.e. do CK centres generally score toxicity differently): a -4% difference in worst RTOG GU 2+ scores for CMFHRT patients (used as a control) between CK-equipped centres and non-CK-equipped centres but this was not a statistically significant difference (Recall **4.5.7.2**).

The signal of a protective effect from CK for acute CRO GU toxicity is interesting, but it would be far stronger with a confirmatory finding in the late toxicity data. Should CK also be a significant protective factor for late GU toxicity, then the argument for its more widespread use would be strengthened. In particular, men who have other risk factors for GU toxicity might benefit from this modality more than others. Any protective effect size seen in the acute (and potentially late) toxicity data must be balanced against the additional time taken for treatment delivery.

Fiducials show benefit in EPIC-26 bowel and UI. Here, the first assessment was at 2 weeks post-RT, potentially missing the acute effects of insertion. For the SBRT-Only GU CRO models they fit towards adverse effect, although estimates cross no-effect. We note that the effect of CK in those models would be slightly attenuated, since 100% of such patients had fiducials.

7.4.2.3 Overall Treatment Time

The most consistent signal is for a protective effect through >1-week overall treatment time in the SBRT-Only models. This is largely a one-week (20.7%) vs two-week (73.5%) comparison, per protocol recommended options. This data complements the findings of the PATRIOT study, comparing two-week versus 5-week delivery of prostate SBRT [100]. Through EPIC-50 assessment, they found more two-week patients reporting a minimal clinically important change (MCID; 5.8 points bowel, 5.5 points urinary) in bowel (90% versus 68%, p=0.002) and urinary (94% vs 78, p=0.006%) domains. Although the different EPIC versions prevent precise comparison, it is interesting that our data suggest using 2-week vs 1-week SBRT confers a 5-point benefit to worst EPIC-26 Bowel and EPIC-26 UO domains – about the width of one MCID in the PATRIOT study. The PATRIOT study was not powered for efficacy, but observed a 7.2% vs 3% 5-year failure rate with 5-week vs 2-week. For UK patients, where home to RT centre travel distances are generally manageable, I would favour a two-week schedule being recommended for SBRT going forward. This would be similar to HYPO-RT-

PC (2.5 weeks) and would be less detrimental to acute toxicity than a one-week regimen.

7.4.2.4 Age and Performance Status

EPIC-26 Bowel Whole-Trial fitted age as a protective factor, with a fairly modest coefficient (+0.2 points/year). Protective coefficient trends for age were also seen in the Whole-Trial EPIC-26 Bowel & EPIC-26 UO models, although it was detrimental in CTCAE GU 2+ Whole-Trial. It is possible that the protective effect is a result of controlling for performance status and baseline toxicity; potentially suggesting age as a predictor of lower likelihood to report toxicity on PRO instruments. The results do not appear to warrant considering age a limitation in treatment decision making.

Performance status ≥1 is fitted as a significant negative factor in the SBRT-Only CTCAE GI G2+ model and EPIC-26 UI Whole-Trial (-8.9 points) plus SBRT-Only (-9.7). This increasing urinary incontinence is independent of baseline function (EPIC-26 UI fitted to both models). Interestingly fiducials were the only fitted protective predictor in such models as a potentially modifiable factor. This suggests that practitioners should not be overly concerned about utilising fiducials in those with pre-existent urinary incontinence.

7.4.2.5 Prostate Volume

Whether larger prostate volumes contribute to the acute toxicity of SBRT is an area of interest to the urological radiation community [239], though

retrospective data of 57 men with prostate volumes >50cc receiving SBRT suggests reasonable tolerance [240]. We find prostate volume a significant risk factor in Whole-Trial RTOG and SBRT-Only RTOG and CTCAE models. The absence from PRO endpoints may reflect later collection for those instruments – i.e. a large prostate causes toxicity earlier than 2-weeks post-RT. Comparing the additional risk of a 90th centile prostate (73cc) compared to a 10th centile prostate (25cc), it would range from 1.8x (CTCAE GU G2+ Whole-Trial) to 2.7x (RTOG GU G2+ SBRT-Only). These are modest odds ratios for large variations in prostate size. The PACE-B data therefore does not appear to suggest limits for SBRT based on prostate volume, although this may not hold at extremely large sizes (e.g. 100cc) that were not represented in the trial.

7.4.2.6 Baseline Toxicity

Baseline toxicity predictors feature as strong and significant in all models except for SBRT-Only RTOG & CTCAE GI models and SBRT-Only CTCAE GU G2+. While this is of course unsurprising, it is reassuring that these expected predictors were included to control other predictor coefficients. Interestingly, some GU models fitted both CRO and PRO baseline measures, suggesting independence of data from these metrics.

7.4.2.7 Dosimetry & Margins

Dosimetry is included in most GI models and some GU models, always in a detrimental capacity as might be expected. The dose constraints applied in the PACE-B trial have clearly produced acceptable acute toxicity at a trial

level (**Chapter 4 & Chapter 5**). Direct application of the coefficients can only be considered for the SBRT-Only models, due to the percentage dose scaling applied across regimens in the Whole-Trial. Effects are modest; for example, an increase of 5% at V27Gy (80% dose) for the rectum suggests a -3.77 point on worst acute EPIC-26 bowel. These effects are seen despite the application of protocolised dose constraints for the trial – suggesting that tighter secondary rectal dose constraints might be considered for improved SBRT optimisation.

Non-posterior margin size was a detrimental factor in several GU models. Conversely posterior margin size was not fitted in any GI model. Given the differential sizes used in PACE-B (smaller at posterior edge), this suggests that further posterior margin reduction may have limited benefit, while non-posterior margin reduction may attenuate some acute GU symptoms. With improvements in GTV delineation on MRI, it might be considered to have smaller margins on all sides except the dominant nodule.

7.4.2.8 Other Predictors

There is no evidence of a consistent signal for protective effect from statins, with fitting only to two models, with both coefficients crossing no-effect. GU medication use at baseline (alpha-blockers or anti-cholinergics) was associated with worse EPIC-UO in both Whole-Trial and SBRT-Only models. It is perhaps surprising this was not replicated in the baseline IPSS score, where it was feature in a few of the top 20 IPSS models at Whole-Trial level and none of the top 20 SBRT-Only models.

PTV Dmax was not selected in any GU model. This is a good finding, given the presence of heterogenous SBRT plans in the dataset, with the potential for high urethral doses (optional dose constraint only).

7.4.3 Prior Studies in the Field

The HYPRO trial (Recall **Table 1**) reported multivariate models for RTOG Acute G2+ GI and GU toxicity [206].The modelling approach is simply described as multivariate logistic regression. No dosimetry information was included. No info was given on variable curation steps or variable selection. Model performance was not assessed by either CV or hold-out set. The strongest RTOG G2+ GI multivariate model predictor was RTOG GI G2+ at baseline (OR 5.5, 95% CI 2.1-14.3), a predictor corroborated by our study. Moderate hypofractionation was selected as a GI predictor in their study (OR 1.6, 95% CI 1.2 – 2.1), but not in our GI models. However, SBRT was selected for both CTCAE and EPIC-26 Bowel, which together does suggest increased acute toxicity with hypofractionation. Other important predictors from the HYPRO trial were 2+ months of ADT (OR 0.6, 95%CI 0.4 – 0.9), which the PACE-B data cannot corroborate. Effects they saw for CTV size may have been superseded by inclusion of dosimetry data in the models here.

The HYPRO final multivariate acute RTOG G2+ GU model included: RTOG GU G2+ at baseline (OR 14.5, 95% CI 7.7-27.3), an odds ratio far higher than seen in our data. Age ≥70 years (OR 1.4, 95%CI 1.0 – 1.9) was not

strongly selected as a detrimental factor across our GU models. Prostate volume ≥50cc (OR 1.4, 95% CI 1.0 – 1.9), which is corroborated by inclusion of prostate size in multiple GU CRO models. Other included predictors had odds ratios spanning no-effect.

Wang *et al* have reported a retrospective study of 259 men receiving SBRT 38 Gy in 4 fractions; modelling multivariate linear regression of 1-month post-RT EPIC-26 bowel, UI and UO scores [232]. Predictor selection was by simple p-value methodology, $p<0.15$, no cross-validation was performed. Baseline symptoms score was a selected predictor in all models, as has been the case in our EPIC-26 models. Age was a risk factor for EPIC-26 UI (coefficient = -0.13). We see no similar effect, although their study fitted without performance status, for which age may be a surrogate. Their EPIC-UO model also included prior-TURP as protective (coefficient 0.11), which we unfortunately lack the data to model. For EPIC-Bowel, they reported rectum D25% (coefficient -0.57) and Dmax as deleterious (coefficient -0.12), but D50% as protective (coefficient 0.4) which defies bioplausibility and may represent overfitting.

Palumbo *et al* reported a retrospective study of 195 patients with nmPca receiving EBRT 74.25 Gy in 33 fractions (2.25 Gy per fraction) to the prostate ± seminal vesicles (if >15% risk of involvement) [231]. Acute CTCAE GI toxicity (≤3 months) was assessed at 1- and 3-months post-irradiation, potentially missing some acute toxicity (recall rapid decline in acute toxicity in

Figure 22). Their multivariate logistic regression model was constructed by including all variables with p<0.25 on univariate regression, a method subject to debate [160]. They reported age (OR 1.1, 95%CI 1.0 – 1.16) and statin use (OR 0.46, 95% CI 0.21 – 0.98) as significant p<0.05. Our models do not strongly corroborate either of these. No efforts at CV or holdout validation were made, so external generalisability may be limited.

Delobel *et al* retrospectively[48] reported on 972 nmPCa patients treated with 70-80 Gy, 2-2.5 Gy per fraction, modelling acute RTOG/CTCAE[49] grade 2+ rectal toxicity, occurring ≤3 months from radiotherapy. They fitted a multivariate logistic regression to predictors with p<0.2 on univariate testing. They reported only RT technique (3DCRT, IMRT or IMRT+IGRT) as significant (p<0.001) in the multivariate model, although failed to report odds ratios for the levels of this nominal variable. No efforts at cross-validation or holdout validation were made. We do see inclusion of Cyberknife as a protective factor improving several GI models, although 95% CI for odds ratios do cross no effect.

Away from hypofractionation, Keyes *et* al reported a retrospective cohort of 932 low & favourable intermediate risk men undergoing LDRBT, 144Gy I-125. A multivariate model was reported for acute RTOG G2+ toxicity (defined ≤6 months from implant), which occurred in around 45% of patients[50] [234]. They used a backward elimination logistic regression, with cut off p-value of

[48] 487 had been prospectively collected within GETUG-06 and STIC-IGRT trials
[49] Unclear on what proportion of patients were assessed by either method.
[50] Proportion of acute RTOG G2+ only reported in graphical form.

0.05, a figure which is fairly conservative [160]. No efforts at cross-validation or holdout validation were made. They reported pre-treatment ADT (OR 1.42, 95% CI 1.01-2.02) and Baseline IPSS (OR 1.08, 95% CI 1.05-1.12) as significant predictors, along with needle number and needle order, which are not relevant to our SBRT study. It should be noted that treatment volume was dropped for the multivariate model, so ADT in this context may be a prostate volume surrogate (since it is given to men with larger prostate pre-brachytherapy). Our study validates the finding here that a PRO instrument (IPSS) may be a useful predictor of CRO (RTOG & CTCAE) outcomes for GU endpoints.

7.4.4 Strengths of this Study

As a large, prospectively collected dataset, the data quality of PACE-B is higher than several of the preceding studies into determinants of acute toxicity. Specifically, to my knowledge, it is the first prospectively collected study incorporating patients receiving SBRT dose schedules. Through the development of a comprehensive DICOM library, it includes dosimetry, which is an improvement on several prior models. I have investigated as many previously identified predictors as possible (not ADT usage for example), allowing investigation of prior reported predictors that might modulate the acute toxicity response – for example statin usage.

I believe that the inclusion of both CRO and PRO endpoints is a significant strength of this study. As outlined in **Chapter 5**, CRO and PRO data is not in perfect agreement, meaning that reliance on one may miss some forms of

toxicity. The inclusion of both CRO and PRO baseline toxicity as significant predictors in several models is testament to this. The PRO instruments reported are without significant copyright restrictions[51], meaning that they will hopefully continue to see regular use in prostate EBRT practice.

Previously reported acute toxicity multivariate frameworks have fitted only to the training data, without use of CV or holdout methods. This leaves them at high risk of overfitting, since a model optimised to a single training dataset may perform poorly when exposed to novel data. A significant strength of this study is that candidate model selection has been made through cross-validation, with performance assessed on the 10th fold of data unseen in the fitting. Such a method should increase the external reliability of the selected model. It should be noted however, that differences seen in the predictor selections for top-20 backward selected models and top-20 models following cross-validation were few (**Table 54 & Table 56**).

7.4.5 Limitations of this Study

This is a study of acute toxicity, one of the several key components that would determine the choice of a radiotherapy regimen. The ideal model to determine optimal treatment schedule might also include late toxicity, efficacy, cost, patient preference and system resource availability. While this is some way off, I recognise that acute toxicity in isolation may not sway a treatment decision. However, the data here may hopefully reassure clinicians

[51] EPIC-26 and IPSS are free to use. UCLA-PCI (for example) contains SF-36, which is under restrictive copyright.

concerned about the magnitude of excess acute toxicity expected in certain situations; for example, larger prostates [239].

In any study fitting events into a logistic regression, a high event rate is desirable in order to allow more precise modelling. The low toxicity seen in the PACE-B trial therefore limits investigation of RTOG and CTCAE models, especially GI, since the event rates were low (10-20%). While limiting to the modelling undertaking, it is obviously a positive fact for patients who have participated in the trial. It must also be acknowledged that the quality of fits are far from perfect; AUCs generally <0.7, average errors 2x MCID or more in continuous variables. As discussed in Chapter 6, there may be a limit to the performance of dosimetric and clinical data.

A criticism might also be made of the use of logistic regression. Regularisation in the form of the LASSO (L1 regularisation) would have been attractive in providing variable selection through hyperparameter fitting [241]. However, my strong desire in this study was to provide odds ratios or coefficients for inferential usage. The use of LASSO for inference allows some predictors to have confidence intervals, while other known predictors act as control variables [242]. Given the weak pre-existent evidence for any predictor in the literature, this approach was not undertaken.

Logistic regression is also a less powerful modelling architecture than more recent methods such as support vector machines (SVM) or neural networks (NNs). Such approaches have been reported in a retrospective study of 321

men for the prediction of acute GI and/or GU (i.e. combined) toxicity following prostate radiotherapy [243]. Using clinical and treatment predictors, they found similar AUCs of 0.7 for the SVM and NN approaches. These are slightly higher than the test AUCs found in our own work, however such methods are perceived as “black-box”, failing to provide odds ratios or coefficients for the predictors. Work is ongoing to improve interpretability (e.g. integrated gradients [244]), but, to my knowledge, has not been applied in clinical research. Such coefficients may help in individualising risk prediction, since probable sources of excess toxicity might be determined.

7.5 Conclusions

This study provides the first large, prospectively collected models for the prediction of acute toxicity in patients receiving ultrahypofractionated SBRT schedules. The CRO and PRO models presented may assist clinicians in considering the relative likelihood of acute toxicity for patients considering EBRT for nmPCa. Baseline toxicity and dosimetry are common predictors within the models presented, which reinforces their importance in the determination of acute toxicity. There is a strong protective effect of Cyberknife on CRO related acute GU toxicity, although not replicated in PRO models. Prostate volume confers increased GU acute toxicity risk, but the risk elevation is relatively modest (2.7x) from the 10th - 90th centile (25 - 73cc). We note further the protective effect of a two-week SBRT regimen over a one-week regimen, an implementation easily put into clinical practice with every-other-day dosing. The datasets established for this study will allow rapid production of late toxicity and efficacy models as PACE-B matures.

Chapter 8.Summary and Conclusions

8.1 Overview

Hypofractionationated EBRT, as a radical modality for nmPCa, is now widely accepted [20]. With phase III efficacy evidence for ultrahypofractionation [95], it is likely that this too will soon become standard-of-care. This thesis has set out to improve our knowledge regarding the toxicities seen with hypofractionation, and the prediction thereof. I believe that the thesis makes novel contributions to our current understanding, which will be summarised below. Ideas for future research that might extend the work in this thesis will then be discussed.

8.2 Contributions to Current Knowledge

8.2.1 Fraction Size Sensitivity of Individual Rectal Side Effects

Individual α/β ratio estimates have not previously been provided for individual late rectal endpoints (Recall **2.3.4.4**). In **Chapter 3**, modelled estimates for such endpoints have been provided, covering common rectal toxicities such as bleeding, proctitis and stool frequency. These are the most robust late rectal α/β estimates produced to date, with 95% CIs that are not artificially reduced by fixing model parameters [62]. The pooled estimate of these is low (2.3-2.4 Gy), which is less than the late rectal α/β ratio assumed in HYPRO (4-6 Gy), a recent moderate hypofractionation era trial [75]. The data presented suggest a maximum late rectal α/β = 3 Gy. By guiding clinicians to avoid higher α/β ratio estimates, this should help safeguard future trial

patients from the possibility of higher than expected late rectal toxicity events.

8.2.2 Differences in Clinician Reported Acute Toxicity Between 5-fraction SBRT and CFMHRT Radiotherapy

Given the abbreviated overall treatment time, there exists the possibility of increased acute toxicity with ultrahypofractionated regimens. **Chapter 4** explores CROs of acute toxicity in the PACE-B trial, finding no difference in RTOG toxicity, but slightly increased worst CTCAE GI acute toxicity, non-significantly different by 12 weeks. Although the acute toxicity data from HYPO-RT-PC was published three months prior to the publication of this work [95,188], the PACE-B data provides substantially more granular overview, with far more toxicity assessments during and ≤3 months from RT. Acute toxicity is known to follow a peaked time course, so the graphical representations in **Chapter 4 (Figure 22 & Figure 25)** provide the best overview of GI and GU acute toxicity time courses for SBRT patients. Patients considering SBRT might usefully be reassured through such graphs, demonstrating toxicity has abated by three months, for the vast majority. Furthermore, sub-grouped graphical analysis of SBRT patients showing less pronounced toxicity with 2-week schedule and CK usage has provided useful hypothesis generation on acute toxicity risk factors; explored in **Chapter 7**.

8.2.3 Differences in Patient Reported Acute Toxicity Between 5-fraction SBRT and CFMHRT Radiotherapy

In **Chapter 5**, analysis of the differences in PRO acute toxicity between CFMHRT and SBRT is presented. The multiple scales analysed are presented in far greater detail than the PRO data reported, at the time of writing, from HYPO-RT-PC [95]. There, only two timepoints were presented for acute toxicity (end of RT and 3 months post RT), for two PRO questions (bowel bother and urinary bother). No other phase III data for the PRO toxicity of ultrahypofractionation has previously been published.

The data in this thesis shows that, across a wide range of patient-answered questions, there were no significant differences between CFMHRT and SBRT. This covered all relevant organ systems: GI, GU and sexual. No signal was seen to corroborate the worse CTCAE G2+ CRO for SBRT. Patients considering SBRT should be reassured by this data, with similar peaks of toxicity and recovery by 12-weeks.

8.2.4 Rectal Contour Delineation for Rectal Toxicity Prediction

Work for the RT01 trial [245] and CHHiP trial (**Chapter 3**) have involved the central accrual and editing of almost 3000 rectal contours for the development of predictive models. In **Chapter 6**, I have shown that such edited contours have no additional predictive power for rectal toxicity above the original contours. This suggests that recontouring of the rectum for dosimetric analyses of prostate EBRT trials need no longer be undertaken.

Given it is a time-consuming and expensive process, this finding will be of immediate benefit.

In addition, **Chapter 6** provides a prospective analysis, larger than all preceding studies combined [216,218,220,221], of relative versus absolute volumes for the prediction of rectal toxicity, showing no difference. Furthermore, that procedurally truncated rectal volumes do not provide additional benefit to rectal toxicity prediction, using a sample far larger than previous study in this field [12]. Although these results are negative, they provide strong evidence against change in current practice of whole rectum, with relative volume DVH parameters.

8.2.5 Defining Risk Factors for Acute Toxicity After SBRT

Prior models of acute toxicity have generally been fitted to patients receiving CFMHRT regimens (recall **7.4.3**). In **Chapter 7**, I provide models for a range of CRO and PRO toxicity measures, all fitted to both a combination of CFMHRT/SBRT and to SBRT-Only. Additionally, these models have been internally validated through cross-validation, hopefully providing models with better external generalisability. The finding of most immediate utility is confirmation of a consistent increase in GI and GU acute toxicity risk with a 1-week versus ≥2-week SBRT schedule. Prostate volume has been considered a risk factor for SBRT, however the data here suggests that for typical PACE-B patients (25-73cc), the risk effect is moderate and not a contraindication to SBRT. Dosimetry featured in several GI models, despite

strict dose constraints being applied. This suggests that lower rectal doses than the current dose constraints may be desirable, to minimise toxicity.

8.3 Recommendations for Further Study

8.3.1 Late Toxicity of Ultra-hypofractionation

8.3.1.1 The Late Toxicity of a 5-Fraction SBRT Regimen

Through reason of data availability, this thesis was limited to analysis of the acute toxicity from PACE-B. Late toxicity data (min follow-up 2 years) is approaching maturity, so extension of the work in **Chapters 4, 5 and 7** to the late toxicity setting would be of interest. For late toxicity, the granularity of data will be similar to HYPO-RT-PC [95], however the late toxicity from a more abbreviated 5-fraction regimen may differ. The ultimate global choice of a 5 or 7 fraction ultrahypofractionated regimen will depend on efficacy, late toxicity and, to some extent, acute toxicity. Assuming PACE-B achieves non-inferiority for efficacy, with proven acceptable acute toxicity, only an elevated late toxicity risk with a 5-fraction regimen would prevent adoption over a less convenient 7-fraction regimen.

8.3.1.2 Confirmation of α/β Ratio Estimates in Ultra-Hypofractionation

As discussed in the literature review (**2.3.3.4.3**), there is some uncertainty as to the appropriateness of the LQ-model in the ultrahypofractionation setting [246]. The PACE-B late toxicity data would provide an ideal opportunity to validate the α/β ratio estimates provided in **Chapter 3**. For this thesis, I have overseen curation of a complete DICOM set for PACE-B as part of **Chapter**

7, meaning dose data are readily available. Although the rectums for PACE-B have not been quality assured in the manner of CHHiP, the results of **Chapter 6** suggest this would not significantly improve toxicity prediction of the DVH information. It would therefore be trivial to apply similar LKB-EQD2 methodology and provide α/β ratio estimates with the inclusion of ultrahypofractionated patients. The critical step for such a study will be the choice of endpoints, since the CRO scales collected in PACE-B differ to CHHiP, meaning the same amalgamated endpoints cannot be replicated. It might be interesting to model the PACE-B scales separately to see if α/β ratio estimates for similar endpoints agree; for example, CTCAE rectal bleeding and the EPIC-26 bloody stools question.

Future studies are going to hypofractionate below 5-fractions, though one fraction may be avoided given the difficulties with single fraction HDR brachytherapy [67]. Should the α/β ratio estimate be confirmed in PACE-B then this would give confidence in the EQD2 predictions for rectal late effects in these future trials. The addition of genitourinary α/β ratios would also help, although the modelling of such data will be more difficult owing to the subjective nature of urinary symptoms.

8.3.1.3 Combining Acute and Late Toxicity Models

Should predictive models for late toxicity identify similar risk factors to the acute toxicity models of **Chapter 7**, there may be a case to consider individualised fractionation depending on patient risk factors. CFMHRT regimens are efficacious and safe, so an elevated individual risk of toxicity

with SBRT might not warrant the less convenient overall treatment time. The toxicity models in this thesis have been created with dosimetry and clinical information, however genomic information may further improve these [247]. In particular, the identification of tumour and normal tissue genomic factors influencing fraction size sensitivity would be useful [248].

8.3.2 Further Development of SBRT Acute Toxicity Models

8.3.2.1 Validation of Acute Toxicity Models for SBRT in an Independent Cohort

The models presented in **Chapter 7** might be used to predict the risk of acute GI and GU side effects for patients undergoing SBRT. While cross-validated estimates of model performance have been generated, these are only fitted to 80% of the data. The final models are fitted on the complete dataset, meaning external validation would be optimal; in line with TRIPOD methodology [249]. An easy candidate dataset would be the PACE-A trial, where precisely matching data has been collected. The trial is targeting a 234 patient accrual, of which 50% will have received SBRT, allowing for >100 patients to validate findings. It would of course also be desirable to validate on a fully external dataset; this might be considered by institutions with large prospectively collected datasets of 5-fraction SBRT [94].

8.3.2.2 Acute Toxicity Models for Routine Clinical Practice

The models presented in **Chapter 7** have utilised the full spectrum of available predictors from the DICOM files and PACE-B database. However, the large number of CRO and PRO instruments collected may not be

feasible in routine clinical practice. Ideally, the UK would agree a standardised set of instruments to routinely collect for prostate SBRT patients. This might include a single CRO measure and both IPSS and EPIC-26, since these are without licencing fees. For the CRO, it would probably be best to use CTCAE, homogenising practice with systemic therapy trials. A utopian view might anticipate a universal NHS electronic patient record, which would then permit easy roll-out of electronic PRO collection software. With this standardisation, the models could then be refitted to include only the agreed routinely collected predictors, improving external generalisability.

8.3.3 Alternative Rectal Definitions

8.3.3.1 The Rectal Wall

As discussed in **Chapter 3 & Chapter 6**, the definition of a rectum in this thesis has been the solid rectum, as is typical UK practice. However, in other countries (e.g. Italy), the rectal wall is used as the standard-of-care OAR for the rectum [155]. This international discrepancy might be tackled using the CHHiP dataset. It would be straightforward to generate rectal wall OARs and DVHs through algorithmic approaches, allowing ≈2000 patients to be included. Modelling approaches similar to Chapter 6 would allow interrogation of the modelling performance for each method. Additionally, the potential utility of absolute-volume DVH analysis for the rectal wall could be analysed, for which signals have been reported previously [218].

8.3.3.2 Auto-Contoured Rectums

Algorithms already exist to permit the auto-contouring of the rectum, amongst other normal pelvic organ tissues [230]. The CHHiP trial dataset

would provide an excellent opportunity to investigate morphological differences between different algorithms and human defined contours. This would allow testing of whether algorithmically-defined rectums are outperformed by human-defined rectums for toxicity prediction. This would be an important step in demonstrating the safety of such algorithms. A particular attention must obviously be paid to outlier cases, since a key question would be to define the proportion of auto-contoured rectums which would require human intervention for “correction”.

Of course, it might even be preferable to use the CHHiP dataset to develop such an algorithm, given its large size and robust QA. Excellent auto-contouring has been seen through do-novo training of deep convolutional 3-D segmentation neural network approaches [250]. This could then be validated using PACE-B. Usage of auto-contouring software could reduce the cost and time involved in the radiotherapy workflow. For patients not receiving ADT (such as those in PACE-B) lead time to radiotherapy is an important consideration for UK national cancer targets.

Such algorithms might also help with the implementation of dose-accumulation mapping. The models in this thesis have relied upon the planned dose, not accounting for the possibility of inter- and intra-fraction motion. MRI-guided radiotherapy permits real-time rectal imaging during treatment. The accumulated dose to the rectum during treatment might be determined in a clinical setting by rectal auto-contouring of the 4D-MRI and deformation to a reference rectum.

8.3.4 Improving Model Predictions

8.3.4.1 Use of Neural Networks

The models presented in this thesis are simple parametric models: Lyman model (probit) in **Chapter 3**, multivariate linear and logistic in **Chapter 7**. Application of complex dosimetric radiotherapy data to such models is challenging, due to the very large number of datapoints in a radiotherapy plan and the existence of severe dose-bin multicollinearity. Additionally, there is substantial spatial dose distribution data, discarded in the use of DVH curves as predictors. For predictive modelling of such data, a convolutional neural network (CNN) would hold great appeal, since this would permit non-human-curated derivation of spatial relationships between dose and toxicity. A hybrid model would be needed to additionally incorporate clinical/genomic factors.

Transfer learning describes the repurposing of a pretrained CNN, using new data to retrain only a few of the final layers. This has been applied to radiotherapy toxicity prediction for: pneumonitis [251], rectal toxicity after cervical radiotherapy [252] & liver toxicity post-SBRT [253]. The original models used for the transfer learning were derived from image and video recognition networks. It is possible that a 3D-CNN, trained from scratch on large dataset, may offer improved toxicity and/or failure prediction. This goal

will become easier with the rapidly reducing cost per unit of graphics card processing power, video random access memory and memory speed[52].

8.3.4.2 Dataset Size

Critically though, such approaches will rely upon large datasets, given the low event rates seen. International collaboration through pooling of radiotherapy datasets would seem the best approach currently possible.

Real-world approaches could also be utilised; groups have already reported surrogate toxicity outcomes from prostate radiotherapy treatments, generated from routinely collected UK Hospital Episodes Statistics (HES) [254]. It would be desirable to corroborate these surrogates – something which could be done by checking for CHHiP patients if there is concordance of (e.g.) G2+ rectal bleeding HES surrogate with actual recorded toxicity. To supplement this, if the UK national Radiotherapy Dataset (RTDS) began collecting full DICOM data (rather than summaries) then the combination of RTDS and HES might permit toxicity modelling on tens of thousands of patients. The biggest hurdles would be in funding/motivating RTDS and the regulatory issues surrounding research on routinely collected data.

Such a dataset would have application in non-toxicity research too. If disease characteristics and reliable surrogates of treatment failure patterns can be

[52] For example, two NV-linked Q3 2020 Nvidia RTX 3090 cards (2 x $1499 ≈ $3000) should offer equal VRAM memory (48GB) and faster memory speeds compared to a single 2018 Nvidia Quadro RTX8000 ($10000).

pulled from HES (or other databases) then it would be possible to train a network to examine plans to predict treatment failure. In the future, all proposed plans in the UK might be examined by such an algorithm to ensure any with a high risk of failure are reviewed by an external centre. This could help to reduce regional differences in radiotherapy quality.

8.3.4.3 Addition of Genomic Data

Large consortium radiobiological studies are beginning to determine gene single nucleotide polymorphism predictors for an individual's genomic risk of radiotherapy toxicity [247]. This may in future include genetic factors determining response to altered fractionation [255]. Future models of normal tissue toxicity would ideally include genetic factor, as they likely account for some of the inter-individual variation in toxicity responses. While implementation on the whole PACE-B dataset is not possible, due to absence of routine genetic testing, some data may come in future from those recruited to the RAPPER UK radiogenomics study [256]. Uniform collection of genetic data would be an ideal component of forthcoming trials testing more extremely hypofractionated schedules (<5 fractions). Both germline and tumour genetic information will likely strengthen future "big data" approaches to radiotherapy outcome prediction.

8.4 Concluding Remarks

Moderately hypofractionated radiotherapy for nmPCa is safe, convenient and efficacious [1]. This thesis has focussed on extending our knowledge to better understand and predict the toxicity resulting from ultra-hypofractionation. I hope that data and models generated will be of use to patients in making informed treatment decisions, and to clinicians, both for delivering such treatments and for planning future ultrahypofractionation trials.

Bibliography

[1] Dearnaley D, Syndikus I, Mossop H, Khoo V, Birtle A, Bloomfield 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. Lancet Oncol 2016;17:1047–60. doi:10.1016/S1470-2045(16)30102-4.

[2] Lee WR, Dignam JJ, Amin MB, Bruner DW, Low D, Swanson GP, et al. Randomized phase III noninferiority study comparing two radiotherapy fractionation schedules in patients with low-risk prostate cancer. J Clin Oncol 2016;34:2325–32. doi:10.1200/JCO.2016.67.0448.

[3] Catton CN, Lukka H, Gu C-S, Martin JM, Supiot S, Chung PWM, et al. Randomized Trial of a Hypofractionated Radiation Regimen for the Treatment of Localized Prostate Cancer. J Clin Oncol 2017;35:1884–90. doi:10.1200/JCO.2016.71.7397.

[4] Incrocci L, Wortel RC, Alemayehu WG, Aluwini S, Schimmel E, Krol S, 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. Lancet Oncol 2016;17:1061–9. doi:10.1016/S1470-2045(16)30070-5.

[5] Fowler JF. 21 Years of biologically effective dose. Br J Radiol 2010;83:554–68. doi:10.1259/bjr/31372149.

[6] Fowler J, Chappell R, Ritter M. Is α/β for prostate tumors really low? Int J Radiat Oncol Biol Phys 2001;50:1021–31. doi:10.1016/S0360-3016(01)01607-8.

[7] Brenner DJ. Fractionation and late rectal toxicity. Int J Radiat Oncol 2004;60:1013–5. doi:10.1016/j.ijrobp.2004.04.014.

[8] Fowler JF, Ritter MA, Chappell RJ, Brenner DJ. What hypofractionated protocols should be tested for prostate cancer? Int J Radiat Oncol Biol Phys 2003;56:1093–104. doi:10.1016/S0360-3016(03)00132-9.

[9] Lyhne NM, Primdahl H, Kristensen CA, Andersen E, Johansen J, Andersen LJ, et al. The DAHANCA 6 randomized trial: Effect of 6 vs 5 weekly fractions of radiotherapy in patients with glottic squamous cell carcinoma. Radiother Oncol 2015;117:91–8. doi:10.1016/j.radonc.2015.07.004.

[10] Rammant E, Ost P, Swimberghe M, Vanderstraeten B, Lumen N, Decaestecker K, et al. Patient- versus physician-reported outcomes in prostate cancer patients receiving hypofractionated radiotherapy within a randomized controlled trial. Strahlentherapie Und Onkol 2019;195:393–401. doi:10.1007/s00066-018-1395-y.

[11] Gulliford SL, Partridge M, Sydes MR, Webb S, Evans PM, Dearnaley DP. Parameters for the Lyman Kutcher Burman (LKB) model of Normal Tissue Complication Probability (NTCP) for specific rectal complications observed in clinical practise. Radiother Oncol 2012;102:347–51. doi:10.1016/j.radonc.2011.10.022.

[12] Nitsche M, Brannath W, Brückner M, Wagner D, Kaltenborn A, Temme N, et al.

Comparison of different contouring definitions of the rectum as organ at risk (OAR) and dose-volume parameters predicting rectal inflammation in radiotherapy of prostate cancer: Which definition to use? Br J Radiol 2017;90. doi:10.1259/bjr.20160370.

[13] Liu X, Li J. Patient Specific Characteristics Are an Important Factor That Determines the Risk of Acute Grade ≥ 2 Rectal Toxicity in Patients Treated for Prostate Cancer with IMRT and Daily Image Guidance Based on Implanted Gold Markers. Omi J Radiol 2016;5. doi:10.4172/2167-7964.1000225.

[14] Cancer Research UK. Prostate Cancer Incidence Statistics. Cancer Res UK 2013:1. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-One (accessed August 1, 2018).

[15] Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: A population-based study. Lancet 2015;385:1206–18. doi:10.1016/S0140-6736(14)61396-9.

[16] Moore AL, Dimitropoulou P, Lane A, Powell PH, Greenberg DC, Brown CH, et al. Population-based prostate-specific antigen testing in the UK leads to a stage migration of prostate cancer. BJU Int 2009;104:1592–8. doi:10.1111/j.1464-410X.2009.08652.x.

[17] NICE. Prostate cancer: diagnosis and treatment. CG 175. NICE; 2014.

[18] National Comprehensive Cancer Network. Prostate Cancer (Version 1.2019) 2019. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf (accessed April 2, 2019).

[19] Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med 2016;375:1415–24. doi:10.1056/NEJMoa1606220.

[20] NICE. Prostate cancer: diagnosis and management. NICE guideline [NG131] 2019.

[21] Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, et al. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. N Engl J Med 2016;375:1425–37. doi:10.1056/NEJMoa1606221.

[22] Rodda S, Tyldesley S, Morris WJ, Keyes M, Halperin R, Pai H, et al. ASCENDE-RT: An Analysis of Treatment-Related Morbidity for a Randomized Trial Comparing a Low-Dose-Rate Brachytherapy Boost with a Dose-Escalated External Beam Boost for High- and Intermediate-Risk Prostate Cancer. Int J Radiat Oncol Biol Phys 2017;98:286–95. doi:10.1016/j.ijrobp.2017.01.008.

[23] Morris WJ, Tyldesley S, Rodda S, Halperin R, Pai H, McKenzie M, et al. Androgen Suppression Combined with Elective Nodal and Dose Escalated Radiation Therapy (the ASCENDE-RT Trial): An Analysis of Survival Endpoints for a Randomized Trial Comparing a Low-Dose-Rate Brachytherapy Boost to a Dose-Escalated External Beam Boost f. Int J Radiat Oncol Biol Phys 2017;98:275–85.

doi:10.1016/j.ijrobp.2016.11.026.

[24] Giberti C, Chiono L, Gallo F, Schenone M, Gastaldi E. Radical retropubic prostatectomy versus brachytherapy for low-risk prostatic cancer: A prospective study. World J Urol 2009;27:607–12. doi:10.1007/s00345-009-0418-9.

[25] Demanes DJ, Ghilezan MI. High-dose-rate brachytherapy as monotherapy for prostate cancer. Brachytherapy 2014;13:529–41. doi:10.1016/j.brachy.2014.03.002.

[26] Morton G, McGuffin M, Chung HT, Ravi A, Tseng CL, Hasan Y, et al. High Dose-Rate Prostate Brachytherapy as Monotherapy: a Single Fraction May Not be Enough. Int J Radiat Oncol 2019;105:S56–7. doi:10.1016/j.ijrobp.2019.06.492.

[27] Hoskin PJ, Rojas AM, Bownes PJ, Lowe GJ, Ostler PJ, Bryant L. Randomised trial of external beam radiotherapy alone or combined with high-dose-rate brachytherapy boost for localised prostate cancer. Radiother Oncol 2012;103:217–22. doi:10.1016/j.radonc.2012.01.007.

[28] Wise J. Prostate cancer: only one in four men with metastatic disease gets recommended chemotherapy. BMJ 2020;368:m120. doi:10.1136/bmj.m120.

[29] Harisiadis L, Senyszyn JJ, Tretter P, Chang CH, Veenema RJ, Puchner PJ, et al. Carcinoma of the prostate: Treatment with external radiotherapy. Cancer 1978;41:2131–42. doi:10.1002/1097-0142(197806)41:6<2131::AID-CNCR2820410609>3.0.CO;2-R.

[30] Rosen EM, Cassady JR, Connolly J, Chaffey JT. Radiotherapy. for localized prostate carcinoma. Int J Radiat Oncol Biol Phys 1984;10:2201–10. doi:10.1016/0360-3016(84)90224-4.

[31] Goitein M, Abrams M. Multi-dimensional treatment planning: I. Delineation of anatomy. Int J Radiat Oncol Biol Phys 1983;9:777–87.

[32] Matsuoka A, Inamura K, Furukawa T. [Computer-controlled multi-leaf conformation radiotherapy system]. Iyodenshi To Seitai Kogaku 1981;19:393–4.

[33] Dearnaley DP, Khoo VS, Norman AR, Meyer L, Nahum A, Tait D, et al. Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial. Lancet 1999;353:267–72. doi:10.1016/S0140-6736(98)05180-0.

[34] Hou Z, Li G, Bai S. High dose versus conventional dose in external beam radiotherapy of prostate cancer: a meta-analysis of long-term follow-up. J Cancer Res Clin Oncol 2015;141:1063–71. doi:10.1007/s00432-014-1813-1.

[35] Nutting C, Dearnaley DP, Webb S. Intensity modulated radiation therapy: A clinical review. Br J Radiol 2000;73:459–69. doi:10.1259/bjr.73.869.10884741.

[36] Quan EM, Li X, Li Y, Wang X, Kudchadker RJ, Johnson JL, et al. A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment. Int J Radiat Oncol Biol Phys 2012;83:1169–78. doi:10.1016/j.ijrobp.2011.09.015.

[37] Yu T, Zhang Q, Zheng T, Shi H, Liu Y, Feng S, et al. The effectiveness of intensity modulated radiation therapy versus three-dimensional radiation therapy in prostate

cancer: A meta-analysis of the literatures. PLoS One 2016;11:e0154499. doi:10.1371/journal.pone.0154499.

[38] Ten Haken RK, Forman JD, Heimburger DK, Gerhardsson A, McShan DL, Perez-Tamayo C, et al. Treatment planning issues related to prostate movement in response to differential filling of the rectum and bladder. Int J Radiat Oncol Biol Phys 1991;20:1317–24. doi:10.1016/0360-3016(91)90244-X.

[39] Alasti H, Petric MP, Catton CN, Warde PR. Portal imaging for evaluation of daily on-line setup errors and off-line organ motion during conformal irradiation of carcinoma of the prostate. Int J Radiat Oncol Biol Phys 2001;49:869–84. doi:10.1016/S0360-3016(00)01446-2.

[40] Pathmanathan AU, van As NJ, Kerkmeijer LGW, Christodouleas J, Lawton CAF, Vesprini D, et al. Magnetic Resonance Imaging-Guided Adaptive Radiation Therapy: A “Game Changer” for Prostate Treatment? Int J Radiat Oncol Biol Phys 2018;100:361–73. doi:10.1016/j.ijrobp.2017.10.020.

[41] Leksell L. Stereotactic radiosurgery. J Neurol Neurosurg Psychiatry 1983;46:797–803. doi:10.1136/jnnp.46.9.797.

[42] Blomgren H, Lax I, Näslund I, Svanström R. Stereotactic high dose fraction radiation therapy of extracranial tumors using an accelerator: Clinical experience of the first thirty-one patients. Acta Oncol (Madr) 1995;34:861–70. doi:10.3109/02841869509127197.

[43] Yan Y, Yadav P, Bassetti M, Du K, Saenz D, Harari P, et al. Dosimetric differences in flattened and flattening filter-free beam treatment plans. J Med Phys 2016;41:92–9. doi:10.4103/0971-6203.181636.

[44] Alongi F, Fersino S, Mazzola R, Levra NG, Ricchetti F, Fiorentino A, et al. Phase 2 Study With FFF LINAC-Based SBRT in Five Consecutive Fractions for Localized Prostate Cancer. Int. J. Radiat. Oncol., vol. 99, 2017, p. E210–1. doi:10.1016/j.ijrobp.2017.06.1107.

[45] Vargas C, Saito AI, Hsi WC, Indelicato D, Falchook A, Zengm Q, et al. Cine-magnetic resonance imaging assessment of intrafraction motion for prostate cancer patients supine or prone with and without a rectal balloon. Am J Clin Oncol Cancer Clin Trials 2010;33:11–6. doi:10.1097/COC.0b013e31819fdf7c.

[46] Ferreira BC, Svensson R, Löf J, Brahme A. The Clinical Value of Non-Coplanar Photon Beams in Biologically Optimized Intensity Modulated Dose Delivery on Deep-Seated Tumours. Acta Oncol (Madr) 2003;42:852–64. doi:10.1080/02841860310013120.

[47] Adler JR, Chang SD, Murphy MJ, Doty J, Geis P, Hancock SL. The Cyberknife: A frameless robotic system for radiosurgery. Stereotact. Funct. Neurosurg., vol. 69, Karger Publishers; 1997, p. 124–8. doi:10.1159/000099863.

[48] King CR, Lehmann J, Adler JR, Hai J. CyberKnife radiotherapy for localized prostate cancer: Rationale and technical feasibility. Technol Cancer Res Treat 2003;2:25–9.

doi:10.1177/153303460300200104.

[49] Lin YW, Lin KH, Ho HW, Lin HM, Lin LC, Lee SP, et al. Treatment plan comparison between stereotactic body radiation therapy techniques for prostate cancer: Non-isocentric CyberKnife versus isocentric RapidArc. Phys Medica 2014;30:654–61. doi:10.1016/j.ejmp.2014.03.008.

[50] Fowler F. THE PLACE OF X RAYS IN THE TREATMENT OF CANCER. Br Med J 1914;1:1284–6.

[51] Strandqvist M. Studien über die kumulative Wirkung der Röntgenstrahlen bei Fraktionierung. 1944.

[52] Douglas BG, Fowler JF. The Effect of Multiple Small Doses of X Rays on Skin Reactions in the Mouse and a Basic Interpretation. Radiat Res 1976;66:401. doi:10.2307/3574407.

[53] Joiner M, Kogel A van der. Basic Clinical Radiobiology. Fifth edit. Taylor and Francis; 2018.

[54] Fujisawa H, Genik PC, Kitamura H, Fujimori A, Uesaka M, Kato TA. Comparison of human chordoma cell-kill for 290 MeV/n carbon ions versus 70 MeV protons in vitro. Radiat Oncol 2013;8:91. doi:10.1186/1748-717X-8-91.

[55] Brenner DJ, Hall EJ. Fractionation and protraction for radiotherapy of prostate carcinoma. Int J Radiat Oncol Biol Phys 1999;43:1095–101. doi:10.1016/S0360-3016(98)00438-6.

[56] van Leeuwen CM, Oei AL, Crezee J, Bel A, Franken NAPP, Stalpers LJAA, et al. The alfa and beta of tumours: A review of parameters of the linear-quadratic model, derived from clinical radiotherapy studies. vol. 13. BioMed Central; 2018. doi:10.1186/s13014-018-1040-z.

[57] Brand DH, Yarnold JR. The Linear–Quadratic Model and Implications for Fractionation. Clin Oncol 2019;31:673–7. doi:10.1016/j.clon.2019.06.007.

[58] Fowler JF. The linear-quadratic formula and progress in fractionated radiotherapy. Br J Radiol 1989;62:679–94. doi:10.1259/0007-1285-62-740-679.

[59] Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. Int J Radiat Oncol Biol Phys 1989;16:1623–30. doi:10.1016/0360-3016(89)90972-3.

[60] Lyman JT. Complication probability as assessed from dose-volume histograms. Radiat Res Suppl 1985;8:S13-9.

[61] Wang Z, Chen M, Sun J, Jiang S, Wang L, Wang X, et al. Lyman–Kutcher–Burman normal tissue complication probability modeling for radiation-induced esophagitis in non-small cell lung cancer patients receiving proton radiotherapy: LKB NTCP modeling for radiation-induced esophagitis in NSCLC patients receiving pr. Radiother Oncol 2020;146:200–4. doi:10.1016/j.radonc.2020.03.003.

[62] Marzi S, Saracino B, Petrongari M, Arcangeli S, Gomellini S, Arcangeli G, et al. Modeling of alpha/beta for late rectal toxicity from a randomized phase II study:

conventional versus hypofractionated scheme for localized prostate cancer. J Exp Clin Cancer Res 2009;28:117. doi:10.1186/1756-9966-28-117.

[63] Tucker SL, Liu HH, Liao Z, Wei X, Wang S, Jin H, et al. Analysis of Radiation Pneumonitis Risk Using a Generalized Lyman Model. Int J Radiat Oncol Biol Phys 2008;72:568–74. doi:10.1016/j.ijrobp.2008.04.053.

[64] Joiner MC, Marples B, Lambin P, Short SC, Turesson I. Low-dose hypersensitivity: Current status and possible mechanisms. Int. J. Radiat. Oncol. Biol. Phys., vol. 49, 2001, p. 379–89. doi:10.1016/S0360-3016(00)01471-1.

[65] Macià i Garau M. Radiobiology of stereotactic body radiation therapy (SBRT). Reports Pract Oncol Radiother 2017;22:86–95. doi:10.1016/j.rpor.2017.02.010.

[66] Brenner DJ. The Linear-Quadratic Model Is an Appropriate Methodology for Determining Isoeffective Doses at Large Doses Per Fraction. Semin Radiat Oncol 2008;18:234–9. doi:10.1016/j.semradonc.2008.04.004.

[67] Morton G, McGuffin M, Chung HT, Tseng CL, Helou J, Ravi A, et al. Prostate high dose-rate brachytherapy as monotherapy for low and intermediate risk prostate cancer: Efficacy results from a randomized phase II clinical trial of one fraction of 19 Gy or two fractions of 13.5 Gy. Radiother Oncol 2020;146:90–6. doi:10.1016/j.radonc.2020.02.009.

[68] Brown JM, Brenner DJ, Carlson DJ. Dose Escalation, Not “New Biology,” Can Account for the Efficacy of Stereotactic Body Radiation Therapy With Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2013;85:1159–60. doi:10.1016/j.ijrobp.2012.11.003.

[69] Datta NR, Stutz E, Rogers S, Bodis S. Conventional Versus Hypofractionated Radiation Therapy for Localized or Locally Advanced Prostate Cancer: A Systematic Review and Meta-analysis along with Therapeutic Implications. Int J Radiat Oncol Biol Phys 2017;99:573–89. doi:10.1016/j.ijrobp.2017.07.021.

[70] Royce TJ, Lee DH, Keum N, Permpalung N, Chiew CJ, Epstein S, et al. Conventional Versus Hypofractionated Radiation Therapy for Localized Prostate Cancer: A Meta-analysis of Randomized Noninferiority Trials. Eur Urol Focus 2017:1–8. doi:10.1016/j.euf.2017.10.011.

[71] Pollack A, Walker G, Horwitz EM, Price R, Feigenberg S, Konski AA, et al. Randomized trial of hypofractionated external-beam radiotherapy for prostate cancer. J Clin Oncol 2013;31:3860–8. doi:10.1200/JCO.2013.51.1972.

[72] Yeoh EE, Botten RJ, Butters J, Di Matteo AC, Holloway RH, Fowler J. Hypofractionated Versus Conventionally Fractionated Radiotherapy for Prostate Carcinoma: Final Results of Phase III Randomized Trial. Int J Radiat Oncol 2011;81:1271–8. doi:10.1016/j.ijrobp.2010.07.1984.

[73] Hoffman KE, Voong KR, Levy LB, Pugh TJ, Choi S, Du W, et al. Randomized Trial of Hypofractionated Dose-Escalated Intensity Modulated Radiation Therapy Versus Conventionally Fractionated Intensity Modulated Radiation Therapy for Localized

Prostate Cancer. Int J Radiat Oncol 2016;96:S32. doi:10.1016/j.ijrobp.2016.06.091.

[74] Arcangeli G, Saracino B, Arcangeli S, Gomellini S, Petrongari MG, Sanguineti G, et al. Moderate Hypofractionation in High-Risk, Organ-Confined Prostate Cancer: Final Results of a Phase III Randomized Trial. J Clin Oncol 2017;35:1891–7. doi:10.1200/JCO.2016.70.4189.

[75] Aluwini S, Pos F, Schimmel E, Krol S, van der Toorn PP, de Jager H, et al. Hypofractionated versus conventionally fractionated radiotherapy for patients with prostate cancer (HYPRO): late toxicity results from a randomised, non-inferiority, phase 3 trial. Lancet Oncol 2016;17:464–74. doi:10.1016/S1470-2045(15)00567-7.

[76] Kogel AJ Van Der, Jarrett KA. Radiation tolerance of the rat rectum to fractionated X-rays and pi-mesons. Animals 1988;12:225–32.

[77] Dewit L, Oussoren Y, Bartelink H, Thames HD. The effect of cis-diamminedichloroplatinum(II) on radiation damage in mouse rectum after fractionated irradiation. Radiother Oncol 1989;16:121–8. doi:10.1016/0167-8140(89)90029-7.

[78] Gasinska A, Dubray B, Hill SA, Denekamp J, Thames HD, Fowler JF. Early and late injuries in mouse rectum after fractionated X-ray and neutron irradiation. Radiother Oncol 1993;26:244–53. doi:10.1016/0167-8140(93)90266-B.

[79] Terry NHA, Denekamp J. RBE values and repair characteristics for colo-rectal injury after caesium 137 gamma-ray and neutron irradiation. II. Fractionation up to ten doses. Br J Radiol 1984;57:617–29. doi:10.1259/0007-1285-57-679-617.

[80] Brenner D, Armour E, Corry P, Hall E. Sublethal damage repair times for a late-responding tissue relevant to brachytherapy (and external-beam radiotherapy): Implications for new brachytherapy protocols. Int J Radiat Oncol Biol Phys 1998;41:135–8. doi:10.1016/S0360-3016(98)00029-7.

[81] Dubray BM, Thames HD. Chronic radiation damage in the rat rectum: an analysis of the influences of fractionation, time and volume. Radiother Oncol 1994;33:41–7. doi:10.1016/0167-8140(94)90084-1.

[82] Kogel AJ Van Der, Jarrett KA, van der Kogel AJ, Jarrett KA, Paciotti MA, Raju MR. Radiation tolerance of the rat rectum to fractionated X-rays and pi-mesons. Animals 1988;12:225–32. doi:10.1016/0167-8140(88)90265-4.

[83] Skwarchuk MW, Jackson A, Zelefsky MJ, Venkatraman ES, Cowen DM, Levegrün S, et al. Late rectal toxicity after conformal radiotherapy of prostate cancer (I): Multivariate analysis and dose-response. Int J Radiat Oncol Biol Phys 2000;47:103–13. doi:10.1016/S0360-3016(99)00560-X.

[84] Kuban D, Pollack A, Huang E, Levy L, Dong L, Starkschall G, et al. Hazards of dose escalation in prostate cancer radiotherapy. Int J Radiat Oncol Biol Phys 2003;57:1260–8. doi:10.1016/S0360-3016(03)00772-7.

[85] Michalski JM, Winter K, Purdy JA, Wilder RB, Perez CA, Roach M, et al. Preliminary evaluation of low-grade toxicity with conformal radiation therapy for prostate cancer on RTOG 9406 dose levels I and II. Int J Radiat Oncol Biol Phys 2003;56:192–8.

doi:10.1016/S0360-3016(03)00072-5.

[86] Akimoto T, Muramatsu H, Takahashi M, Saito JI, Kitamoto Y, Harashima K, et al. Rectal bleeding after hypofractionated radiotherapy for prostate cancer: Correlation between clinical and dosimetric parameters and the incidence of grade 2 or worse rectal bleeding. Int J Radiat Oncol Biol Phys 2004;60:1033–9. doi:10.1016/j.ijrobp.2004.07.695.

[87] Burman C, Kutcher GJ, Emami B, Goitein M. Fitting of normal tissue tolerance data to an analytic function. Int J Radiat Oncol Biol Phys 1991;21:123–35. doi:10.1016/0360-3016(91)90172-Z.

[88] Tucker SL, Thames HD, Michalski JM, Bosch WR, Mohan R, Winter K, et al. Estimation of α/β for late rectal toxicity based on RTOG 94-06. Int J Radiat Oncol Biol Phys 2011;81:600–5. doi:10.1016/j.ijrobp.2010.11.080.

[89] Zhu J, Simon A, Bossi A, Messai T, Ospina JD, Li B, et al. PO-0961: Estimation of a/þ ratio for late rectal toxicity from localized prostate cancer irradiation. Radiother Oncol 2014;111:S131. doi:10.1016/S0167-8140(15)31079-3.

[90] Lloyd-Davies RW, Collins CD, Swan AV. Carcinoma of prostate treated by radical external beam radiotherapy using hypofractionation twenty-two years' experience (1962–1984). Urology 1990;36:107–11. doi:10.1016/0090-4295(90)80207-4.

[91] Collins CD, Lloyd-Davies RW, Swan A V. Radical external beam radiotherapy for localised carcinoma of the prostate using a hypofractionation technique. Clin Oncol 1991;3:127–32. doi:10.1016/S0936-6555(05)80831-3.

[92] Madsen BL, Hsi RA, Pham HT, Fowler JF, Esagui L, Corman J. Stereotactic hypofractionated accurate radiotherapy of the prostate (SHARP), 33.5 Gy in five fractions for localized disease: First clinical trial results. Int J Radiat Oncol Biol Phys 2007;67:1099–105. doi:10.1016/j.ijrobp.2006.10.050.

[93] King CR, Brooks JD, Gill H, Pawlicki T, Cotrutz C, Presti JC. Stereotactic Body Radiotherapy for Localized Prostate Cancer: Interim Results of a Prospective Phase II Clinical Trial. Int J Radiat Oncol Biol Phys 2009;73:1043–8. doi:10.1016/j.ijrobp.2008.05.059.

[94] Kishan AU, Dang A, Katz AJ, Mantz CA, Collins SP, Aghdam N, et al. Long-term Outcomes of Stereotactic Body Radiotherapy for Low-Risk and Intermediate-Risk Prostate Cancer. JAMA Netw Open 2019;2:e188006. doi:10.1001/jamanetworkopen.2018.8006.

[95] Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, 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. Lancet 2019;394:385–95. doi:10.1016/S0140-6736(19)31131-6.

[96] Overgaard J, Hjelm-Hansen M, Johansen LV, Andersen AP. Comparison of conventional and Split-course radiotherapy as primary treatment in carcinoma of the

larynx. Acta Oncol (Madr) 1988;27:147–52. doi:10.3109/02841868809090334.

[97] Meier RM, Bloch DA, Cotrutz C, Beckman AC, Henning GT, Woodhouse SA, et al. Multicenter Trial of Stereotactic Body Radiation Therapy for Low- and Intermediate-Risk Prostate Cancer: Survival and Toxicity Endpoints. Int J Radiat Oncol Biol Phys 2018;102:296–303. doi:10.1016/j.ijrobp.2018.05.040.

[98] Katz AJ, Santoro M, Diblasio F, Ashley R. Stereotactic body radiotherapy for localized prostate cancer: Disease control and quality of life at 6 years. Radiat Oncol 2013;8:118. doi:10.1186/1748-717X-8-118.

[99] Fuller DB, Falchook AD, Crabtree T, Kane BL, Medbery CA, Underhill K, et al. Phase 2 Multicenter Trial of Heterogeneous-dosing Stereotactic Body Radiotherapy for Low- and Intermediate-risk Prostate Cancer: 5-year Outcomes. Eur Urol Oncol 2018;1:540–7. doi:10.1016/j.euo.2018.06.013.

[100] Quon HC, Ong A, Cheung P, Chu W, Chung HT, Vesprini D, et al. Once-weekly versus every-other-day stereotactic body radiotherapy in patients with prostate cancer (PATRIOT): A phase 2 randomized trial. Radiother Oncol 2018;127:206–12. doi:10.1016/j.radonc.2018.02.029.

[101] Zelefsky MJ, Kollmeier M, McBride S, Varghese M, Mychalczak B, Gewanter R, et al. 5-Year Outcomes of a Phase I Dose Escalation Study Using Stereotactic Body Radiosurgery for Patients with Low and Intermediate Risk Prostate Cancer. Int J Radiat Oncol 2019. doi:10.1016/J.IJROBP.2018.12.045.

[102] Mantz C. A Phase II Trial of Stereotactic Ablative Body Radiotherapy for Low-Risk Prostate Cancer Using a Non-Robotic Linear Accelerator and Real-Time Target Tracking: Report of Toxicity, Quality of Life, and Disease Control Outcomes with 5-Year Minimum Follow-Up. Front Oncol 2014;4:279. doi:10.3389/fonc.2014.00279.

[103] Hannan R, Tumati V, Xie XJ, Cho LC, Kavanagh BD, Brindle J, et al. Stereotactic body radiation therapy for low and intermediate risk prostate cancer - Results from a multi-institutional clinical trial. Eur J Cancer 2016;59:142–51. doi:10.1016/j.ejca.2016.02.014.

[104] Loblaw A, Cheung P, D'Alimonte L, Deabreu A, Mamedov A, Zhang L, et al. Prostate stereotactic ablative body radiotherapy using a standard linear accelerator: Toxicity, biochemical, and pathological outcomes. Radiother Oncol 2013;107:153–8. doi:10.1016/j.radonc.2013.03.022.

[105] Jackson WC, Dess RT, Litzenberg DW, Li P, Schipper M, Rosenthal SA, et al. A multi-institutional phase 2 trial of prostate stereotactic body radiation therapy (SBRT) using continuous real-time evaluation of prostate motion with patient-reported quality of life. Pract Radiat Oncol 2018;8:40–7. doi:10.1016/j.prro.2017.08.004.

[106] Boyer MJ, Papagikos MA, Kiteley R, Vujaskovic Z, Wu J, Lee WR. Toxicity and quality of life report of a phase II study of stereotactic body radiotherapy (SBRT) for low and intermediate risk prostate cancer. Radiat Oncol 2017;12:14. doi:10.1186/s13014-016-0758-8.

[107] McBride SM, Wong DS, Dombrowski JJ, Harkins B, Tapella P, Hanscom HN, et al. Hypofractionated stereotactic body radiotherapy in low-risk prostate adenocarcinoma: Preliminary results of a multi-institutional phase 1 feasibility trial. Cancer 2012;118:3681–90. doi:10.1002/cncr.26699.

[108] Bolzicco G, Favretto MS, Scremin E, Tambone C, Tasca A, Guglielmi R. Image-guided stereotactic body radiation therapy for clinically localized prostate cancer: Preliminary clinical results. Technol Cancer Res Treat 2010;9:473–7. doi:10.1177/153303461000900505.

[109] Alongi F, Cozzi L, Arcangeli S, Iftode C, Comito T, Villa E, et al. Linac based SBRT for prostate cancer in 5 fractions with VMAT and flattening filter free beams: Preliminary report of a phase II study. Radiat Oncol 2013;8:171. doi:10.1007/978-3-319-27784-4_12.

[110] Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, et al. Extreme Hypofractionation versus Conventionally Fractionated Radiotherapy for Intermediate Risk Prostate Cancer: Early Toxicity Results from the Scandinavian Randomized Phase III Trial “HYPO-RT-PC.” Int J Radiat Oncol 2016;96:938–9.

[111] Dörr W, Hendry JH. Consequential late effects in normal tissues. Radiother Oncol 2001;61:223–31. doi:10.1016/S0167-8140(01)00429-7.

[112] Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: Retrospective cohort study of drug approvals 2009-13. BMJ 2017;359. doi:10.1136/bmj.j4530.

[113] NICE. 7 Assessing cost effectiveness | The guidelines manual | Guidance | NICE n.d. https://www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness (accessed June 17, 2020).

[114] Sweeney C, Nakabayashi M, Regan M, Xie W, Hayes J, Keating N, et al. The Development of Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP). J Natl Cancer Inst 2015;107:djv261. doi:10.1093/jnci/djv261.

[115] Cohen L, Kerrich JE. Estimation of biological dosage factors in clinical radiotherapy. Br J Cancer 1951;5:180–94. doi:10.1038/bjc.1951.20.

[116] World Health Organization. Handbook for Reporting Results of Cancer Treatment. 1979.

[117] NCI. Guidelines for reporting of adverse drug reactions. 1988.

[118] RTOG, Rubin P, Constine LS, Fajardo LF, Phillips TL, Wasserman TH, et al. Late effects consensus conference: RTOG/EORTC. Radiother Oncol 1995;35:5–7. doi:10.1016/0167-8140(95)01548-U.

[119] Cox JD, Stetz J, Pajak TF. Toxicity criteria of the radiation therapy oncology group (RTOG) and the european organization for research and treatment of cancer (EORTC). Pergamon Int J Radi Ati Oncol Biol Phys 1995;3:34–1346.

doi:10.1016/0360-3016(95)00060-C.

[120] Dische S, Warburton MF, Jones D, Lartigau E. The recording of morbidity related to radiotherapy. Radiother Oncol 1989;16:103–8. doi:10.1016/0167-8140(89)90026-1.

[121] LENT SOMA tables. Radiother Oncol 1995;35:1–81. doi:10.1093/carcin/bgt460.

[122] Zhen Y, Jiang Y, Yuan L, Kirkpartrick J, Wu J, Ge Y. Analyzing the usage of standards in radiation therapy clinical studies. 2017 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2017, vol. 2017, Institute of Electrical and Electronics Engineers Inc.; 2017, p. 349–52. doi:10.1109/BHI.2017.7897277.

[123] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. JNCI J Natl Cancer Inst 1993;85:365–76. doi:10.1093/jnci/85.5.365.

[124] Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: Development and validation of the general measure. J Clin Oncol 1993;11:570–9. doi:10.1200/JCO.1993.11.3.570.

[125] Ware JEJ, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

[126] Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: Scoping review. SAGE Open Med 2016;4:205031211667172. doi:10.1177/2050312116671725.

[127] Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: A history and review of methods for the cultural adaptation of the euroqol five-dimensional questionnaire. Value Heal 2014;17:70–6. doi:10.1016/j.jval.2013.10.006.

[128] Position statement on use of the EQ-5D-5L value set for England (updated October 2019) | NICE technology appraisal guidance | NICE guidance | Our programmes | What we do | About | NICE n.d.

[129] Franks P, Lubetkin EI, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQol EQ-5D index in a national US sample. Med Decis Mak 2004;24:247–54. doi:10.1177/0272989X04265477.

[130] van Andel G, Bottomley A, Fosså SD, Efficace F, Coens C, Guerif S, et al. An international field study of the EORTC QLQ-PR25: A questionnaire for assessing the health-related quality of life of patients with prostate cancer. Eur J Cancer 2008;44:2418–24. doi:10.1016/j.ejca.2008.07.030.

[131] Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology 1997;50:920–8. doi:10.1016/S0090-4295(97)00459-7.

[132] Litwin MS, Hays RD, 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. Med Care 1998;36:1002–12. doi:10.1097/00005650-199807000-00007.

[133] Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology 2000;56:899–905. doi:10.1016/S0090-4295(00)00858-X.

[134] Szymanski KM, Wei JT, Dunn RL, Sanda MG. 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;76:1245–50. doi:10.1016/j.urology.2010.01.027.

[135] Fransson P, Tavelin B, Widmark A. Reliability and responsiveness of a prostate cancer questionnaire for radiotherapy-induced side effects. Support Care Cancer 2001;9:187–98. doi:10.1007/s005200000146.

[136] Barry MJ, Fowler FJ, O'leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, et al. The American Urological Association Symptom Index for Benign Prostatic Hyperplasia. J Urol 1992;148:1549–57.

[137] Terk MD, Stock RG, Stone NN. Identification of patients at increased risk for prolonged urinary retention following radioactive seed implantation of the prostate. J Urol 1998;160:1379–82. doi:10.1016/S0022-5347(01)62542-4.

[138] Tree AC, Ostler P, Hoskin P, Dankulchai P, Nariyangadu P, Hughes RJ, et al. Prostate Stereotactic Body Radiotherapy - First UK Experience. Clin Oncol 2014;26:757–61. doi:10.1016/j.clon.2014.08.007.

[139] Olopade FA, Norman A, Blake P, Dearnaley DP, Harrington KJ, Khoo V, et al. A modified Inflammatory Bowel Disease questionnaire and the Vaizey incontinence questionnaire are simple ways to identify patients with significant gastrointestinal symptoms after pelvic radiotherapy. Br J Cancer 2005;92:1663–70. doi:10.1038/sj.bjc.6602552.

[140] Khalid U, McGough C, Hackett C, Blake P, Harrington KJ, Khoo VS, et al. A modified inflammatory bowel disease questionnaire and the Vaizey Incontinence questionnaire are more sensitive measures of acute gastrointestinal toxicity during pelvic radiotherapy than RTOG grading. Int J Radiat Oncol Biol Phys 2006;64:1432–41. doi:10.1016/j.ijrobp.2005.10.007.

[141] Rosen RC, Cappelleri JC, Smith MD, Lipsky J, Peñ BM. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res 1999;11:319–26. doi:10.1038/sj.ijir.3900472.

[142] Zagar TM, Stock RG, Cesaretti JA, Stone NN. Assessment of postbrachytherapy sexual function: A comparison of the IIEF-5 and the MSEFS. Brachytherapy 2007;6:26–33. doi:10.1016/j.brachy.2006.11.003.

[143] Wilkins A, Mossop H, Syndikus I, Khoo V, Bloomfield D, Parker C, 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. Lancet Oncol 2015;16:1605–16. doi:10.1016/S1470-2045(15)00280-6.

[144] MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods 2002;7:19–40. doi:10.1037/1082-989X.7.1.19.

[145] Altman DG, Royston P. The cost of dichotomising continuous variables. Br Med J 2006;332:1080. doi:10.1136/bmj.332.7549.1080.

[146] Heinze G. CeMSIIS-Section for Clinical Biometrics Logistic regression with rare events: problems and solutions. n.d.

[147] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics; 2013.

[148] Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. Digit Signal Process A Rev J 2018;73:1–15. doi:10.1016/j.dsp.2017.10.011.

[149] Kuhn M. 3 Pre-Processing | The caret Package n.d. https://topepo.github.io/caret/pre-processing.html (accessed July 24, 2020).

[150] Marks LB, Bentzen SM, Deasy JO, Kong F-M (Spring), Bradley JD, Vogelius IS, et al. Radiation Dose–Volume Effects in the Lung. Int J Radiat Oncol 2010;76:S70–6. doi:10.1016/j.ijrobp.2009.06.091.

[151] Tucker SL, Zhang M, Dong L, Mohan R, Kuban D, Thames HD. Cluster model analysis of late rectal bleeding after IMRT of prostate cancer: A case-control study. Int J Radiat Oncol Biol Phys 2006;64:1255–64. doi:10.1016/j.ijrobp.2005.10.029.

[152] Buettner F, Gulliford SL, Webb S, Partridge M. Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach. Phys Med Biol 2009;54:5139–53. doi:10.1088/0031-9155/54/17/005.

[153] Dean JA, Wong KH, Welsh LC, Jones AB, Schick U, Newbold KL, et al. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. Radiother Oncol 2016;120:21–7. doi:10.1016/j.radonc.2016.05.015.

[154] Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO. Radiation Dose-Volume Effects in Radiation-Induced Rectal Injury. Int J Radiat Oncol Biol Phys 2010;76:123–9. doi:10.1016/j.ijrobp.2009.03.078.

[155] Sanguineti G, Faiella A, Farneti A, D'Urso P, Fuga V, Olivieri M, et al. Refinement & validation of rectal wall dose volume objectives for prostate hypofractionation in 20 fractions. Clin Transl Radiat Oncol 2020;21:91–7. doi:10.1016/j.ctro.2020.01.006.

[156] Scaife JE, Thomas SJ, Harrison K, Romanchikova M, F Sutcliffe MP, Forman JR, 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;88. doi:10.1259/bjr.20150243.

[157] Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24:498–520. doi:10.1037/h0070888.

[158] Vivekanandan S, Landau DB, Counsell N, Warren DR, Khwanda A, Rosen SD, et al. The Impact of Cardiac Radiation Dosimetry on Survival After Radiation Therapy for Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2017;99:51–60. doi:10.1016/j.ijrobp.2017.04.026.

[159] Dean JA, Wong KH, Gay H, Welsh LC, Jones AB, Schick U, et al. Functional Data Analysis Applied to Modeling of Severe Acute Mucositis and Dysphagia Resulting From Head and Neck Radiation Therapy. Int J Radiat Oncol Biol Phys 2016;96:820–31. doi:10.1016/j.ijrobp.2016.08.013.

[160] Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. Biometrical J 2018;60:431–49. doi:10.1002/bimj.201700067.

[161] Alsadius D, Hedelin M, Lundstedt D, Pettersson N, Wilderäng U, Steineck G. Mean absorbed dose to the anal-sphincter region and fecal leakage among irradiated prostate cancer survivors. Int J Radiat Oncol Biol Phys 2012;84:e181-5. doi:10.1016/j.ijrobp.2012.03.065.

[162] Buettner F, Gulliford SL, Webb S, Partridge M. Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution. Phys Med Biol 2011;56:2103–18. doi:10.1088/0031-9155/56/7/013.

[163] Stenmark MH, Conlon ASC, Johnson S, Daignault S, Litzenberg D, Marsh R, et al. Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer. Radiother Oncol 2014;110:291–7. doi:10.1016/j.radonc.2014.01.007.

[164] Thor M, Olsson CE, Oh JH, Petersen SE, Alsadius D, Bentzen L, et al. Relationships between dose to the gastro-intestinal tract and patient-reported symptom domains after radiotherapy for localized prostate cancer. Acta Oncol (Madr) 2015;54:1326–34. doi:10.3109/0284186X.2015.1063779.

[165] Schaake W, van der Schaaf A, van Dijk L V., Bongaerts AHH, van den Bergh ACM, Langendijk JA. Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer NTCP models for anorectal side effects patients. Radiother Oncol 2016;119:381–7. doi:10.1016/j.radonc.2016.04.005.

[166] Cicchetti A, Rancati T, Ebert M, Fiorino C, Palorini F, Kennedy A, et al. Modelling late stool frequency and rectal pain after radical radiotherapy in prostate cancer patients: Results from a large pooled population. Phys Medica 2016;32:1690–7. doi:10.1016/j.ejmp.2016.09.018.

[167] Arcangeli G, Fowler J, Gomellini S, Arcangeli S, Saracino B, Petrongari MG, et al. Acute and late toxicity in a randomized trial of conventional versus hypofractionated three-dimensional conformal radiotherapy for prostate cancer. Int J Radiat Oncol Biol

Phys 2011;79:1013–21. doi:10.1016/j.ijrobp.2009.12.045.

[168] Thor M, Deasy JO, Paulus R, Robert Lee W, Amin MB, Bruner DW, et al. Tolerance doses for late adverse events after hypofractionated radiotherapy for prostate cancer on trial NRG Oncology/RTOG 0415. Radiother Oncol 2019;135:19–24. doi:10.1016/j.radonc.2019.02.014.

[169] Ebert MA, Foo K, Haworth A, Gulliford SL, Kennedy A, Joseph DJ, 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;91:595–603. doi:10.1016/j.ijrobp.2014.11.015.

[170] Moulton CR, House MJ, Lye V, Tang CI, Krawiec M, Joseph DJ, et al. Prostate external beam radiotherapy combined with high-dose-rate brachytherapy: Dose-volume parameters from deformably-registered plans correlate with late gastrointestinal complications. Radiat Oncol 2016;11:144. doi:10.1186/s13014-016-0719-2.

[171] Okamoto M, Ishikawa H, Ebara T, Kato H, Tamaki T, Akimoto T, et al. Rectal bleeding after high-dose-rate brachytherapy combined with hypofractionated external-beam radiotherapy for localized prostate cancer: The relationship between dose-volume histogram parameters and the occurrence rate. Int J Radiat Oncol Biol Phys 2012;82:e211–7. doi:10.1016/j.ijrobp.2011.03.041.

[172] Andreyev HJN, Vlavianos P, Blake P, Dearnaley D, Norman AR, Tait D. Gastrointestinal symptoms after pelvic radiotherapy: Role for the gastroenterologist? Int J Radiat Oncol Biol Phys 2005;62:1464–71. doi:10.1016/j.ijrobp.2004.12.087.

[173] Flickinger JC, Kondziolka D, Maitz AH, Lunsford LD. Linear-quadratic alpha/beta values for normal tissue injury dose-responses from acoustic neuroma and arteriovenous malformation radiosurgery. Int J Radiat Oncol 2001;51:255. doi:10.1016/S0360-3016(01)02290-8.

[174] Dearnaley DP, Sydes MR, Langley RE, Graham JD, Huddart RA, Syndikus I, et al. The early toxicity of escalated versus standard dose conformal radiotherapy with neo-adjuvant androgen suppression for patients with localised prostate cancer: Results from the MRC RT01 trial (ISRCTN47772397). Radiother Oncol 2007;83:31–41. doi:10.1016/j.radonc.2007.02.014.

[175] Gulliford SL, Foo K, Morgan RC, Aird EG, Bidmead AM, Critchley H, 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;76:747–54. doi:10.1016/j.ijrobp.2009.02.025.

[176] Tucker SL. Incorporating single nucleotide polymorphisms into the Lyman model to improve prediction of radiation pneumonitis. Int J Radiat Oncol Biol Phys 2013;85:251–7. doi:10.1016/j.ijrobp.2012.02.021.INCORPORATING.

[177] Niemierko A. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. Med Phys 1997;24:103–10. doi:10.1118/1.598063.

[178] Efron B, Tibshirani R. Improvements on cross-validation: The .632+ bootstrap method. J Am Stat Assoc 1997;92:548–60. doi:10.1080/01621459.1997.10474007.

[179] Dong Y, Egleston B, Price RA, Rosenstein BS, Meyer JE. The History of the Lyman-Kutcher-Burman Model and Its Application to the Alpha-Beta Ratio Modeling Using Modern Statistics. Int J Radiat Oncol 2017;99:E122. doi:10.1016/j.ijrobp.2017.06.888.

[180] Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim 1998;9:112–47. doi:10.1137/S1052623496303470.

[181] D'Errico J. fminsearchbnd, fminsearchcon 2012. https://uk.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon (accessed January 8, 2020).

[182] Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. J Am Stat Assoc 1983;78:316. doi:10.2307/2288636.

[183] Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità, 1936.

[184] Higgins JPT, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. Stat Med 2008;27:6072–92. doi:10.1002/sim.3427.

[185] Peeters STH, Hoogeman MS, Heemsbergen WD, Hart AAM, Koper PCM, Lebesque J V. 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;66:11–9. doi:10.1016/j.ijrobp.2006.03.034.

[186] Defraene G, Van Den Bergh L, Al-Mamgani A, Haustermans K, Heemsbergen W, Van Den Heuvel F, 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;82:1233–42. doi:10.1016/j.ijrobp.2011.03.056.

[187] Rancati T, Fiorino C, Gagliardi G, Cattaneo GM, Sanguineti G, Borca VC, et al. Fitting late rectal bleeding data using different NTCP models: Results from an Italian multi-centric study (AIROPROS0101). Radiother Oncol 2004;73:21–32. doi:10.1016/j.radonc.2004.08.013.

[188] Brand DH, Tree AC, Ostler P, van der Voet H, Loblaw A, Chu W, 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. Lancet Oncol 2019;20:1531–43. doi:10.1016/S1470-2045(19)30569-8.

[189] Rancati T, Fiorino C. Modelling Radiotherapy Side Effects. Boca Raton, FL : CRC Press, Taylor & Francis Group, [2019] |: CRC Press; 2019. doi:10.1201/b21956.

[190] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55. doi:10.1097/00000421-198212000-00014.

[191] Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours. Wiley-Liss; 2003. doi:10.1002/ejoc.201200111.

[192] Malone S, Roy S, Eapen L, E C, MacRae R, Perry G, et al. Sequencing of Androgen-Deprivation Therapy With External-Beam Radiotherapy in Localized Prostate Cancer: A Phase III Randomized Controlled Trial. J Clin Oncol 2019:JCO.19.01904. doi:10.1200/JCO.19.01904.

[193] Dearnaley D, Hall E. How will the CHHiP trial affect the future of prostate radiotherapy? Expert Rev Anticancer Ther 2018;18:607–9. doi:10.1080/14737140.2018.1477595.

[194] Schwarz F, Christie D. Use of “sham” radiotherapy in randomized clinical trials. J Med Imaging Radiat Oncol 2008;52:269–77. doi:10.1111/j.1440-1673.2008.01936.x.

[195] Lips IM, van der Heide UA, Haustermans K, van Lin EN, Pos F, Franken SP, et al. Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial. Trials 2011;12:255. doi:10.1186/1745-6215-12-255.

[196] Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas 2018;Volume 9:353–67. doi:10.2147/prom.s156279.

[197] Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols the spirit-pro extension. JAMA - J Am Med Assoc 2018;319:483–94. doi:10.1001/jama.2017.21903.

[198] Flores LT, Bennett A V, Law EB, Hajj C, Griffith MP, Goodman KA. Patient-Reported Outcomes vs. Clinician Symptom Reporting During Chemoradiation for Rectal Cancer. Gastrointest Cancer Res 2012;5:119–24.

[199] Haviland JS, Hopwood P, Mills J, Sydenham M, Bliss JM, Yarnold JR. Do Patient-reported Outcome Measures Agree with Clinical and Photographic Assessments of Normal Tissue Effects after Breast Radiotherapy? The Experience of the Standardisation of Breast Radiotherapy (START) Trials in Early Breast Cancer. Clin Oncol 2016;28:345–53. doi:10.1016/j.clon.2016.01.011.

[200] Bhattacharya IS, Haviland JS, Hopwood P, Coles CE, Yarnold JR, Bliss JM, et al. Can patient-reported outcomes be used instead of clinician-reported outcomes and photographs as primary endpoints of late normal tissue effects in breast radiotherapy trials? Results from the IMPORT LOW trial. Radiother Oncol 2019;134:220–30. doi:10.1016/j.radonc.2019.01.036.

[201] Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: Agreement between patient and physician reporting in three randomized trials. J Clin Oncol 2015;33:910–5. doi:10.1200/JCO.2014.57.9334.

[202] Thurtle DR, Greenberg DC, Lee LS, Huang HH, Pharoah PD, Gnanapragasam VJ. Individual prognosis at diagnosis innonmetastatic prostate cancer: Developmentand external validation of the predictprostate multivariable model. PLoS Med 2019;16. doi:10.1371/journal.pmed.1002758.

[203] Barry MJ, Fowler FJ, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol 1992;148:1549–57; discussion 1564.

[204] Skolarus TA, Dunn RL, Sanda MG, Chang P, Greenfield TK, Litwin MS, et al. Minimally important difference for the expanded prostate cancer index composite short form. Urology 2015;85:101–6. doi:10.1016/j.urology.2014.08.044.

[205] Bruner DW, Pugh SL, Lee WR, Hall WA, Dignam JJ, Low D, et al. Quality of Life in Patients with Low-Risk Prostate Cancer Treated with Hypofractionated vs Conventional Radiotherapy: A Phase 3 Randomized Clinical Trial. JAMA Oncol 2019;5:664–70. doi:10.1001/jamaoncol.2018.6752.

[206] Aluwini S, Pos F, Schimmel E, van Lin E, Krol S, van der Toorn PP, et al. Hypofractionated versus conventionally fractionated radiotherapy for patients with prostate cancer (HYPRO): Acute toxicity results from a randomised non-inferiority phase 3 trial. Lancet Oncol 2015;16:274–83. doi:10.1016/S1470-2045(14)70482-6.

[207] Liu T. Route of placebo administration: Robust placebo effects in laboratory and clinical settings. Neurosci Biobehav Rev 2017;83:451–7. doi:10.1016/j.neubiorev.2017.09.018.

[208] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol 2016;121:169–79. doi:10.1016/j.radonc.2016.09.009.

[209] Foppiano F, Fiorino C, Frezza G, Greco C, Valdagni R. The impact of contouring uncertainty on rectal 3D dose-volume data: Results of a dummy run in a multicenter trial (AIROPROS01-02). Int J Radiat Oncol Biol Phys 2003;57:573–9. doi:10.1016/S0360-3016(03)00659-X.

[210] Fiorino C, Vavassori V, Sanguineti G, Bianchi C, Cattaneo GM, Piazzolla A, 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;63:249–55. doi:10.1016/S0167-8140(01)00469-8.

[211] Roach D, Jameson MG, Dowling JA, Ebert MA, Greer PB, Kennedy AM, et al. Correlations between contouring similarity metrics and simulated treatment outcome for prostate radiotherapy. Phys Med Biol 2018;63:35001. doi:10.1088/1361-6560/aaa50c.

[212] Wright JL, Yom SS, Awan MJ, Dawes S, Fischer-Valuck B, Kudner R, et al. Standardizing Normal Tissue Contouring for Radiation Therapy Treatment Planning: An ASTRO Consensus Paper. Pract Radiat Oncol 2019;9:65–72.

doi:10.1016/j.prro.2018.12.003.

[213] Yahya N, Ebert MA, Bulsara M, Haworth A, Kearvell R, Foo K, 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;9:282. doi:10.1186/s13014-014-0282-7.

[214] Heemsbergen WD, Incrocci L, Pos FJ, Heijmen BJM, Witte MG. Local Dose Effects for Late Gastrointestinal Toxicity After Hypofractionated and Conventionally Fractionated Modern Radiotherapy for Prostate Cancer in the HYPRO Trial. Front Oncol 2020;10. doi:10.3389/fonc.2020.00469.

[215] Dearnaley DP, Jovic G, Syndikus I, Khoo V, Cowan R a., Graham JD, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: Long-term results from the MRC RT01 randomised controlled trial. Lancet Oncol 2014;15:464–73. doi:10.1016/S1470-2045(14)70040-3.

[216] Kupelian PA, Reddy CA, Carlson TP, Willoughby TR. Dose/volume relationship of late rectal bleeding after external beam radiotherapy for localized prostate cancer: Absolute or relative rectal volume? Cancer J 2002;8:62–6. doi:10.1097/00130404-200201000-00011.

[217] Huang EH, Pollack A, Levy L, Starkschall G, Dong L, Rosen I, et al. Late rectal toxicity: Dose-volume effects of conformal radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2002;54:1314–21. doi:10.1016/S0360-3016(02)03742-2.

[218] Koper PCM, Heemsbergen WD, Hoogeman MS, Jansen PP, Hart GAM, Wijnmaalen AJ, et al. Impact of volume and location of irradiated rectum wall on rectal blood loss after radiotherapy of prostate cancer. Int J Radiat Oncol Biol Phys 2004;58:1072–82. doi:10.1016/j.ijrobp.2003.08.008.

[219] Vargas C, Martinez A, Kestin LL, Yan D, Grills I, Brabbins DS, et al. Dose-volume analysis of predictors for chronic rectal toxicity after treatment of prostate cancer with adaptive image-guided radiotherapy. Int J Radiat Oncol Biol Phys 2005;62:1297–308. doi:10.1016/j.ijrobp.2004.12.052.

[220] Kotabe K, Nakayama H, Takashi A, Takahashi A, Tajima T, Kume H. Association between rectal bleeding and the absolute dose volume of the rectum following image-guided radiotherapy for patients with prostate cancer. Oncol Lett 2018;16:2741–9. doi:10.3892/ol.2018.8888.

[221] Paleny R, Bremer M, Walacides D, Mainwaring S, Weber K, Henkenberens C. Comparison of relative and absolute rectal dose–volume parameters and clinical correlation with acute and late radiation proctitis in prostate cancer patients. Strahlentherapie Und Onkol 2019;195:103–12. doi:10.1007/s00066-018-1365-4.

[222] Seddon B, Bidmead M, Wilson J, Khoo V, Dearnaley D. Target volume definition in conformal radiotherapy for prostate cancer: Quality assurance in the MRC RT-01 trial. Radiother Oncol 2000;56:73–83. doi:10.1016/S0167-8140(00)00191-2.

[223] Onal C, Topkan E, Efe E, Yavuz M, Sonmez S, Yavuz A. 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;4:14. doi:10.1186/1748-717X-4-14.

[224] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. Med Phys 2003;30:979–85. doi:10.1118/1.1568978.

[225] Wilkins A, Naismith O, Brand D, Fernandez K, Hall E, Dearnaley D, et al. Derivation of Dose/Volume Constraints for the Anorectum from Clinician- and Patient-Reported Outcomes in the CHHiP Trial of Radiation Therapy Fractionation. Int J Radiat Oncol Biol Phys 2020;106:928–38. doi:10.1016/j.ijrobp.2020.01.003.

[226] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.

[227] Big-RT: Big Data analysis to identify combinatorial predictors of radiotherapy toxicity for personalised treatment in prostate cancer patients. - NCRI Cancer Conference abstracts n.d. https://abstracts.ncri.org.uk/abstract/big-rt-big-data-analysis-to-identify-combinatorial-predictors-of-radiotherapy-toxicity-for-personalised-treatment-in-prostate-cancer-patients/ (accessed October 12, 2020).

[228] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23:903–21. doi:10.1109/TMI.2004.828354.

[229] Rasch C, Remeijer P, Koper PCM, Meijer GJ, Stroom JC, Van Herk M, et al. Comparison of prostate cancer treatment in two institutions: A quality control study. Int J Radiat Oncol Biol Phys 1999;45:1055–62. doi:10.1016/S0360-3016(99)00280-1.

[230] Zabel WJ, Conway JL, Gladwish A, Skliarenko J, Didiodato G, Goorts-Matthews L, et al. Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring of Bladder and Rectum for Prostate Radiation Therapy. Pract Radiat Oncol 2020;11:e80–9. doi:10.1016/j.prro.2020.05.013.

[231] Palumbo I, Matrone F, Montesi G, Bellavita R, Lupattelli M, Saldi S, et al. Statins protect against acute rt-related rectal toxicity in patients with prostate cancer: An observational prospective study. Anticancer Res 2017;37:1453–7. doi:10.21873/anticanres.11469.

[232] Wang K, Chen RC, Kane BL, Medbery CA, Underhill KJ, Gray JR, et al. Patient and Dosimetric Predictors of Genitourinary and Bowel Quality of Life After Prostate SBRT: Secondary Analysis of a Multi-institutional Trial. Int J Radiat Oncol Biol Phys 2018;102:1430–7. doi:10.1016/j.ijrobp.2018.07.191.

[233] Delobel JB, Gnep K, Ospina JD, Beckendorf V, Chira C, Zhu J, et al. Nomogram to predict rectal toxicity following prostate cancer radiotherapy. PLoS One 2017;12. doi:10.1371/journal.pone.0179845.

[234] Keyes M, Miller S, Moravan V, Pickles T, McKenzie M, Pai H, et al. Predictive Factors for Acute and Late Urinary Toxicity After Permanent Prostate Brachytherapy:

Long-Term Outcome in 712 Consecutive Patients. Int J Radiat Oncol Biol Phys 2009;73:1023–32. doi:10.1016/j.ijrobp.2008.05.022.

[235] Brenner DJ, Hall EJ. Hypofractionation in prostate cancer radiotherapy. Transl Cancer Res 2018;7:S632–9. doi:10.21037/tcr.2018.01.30.

[236] Shankar PM. Tutorial overview of simple, stratified, and parametric bootstrapping. Eng Reports 2020;2. doi:10.1002/eng2.12096.

[237] Zhang Y, Yang Y. Cross-validation for selecting a model selection procedure. J Econom 2015;187:95–112. doi:10.1016/j.jeconom.2015.02.006.

[238] Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014;6:10. doi:10.1186/1758-2946-6-10.

[239] Kishan AU, Collins SP, Nickols NG. Optimal patient selection for stereotactic body radiotherapy. Lancet Oncol 2019;20:e661. doi:10.1016/S1470-2045(19)30761-2.

[240] Janowski E, Chen LN, Kim JS, Lei S, Suy S, Collins B, et al. Stereotactic body radiation therapy (SBRT) for prostate cancer in men with large prostates (≥50 cm(3)). Radiat Oncol 2014;9:241. doi:10.1186/s13014-014-0241-3.

[241] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B 1996;58:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

[242] Stata Corp. Lasso inference intro-Introduction to inferential lasso models. n.d.

[243] Pella A, Cambria R, Riboldi M, Jereczek-Fossa BA, Fodor C, Zerini D, et al. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. Med Phys 2011;38:2859–67. doi:10.1118/1.3582947.

[244] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. 34th Int. Conf. Mach. Learn. ICML 2017, vol. 7, International Machine Learning Society (IMLS); 2017, p. 5109–18.

[245] Gulliford SL, Partridge M, Sydes MR, Andreyev J, Dearnaley DP. A comparison of dose-volume constraints derived using peak and longitudinal definitions of late rectal toxicity. Radiother Oncol 2010;94:241–7. doi:10.1016/j.radonc.2009.12.019.

[246] Brown JM, Carlson DJ, Brenner DJ. The tumor radiobiology of SRS and SBRT: Are more than the 5 Rs involved? Int J Radiat Oncol Biol Phys 2014;88:254–62. doi:10.1016/j.ijrobp.2013.07.022.

[247] Kerns SL, Fachal L, Dorling L, Barnett GC, Baran A, Peterson DR, et al. Radiogenomics Consortium Genome-Wide Association Study Meta-Analysis of Late Toxicity After Prostate Cancer Radiotherapy. J Natl Cancer Inst 2020;112:179–90. doi:10.1093/jnci/djz075.

[248] Bergom C, West CM, Higginson DS, Abazeed ME, Arun B, Bentzen SM, et al. The Implications of Genetic Testing on Radiation Therapy Decisions: A Guide for Radiation Oncologists. Int J Radiat Oncol Biol Phys 2019;105:698–712. doi:10.1016/j.ijrobp.2019.07.026.

[249] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ 2015;350. doi:10.1136/bmj.g7594.

[250] Nikolov S, Blackwell S, Mendes R, de Fauw J, Meyer C, Hughes C, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. ArXiv 2018.

[251] Liang B, Tian Y, Chen X, Yan H, Yan L, Zhang T, et al. Prediction of Radiation Pneumonitis With Dose Distribution: A Convolutional Neural Network (CNN) Based Model. Front Oncol 2020;9:1500. doi:10.3389/fonc.2019.01500.

[252] Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. Phys Med Biol 2017;62:8246–63. doi:10.1088/1361-6560/aa8d09.

[253] Ibragimov B, Toesca D, Chang D, Yuan Y, Koong A, Xing L. Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT. Med Phys 2018;45:4763–74. doi:10.1002/mp.13122.

[254] Sujenthiran A, Parry M, Nossiter J, Berry B, Cathcart PJ, Clarke NW, et al. Comparison of Treatment-Related Toxicity With Hypofractionated or Conventionally Fractionated Radiation Therapy for Prostate Cancer: A National Population-Based Study. Clin Oncol 2020;32:501–8. doi:10.1016/j.clon.2020.02.004.

[255] Herskind C, Talbot CJ, Kerns SL, Veldwijk MR, Rosenstein BS, West CML. Radiogenomics: A systems biology approach to understanding genetic risk factors for radiotherapy toxicity? Cancer Lett 2016;382:95–109. doi:10.1016/j.canlet.2016.02.035.

[256] Burnet NG, Barnett GC, Elliott RM, Dearnaley DP, Pharoah PDP, Dunning AM, et al. RAPPER: The Radiogenomics of Radiation Toxicity. Clin Oncol 2013;25:431–4. doi:10.1016/j.clon.2013.04.001.

[257] National Prostate Cancer Audit. Results of the NPCA Prospective Audit in England and Wales for men diagnosed from 1 April 2017 to 31 March 2018 (published January 2020). 2019.

[258] Heng S, Low S, Sivamany K. The influence of the bowel and bladder preparation protocol for radiotherapy of prostate cancer using kilo-voltage cone beam CT: Our experience. Indian J. Cancer, vol. 52, Medknow Publications; 2015, p. 639–44. doi:10.4103/0019-509X.178386.

Appendices

Appendix 1. PACE-B Dose Constraint History

Over the course of trial recruitment, several changes were made to the normal tissue dose constraints applied during radiotherapy planning. The final constraints used are detailed in the final protocol (version 9), (**Appendix p97-98**). Below are the original dose constraints used in version 1 of the protocol and dates of changes made. From Protocol version 7, (24/03/2016), patients could be treated with 62Gy in 20 fractions. The dose constraints for these patients were proportionally scaled to those for 74 Gy in 37 fractions listed below. E.g. V74 = V62, V70 = V57 etc.

Protocol Version 3 (19/07/2012). The first clinically used protocol version.
First patient randomised 07/08/12. Total 118/847 patients randomised with these constraints

Supplementary Table 1. Original CFMHRT Dose Constraints

OAR	Dose Constraint (2 Gy per fraction)	Max Vol (% or cc)	Notes
Rectum	V30	80%	Recommended
	V40	70%	Recommended
	V50	60%	
	V60	50%	
	V65	30%	
	V70	25%	
	V74	15%	Mandatory
	V74	5%	Recommended
Bladder	V50	50%	
	V60	25%	
	V74	5%	
Femoral Heads	V50	50%	
Bowel	V50	17cc	

Supplementary Table 2. Original SBRT Dose Constraints

OAR	Dose constraint	Max Vol (% or cc)
Rectum	V18.1 V29 V36	50% 20% 1 cc
Bladder	V18.1 V37	40% 10 cc
Prostatic urethra (if visualized)	V44	20%
Neurovascular bundle (if seen)	V38	50%
Femoral head	V14.5	5%
Penile Bulb	V29.5	50%
Testicular	Blocking structure	
Bowel	V18.1 V30	5 cc 1 cc

Version 5 (05/08/2014)

Total 58/847 patients randomised with these constraints

SBRT

- Bladder. V37<5cc optimal constraint added
- Prostatic urethra. V44<20% (v3) changed to V42<50% (optional)

Version 6 (22/06/2015)

Total 67/847 patients randomised with these constraints

SBRT

- REMOVED OAR: Neurovascular bundle constraint

Version 7 (24/03/2016) onwards (including version 8 and 9)

Total 604/847 patients randomised with these constraints

CFMHRT

- Rectum. V30<80% re-termed “optimal”
- Rectum. V40<70% changed to V40<65% (optimal)
- Rectum. Added V50<50% (optimal)
- Rectum. Added V60<35% (optimal)
- Rectum. Added V70<15% (optimal)
- Rectum. Changed V75<15% mandatory to V75<5% mandatory
- Rectum. Changed V75<5% recommended to V75<3% optimal
- Bladder. Added V74<5% (optimal)
- NEW OAR: Penile bulb V50<50% (optimal)
- NEW OAR: Penile bulb V60<10% (optimal)

Appendix 2. PACE-B Acute Toxicity Statistical Analysis Plan

ST.02.G1.F1 Template for Statistical Analysis Plan v4

The PACE Trial (Prostate Advances in Comparative Evidence)

Statistical Analysis Plan

PACE B - Acute Toxicity Sub-study

Version 5.2
30.05.19

This statistical analysis plan is based on protocol version 8

Dated: 16th November 2016

ISRCTN Number: 17627211

ICR-CTSU Number:

REC reference number:

ClinicalTrials.gov reference number: NCT01584258

CRUK (other funder) number: **Sponsor reference number:**

CRUKE/12/025

The current draft version of SAP if prior to principal analysis of primary endpoint, any previous draft versions used for formal presentation, i.e. peer reviewed conference posters/presentations prior to the principal analysis, and any final version(s) of the SAP will be stored in the Statistical Section of the Trial Master File

This statistical analysis plan is a framework to guide statistical analysis and may be supplemented by additional and exploratory analyses. Trial statisticians reserve the right to amend analysis methods as appropriate after discussion with the ICR-CTSU Scientific Lead.

For final versions only:

This statistical analysis plan has been approved by the following personnel:

PACE ICR Clinical Fellow: Douglas Brand

Signed:_____ **Date: __/__/____**

Trial Statistician: Vicki Hinder/Clare Griffin

Signed: _____ **Date: __/__/____**

Signed: _____ **Date: __/__/____**

ICR-CTSU Scientific Lead: Prof Emma Hall

Signed: _____ **Date: __/__/____**

Chief Investigator: Dr Nicholas van As

Signed: _____ **Date: __/__/____**

Document history

Version	Date	Changes made (including justifications)
1.0	12/12/17	N/A
2.0	05/03/18	Power calculation added Primary and secondary endpoints added At request Prof Hall
2.1	18/10/18	Amended primary endpoint analysis to be chi-square unless assumptions not met
3.0	08/11/18	Clarified purpose. Included information regarding the IDMC recommendations. Added Cyberknife exploratory work
3.1	15/11/18	Amended alpha to one sided following discussion with Dr van As.
3.11	15/11/18	Added re exploratory presentation of data comparing the different durations of each modality.
3.12	22/11/18	Reversion to two-sided test per discussion NvA and EH
4.0	04/12/18	Following examination of data, an exceptional change in co-primary endpoints made. Instead of emergent (above baseline) RTOG G2+ GU and GI toxicity, it will be G2+ RTOG toxicity for GI and GU without reference to baseline. This is due to 10% of baseline forms being done after fiducial insertion, with more SBRT patients receiving fiducials and this introducing a strong risk of bias.
5.0	16/05/19	Major changes to include specific p-value adjusted hypothesis testing of the secondary endpoints of interest. This is to bring in line with the methods used on CHHiP QoL analysis. Specific outline of probable content of paper included in the appendix 3.

5.1 FINAL	23/05/19	Mann-whitney to compare baseline distributions for RTOG and CTCAE. Minor amendments to references. MCIDs more clearly defined. Missing data approach better defined, without bias to direction of effect.
5.2 [minor edits to final version]	30/05/19	Sensitivity analysis added for CTCAE – restricting to only the forms common to CFMHRT and SBRT. IPSS category comparison changed to chi-square, not chi-square test for trend, since the categories are dependent variable, so linear trend not anticipated.

Introduction

The PACE (Prostate Advances in Comparative Evidence) trial had two initial main comparisons (A & B), with a third (PACE C) commencing soon. PACE B compares experimental stereotactic body radiotherapy (SBRT) to conventionally fractionated or moderately hypofractionated radiotherapy (CMFHRT) radiotherapy for the treatment of localised prostate cancer. This statistical analysis plan describes the methods that will be used to analyse acute toxicity data from the trial, defined as the toxicity data accruing up to the 12 week follow-up point. This sub-study will hereafter be referred to as the PACE B Acute Toxicity Study. This document is referred to as the PACE B Acute Toxicity SAP. The SAP can be viewed at request, subject to approval by the CI and ICR-CTSU.

The purpose of this SAP is to outline clearly the planned methods for analysis and avoid data-driven approaches. Any derivations from this analysis plan will be documented in the statistical analysis report. Procedures for monitoring data accuracy and data entry quality are provided in the PACE trial main statistical monitoring plan.

1.1 Trial design

The PACE (Prostate Advances in Comparative Evidence) trial is an international, multi-centre, randomised, open label, phase III, non-inferiority clinical trial addressing the comparative effectiveness of treatments for early prostate cancer. There are two halves to the trial, with both halves comparing SBRT (36.25Gy in 5 fractions) against, respectively, surgery [PACE A] or CMFHRT [PACE B]. This analysis will focus on the patients recruited to PACE B, for whom the recruitment completed in December 2017.

The primary endpoint is biochemical/clinical progression free survival. Biochemical progression is defined, using Phoenix consensus guidelines, as PSA nadir + 2 ng/mL, confirmed on a second PSA reading. For patients within 24 months of radiotherapy, 3 successive PSA rises are required due to the PSA bounce phenomenon. Clinical progression was defined as time of commencing androgen deprivation therapy. PACE B will need to recruit 858 patients (1:1 randomisation) in order to have 80% power to exclude 6% or more absolute detriment (non-inferiority margin) to biochemical or clinical progression at 5 years. An assumption is that 85% of control patients will be biochemical/clinical progression free at 5 years.

Secondary endpoints are (*Adapted from PACE Protocol Version 8, 16th November 2016*):

1. Clinician reported acute toxicity, assessed using CTCAEv4.03, RTOG
2. Clinician reported late toxicity, assessed using CTCAEv4.03 and RTOG

3. Patient reported outcomes and quality of life assessment for all treatment patients: Assessed using International Index of Erectile Function-5 (IIEF-5), International Prostate Symptom Score (IPSS), Vaizey score, Expanded Prostate Index Composite-26 (EPIC-26).
4. Disease-specific and overall survival.
5. Progression-free survival– radiographic, clinical or biochemical evidence of local or distant failure.
6. Commencement of androgen deprivation therapy (LHRH analogues, anti-androgens, orchidectomy).

No formal interim analysis will be performed. The Independent Data Monitoring Committee (IDMC) reviews safety and efficacy data approximately 6 monthly. The standard linear accelerator (LINAC) based SBRT patients were examined after 30 patients had been treated to ensure there was not excess toxicity, with no recommendation to halt trial given by the IDMC.

This analysis will concern the evaluation of acute toxicity, as covered in the secondary endpoints 1 and 3. This will be analysed according to treatment received (not intention-to-treat), with only patients receiving at least one fraction of SBRT or CFMHRT radiotherapy included.

The control treatment in PACE B is CFMHRT with daily fractionation. In the initial protocol, this was mandated as 78 Gy in 39 fractions, however following the results of the CHHiP trial [1], a protocol amendment was made to allow 62 Gy in 20 fractions, over at least 27 days, as a control treatment option. The experimental treatment is SBRT, delivered as 36.25 Gy in 5 fractions, delivered either daily or every other day (max 14 days total treatment time). The treatment could be delivered on either Cyberknife (Accuray, USA) system or on a standard LINAC gantry.

For both arms, it was recommended to start radiotherapy within 8 weeks of randomisation and strictly not more than 12 weeks. No patients were treated with androgen deprivation therapy. For all patients undergoing radiotherapy, image guidance fiducials (3 or more) were strongly advised, bowel preparation was strongly advised, and partial bladder filling was recommended. All patients had a planning CT scan prior to radiotherapy and it was strongly advised to also perform a planning MRI scan (without endorectal coil).

1.2 Study population

All patients in the PACE B trial are eligible for this acute toxicity analysis, provided they received at least a single fraction of radiotherapy, either CFMHRT or SBRT. In the unlikely event that a patient started one form of radiotherapy (SBRT or CFMHRT) and then switched to the other, they will be excluded from this analysis. Reasons for all exclusions will be clearly documented.

The inclusion and exclusion criteria are otherwise those of the main PACE trial:

Per PACE Protocol Version 8 (16th November 2016):

1.2.1 Inclusion Criteria

1.2.1.1 Histological confirmation of prostate adenocarcinoma with a minimum of 10 biopsy cores taken within 18 months of randomisation.

1.2.1.1.1 This requirement for biopsy within 18 months of randomisation may be omitted (unless clinically indicated) if the patient has become a candidate for radical treatment (e.g. due to patient choice or PSA/MRI progression) while being followed up in an active surveillance programme. The patient's most recent biopsy must satisfy all other relevant PACE trial eligibility criteria. Patients progressing on active surveillance will be considered as having intermediate risk disease, and treated accordingly.

1.2.1.2 Gleason score ≤ 3+4

1.2.1.3 Men aged ≥18 years

1.2.1.4 Clinical and/or MRI stage T1c –T2c, N0-X, M0-X (TNM 6th Edition)

1.2.1.5 PSA ≤ 20 ng/ml

1.2.1.6 Pre-enrolment PSA must be completed within 60 days of randomisation

1.2.1.7 Patients belong in one of the following risk groups according to the National Comprehensive Cancer Network (www.nccn.org):

1.2.1.7.1 Low risk: Clinical stage T1-T2a and Gleason ≤ 6 and PSA < 10 ng/ml, or

1.2.1.7.2 Intermediate risk includes any one of the following:

- Clinical stage T2b orT2c
- PSA 10-20 ng/ml or
- Gleason 3+4

1.2.1.8 WHO performance status 0 – 2

1.2.1.9 Ability of the research subject to understand and the willingness to sign a written informed consent document

1.2.2 Exclusion Criteria

1.2.2.1 Clinical stage T3 or greater

1.2.2.2 Gleason score ≥ 4 + 3

1.2.2.3 High-risk disease defined by National Comprehensive Cancer Network (www.nccn.org)

1.2.2.4 Previous malignancy within the last 2 years (except basal cell carcinoma or squamous cell carcinoma of the skin), or if previous malignancy is expected to significantly compromise 5-year survival

1.2.2.5 Prior pelvic radiotherapy

1.2.2.6 Prior androgen deprivation therapy (including LHRH agonists and antagonists and anti-androgens)

1.2.2.7 Any prior active treatment for prostate cancer. Patients previously on active surveillance are eligible if they continue to meet all other eligibility criteria.

1.2.2.8 Life expectancy <5 years

1.2.2.9 Bilateral hip prostheses or any other implants/hardware that would introduce substantial CT artefacts

1.2.2.10 Medical conditions likely to make radiotherapy inadvisable e.g. inflammatory bowel disease, significant urinary symptoms

1.2.2.11 For patients having fiducials inserted. Anticoagulation with warfarin/ bleeding tendency making fiducial placement or surgery unsafe in the opinion of the clinician (see section 11, Treatment).

1.2.2.12 Participation in another concurrent treatment protocol for prostate cancer

2.0 Study objectives

The objective of this PACE B Acute toxicity study will be to compare acute toxicity data (up to week 12 post treatment) by treatment received (per protocol) for PACE B trial participants.

The primary endpoints for this study will be worst grade RTOG G2+ bowel toxicity and worst grade RTOG G2+ bladder toxicity during this period.

Secondary objectives will be to assess differences between SBRT and CFMHRT :

- Baseline, worst, worst exceeding baseline, 12 week scores of all of the following [differences from this in square brackets]:
- RTOG GI G2+
- RTOG GU G2+
- RTOG GI G3+
- RTOG GU G3+
- CTCAE GI G2+
- CTCAE GU G2+
- CTCAE GI G3+
- CTCAE GU G3+
- IPSS total
- IPSS QoL
- IPSS severity distributions [at timepoints collected]
- EPIC-26 Urinary Incontinence Subdomain Score [plus minimal clinically important difference (8 point reduction at any time relative to baseline)]

- EPIC-26 Urinary Obstructive Subdomain Score [plus minimal clinically important difference (6 point reduction at any time relative to baseline)]
- EPIC-26 Urinary Bother
- EPIC-26 Bowel Subdomain Score [plus minimal clinically important difference (5 point reduction at any time relative to baseline)]
- EPIC-26 Bowel Bother
- EPIC-26 Sexual Subdomain Score [plus minimal clinically important difference (11 point reduction at any time relative to baseline)]
- IIEF
- Vaizey

n.b. Minimal clinically important differences for EPIC-26 defined per Skolarus et al [204]

Exploratory analysis

- Compare toxicities between Cyberknife and non-Cyberknife treated patients for patients receiving SBRT [per IDMC request].
- Examine effect of other rational possible predictors on this toxicity, including:
  - International effect (UK&Ireland vs Canada)
  - Learning effect (by centre volume of PACE patients)
  - Margin effects
    - Absolute margins used
    - Deviations from protocol margins
  - Dosimetry information. (Reported dose-volume constraints)
    - Potentially incorporate as binary predictor (planned to protocol, not planned to protocol)
  - Fiducials vs non-fiducials
  - Other plausible predictors (If any)
  - Alpha-blockers/cholinergics at diagnosis
- Dosimetric analyses of individual patient DICOMs in the study, with reference to resultant toxicities, including analysis of differences in contouring,

3.0 Randomisation/recruitment procedures

Central randomisation is performed at ICR-CTSU, via telephone for UK sites and fax for non-UK sites. There is a 1:1 random allocation, by computer generated random permutated block, between CFMHRT and SBRT.

The randomisation will be stratified by:

- Treatment centre
- Risk group (Low, Intermediate)

Treatment allocation is open label.

4.0 Endpoints

Co-Primary Endpoints

- Worst grade RTOG G2+ for bowel toxicity (GI)
- Worst grade RTOG G2+ for bladder toxicity (GU)
    - For each of these, a patient fulfils the endpoint if they have:
        1. GU or GI toxicity at grade 2 or more in follow-up (during and up to 12 weeks post RT treatment)

Secondary Endpoints

All secondary endpoints will be calculated from commencement of RT to 12 weeks post RT

- Graphically report data at baseline and up to 12 weeks follow-up for (details below in graphical section):
    - RTOG -GI & GU
    - CTCAE -GI & GU
    - IPSS – each subdomain and total score. Plus categories by time
    - EPIC-26 – by domain. EPIC bowel and urine bother scores
    - IIEF
    - Vaizey
- Comparison between CFMHRT and SBRT for secondary objectives in **2.0 Study Objectives**

5.0 Sample size and power

The analysis of acute toxicity was not considered as part of the sample size calculations for the main PACE trial. The best comparator for expected acute toxicity in the control arm is the PROFIT trial [3]. This compared 78Gy in 39 fractions to 60 Gy in 20 fractions, both delivered as daily treatments. Similar to PACE B, androgen deprivation therapy was not permitted. The CHHiP trial (mentioned earlier) is also a reasonable comparator, although it permitted higher risk patients and androgen deprivation was administered. The control arm of HYPRO, another trial of the Hypofractionation era, is also of interest, treating patients with 78 Gy / 39# [206]. The table below summarises the relevant data available for cumulative acute toxicity.

Acute Toxicity	PROFIT (without ADT)		CHHiP (with ADT)			HYPRO (+/- ADT)
RTOG G2+	78 Gy / 39 #	60 Gy / 20#	74 Gy / 37 #	60 Gy / 20#	57 Gy / 19#	78 Gy / 39 #
Bowel	10.5%	16.7%	25%	38%	38%	31.2 %
Bladder	31%	30.9%	46%	49%	46%	57.8%

Based on data from ICR-CTSU, the proportion of patients in PACE B receiving the two different acceptable conventional treatment arms are:

- 62 Gy / 20# = 90-95%
- 78 Gy / 39# = 5-10%

Therefore, accounting for the above and the disparate nature of available data, the estimates of acute RTOG G2+ cumulative toxicity in the conventional arm of PACE B will be:

- Bowel = 25%
- Bladder = 40%

These are weighted closer to findings in the PROFIT trial, which had more similar patient risk groups (intermediate) and did not allow ADT.

Power calculations were performed using Stata 15: (Table below)
power twoproportions [control proportion] [test proportion] , n(871) alpha(0.025)

- Note alpha set at 2.5% (two-way) to split a total alpha 0.05 across the two endpoints of bowel and bladder.

Power Calculation	Conventional	SBRT		
G2+ RTOG	Proportion	Proportion 1	Proportion 2	Proportion 3
Bowel	**0.25**	0.30	0.34	**0.35**
POWER		27%	75%	**83%**
Bladder	**0.40**	0.45	0.50	**0.51**
POWER		22%	76%	**84.5%**

Final selections for acute G2+ cumulative RTOG acute toxicity are made to preserve an 80% power (or more), with two way alpha of 0.025 for each endpoint:

- Bowel: 83% power to detect 10% increase in toxicity from 25% control arm

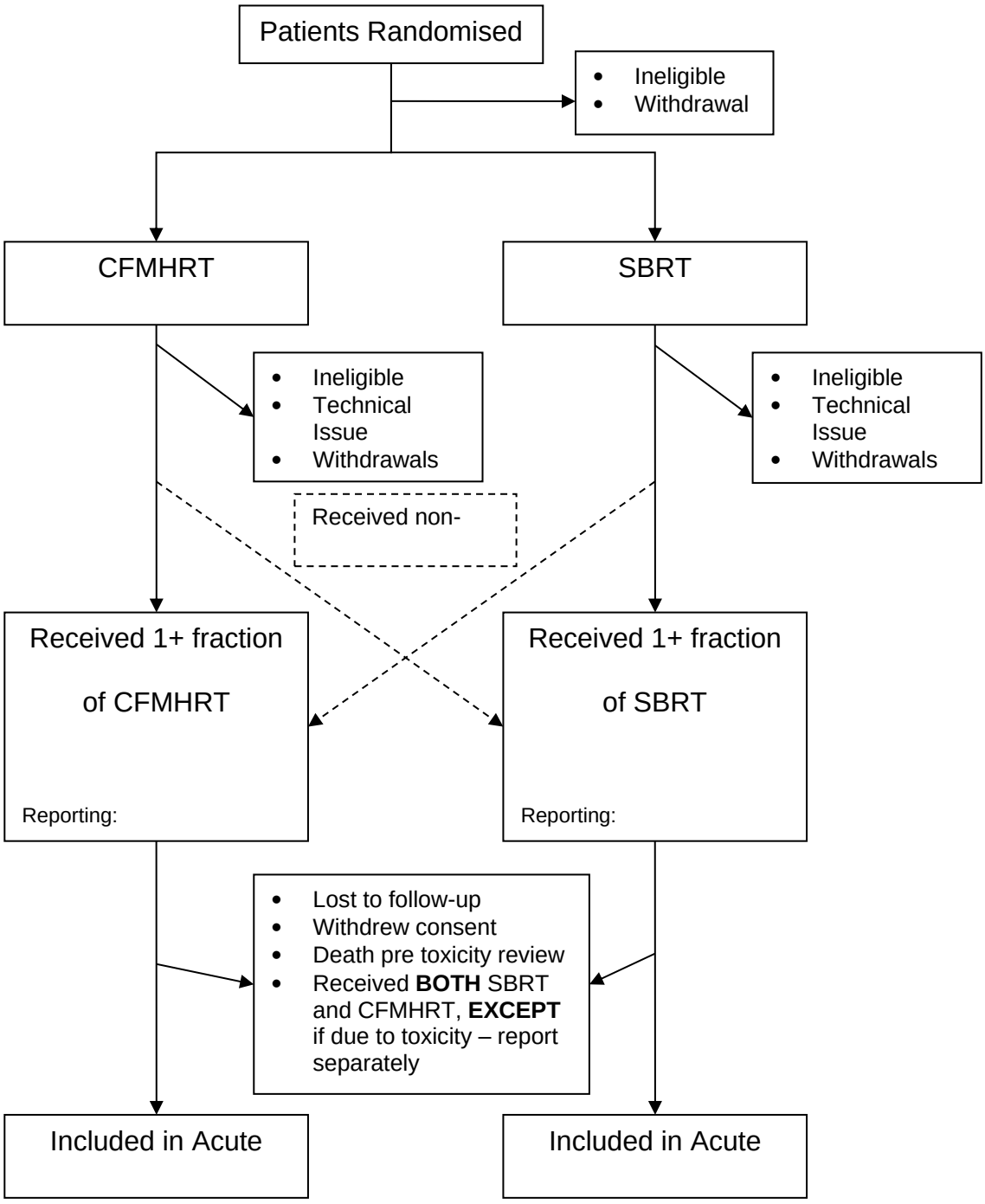
- Bladder: 84.5% power to detect 11% increase in toxicity from 40% control arm.

6.0 Data completeness and consistency

Data completeness and quality will be summarised descriptively, discussing measures taken to clean the data, monitoring checks performed on the data and the findings of these. The Central Statistical Monitoring Plan outlines the data consistency and accuracy checks routinely carried our prior to analyses.

6.1 Patient flow through trial

For the acute toxicity study analysis, patient flow through the trial will be presented using a CONSORT diagram as below:



6.2 Compliance with assessments

Compliance with the radiotherapy treatment protocols and acute toxicity follow-up assessments will be reported as part of the acute toxicity study. Treatment protocol compliance will be visually identifiable in the CONSORT diagram (see section 6.1). Where there appear to be differences in compliance between arms, sensitivity analysis will be performed, with reference to date of CRF returns, should the quality of submitted data enable this. Returns rates for all forms will be reported, by per protocol assignment.

6.3 Consistency of reporting

In the event of non-returned CRFs, ICR-CTSU will contact participating centres in order to maximise return rate. In the unusual event (given electronic reporting) of two forms being returned for a single time point, then for each question, the worse outcome score of the two forms will be used in analysis. Forms that have not been dated will be assumed to have been completed on the closest previous clinical assessment date. The handling of any other inconsistencies in data, that data cleaning and querying have not resolved will be recorded in the statistical analysis report.

6.4 Missing data

Any missing data will be requested by the PACE trial manager/data manager. Levels of missing data will be summarised by form type. Should any patterns emerge then these will be examined. It is expected that the majority of missing data will be missing at random and therefore, it is not expected that imputation methods will be used for individual patients or variables.

Should there be a substantial (e.g. 10%) difference in missing data seen between the two arms then a sensitivity analysis will be performed, restricting to those patients with 1 or fewer missing data.

EPIC-26 is the only PRO questionnaire with specific minimum domain completion fraction for the domain sub-score to be valid:

- Urinary Obstructive = 4/4, 100% of items
- Urinary Incontinence = 4/4, 100% of items
- Bowel = 5/6, 83% of items
- Sexual = 5/6, 83% of items
- Hormonal = 4/5, 80% of items

7.0 Analysis methods

7.1 Analysis populations

This PACE acute toxicity analysis will include all patients who received at least a single fraction of radiotherapy. As this is a toxicity assessment, analysis will be performed by treatment received, rather than by intention to treat. Patients receiving both conventional and SBRT fractions of radiotherapy will be excluded from this analysis, by identification from deviation forms and radiotherapy delivery forms.

Patients will then be included in the analysis of each endpoint/timepoint if they have completed the relevant follow-up assessments/questionnaires. In the case of the analysis of EPIC-26, domain scores will be void unless the completion fraction meets the cut-points as stated in section 6.4.

7.2 Baseline characteristics

Baseline Data Collected:

- Age (40-49, 50-59, 60-69, 70-79, 80+)
- Family history of prostate cancer (yes, no)
- WHO Performance status (0, 1, 2, 3+)
- PSA [ng/mL] – (0-5, 5-10, 10-15, 15-20)
- Testosterone [μmol/L] – (<1.5, ≥1.5)
- DRE (T1c, T2a, T2b, T2c) {**non-mandatory**}
- TRUS (T1c, T2a, T2b, T2c) {**non-mandatory**}
- MRI T Stage (T1c, T2a, T2b, T2c)
- MRI N Stage (N0, N1, NX)
- MRI M Stage (M0, M1, MX)
- Prostate Volume [cc]
- Gleason Score (≤3+2, 3+3, 3+4, ≥4+3)
- Percentage Positive Cores [Positive Cores / Total Cores]
- Total length of cores [mm] {**non-mandatory**}
- Total linear extent positive [mm] {**non-mandatory**}
- Concomitant medications:
    - Alpha blockers (yes, no)
    - Aspirin (yes, no)
    - Statins (yes, no)
    - Anti-cholinergics for bladder symptoms (yes, no)

Square brackets [] show units of measurement

Round brackets () show categories for categorical data

All baseline characteristics will be tabulated by treatment group. It is not planned to conduct formal tests between the groups as, by virtue of the randomisation, the demographics and characteristics should be well balanced between the four groups. However, if, by chance, there appears to be a large difference between the treatment groups with regard to a particular baseline variable a formal test (e.g. chi squared test) will be performed.

For nominal/ordinal data (round brackets show categories), percentages will be reported by group. Median average age will also be presented for age, alongside distribution by grouping. For T-stage, the highest value of the DRE, TRUS and MRI will be reported. Median average will be reported for continuous data. Interquartile range and range will also be presented.

7.3 Analyses of defined endpoints

Primary Analysis (Alpha 0.05 two-sided, split as 0.025 per comparison)

- Chi-square test will be performed (Unless assumptions failure thus requires Fisher's exact test) for:
    - Worst bowel RTOG G2+ acute toxicity
    - Worst bladder RTOG G2+ acute toxicity
- Comparing conventional to SBRT. Alpha 0.025 for each comparison.
- Proportions and corresponding confidence intervals will also be presented

Secondary Analyses (Significant p-value = 0.001)

1. RTOG GI baseline (mann-whitney)
2. RTOG GI G2+ at 12 weeks (chi-square – compare proportions)
3. RTOG GI G2+ worst exceeding baseline (chi-square – compare proportions)
4. RTOG GU baseline (mann-whitney)
5. RTOG GU G2+ at 12 weeks (chi-square – compare proportions)
6. RTOG GU G2+ worst exceeding baseline (chi-square – compare proportions)
7. RTOG GI G3+ worst (chi-square – compare proportions)
8. RTOG GI G3+ worst exceeding baseline (chi-square – compare proportions)
9. RTOG GI G3+ at 12 weeks (Chi-square – compare proportions)
10. RTOG GU G3+ worst (chi-square – compare proportions)
11. RTOG GU G3+ worst exceeding baseline (chi-square – compare proportions)
12. RTOG GU G3+ at 12 weeks (Chi-square – compare proportions)

13. CTCAE GU baseline (mann-whitney)
14. CTCAE GU G2+ worst (chi-square – compare proportions)
15. CTCAE GU G2+ worst exceeding baseline (chi-square – compare proportions)
16. CTCAE GU G2+ at 12 weeks (Chi-square – compare proportions)
17. CTCAE GU G3+ worst (chi-square – compare proportions)
18. CTCAE GU G3+ worst exceeding baseline (chi-square – compare proportions)
19. CTCAE GU G3+ at 12 weeks (Chi-square – compare proportions)
20. CTCAE GI baseline (mann-whitney)
21. CTCAE GI G2+ worst (chi-square – compare proportions)
22. CTCAE GI G2+ worst exceeding baseline (chi-square – compare proportions)
23. CTCAE GI G2+ at 12 weeks (Chi-square – compare proportions)
24. CTCAE GI G3+ worst (chi-square – compare proportions)
25. CTCAE GI G3+ worst exceeding baseline (chi-square – compare proportions)
26. CTCAE GI G3+ at 12 weeks (Chi-square – compare proportions)
27. IPSS total baseline (Mann-whitney – compare scores)
28. IPSS total worst (Mann-whitney – compare worst total scores)
29. IPSS total worst change (Mann-whitney – compare worst total scores worsening)
30. IPSS total 12 week (Mann-whitney – compare scores)
31. IPSS QoL baseline (Mann-whitney – compare scores)
32. IPSS QoL worst (Mann-whitney – compare worst QoL scores)
33. IPSS QoL worst change (Mann-whitney – compare worst QoL drop)
34. IPSS QoL 12 week (Mann-whitney – compare scores)
35. IPSS Categories: Baseline (Chi-square test)
36. IPSS Categories: Worst (Chi-square test)
37. IPSS Categories: 12-weeks (Chi-square test)
38. EPIC-26 Urinary Incontinence baseline (Mann-whitney – compare scores)
39. EPIC-26 Urinary Incontinence worst (Mann-whitney – compare worst QoL scores)
40. EPIC-26 Urinary Incontinence worst change (Mann-whitney – worst drop cf baseline)
41. EPIC-26 Urinary Incontinence 12 week (Mann-whitney – compare scores)
42. EPIC-26 Urinary Incontinence MCID drop Any time – Chi-square
43. EPIC-26 Urinary Incontinence MCID drop 12 week – Chi-square
44. EPIC-26 Urinary Obstructive baseline (Mann-whitney – compare scores)
45. EPIC-26 Urinary Obstructive worst (Mann-whitney – compare worst QoL scores)
46. EPIC-26 Urinary Obstructive worst change (Mann-whitney – worst drop cf baseline)
47. EPIC-26 Urinary Obstructive 12 week (Mann-whitney – compare scores)
48. EPIC-26 Urinary Obstructive MCID drop Any time – Chi-square

49. EPIC-26 Urinary Obstructive MCID drop 12 week – Chi-square
50. EPIC-26 Urinary Bother Baseline – Mann whitney
51. EPIC-26 Urinary Bother Worst – Mann whitney
52. EPIC-26 Urinary Bother Worst change – Mann whitney (cf baseline)
53. EPIC-26 Urinary Bother 12 weeks– Mann whitney
54. EPIC-26 Bowel baseline (Mann-whitney – compare scores)
55. EPIC-26 Bowel worst (Mann-whitney – compare scores)
56. EPIC-26 Bowel worst change cf baseline (Mann-whitney – compare scores)
57. EPIC-26 Bowel 12 week (Mann-whitney – compare scores)
58. EPIC-26 Bowel MCID drop Any time – Chi-square
59. EPIC-26 Bowel MCID drop 12 week – Chi-square
60. EPIC-26 Bowel Bother Baseline – Mann whitney
61. EPIC-26 Bowel Bother Worst – Mann whitney
62. EPIC-26 Bowel Bother Worst change – Mann whitney (cf baseline)
63. EPIC-26 Bowel Bother 12 weeks– Mann whitney
64. EPIC-26 Sexual baseline (Mann-whitney – compare scores)
65. EPIC-26 Sexual worst (Mann-whitney – compare scores)
66. EPIC-26 Sexual worst cf baseline (Mann-whitney – compare scores)
67. EPIC-26 Sexual 12 week (Mann-whitney – compare scores)
68. EPIC-26 Sexual MCID drop Any time – Chi-square
69. EPIC-26 Sexual MCID drop 12 week – Chi-square
70. EPIC-26 Hormonal baseline (Mann-whitney – compare scores)
71. EPIC-26 Hormonal worst (Mann-whitney – compare scores)
72. EPIC-26 Hormonal worst cf baseline (Mann-whitney – compare scores)
73. EPIC-26 Hormonal 12 week (Mann-whitney – compare scores)
74. EPIC-26 Hormonal MCID drop Any time – Chi-square
75. EPIC-26 Hormonal MCID drop 12 week – Chi-square
76. IIEF baseline (Mann-whitney – compare scores)
77. IIEF 12 week (Mann-whitney – compare scores)
78. Vaizey Total baseline (Mann-whitney – compare scores)
79. Vaizey Total worst (Mann-whitney – compare scores)
80. Vaizey Total worst cf baseline (Mann-whitney – compare scores)
81. Vaizey 12 week score (Mann-whitney – compare scores)

N.b. Fisher's exact will be used if assumptions of Chi-Square not met.

7.4 Treatment compliance

Treatment compliance will be reported in the CONSORT diagram, indicating how many patients went on to receive at least one fraction of their assigned radiotherapy treatment arm. Additionally it will indicate the number of patients failing to complete the intended course and for what reasons.

7.5 Exploratory analyses

Sensitivity analysis of CTCAE endpoints will be performed, restricting analysis to only the forms common to both CFMHRT and SBRT (i.e. weeks 2,4,8,12 of follow-up). This is to explore whether the additional end-of-treatment assessment for SBRT might upwardly bias the toxicity.

Compare toxicities between Cyberknife and non-Cyberknife treated patients for patients receiving SBRT [per IDMC request]:

1. The margins set and achieved for different techniques – It was felt that SBRT - conventional LINAC was the hardest technique to achieve the margins required.

2. Centre effect – some sites may have more experience with techniques, may be reporting CTCAE and RTOG differently.

3. There may be a learning curve, therefore changes in toxicity patterns over time may be seen.

Differences will be examined by:

- Chi-square (or Fishers) for worst GU and GI RTOG G2+ toxicity
- Examination of effect in other data collected:
    - CTCAE
    - EPIC-26
    - IPSS

Logistic Regression Model

- Treatment platform
- 1 week vs 2 week administration
- International effect (UK&Ireland vs Canada)
- Learning effect (by centre volume of PACE patients)
- Margin effects
    - Absolute margins used
    - Deviations from protocol margins
- Dosimetry information. (Reported dose-volume constraints)
    - Potentially incorporate as binary predictor (planned to protocol, not planned to protocol)
- Fiducials vs non-fiducials

  - Alpha-blockers/cholinergics at diagnosis
  - Other biologically plausible predictors (If any)

Additional presented data for hypothesis generation

- Graph of RTOG against time, separated by:
  - 1 week SBRT
  - 2 week SBRT
  - 4 week CFMHRT
  - 8 week CFMHRT
- Tabular presentation of worst RTOG, CTCAE, Worst EPIC GI & GU subdomains, Worst IPSS
  - Separated by:
  - 1 week SBRT
  - 2 week SBRT
  - 4 week CFMHRT
  - 8 week CFMHRT

Dosimetry data will be obtained for as many PACE B patients as possible. Ideally for each patient the following data items will be collected.:

- Planning CT DICOM files
- Planning MRI DICOM files (if performed)
  - DICOM Registration Files where MRI employed
- Dose Cube data (DICOM Dose)
- DICOM Planning File
- DICOM Structure Sets

Acceptability of structure contouring for normal organs will be checked, with re-contouring if needed. The data can then be used to generate finalised Dose Volume Histogram (DVH) files. These DVH files can be used to make predictions about the expected toxicity of PACE, based upon normal tissue alpha/beta ratios and time factor estimates derived from the CHHiP data. This will make use of an EQD2 corrected LKB model, potentially incorporating time factor for recovery. This will be done for at least bladder and rectal endpoints. Other exploratory analyses of the methods of contouring and dosimetry may undertaken as appropriate.

8.0 General considerations

PACE-B complete recruitment in late December 2017. All patients have therefore reached the required duration of follow-up to collect 12 week post-RT forms.

8.1 Subgroup analyses

See section 7.5 exploratory analyses

9.0 Independent Data Monitoring Committee (IDMC) and interim analyses

From PACE protocol version 8:
"It is planned that an Independent Data Monitoring Committee (IDMC) will meet at approximately 6 monthly intervals to review the accumulating safety and emerging efficacy data.

Once 30 patients have been treated with SBRT on a conventional linac (ie non-Cyberknife systems), the toxicity, acute and late, will be reviewed by the IDMC to ensure there is not an augmented rate of side effects in this cohort. After this, conventional linac SBRT vs Cyberknife SBRT toxicity and outcomes will continue to be monitored by the IDMC separately and together to ensure ongoing safety of this technique."

The trial has not been halted following the 30 patient review and no interim issues with safety (toxicity) have been raised to date.

IDMC approval of release of the PACE-B acute toxicity data ahead of primary analysis will be required. The IDMC will asked to consider potential impact of knowledge of acute PACE-B toxicity data by investigators on the continued integrity of PACE-A (open to recruitment) and follow-up/reporting of late toxicity in PACE B.

Per the IDMC Recommendations, we are examining:

1. *The margins set and achieved for different techniques – It was felt that SBRT - conventional LINAC was the hardest technique to achieve the margins required.*
2. *Centre effect – some sites may have more experience with techniques, may be reporting CTCAE and RTOG differently.*
3. *There may be a learning curve, therefore changes in toxicity patterns over time may be seen.*

10.0 Analysis Programs

Analysis will be conducted using Stata version 15 (or prior versions). Exploratory analyses may require additional programs, such as R or MATLAB.

11.0 Analysis program locations

All programs will be stored in the analyses folder for PACE on the ICR-CTSU server. Only the PACE trial statistician(s), PACE ICR Clinical Fellow, ICR-CTSU IT staff, Director and Deputy Directors of ICR-CTSU will be able to see the analysis folder. Programmes will be stored under the type of analysis e.g. Acute Toxicity Analysis. All official analysis reports that are to be circulated externally of ICR-CTSU will be password protected. Hard copies of reports will be stored securely in the statistical section of the PACE trial master file held in a locked fire proof cupboard with restricted access.

Appendix 1
Summary of the Routinely Collected Clinician Reported Outcome Measures

Data Collection Timepoints

Pre-treatment
CTCAE **and**
RTOG Bladder and Bowel

During Treatment
RTOG Bladder and Bowel **only** at:
- Week 2,4,6,8 for CFMHRT radiotherapy arm
- **OR**
- Last fraction only for SBRT arm

Post-Treatment
CTCAE **and**
RTOG Bladder and Bowel **at**
- Week 2,4,8,12 post treatment
- Month 3,6,9
- Annually thereafter to 10 years follow-up

Clinician Reported Outcome Metrics

1) **Urinary**

   a. CTCAE Genito-Urinary Scores
      i. Haematuria (0-5)
      ii. Pain/Dysuria (0-3)
      iii. Frequency (0-2)
      iv. Incontinence (0-3)
      v. Urinary Retention (0-5)
      vi. Urgency (0-2)
   b. RTOG Genito-urinary
      i. Cystitis
      ii. Haematuria
      iii. Urethral Stricture

2) **Bowel**

   a. CTCAE Gastrointestinal Scores
      i. Colitis (0-5)
      ii. Constipation (0-5)
      iii. Diarrhoea (0-5)
      iv. GI Fistula (0-5)
      v. Nausea (0-3)
      vi. Proctitis (0-5)
      vii. GI Haemorrhage Anus/rectum (0-5)
      viii. Rectal Pain (0-2)

b. RTOG Gastrointestinal
   i. Diarrhoea
   ii. Proctitis
   iii. Rectal-anal Stricture
   iv. Rectal Ulcer
   v. Bowel Obstruction

3) Sexual

a. CTCAE
   i. Erectile dysfunction (0-3)

4) Hormonal/General

a. CTCAE
   i. Hot flushes (0-3)
   ii. Pain (0-3) (Specified)
   iii. Fatigue (0-3)
   iv. Anorexia (0-5)
   v. Weight loss (0-3)
   vi. Radiation Dermatitis (0-3)

Appendix 2
Summary of the Routinely Collected Patient Reported Outcome Measures

Data Collection Timepoints

All proformas have been collected at:
- Pre-randomisation
- Week 2,4,8,12 post treatment (acute)
- Month 6, 9 ,12 post treatment
- Annually thereafter to year 5

Patient Reported Outcome Metrics

1) **Urinary**

a. EPIC-26 Urinary Incontinence Domain – *Need all 4 or score as missing data*
   i. EPIC 26.23 How often leaked urine (1-5) [0,25,50,75,100]
   ii. EPIC 26.26 Urine control (1-4) [0,33,67,100]
   iii. EPIC 26.27 Pads per day (0-3) [100,67,33,0]
   iv. EPIC 26.28 Dripping/leaking urine (0-4) [100,75,50,25,0]

b. EPIC-26 Urinary Obstructive Domain – *Need all 4 or score as missing data*
   i. EPIC 26.29 Pain/Burning or urination (0-4) [100,75,50,25,0]
   ii. EPIC 26.30 Bleeding on urination (0-4) [100,75,50,25,0]
   iii. EPIC 26.31 Weak Stream/Incomplete emptying (0-4) [100,75,50,25,0]
   iv. EPIC 26.33 Frequent daytime urination (0-4) [100,75,50,25,0]

c. EPIC-26 Urinary Non-subscale (Over last 4 weeks)
   i. EPIC 26.34 Overall urine function problem (1-5) [100,75,50,25,0]

d. I-PSS Composite Score (0-35)
   i. Incomplete emptying (0-5)
   ii. Frequency <2 hours (0-5)
   iii. Intermittency (0-5)
   iv. Urgency (0-5)
   v. Weak stream (0-5)
   vi. Straining (0-5)
   vii. Nocturia (0-5)

e. IPSS Quality of Life Score (0-6)

2) **Bowel**

a. EPIC-26 Bowel Domain – *Need 5 or score as missing data*

i. EPIC 26.49 Urgency of bowel movement (0-4) [100,75,50,25,0]
ii. EPIC 26.50 Inc. frequency of bowel movements (0-4) [100,75,50,25,0]
iii. EPIC 26.52 Loss of bowel control (0-4) [100,75,50,25,0]
iv. EPIC 26.53 Bloody stools (0-4) [100,75,50,25,0]
v. EPIC 26.54 Abdominal/pelvic/rectal pain (0-4) [100,75,50,25,0]
vi. EPIC 26.55 How big a problem is bowel habits (1-5) [100,75,50,25,0]

b. Vaizey Score Composite (0-24)
i. Incontinence solid stool (0-4)
ii. Incontinence of liquid stool (0-4)
iii. Incontinence of gas (0-4)
iv. Alteration in Lifestyle (0-4)
v. Wear pad or plug (0=no, 2=yes)
vi. Constipating medications (0=no, 2=yes)
vii. Inability to defer defaecation 15 mins (0=no, 4=yes)

3) Sexual

a. EPIC-26 Sexual Domain – *Need 5 or score as missing data*
i. EPIC 26.57 Ability to have an erection (1-5) [0,25,50,75,100]
ii. EPIC 26.58 Ability to orgasm (1-5) [0,25,50,75,100]
iii. EPIC 26.59 Quality of erections (1-4) [0,33,67,100]
iv. EPIC 26.60 Frequency of erections (1-5) [0,25,50,75,100]
v. EPIC 26.64 Ability to function sexually (1-5) [0,25,50,75,100]
vi. EPIC 26.68 How big a problem is sexual function (1-5) [100,75,50,25,0]

b. IIEF-5 Composite Score (5-25)
i. Confidence of erection (1-5)
ii. How often erections suitable for penetrative sex (1-5)
iii. How often maintain erection after penetration (1-5)
iv. How difficult to maintain erection to completion (1-5)
v. How frequently sex satisfactory (1-5)

4) General/Hormonal

a. EPIC-26 Hormonal Domain – *Need 4 or score as missing data*
i. EPIC 26.74 Hot flushes (0-4) [100,75,50,25,0]
ii. EPIC 26.75 Breast tenderness/enlargement (0-4) [100,75,50,25,0]
iii. EPIC 26.77 Depression (0-4) [100,75,50,25,0]
iv. EPIC 26.78 Lack of Energy (0-4) [100,75,50,25,0]
v. EPIC 26.79 Change in body weight (0-4) [100,75,50,25,0]

Appendix 3
Plan of Results to Report in Paper

Results Section Plan – By Paragraph

1. Trial Details
    a. Men, centres, dates
    b. CONSORT – explanation of exceptional SBRT + CFMHRT patient to per protocol
        i. **Figure 1** CONSORT Flow Diagram

2. Baseline/Disease/Treatment Characteristics
    a. **Table 1**. By Per Protocol.
        i. Age, Ethnic origin, FHx Prostate ca, WHO PS, NCCN risk, T-score, Gleason, Pre-treatment PSA, Pre-treatment testosterone, Prostate volume, con meds
    b. **Supplementary Table 1**. By Per Protocol
        i. Fid marks, fid mark numbers, RT method, IGRT method, overall treatment times, margins
    c. Brief text explanation of Supp table 1 key points (quicker, more fiducials, smaller margins, more non co-planar in SBRT)

3. Return Rates
    a. Brief comment on return rates for each instrument (perhaps a single number average for each).
    b. Supplementary Table 2. RTOG Return Rates
    c. Supplementary Table 3. CTCAE Return Rates
    d. Supplementary Table 4. EPIC Return Rates
    e. Supplementary Table 5. IPSS Return Rates
    f. Supplementary Table 6. IIEF-5 Return Rates
    g. Supplementary Table 7. Vaizey Return Rates

4. Primary Endpoint Analysis + Other RTOG
    a. **Figure 2, Panel A** RTOG GI toxicity G1+,2+,3+ (point prevalence, all patients) vs time
    b. **Figure 2, Panel B** RTOG GU toxicity G1+,2+,3+ (point prevalence, all patients) vs time
    c. **Supplementary Figure 1.** RTOG GI and GU toxicity separated as 1 week SBRT vs 2 week SBRT vs 4 week MHRT vs 7.8 week CFRT.
    d. Text: Worst acute G2+ proportions GI compared by chi-square (interpreted to sig p-value 0.025)
    e. Text: Worst acute G2+ proportions GU compared by chi-square (interpreted to sig p-value 0.025)
    f. **Supplementary Table 8**. Split by per protocol analysis:
        i. Worst RTOG toxicities (not referencing baseline)
            1. G2+ G3+ GI compared by chi-square (n.b G2+ is also the primary comparison)
            2. G2+ G3+ GU compared by chi-square (n.b G2+ is also the primary comparison)
    g. **Supplementary Table 9.** Split by per protocol analysis:
        i. Worst RTOG toxicities (above baseline) – exclude those without baseline
            1. G2+ G3+ GI compared by chi-square

        2. G2+ G3+ GU compared by chi-square
    ii. Baseline RTOG
        1. GI – compare by mann-whitney
        2. GU – compare by mann-whitney
    iii. 12 week RTOG
        1. GI – compare G2+ and G3+ by chi square
        2. GU – compare G2+ and G3+ by chi square

5. CTCAE
    a. **Figure 3, Panel A** CTCAE GI toxicity G1+,2+,3+ (point prevalence, all patients) vs time
    b. **Figure 3, Panel B** CTCAE GU toxicity G1+,2+,3+ (point prevalence, all patients) vs time
    c. In text: state p-values of worst G2+ and G3+ GI and GU CTCAE comparisons (Chi-square)
    d. **Supplementary Table 10** - Worst Acute CTCAE GI Toxicity Items
        i. **Composite**
        ii. Individual items: Anal Pain, Colitis, Constipation, Diarrhoea, Diverticulitis, Fecal incontinence, Fistula, GI Pain, Haemorrhoids, GI haemorrhage , Proctitis, GI Unspecified, Rectal Prolapse
    e. **Supplementary Table 11** - Worst Acute CTCAE GU Toxicity Items
        i. **Composite**
        ii. Individual items: Bladder Spasm, Cystitis, Haematuria, Prostatic Obstruction, Urinary Frequency, Urinary incontinence, Urinary retention, Urinary urgency
    f. **Supplementary Table 11** - Worst Acute CTCAE GI Toxicity Items ABOVE BASELINE
        i. **Composite**
        ii. Individual items: Anal Pain, Colitis, Constipation, Diarrhoea, Diverticulitis, Fecal incontinence, Fistula, GI Pain, Haemorrhoids, GI haemorrhage, Proctitis, GI Unspecified, Rectal Prolapse
    g. **Supplementary Table 12** - Worst Acute CTCAE GU Toxicity Items ABOVE BASELINE
        i. **Composite**
        ii. Individual items: Bladder Spasm, Cystitis, Haematuria, Prostatic Obstruction, Urinary Frequency, Urinary incontinence, Urinary retention, Urinary urgency
    h. Supplementary Table 13 – CTCAE GI Baseline
        i. Composite – compare distributions by mann-whitney
        ii. Individual (as above)
    i. Supplementary Table 14 – CTCAE GU Baseline
        i. Composite – compare distributions by mann-whitney
        ii. Individual (as above)
    j. Supplementary Table 15 – CTCAE GI 12 week
        i. Composite - compare distributions by chi-square
        ii. Individual (as above)
    k. Supplementary Table 16 – CTCAE GU 12 week
        i. Composite - compare distributions by chi-square
        ii. Individual (as above)

6. EPIC-26 all By 5 subdomains (UI, UO, Bowel, Sexual, Hormonal) *n.b. urine overall bother is separate – outside the subdomain scores, but bowel overall bother is within bowel subdomain*
    a. **Figure 4** — mean change in score from baseline with confidence intervals + separate urine QoL (not included in subdomain)
    b. **Supplementary Figure 2** – mean scores with confidence intervals

c. **Supplementary Table 17** 5x +urine bother Average, CI for baseline, worst, 12-week scores and comparison by Mann-Whitney
d. **Supplementary Table 18** 5x + urine bother Average, CI for worst change from baseline and comparison by Mann-Whitney
e. **Supplementary table 19** – Proportions experiencing clinically important differences at any time.
f. **Supplementary table 20** – Proportions experiencing clinically important differences at 12 weeks
g. **Supplementary table 21** – Table showing urinary and bowel bother question distributions at 0,4,12 weeks. – I.e. comparative to CHHiP

7. IPSS
    a. **Figure 5** – by question + total + QoL (10 plots) – change from baseline with confidence intervals
    b. **Supplementary Figure 3** - by question + total + QoL (10 plots) – average scores at each timepoint with confidence intervals
    c. **Supplementary Figure 4** – Stacked bar charts showing IPSS severity category by time.
    d. Supplementary Table 22 – IPSS categories – baseline, worst, 12 weeks with chi-square test for the distribution of severity grades at each time point between SBRT and CFMHRT
    e. Test: Mann Whitney x3 for IPSS total, at baseline, worst, 12 week
    f. Test: Mann Whitney x3 for IPSS QoL, at baseline, worst, 12 week
    g. Test: Mann Whitney for IPSS worst score change from baseline
    h. Test: Mann Whitney for IPSS worst QoL change from baseline

8. IIEF-5
    a. Text. Score average and CI at 0, 12 weeks. Compare by mann whitney at each timepoint
    b. Text. Change in IIEF-5 from baseline average and CI. Compare by Mann Whitney

9. Vaizey
    a. **Supplementary Table 22**. Score average and CI at 0,4,12 weeks. Compare 0 and 12 weeks by Mann Whitney
    b. Text. **Compare worst Vaizey** – Mann-Whitney
    c. Text. **Compare worst Vaizey score change from baseline** – Mann-Whitney

Appendix 3. Interpolation Method for PACE-B Graphs

There is difficulty in producing graphs showing toxicity over time for two arms (CFMHRT and SBRT), with the x-axis beginning at the start of radiotherapy (represented by the baseline data). This is caused by each arm having two schedules of different durations:

- CFMHRT has:
  - 78 Gy in 39 fractions over 7.8 weeks
  - 62 Gy in 20 fractions over 4 weeks
- SBRT has
  - 36.25 Gy in 5 fractions over 1 week
  - 36.25 Gy in 5 fractions over 2 weeks

Therefore, the follow-up assessments do not necessarily fall at the same time for each schedule.
For example, week 2 follow-up post RT for 1-week SBRT occurs at 3 weeks from start of RT, whereas it will occur 4 weeks from start of RT for 2-week SBRT patients.

For a given grade (e.g. G1+), each patient is scored as a 1 (toxicity of that grade or more) or 0 (toxicity less than that grade). We wish to show at each timepoint the proportion of patients with grade 1+, grade 2+ and grade 3+ toxicity. For example, a patient with Grade 2 toxicity at a timepoint would be grade 1+ = 1 (yes), grade 2+ = 1 (yes), grade 3+ = 0 (no)

To obtain interpolated score for (e.g.) Grade 1+ at week 6 from start of RT, for a given patient not assessed at that timepoint:

  - Take G1+ toxicity status (0/1) at week 4.
  - Add G1+ toxicity status (0/1) at week 8.
  - Multiply by 0.5 (since the timepoint of interest is halfway between the known measurements)

The final multiplier could be altered if a different week of interest required interpolated data. In the above example, if week 5 interpolated data were required, then a final multiplier of 0.25 would be applied.

This of course assumes that patients' probability of having a toxicity or not changes in a linear fashion between timepoints.

The final point of each line contains only data from the longer of the two schedules to avoid extrapolation of data for the shorter schedule.

Appendix 4. PACE-B: Reasons for Non-Protocol Radiotherapy

Patient	Randomised	Per Protocol	Delivered Regimen	Radiotherapy Toxicity Related?	Reason that Non-Protocol Regimen Delivered
1	CFMHRT	CFMHRT	60 Gy in 20 F	No	Patient wanted 4-week regimen but was consented before 4 weekly regimen amendment occurred.
2	CFMHRT	CFMHRT	64 Gy in 32 F	No	Radiotherapy planning issue (small bowel proximity to prostate). Lower dose regimen prescribed.
3	CFMHRT	CFMHRT	74 Gy in 37 F	No	Pre-radiotherapy a protocol deviation to give ADT occurred, so given standard off-trial dose regimen for concurrent ADT usage
4	CFMHRT	CFMHRT	76 Gy in 38 F	No	Radiotherapy planning issue (Dose constraints not met so lower dose used)
5	SBRT	SBRT	14.5 Gy in 2 F then 46 Gy in 23 F	**Yes**	G3 urinary toxicity caused treatment interruption after 2 fractions SBRT. Completed treatment with conventional fractionation
6	SBRT	SBRT	21.75 in 3 F	No	On-treat dosimetry issue. Concerns that normal tissue dose constraints being violated. Decided not to deliver last 2 fractions
7	SBRT	CFMHRT	60 Gy in 20 F	No	Radiotherapy planning issue (bowel volume). Standard-of-care treatment preferred.
8	SBRT	CFMHRT	60 Gy in 20 F	No	Radiotherapy planning issue (bowel proximity to prostate). Standard-of-care treatment preferred.
9	SBRT	CFMHRT	60 Gy in 20 F	No	Radiotherapy planning issue (dosimetry at planning). Standard-of-care treatment preferred.
10	SBRT	CFMHRT	74 Gy in 37 F	No	Significant pre-existent urinary symptoms not recognised until planning CT. Thus had standard of care radiotherapy (with ADT).
11	SBRT	N/A	7.25Gy in 1 F then 55Gy in 20 F	No	On-treat issue. Patient moved during the delivery of first SBRT fraction. Decided to complete course with modified conventional regimen.

Appendix 5. PACE-B CRO Return Rates

Assessments counted as assessed if any useable toxicity data recorded.

RTOG Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
RTOG Baseline						
Assessed	402	93.1%	390	94.0%	792	93.5%
RTOG RT Week 2 (CFMHRT Only)						
Assessed	409	94.7%	N/A	N/A	409	94.7%
RTOG RT Week 4 (CFMHRT Only)						
Assessed	413	95.6%	N/A	N/A	413	95.6%
RTOG RT Week 6 (CFMHRT >25 Fractions Only)						
Assessed	116	89.9%	N/A	N/A	116	89.9%
RTOG RT Week 8 (CFMHRT >35 Fractions Only)						
Assessed	116	90.6%	N/A	N/A	116	90.6%
RTOG End of Treatment (SBRT Only)						
Assessed	N/A	N/A	400	96.4%	400	96.4%
RTOG Post-RT Week 2						
Returned	388	89.8%	389	93.7%	777	91.7%
RTOG Post-RT Week 4						
Returned	409	94.7%	403	97.1%	812	95.9%
RTOG Post-RT Week 8						
Returned	391	90.5%	372	89.6%	763	90.1%
RTOG Post-RT Week 12						
Returned	418	96.8%	402	96.9%	820	96.8%

CTCAE Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
CTCAE Baseline						
Assessed	430	99.5%	413	99.5%	843	99.5%
CTCAE End of Treatment (SBRT Only)						
Assessed	N/A	N/A	399	96.1%	399	96.1%
CTCAE Post-RT Week 2						
Assessed	389	90.0%	390	94.0%	779	92.0%
CTCAE Post-RT Week 4						
Assessed	410	94.9%	403	97.1%	813	96.0%
CTCAE Post-RT Week 8						
Assessed	393	91.0%	374	90.1%	767	90.6%
CTCAE Post-RT Week 12						
Assessed	420	97.2%	405	97.6%	825	97.4%

Appendix 6. PACE-B PRO Return Rates

Assessment for EPIC-26 scored as assessed if any subdomain fully completed, or if overall urinary bother question completed.

EPIC-26 Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
EPIC-26 Baseline						
Assessed	405	93.8%	387	93.3%	792	93.5%
EPIC-26 Post-RT Week 4						
Assessed	354	81.9%	362	87.2%	716	84.5%
EPIC-26 Post-RT Week 12						
Assessed	380	88.0%	382	92.0%	762	90.0%

IPSS assessment counted as assessed if IPSS total score calculable, or if quality of life question completed.

IPSS Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
IPSS Baseline						
Assessed	399	92.4%	384	92.5%	783	92.4%
IPSS Post-RT Week 2						
Assessed	364	84.3%	358	86.3%	722	85.2%
IPSS Post-RT Week 4						
Assessed	347	80.3%	351	84.6%	698	82.4%
IPSS Post-RT Week 8						
Assessed	354	81.9%	346	83.4%	700	82.6%
IPSS Post-RT Week 12						
Assessed	371	85.9%	371	89.4%	742	87.6%

IIEF-5 assessment counted as assessed if IIEF-5 total score calculable.

IPSS Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
IIEF-5 Baseline						
Assessed	322	74.5%	309	74.5%	631	74.5%
IIEF-5 Post-RT Week 12						
Assessed	280	64.8%	286	68.9%	566	66.8%

Vaizey assessment counted as assessed if Vaizey total score calculable.

Vaizey Assessment	Per Protocol Treatment				Total	
	CFMHRT		SBRT			
	n	%	n	%	n	%
Vaizey Baseline						
Assessed	373	86.3%	358	86.3%	731	86.3%
Vaizey Post-RT Week 4						
Assessed	267	61.8%	276	66.5%	543	64.1%
Vaizey Post-RT Week 12						
Assessed	349	80.8%	352	84.8%	701	82.8%