Optimising Dominant Intraprostatic Lesion Outlining for Prostate Radiotherapy

Dr Elin Siân Evans

MBBS MRCP (UK) FRCR

A thesis submitted to Cardiff University for the degree of Doctor of Medicine (MD)

February 2023

Summary

The dominant intraprostatic lesion (DIL) is the most common source of local recurrence following radiotherapy for localised prostate cancer. Recent trial findings suggest dose escalation to the DIL results in a low chance of prostate cancer recurrence at five years and confers minimal excess toxicity compared to standard prostate radiotherapy dose and fractionation schedules. Defining the DIL to outline for radiotherapy planning requires mpMRI interpretation skills and the ability to cognitively transfer the mpMRI defined DIL onto the CT planning scan. Outlining variation amongst radiotherapy outliners across all tumour sites is a well-recognised limitation in the radiotherapy planning pathway. Strategies to minimise inter-observer outlining variation include implementation of tumour site-specific outlining guidelines and educational outlining workshops.

This thesis explores the incidence of outlining variation of the DIL amongst UK prostate oncologists using data from the pre-accrual benchmark case submissions of the PIVOTALboost trial; a phase III randomised control trial of prostate and pelvis versus prostate alone radiotherapy, with or without DIL dose-escalation. Having established a high rate of case resubmission, predominantly due to unacceptable DIL outlining variation, the thesis then explores the role of conformity indices as a semi-automated tool in the assessment of unacceptable versus acceptable outlining variation in the PIVOTALboost pre-accrual benchmark case submissions. Thereafter, this thesis evaluates the impact of an outlining workshop on DIL outlining performance of UK prostate oncologists. Finally, this thesis investigates which step within the DIL outlining process is the cause of outlining variation; i.e. mpMRI interpretation or cognitive transfer, to inform future work aimed at minimising DIL outlining variation.

Contents

Acknowledgements

I would firstly like to thank my supervisors Professor John Staffurth, Dr Sarah Gwynne and Professor Emiliano Spezi for their invaluable input during the progress of this thesis. Professor Staffurth's clinical knowledge has been instrumental to both this work and my clinical practice for which I am incredibly grateful. This has complemented Dr Gwynne's sound knowledge of radiotherapy outlining variation, which has formed the foundation of this work. Both have been incredibly kind, supportive and patient throughout. I am particularly grateful to Professor Spezi for his conformity index calculations and incredible patience when I required just a few more….!

I would like to thank Dr Isabel Syndikus and Dr Alison Tree for their input in this thesis. Their DIL outlining knowledge has been imperative to this project and I am extremely grateful for their specific input in the case development for both chapters 5 and 6. I would also like to thank Dr Syndikus, Dr Tree, Professor Staffurth and the rest of the PIVOTALboost trial's team for allowing me access to the pre-trial data upon which this thesis was developed and to Helen Mayles for preparing and transferring the data for me to analyse.

To Owain Woodley, a huge thank you for your practical input throughout this whole thesis. From assisting with data transfer, to painstakingly trying to help me find a solution to import MRI images into a suitable platform to analyse, I am incredibly grateful for all his help. Thank you also to James Williams, for transferring data from the SWWCC team.

Many thanks to Dr Concetta Piazzese for her conformity index work and teaching me the ins and outs of CERR, how to package data for importing/ exporting and her patience when I needed refresher sessions!

I am also incredibly grateful to the Royal College of Radiologists for allowing me to deliver the prostate outlining workshop with a specific focus on DIL outlining, as well as the delegates for participating and submitting contours. Thank you in particular to Dr Ganesh Radhakrishna and Mr David Christopher for their help in developing and collecting the questionnaire data respectively. I would also like to thank the team at AQUILAB particularly Ms Clarisse Bartau, for technical support before, during and after the workshop.

Thank you to Dr W. John Watkins, Statistician at Cardiff University, for all his input into the statistical analyses within my thesis and the sound advice that 'significance isn't everything'!

I would like to thank the clinicians and non-medical outliners who participated in the outlining exercise within this thesis; I appreciate their patience and dedication in completing all the steps required

Chapter 1: Introduction; Prostate Cancer

1.1 Epidemiology of prostate cancer

Prostate cancer is the most common cancer affecting men in the UK, with 52,254 new cases diagnosed between 2016-18. A rapid increase in prostate cancer incidence has occurred since the 1980's, initially attributed to incidental detection following transurethral resection of the prostate (TURP) and latterly prostate-specific antigen (PSA) testing. Correlating with increasing incidence of prostate cancer in the UK over the past few decades, an initial increase in mortality rate was observed, but in the last ten years mortality rates have decreased by 12%. Survival rates have also improved; 84% of men diagnosed with prostate cancer in England and Wales survived their disease for more than ten years in 2010-2011 compared to only a quarter of men in the 1970's. [1].

1.2 Diagnosis and management of prostate cancer

1.2.1 Diagnosis of prostate cancer

Urgent suspected prostate cancer referrals from primary to secondary care include patients with a malignant feeling prostate on examination and/or a PSA value exceeding the age-specific reference range [2]. Historically, patients with a raised PSA proceeded straight to trans-rectal ultrasound (TRUS) guided biopsy which carries the risk of infection, pain and bleeding. The PROMIS trial [3] evaluated the diagnostic accuracy of multi-parametric magnetic resonance imaging (mpMRI) versus TRUS biopsies in men with a raised serum PSA. MpMRI, which includes functional imaging sequences in addition to the standard T1- and T2-weighted MRI sequences, demonstrated higher sensitivity for detection of clinically significant cancers than TRUS biopsies and 27% of men could be spared unnecessary biopsies using mpMRI as a triage tool. MpMRI followed by TRUS biopsy is currently standard of care in the UK for investigating clinically suspected prostate cancer although only half of eligible men are offered the test prior to biopsy [4] due to mpMRI scan quality and radiology expertise variability across the UK [5].

In addition to its diagnostic role in prostate cancer, mpMRI is also used as a staging modality. Prostate cancer is staged using the universally recognised Union for International Cancer Control (UICC) TNM staging classification which takes into consideration primary tumour (T), regional nodal (N) and distant metastatic (M) features [6] (Table 1.1). One meta-analysis reviewing the accuracy of MRI for local prostate cancer staging, determined that MRI has a high specificity but weak sensitivity for detection of extra-capsular extension (ECE) and seminal vesicle invasion (SVI), features in keeping with a locally advanced primary tumour [7]. This meta-analysis also concluded that the addition of one functional imaging technique to the standard MRI T2 weighted image sequences (T2W) increased sensitivity of

detecting ECE, SVI and overall T3 disease compared to T2WI alone. When two or more functional sequences are employed, ECE detection sensitivity increases further [7].

Classification	Definition
Tumour	
Tx	Tumour cannot be evaluated
T0	No evidence of a primary tumour
T1	Tumour detected incidentally on biopsy
T2	Tumour confined to prostate
T3a **T3b**	Tumour extends beyond prostatic capsule Tumour invades seminal vesicle
T4	Tumour invades adjacent structures (including pelvic wall, bladder, rectum)
Regional lymph nodes (N)	
Nx	Regional lymph nodes not evaluated
N0	No evidence of regional lymph nodes
N1	Pelvic lymph nodes evident
Distant Metastasis (M)	
M0	No distant metastases
M1a **M1b** **M1c**	Non-regional lymph nodes Bone metastases Visceral metastases

Table 1.1: Prostate Cancer TNM Staging [6]

Following mpMRI, patients with suspected localised prostate cancer undergo either transperineal or transrectal prostate biopsies, increasingly guided by the MRI findings. The tumour is graded using five histological growth patterns and the two most dominant tumour grades are combined to give an overall Gleason score and Gleason Grade Group (GGG). The combination of clinical/radiological staging, PSA and GGG risk stratifies prostate cancer to aid decision making regarding further staging and management (Table 1.2) [8].

Risk Group	Clinical/pathological features
Very Low	T1c AND grade group 1 AND PSA <10 AND fewer than 3 prostate cores positive, ≤50% cancer in each core AND PSA density <0.15ng/mL/g
Low	T1-T2a AND grade group 1 AND PSA <10 ng/ml AND no histological features of very low risk
Favourable intermediate	No high or very high risk features Has all of the following: one intermediate risk factor (T2b to T2c, grade group 2 or 3, PSA 10-20 g/ml), grade group 1 or 2, <50% positive biopsy cores
Unfavourable intermediate	No high or very high risk features Has one or more of the following : two or three intermediate risk factors (T2b to T2c, grade group 2 or 3, PSA 10-20 ng/ml), grade group 3, ≥50% positive biopsy cores
High	Has no very high risk features and has one high risk feature: T3a OR grade group 4 or 5 OR PSA > 20 ng/ml
Very high	Has at least one of the following: cT3b-cT4, primary Gleason pattern 5, 2 or 3 high risk features, .4 cores with grade group 4 or 5

Table 1.2: Risk stratification for clinically localised prostate cancer [8]

National Comprehensive Cancer Network (NCCN) prostate cancer guidelines suggest mpMRI alone is suitable for radiological staging in very low, low and favourable intermediate-risk disease [8]. Higher risk cancers require further bone and soft tissue imaging [8]. Given the propensity for prostate cancer to metastasise to bones, nuclear bone imaging is recommended as an additional staging modality in higher risk disease [3]. Although this modality allows visualisation of the whole skeleton, comparative sensitivities for bone metastasis detection by MRI and nuclear bone scan are 79% and 97% and specificity 95% and 82% respectively [9].

Nuclear bone scans however cannot evaluate lymph node metastases thus limiting their staging use. mpMRI has a per-lesion lymph node detection sensitivity and specificity of 44% and 99% respectively [10] and lymph node evaluation is standardly included in MRI reports as part of the diagnostic work up but is limited to the pelvis. Prostate-specific membrane antigen (PSMA) positron-emission computed-tomography (PET-CT) is an imaging modality that exploits the overexpression of the PSMA glycoprotein on prostate cancer cells. Patients are administered a radiolabelled tracer (e.g. 68Ga-PSMA-11) which has a high binding affinity with PSMA permitting whole body imaging. The proPSMA trial compared the use of PSMA PET-CT with conventional imaging modalities (CT and nuclear bone scan) in staging high risk localised prostate cancer and concluded PSMA PET-CT had a higher accuracy by 27% than combined conventional CT and bone scan in detecting pelvic lymph nodes and distant metastatic disease [11]. At present, PSMA PET-CT is not yet included as part of recommended staging

for high risk prostate cancer within the National Institute for Clinical Excellence (NICE) guidelines [4] as patient access remains variable across the UK [12].

1.2.2 Management of localised prostate cancer

Patients diagnosed with localised prostate cancer suitable for radical treatment may be offered either a radical prostatectomy or radiotherapy depending on tumour/ patient factors and choice [3,13]. In low risk disease, active surveillance, i.e. no radical intervention unless regular monitoring raises concern the disease may be progressing, is also a suitable option [3, 13]. The ProtecT trial randomised 1643 men with localised prostate cancer to receive either active monitoring, surgery or radiotherapy [14]. After 10 years of follow up, prostate-cancer- specific mortality was low across all treatment groups, with no significant difference between them. However, the rate of local disease and metastatic progression was less than half in the intervention groups than those in the active surveillance group. Between 2017 and 2018 in England and Wales, 7269 men with prostate cancer had a radical prostatectomy whereas double the number of men (14627) received radical radiotherapy [15] most likely due to the advanced age or stage of patients diagnosed.

1.3 Radiotherapy for localised prostate cancer

Until the 1990s, external beam radiotherapy (EBRT) was delivered non-conformally, using pelvic bony landmarks on x-ray images to estimate prostate position and determine treatment fields accordingly. The introduction of 3D conformal radiotherapy, using CT images to determine beam arrangement, reduced normal tissue irradiation and was associated with significantly lower rates of late radiation–induced proctitis [16]. Bladder and rectal toxicity, using the conventional technique, limited the prostate irradiation dose to 64-70 Gray (Gy) in 1.8 to 2 Gy per fraction. The MRC RT01 trial recruited 813 men to receive either 64 Gy in 32 fractions as standard of care versus 74 Gy in 37 fractions in the experimental dose escalation arm using a conformal technique in both arms [17]. The dose-escalated arm was associated with increased genitourinary and bowel toxicity immediately post treatment [17] and a higher rate of rectal bleeding longer term [18] although patient-reported distress regarding symptoms was generally low. The dose-escalated arm was associated with a better biochemical progression free survival but not overall survival, even at 10-year follow-up [19], a finding consistently demonstrated across dose-escalation studies [20].

Radiotherapy delivery has since evolved with the introduction of intensity-modulated radiotherapy (IMRT). This technique uses computer controlled multi-leaf collimators to move independently within the linear accelerator, thereby shaping the treatment field around the target area. This not only

controls the intensity of the radiation delivered to the target but also conforms the radiotherapy to the target, further sparing organs at risk. IMRT has therefore enabled further evaluation of dose escalation. The Memorial Sloan Kettering group treated 478 men with localised prostate cancer to an 'ultra' dose escalation of 86.4 Gy in 48 fractions and concluded that treatment was well tolerated [21]. A retrospective analysis of 2251 men with localised prostate cancer found radiation dose was an important predictor of long term biochemical control, with doses <70.2Gy and 70-79.2Gy associated with 2.3 and 1.3-fold increased risk of biochemical relapse compared with higher doses [22].

Dose escalation however, when using dose per fraction of 1.8-2 Gy per fractions, necessitates more fractions, placing resource burden on treating centres and increased number of hospital attendances for patients. Hypofractionation, i.e. using fractions sizes of > 2Gy per day, help ease such issues. For prostate cancer, hypofractionation may also be radiobiologically advantageous given the well-supported hypothesis that prostate cancer has a low alpha/beta ratio [23]. Subsequently a number of large randomised clinical trials have focused on the efficacy and toxicity of hypofractionation in prostate cancer. This includes the Conventional or Hypofractionated High Dose Intensity Modulated Radiotherapy for Prostate cancer (CHHiP) trial, which compared 74 Gy in 37 fractions to 60 Gy in 20 fractions and 57 Gy in 19 fractions [24]. Five year follow up confirmed non-inferiority of the 60 Gy in 20 fraction arm with respect to biochemical relapse, although the 57 Gy in 19 fraction arm was deemed inferior. There was no significant difference in proportion or cumulative incidence of side-effects at five years between the groups. Consequently, 60 Gy in 20 fractions is the current standard dose and fractionation schedule in the UK [25].

For patients receiving radiotherapy for unfavourable intermediate- or high-risk localised prostate cancer it is standard of care to offer neoadjuvant androgen deprivation therapy (and adjuvant in higher risk groups) prior to radiotherapy as this significantly improves metastasis free-survival [26].

1.4 Radiotherapy outlining and delivery

1.4.1 Radiotherapy pathway

All patients undergoing radical EBRT for prostate cancer have a planning CT scan of the pelvis. Radiation/clinical oncologists (or specifically trained non-medical outliners), are required to outline the relevant structures (target volumes) on the planning CT as well as the organs at risk (OARs) of toxicity. Pre-determined dose parameters, including the optimal dose to the target volume and maximum dose to OARs, are applied by dosimetrists in order of priority to an optimisation program. This creates a radiotherapy plan that optimally conforms to the dose parameters (inverse planning) using IMRT or volumetric-modulated radiotherapy (VMAT), a form of IMRT in which the linear accelerator rotates around the patient delivering radiotherapy as it rotates. The clinician, or appropriately trained technologist, reviews the radiotherapy plan and doses to target volumes and organs at risk prior to finalising the treatment plan.

1.4.2 Prostate radiotherapy target volumes

Principles of radiotherapy outlining may be applied across all tumour sites and routinely includes delineation of target volumes to be irradiated. The gross tumour volume (GTV) includes all macroscopic disease i.e. the tumour. Clinical target volumes (CTV) encompass regions at risk of microscopic disease. This may include the area directly surrounding the GTV or associated regions at risk of microscopic spread e.g. regional lymph node groups. A planning target volume (PTV) is created by volumetric expansion of either the GTV or CTV to allow for organ motion and set up error (Figure 1.1).
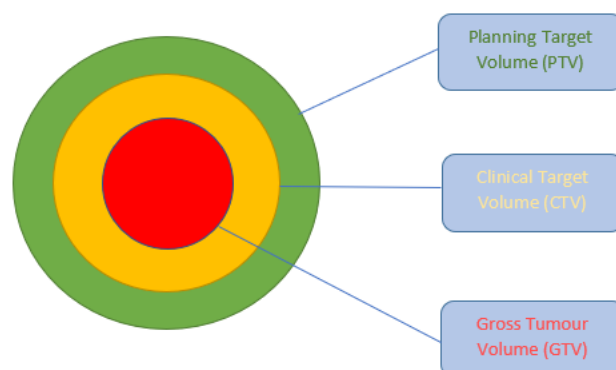
Figure 1.1: Target volumes used in radiotherapy planning

Although target volume definitions for prostate cancer will vary between radiotherapy centres, general principles remain the same. Macroscopic intraprostatic disease is not routinely outlined as a GTV. Instead, the whole prostate gland is outlined as a CTV (commonly denoted as CTVp). An additional CTV may be outlined to incorporate the seminal vesicles (CTVpsv) which may be directly invaded or at risk of invasion from the prostate. An isotropic volumetric expansion (e.g. CTV grown by 5mm in all directions) is applied to the CTVs to create their respective PTVs (Figure 1.2).
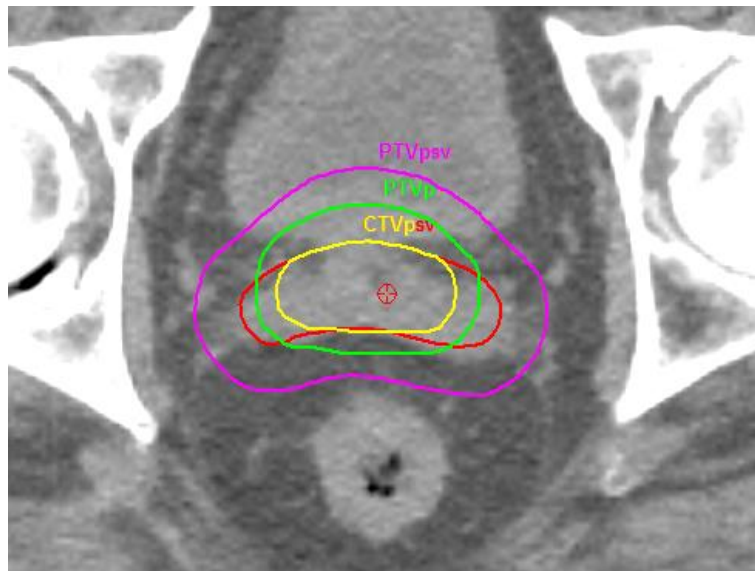
Figure 1.2. Axial section of CT planning scan showing CTVp (yellow) and CTVpsv (red) and their respective PTVs: PTVp (green) and PTVpsv (purple)

The role of elective pelvic nodal irradiation in localised prostate cancer remains controversial, with sufficient evidence supporting its routine inclusion in radiotherapy planning yet to be established [27], although may be considered in high-risk patients [2,25]. This volume, when included, is usually denoted CTVn.

OARs are also delineated (usually by the radiotherapy dosimetrist/planner) to calculate the dose received to these structures so that appropriate constraints may be applied. For prostate radiotherapy, OARs include rectum, bladder, bowel, urethra and penile bulb.

1.4.3 Recurrent disease following prostate radiotherapy

Although dose escalated radiotherapy is associated with decreased local and biochemical failure [22, 28], the most common site of recurrent disease is within the high dose irradiated region [29, 30]. Spratt *et al* [31] reviewed 2694 patients receiving dose-escalated radiotherapy to the prostate, of which 18% of patients developed a clinically detectable recurrence and of those, 41% were localised

to the prostate/seminal vesicles. Other sites of recurrence include lymph nodes and distant metastases.

1.5 Dominant Intraprostatic Lesion (DIL)

1.5.1 Significance of the DIL

Prostate cancer is a multifocal disease with 60 to 90% of men with localised prostate cancer having more than one tumour [32]. The dominant intraprostatic lesion (DIL) is the largest tumour lesion within the prostate and is understood to determine the course of prostate cancer for the individual patient [33]. The histopathological characteristics of the DIL, particularly the grade and presence of extracapsular extension [34], can influence prognosis. Liu *et al* [35] further highlighted the significance of the DIL through tracing metastatic prostate cancer cells to a single genomically aberrant prostate cell. Salvage prostatectomy histopathological analysis following radiotherapy suggests the most likely site of recurrence within the prostate is at the primary tumour [36]. MR imaging studies, which also identified tumour recurrence at the site of the DIL following whole gland radiotherapy [37-39], further supports this finding.

1.5.2 The role of mpMRI in DIL detection

As described above, pre-biopsy pelvic mpMRI is becoming the standard diagnostic imaging for localised prostate cancer in the UK. Turkbey *et al* [40] showed a positive correlation between histopathology tumour volume and MRI tumour volume, particularly in tumours with a volume greater than 0.5cm³. Through the addition of functional sequences, including diffusion weighted imaging (DWI) and dynamic type following administration of contrast (DCE) to the T1- and T2-weighted (T2W) anatomical sequences, sensitivity and specificity prostate cancer increases [41,42].

T2-weighted sequences provide the most detailed anatomical prostate information through high spatial resolution and enable definition of separate prostatic components including the peripheral, central and transitional zones, ejaculatory ducts, seminal vesicles, urethra and anterior fibromuscular stroma [43] (Figure 1.3). On T2W imaging, the dominant tumour typically appears as a region of low signal within normal high signal tissue.
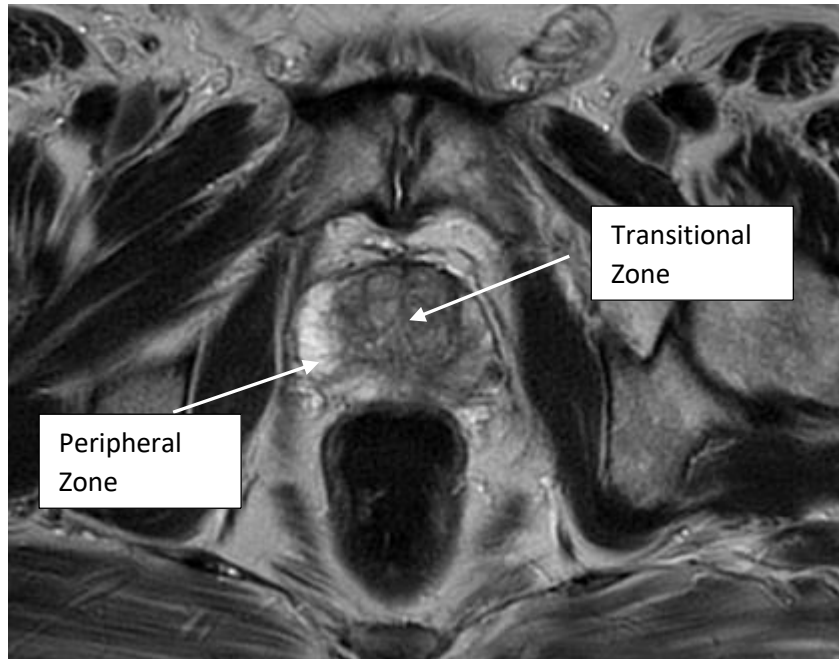
Figure 1.3. Axial image of a T2 weighted image of the prostate differentiating transitional and peripheral zones

Diffusion-weighted imaging (DWI) MR sequences measure the Brownian motion of water molecules in tissues such that highly cellular tissues exhibit lower diffusion coefficients. The apparent diffusion coefficient (ADC) quantifies the restriction of water diffusion and is measured through acquisition of at least two sets of images with differing gradient durations and amplitudes [44]. ADC maps enable visual assessment, illustrating the tumour as an area of decreased signal. Combining DWI and T2W sequences significantly improves sensitivity and specificity of detecting clinically significant intraprostatic lesions than T2W imaging alone [45].

The role of Dynamic Contrast Enhanced (DCE) imaging for evaluating the DIL remains debatable. Given its overlap with detecting benign prostate clinical findings, DCE is mainly used to support DWI/T2W findings [44, 45] particularly when other sequences are sub-optimal. It does however have a role in surveillance following prostatectomy, radiotherapy or focal ablation [46].

1.5.3 PI-RADS scoring system

To gain global consensus regarding prostate cancer imaging reporting, the European Society of Urogenital Radiology (EUSR), American College of Radiology (ACR) and AdMeTech Foundation established a steering committee to develop the current Prostate Imaging Reporting and Data System (PI-RADs) guidance. This includes a 5-point scale based on the probability that mpMRI findings using T2W, DWI and DCE sequences correlates with localisation of a clinically significant tumour for each

intraprostatic lesion identified (Table 1.3). PI-RADSv2.1 does stipulate that where T2W and DWI are of adequate diagnostic quality, DCE plays a minor role in determining PI-RADS assessment. [47]

ADSv2.1 Assessment Categories	
PI-RADS 1	Very low (clinically significant cancer is highly unlikely to be present)
PI-RADS 2	Low (clinically significant cancer is unlikely to be present)
PI-RADS 3	Intermediate (the presence of clinically significant cancer is equivocal)
PI-RADS 4	High (clinically significant cancer is likely to be present)
PI-RADS 5	Very high (clinically significant cancer is highly likely to be present)

Table 1.3: PI-RADSv2.1 Assessment Categories for mpMRI detection of clinically significant cancers within the prostate [47]

The vast majority of prostate cancers are within the peripheral and transitional zones of the prostate. However, the ability of mpMRI sequences to detect potentially significant lesions within these zones differ and as such, PI-RADS outlines different assessment techniques for each location [47].

1.5.4 PI-RADS: Peripheral zone tumours

Approximately 75% of prostate cancers will develop in the peripheral zone (PZ) [48]. On DWI, malignant lesions in the PZ appears as an area of restricted diffusion whereas on T2W imaging they appear as a region of hypo-intensity (Figure 1.4).
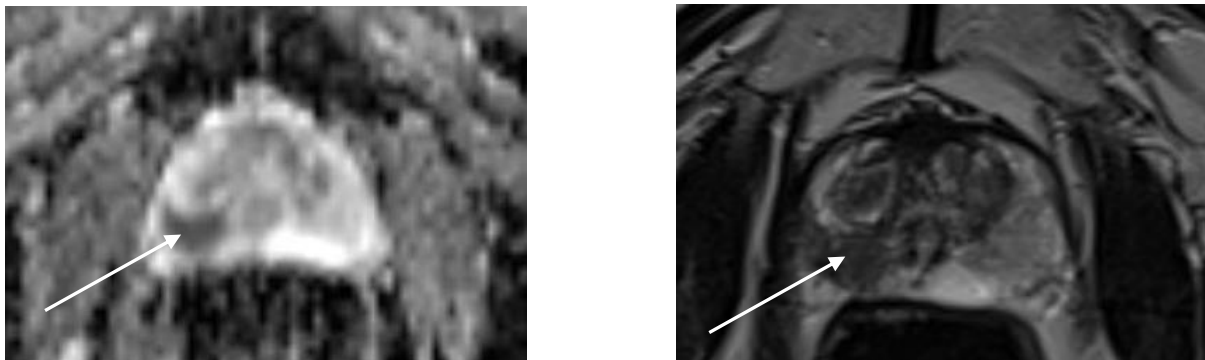
Figure 1.4: MpMRI showing a peripheral zone prostate cancer as a diffusion restricted lesion on an ADC map (left-sided image) and corresponding hypo-intensity on T2W (right-sided image)

Given the difference in features of malignant lesions between T2W and DWI (ADC) in the PZ, PI-RADS v2.1 describes two separate assessment tools for each sequence, to predict the likelihood of a lesion being malignant (Tables 1.4 and 1.5). [47]

Score (T2W)	Peripheral Zone
1	Uniform hyper-intense signal intensity (normal)
2	Linear or wedge-shaped hypo-intensity or diffuse mild hypo-intensity, usually indistinct margin
3	Heterogeneous signal intensity or non-circumscribed, rounded, moderate hypo-intensity Includes others that do not qualify as 2,4 or 5
4	Circumscribed, homogenous moderate hypointense focus/mass confined to prostate and <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extraprostatic extension/ invasive behaviour

Table 1.4: PI-RADS Assessment for T2W imaging of peripheral zone lesions [47]

Score (DWI)	Peripheral Zone or Transitional Zone
1	No abnormality (i.e. normal) on ADC and high b-value DWI
2	Linear or wedge-shaped hypointense on ADC and/or linear/ wedge-shaped hyperintense on high b-value DWI
3	Focal (discrete and different from the background) hypointense on ADC and/or focal hyperintense on high b-value DWI; may be markedly hypointense on ADC or markedly hyperintense on high b-value DWI, but not both
4	Focal markedly hypointense on ADC and markedly hyperintense on high-b-value DWI; <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extraprostatic extension/ invasive behaviour

Table 1.5: PI-RADS Assessment for DWI imaging of peripheral and transitional zone lesions [47]

DWI is the more sensitive sequence to detect PZ cancers [49] with T2W imaging providing supplementary diagnostic information. As such, PI-RADS assessment of peripheral zone lesions places greater emphasis on DWI findings than T2W (Table1.6) [47].

DWI	T2W	DCE	PI-RADS
1	Any (1-5)	Any (1-5)	1
2	Any (1-5)	Any (1-5)	2
3	Any (1-5)	-	3
		+	4
4	Any (1-5)	Any (1-5)	4
5	Any (1-5)	Any (1-5)	5

Table 1.6: PI-RADS assessment of peripheral zone lesions [47]

1.5.5 PI-RADS: Transitional zone tumours

Transitional zone (TZ) tumours may be more challenging to identify on mpMRI, particularly in older patients where benign prostatic hypertrophy (BPH) commonly co-exists causing a heterogeneous appearance on T2W [50]. Features to distinguish malignant TZ tumours from benign prostatic

hypertrophy (BPH) nodules on T2W include homogenous low signal T2 intensity, lack of defined capsule, irregularity, lenticular in shape and invasion of the anterior fibromuscular stroma [51]. TZ lesions on DWI usually appear as a region of restricted diffusion (Figure 1.5). Although PI-RADS v2 uses T2W as the dominant sequence for PI-RADS scoring [47], one sequence alone cannot adequately detect prostate cancer and a combination of T2W and DWI offers the most accurate assessment [52].
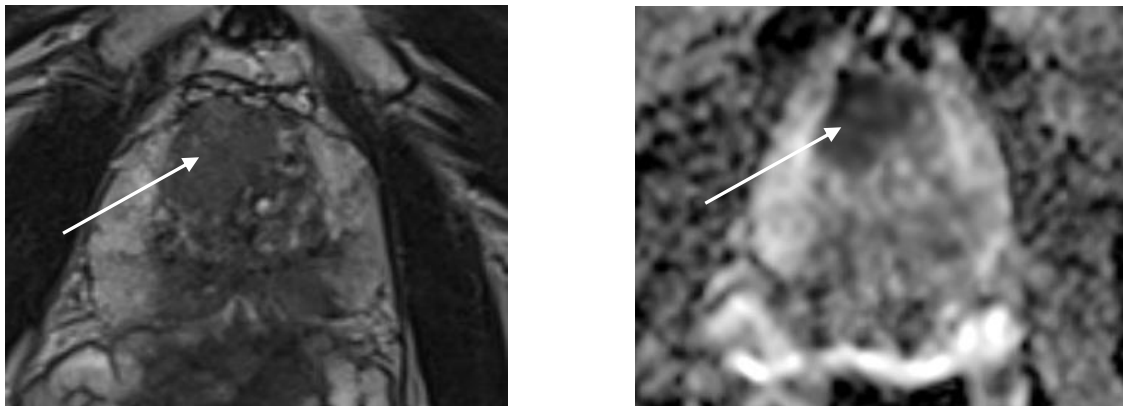
Figure 1.5: MpMRI showing a transitional zone prostate cancer as a non-circumscribed homogenous lesion on T2W (left-sided image) and corresponding hypo-intensity on ADC map (right-sided image)

PI-RADS assessment of TZ lesions uses the same DWI assessment features as for PZ lesions but given the difference in T2W features between the two zones, uses a different assessment description for this sequence (Table 1.7). Given the priority of T2W features over DWI for TZ lesions, an alternative PI-RADS scoring system is used for TZ lesions (Table 1.8). [47]

Score (T2W)	Transitional Zone
1	Normal appearing TZ or a round, completely encapsulated nodule (''typical nodule'')
2	A mostly encapsulated nodule OR a homogenous circumscribed nodule with encapsulation OR a homogenous mildly hypointense area between nodules
3	Heterogeneous signal intensity with obscured margins Includes others that do not qualify as 2,4 or 5
4	Lenticular or non-circumscribed, homogenous, moderately hypointense and prostate and <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extraprostatic extension/ invasive behaviour

Table 1.7: PI-RADS Assessment for T2W imaging of transitional zone lesions [47]

DWI	T2W	DCE	PI-RADS
1	Any (1-5)	Any (1-5)	1
2	≤3	Any (1-5)	2
	≥4	Any (1-5)	3
3	≤ 4	Any (1-5)	3
		Any (1-5)	4
	5		
4	Any (1-5)	Any (1-5)	4
5	Any (1-5)	Any (1-5)	5

Table 1.8: PI-RADS assessment of transitional zone lesions [47]

1.5.6 Sub-volume boost

Given the significance of the DIL in determining prostate cancer outcomes, the concept of therapeutically targeting the lesion specifically is well established. Focal therapies including cryotherapy and high-intensity focused ultrasound (HIFU) offer the advantage of targeted tumour ablation whilst minimising the short and long term effects of whole gland radiotherapy or prostatectomy [53]. Although short-term tumour control outcomes for HIFU are comparable with whole gland treatments [54], a lack of prospective randomised controlled trials and long-term outcome data have hindered focal therapy as an accepted mainstay of treatment [53]. An underlying concern regarding uni-focal therapy is that by targeting the DIL alone, smaller foci of disease within the remaining prostate that may not be apparent on imaging will be omitted from treatment and risk disease recurrence [55]. Haffner *et al* [56] found that in a patient who died from prostate cancer, the lethal clone had not originated as anticipated from the higher grade DIL but from a smaller, lower grade focus of tumour within the prostate.

An alternative approach therefore, has been to consider 'sub-volume boost' of the DIL i.e. simultaneous radiotherapy dose escalation to the DIL whilst treating the whole prostate gland to a standard dose. As early as 1999, Pickett *et al* [57] explored the feasibility of using static field IMRT to treat the DIL to up to 90Gy whilst simultaneously treating the whole prostate to over 70Gy. Since then, a number of patient series treating patients with whole gland EBRT and either a brachytherapy or an external beam boost to the DIL have been evaluated, although from this heterogeneous group of patients, key elements including the optimal dose/fractionation, image guidance for delineation and volume definition are yet to be determined [58].

A number of trials have endeavoured to address these issues. DELINEATE, a single centre phase II multicohort study, reviewed efficacy and toxicity of mp-MRI identified DIL boosting using standard (74Gy/37 fractions) and hypofractionated external beam radiotherapy [59]. Five-year follow up

indicates both fractionation schedules to be safe and tolerable with a low chance of prostate recurrence at five years [60]. Similarly, FLAME, a phase III multicentre randomised controlled trial, found that at five years, biochemical disease free survival was significantly higher at 92% in the cohort that received a simultaneous boost to 95Gy compared to 85% in the standard arm receiving 77Gy in 35 fractions to the prostate alone [61].

1.5.7 PIVOTALboost

The role of DIL boosting in prostate radiotherapy continues to gain momentum. PIVOTALboost is a Cancer Research UK (CRUK) funded multicentre randomized control phase III clinical trial (ISRCTN: 80146950) which aims to evaluate the role of dose escalation to the DIL in prostate radiotherapy with or without pelvic radiotherapy versus standard hypofractionated radiotherapy with or without pelvic radiotherapy [62]. Eligible patients must have either NCCN confirmed high risk or locally advanced node negative adenocarcinoma of the prostate or NCCN intermediate risk prostate cancer with one additional adverse feature (maximum tumour length >6mm and/or >50% positive biopsy cores and/or PI-RADS score 3,4,5 DIL >10mm on diagnostic MRI). The primary aims of the trial are to evaluate the benefits of pelvic lymph node radiotherapy, high dose rate brachytherapy (HDRB) with and without a boost volume and focal boost IMRT identified on mpMRI. Centres may choose which method they wish to deliver the boost i.e. via HDRB or IMRT, depending on local resource availability. The boost volume is determined pre-recruitment through diagnostic MRI findings and must fulfil all specific criteria to be eligible for consideration of a dose escalation boost (Table 1.9).

mpMRI DIL Criteria
PI-RADS (v.2) 4 or 5 lesion
Minimum DIL dimension of 5mm
Total DIL volume <50% of the total prostate volume

Table 1.9: mp-MRI DIL eligibility criteria for PIVOTALboost boost volume

Depending on the patient characteristics and the participating centre's ability to offer brachytherapy or focal boost IMRT, patients may be randomised to one of four arms:

- A: Prostate alone
- B: Prostate and pelvic IMRT
- C: Prostate IMRT and prostate boost
- D: Prostate and pelvic IMRT and prostate boost

Centres unable to offer either HDRB or IMRT boost will still be eligible to randomise patients into arms A and B (Figure 1.6).

The trial is powered to recruit 1952 patients to establish both the role of DIL boosting and elective pelvic nodal irradiation. The primary endpoint is failure free survival (FFS), with failure define as the first of biochemical failure, re-initiation of androgen deprivation therapy (ADT), local, lymph node or pelvic recurrence, distant metastases or death attributed to prostate cancer

Eligible patient group: Localised node negative prostate cancer and:

- PSA <50ng/ml
- NCCN high risk (T3a/b or T4 N0M0 clinical and/or MRI) and/or dominant Gleason 4 or 5 and/or PSA >20
- NCCN intermediate risk (T2b/c N0M0 and/or Gleason 3+4 and/or PSA 10-20 ng/ml and at least 1 adverse feature e.g. maximum tumour length (MTL) >6mm and/or ≥50% biopsy cores positive and/or PI-RADS score 3,4 or 5, DIL lesion >10mm on staging MRI

Determined pre-randomisation:

- Boost volume on staging MRI: none, <50% or ≥50%
- Intended dose escalation RT technique: HDRB whole prostate boost; HDRB or IMRT focal prostate boost)

A: Prostate IMRT

B: Prostate and pelvic IMRT

C: Prostate IMRT and prostate boost*

D: Prostate and pelvic IMRT and prostate boost*

C1: Whole gland HDR*

C2: HDR or IMRT focal boost*

D1: Whole gland HDR*

D2: HDR or IMRT focal boost*

*Arms depend on:
- Availability of technique at centre
- Boost volume seen on staging MRI
- Patient suitability for HDR

Figure 1.6: PIVOTALboost trial randomization schema adapted from PIVOTALboost protocol [64]

Pre-radiotherapy, all patients within PIVOTALboost are required to undergo an mpMRI, ideally pre-biopsy (mandatory for boost patients) which include T1W, T2W and DWI sequences. Patients with a boost volume identified on diagnostic MRI, fulfilling eligibility criteria (Table 1.9), may be randomised to A vs B vs C2 vs D2. Patients should receive androgen deprivation therapy (ADT) for up to six months

if considered intermediate risk and for two years if high risk. Ideally, RT treatment should be initiated during the third month of ADT, but is acceptable to start up to six months after commencing ADT.

1.6 Conclusion

Dose escalation to the DIL is a promising technique but may cause toxicity to critical OARs that lie near the DIL: PZ and rectum, cranial prostate lesions and bladder and anterior prostate lesions and urethra. Accurate outlining of the DIL is thus essential to ensure both tumour control and to minimise late side-effects. As mp-MRI is a relatively new tool for radiologists for which specific training programmes have had to be developed, using data from the PIVOTALboost trial, my thesis aims to explore target volume delineation amongst clinical oncologists in this new era and methods to improve its quality. Target volume delineation accuracy and methods to address this will be discussed in chapter 2.

1.7 References

1. Cancer Research UK: Prostate cancer incidence statistics [Internet]. London. [Cited 2023 Feb 3]. Available from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Zero

2. National Institute for Health and Care Excellence: Urological cancers- recognition and referral [Internet]. 2021 [updated 2021 Feb; cited 2023 Feb 03] Available from: https://cks.nice.org.uk/topics/urological-cancers-recognition-referral/

3. El-Shater Bosaily A, Parker C, Brown LC, Gabe R, Hindley R, Kaplan R et al. PROMIS--Prostate MR imaging study: A paired validating cohort study evaluating the role of multi-parametric MRI in men with clinical suspicion of prostate cancer. Contemp Clin Trials. 2015 May; 42: 26-40.

4. Wise J. NICE recommends MRI for suspected prostate cancer to reduce biopsies. BMJ. 2018, Dec; 363: k5290 doi:10.1136/bmj.k5290.

5. Brizmohun Appayya M, Adshead J, Ahmed HU, Allen C, Bainbridge A, Barrett, T et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection recommendations from a UK consensus meeting. BJU Int. 2018 Jul; 122(1):13-25.

6. Buyyounouski MK, Choyke PL, McKenney JK, Sartor, O, Sandler H, Amin M et al. Prostate cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 2017; 67(3): 245-253.

7. de Rooij M, Hamoen EH, Witjes JA, Barentsz JO, Rovers MM. Accuracy of magnetic resonance imaging for local staging of prostate cancer: A diagnostic meta-analysis. Eur Urol. 2016 Aug; 70(2): 233-245.

8. National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology Prostate Cancer [Internet]. 2022 [updated 2022 Sept; cited 2023 Feb 03] Available from https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf

9. Shen G, Deng H, Hu S, Jia Z. Comparison of choline-PET/CT, MRI, SPECT, and bone scintigraphy in the diagnosis of bone metastases in patients with prostate cancer: a meta-analysis. Skeletal Radiol. 2014 Nov; 43(11): 1503-1513.

10. Yip K, Li Y, Chen S, Ho C & Wartolowska K. ^{68}Ga-PSMA PET/CT and mpMRI for primary lymph node staging of intermediate to high-risk prostate cancer: a systematic review and meta-analysis of diagnostic test accuracy. Clin Transl Imaging. 2021; 9, 523–537.

11. Hofman MS, Lawrentschuk N, Francis RJ, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. Lancet. 2020; 395(10231): 1208-1216.

12. Afaq A, Bomanji J. Prostate-specific membrane antigen positron emission tomography in the management of recurrent prostate cancer. Br Med Bull. 2018 Dec; 128(1): 37-48.

13. European Association of Urology [Internet]. 2022 [updated 2021; cited 2023 Feb 03] Available from: https://uroweb.org/guidelines/prostate-cancer

14. Hamdy F, Donovan J, Lane J, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med. 2016 Oct; 375(15): 1415-1424.

15. National Prostate Cancer Audit: Annual Report 2019 [Internet]. 2020 [cited 2023 Feb 03] Available from: https://www.npca.org.uk/content/uploads/2020/01/NPCA-Annual-Report-2019_090120.pdf

16. Dearnaley D, Khoo V, Norman A, 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(9149): 267-272.

17. Dearnaley D, Sydes M, Graham J, Aird E, Bottomley D, Cowan R, et al. Escalated-dose versus standard-dose conformal radiotherapy in prostate cancer: first results from the MRC RT01 randomised controlled trial. *Lancet Oncol. 2007 June; 8(6):* 475-87.

18. Syndikus I, Morgan RC, Sydes MR, Graham JD, Dearnaley DP; MRC RT01 collaborators. Late gastrointestinal toxicity after dose-escalated conformal radiotherapy for early prostate cancer: results from the UK Medical Research Council RT01 trial (ISRCTN47772397). Int J Radiat Oncol Biol Phys. 2010 Jul; 77(3): 773-783.

19. Dearnaley DP, Jovic G, Syndikus I, Khoo V, Cowan R, Raham J, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: Long term results from the MRC RT01 randomised control trial. Lancet Oncol. 2014, 15(4): 464-473.

20. Viani GA, Stefano EJ, Alfonso SL. Higher-than-conventional radiation doses in localized prostate cancer treatment: a meta-analysis of randomized controlled trials. Int J Radiat Oncol Biol Phys. 2009 Aug 1; 74(5): 1405-1418.

21. Cahlon O, Zelefsky M, Shippy A, et al. Ultra-high dose (86.4 Gy) IMRT for localized prostate cancer: toxicity and biochemical outcomes. Int J Radiat Oncol Biol Phys. 2008; 71(2): 330-337.

22. Zelefsky MJ, Pei X, Chou JF, et al. Dose escalation for prostate cancer radiotherapy: predictors of long-term biochemical tumour control and distant metastases-free survival outcomes. Eur Urol. 2011 Dec; 60(6): 1133-1139.

23. Dasu A, Toma-Dasu I. Prostate alpha/beta revisited -- an analysis of clinical results from 14 168 patients. Acta Oncol. 2012 Nov; 51(8): 963-974.

24. 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 Aug; 17(8): 1047-1060.

25. Royal College of Radiologists: Radiotherapy dose fractionation, third edition. [Internet]. 2019 {updated 2019; cited 2023 Feb 03]. Available from https://www.rcr.ac.uk/publication/radiotherapy-dose-fractionation-third-edition

26. Kishan A, Sun Y, Hartman H, et al. Androgen deprivation therapy use and duration with definitive radiotherapy for localised prostate cancer: an individual patient data meta-analysis [published correction appears in Lancet Oncol. 2022 Jul; 23(7):e319]. Lancet Oncol. 2022;23 (2): 304-316.

27. De Meerleer G, Berghen C, Briganti A, et al. Elective nodal radiotherapy in prostate cancer. Lancet Oncol. 2021; 22(8): e348-e357.

28. Beckendorf V, Guérif S, le Prisé E, Cosset J, Bougnoux A, Chauvet B, et al. (2011). 70 Gy versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial. *Int J Radiat Oncol Biol Phys. 2011 July; 80(4):* 1056-63

29. Jackson WC, Desai NB, Abugharib AE, et al. Anatomical patterns of recurrence following biochemical relapse after post-prostatectomy salvage radiation therapy: a multi-institutional study. BJU Int. 2017 Sep; 120(3): 351-357.

30. Lawal I, Lengana T, Popoola G, Orunmuyi T, Kgatle M, Mokoala K, et al. Pattern of prostate cancer recurrence assessed by ^{68}Ga-PSMA-11 PET/CT in men treated with primary local therapy. J Clin Med. 2021 Aug; 10(17): 3883.

31. Spratt D, Vargas H, Zumsteg Z, Pernicka J, Osborne J, Pei X, et al. Patterns of lymph node failure after dose-escalated radiotherapy: implications for extended pelvic lymph node coverage. Eur Urol. 2017 Jan; 71(1): 37-43.

32. Andreoiu M, Cheng L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. Hum Pathol. 2010 Jun; 41(6): 781-793.

33. Ahmed HU. The index lesion and the origin of prostate cancer. N Engl J Med. 2009 Oct; 361(17): 1704-1706.

34. Stamey T, McNeal J, Wise A, Clayton J. Secondary cancers in the prostate do not determine PSA biochemical failure in untreated men undergoing radical retropubic prostatectomy. Eur Urol. 2001; 39(suppl 4): 22-23

35. Liu W, Laitinen S, Khan S, Vihinen M, Kowalski, J, Guoqiang Y, et al. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer [published correction appears in Nat Med. 2009 Jul; 15(7):819]. Nat Med. 2009, May; 15(5): 559-565.

36. Cellini N, Morganti A, Mattiucci G, Valentini V, Leone M, Luzi S, et al. Analysis of intraprostatic failures in patients treated with hormonal therapy and radiotherapy: implications for conformal therapy planning. Int J Radiat Oncol Biol Phys. 2002 Jul; 53(3): 595-599.

37. Arrayeh E, Westphalen A, Kurhanewicz J, Roach M, Jung A, Carroll P, et al. Does local recurrence of prostate cancer after radiation therapy occur at the site of primary tumor? Results of a longitudinal MRI and MRSI study. Int J Radiat Oncol Biol Phys. 2012; 82(5): e787-e793.

38. Kara T, Akata D, Akyol F, Karcaaltincaba M, Ozmen, M. The value of dynamic contrast-enhanced MRI in the detection of recurrent prostate cancer after external beam radiotherapy: Correlation with transrectal ultrasound and pathological findings. Diagn Interv Radiol. 2009: 17(1); 10.4261/1305-3825.DIR.3079-09.1

39. Rouvière O, Valette O, Grivolat S, Colin_Pangaud C, Bouvier, R, Chapelon J, et al. Recurrent prostate cancer after external beam radiotherapy: value of contrast-enhanced dynamic MRI in localizing intraprostatic tumor--correlation with biopsy findings. Urology. 2004 May; 63(5): 922-927.

40. Turkbey B, Mani H, Aras O, Rastinehad A, Shah V, Bernardo M, et al. Correlation of magnetic resonance imaging tumor volume with histopathology. J Urol. 2012 Oct; 188 (4): 1157-1163.

41. Fang D, Zhao C, Ren D, Yu W, Wang, R, Wang H, et al. Could Magnetic Resonance Imaging Help to Identify the Presence of Prostate Cancer Before Initial Biopsy? The Development of Nomogram Predicting the Outcomes of Prostate Biopsy in the Chinese Population. Ann Surg Oncol. 2016 Dec; 23(13): 4284-4292.

42. Jie C, Rongbo L, Ping T. The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis. Eur Radiol. 2014 Aug; 24(8): 1929-1941.

43. Choi Y, Kim J, Kim N, Kim K, Choi E, Cho K. Functional MR imaging of prostate cancer. Radiographics. 2007 Jan-Feb; 27(1): 63-77.

44. Ghai S, Haider M. Multiparametric-MRI in diagnosis of prostate cancer. Indian J Urol. 2015 Jul-Sep; 31(3): 194-201.

45. Tan C, Wei W, Johnson V, Kundra V. Diffusion-weighted MRI in the detection of prostate cancer: meta-analysis. AJR Am J Roentgenol. 2012 Oct; 199(4): 822-829.

46. Verma S, Turkbey B, Muradyan N, Rajesh A, Cornud F, Haider M, et al. Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management. AJR Am J Roentgenol. 2012 Jun; 198(6): 1277-1288.

47. American College of Radiology: Prostate Imaging- Reporting and Data System Version 2.1. [Internet] 2019. [Updated 2019; cited 2023 Feb 03]. Available from: https://www.acr.org/-/media/ACR/Files/RADS/PI-RADS/PIRADS-V2-1.pdf

48. Lee J, Thomas I, Nolley R, Ferrari M, Brooks J, Leppert J. Biologic differences between peripheral and transition zone prostate cancer. Prostate. 2015 Feb; 75(2): 183-190.

49. Lee H, Hwang S, Lee H, Byun S, Lee S, Hong S. Diagnostic performance of diffusion-weighted imaging for prostate cancer: Peripheral zone versus transition zone. PLoS One. 2018 Jun; 13(6): e0199636.

50. Lewis S, Besa C, Rosen A, Rastinehead A, Semaan S, Hectors S, et al. Multiparametric magnetic resonance imaging for transition zone prostate cancer: essential findings, limitations, and future directions [published correction appears in Abdom Radiol (NY). 2017 Sep 22;:]. Abdom Radiol (NY). 2017; 42(11): 2732-2744

51. Akin O, Sala E, Moskowitz C, Kuroiwa K, Ishill N Pucar D et al. Transition zone prostate cancers: features, detection, localisation and staging at endorectal MR imaging. Radiol. 2006; 239(3): 784-792.

52. Schimmöller L, Quentin M, Arsov C, Hiester A, Buchcender C, Rabenalt R, et al. MR-sequences for prostate cancer diagnostics: validation based on the PI-RADS scoring system and targeted MR-guided in-bore biopsy. Eur Radiol. 2014 Oct; 24(10): 2582-2589.

53. Lebastchi A, Gill I, Abreu A. A Focus on Focal Therapy for Prostate Cancer. *JAMA Surg.* 2021 Jul 28; 156(9): 881–882.

54. Guillaumier S, Peters M, Arya M, Afzal, N, Charman S, Dudderidge T, et al. A Multicentre Study of 5-year Outcomes Following Focal Therapy in Treating Clinically Significant Nonmetastatic Prostate Cancer. Eur Urol. 2018 Oct; 74(4): 422-429.

55. Mearini L, Porena M. Pros and cons of focal therapy for localised prostate cancer. Prostate Cancer. 2011; 2011: 584784.

56. Haffner M, Mosbruger T, Esopi D, Fedor H, Heaphy C, Walker D, et al. Tracking the clonal origin of lethal prostate cancer. *J* Clin Invest. 2013; 123(11): 4918-4922.

57. Pickett B, Vigneault E, Kurhanewicz J, Verhey L, Roach M. Static field intensity modulation to treat a dominant intra-prostatic lesion to 90 Gy compared to seven field 3-dimensional radiotherapy. Int J Radiat Oncol Biol Phys. 1999 Jul; 44(4): 921-929.

58. Bauman G, Haider M, Van der Heide UA, Ménard C. Boosting imaging defined dominant prostatic tumors: a systematic review. Radiother Oncol. 2013 Jun; 107(3): 274-281.

59. Murray J, Tree A, Alexander E, Sohaib, A, Hazell, S, Thomas K, et al. Standard and Hypofractionated Dose Escalation to Intraprostatic Tumor Nodules in Localized Prostate Cancer: Efficacy and Toxicity in the DELINEATE Trial. Int J Radiat Oncol Biol Phys. 2020 Mar 15; 106(4): 715-724.

60. Tree A, Satchwell L, Alexander E, Blasiak_Wal I, deSouza N, Gao A, et al. Standard and Hypofractionated Dose Escalation to Intraprostatic Tumor Nodules in Localized Prostate Cancer: 5-Year Efficacy and Toxicity in the DELINEATE Trial. Int J Radiat Oncol Biol Phys. 2023; 115(2):305-316.

61. Kerkmeijer L, Groen V, Pos F, Haustermans K, Monninkhof E, Smeenk, R, et al. Focal Boost to the Intraprostatic Tumor in External Beam Radiotherapy for Patients With Localized Prostate Cancer: Results From the FLAME Randomized Phase III Trial. J Clin Oncol. 2021 Mar; 39(7): 787-796.

62. Syndikus I, Cruickshank C, Staffurth J, Tree A, Henry A, Naismith O, et al. PIVOTALboost: A phase III randomised controlled trial of prostate and pelvis versus prostate alone radiotherapy with or without prostate boost (CRUK/16/018). Clin Transl Radiat Oncol. 2020 Sep 1; 25: 22-28.

63. Institute of Cancer Research (ICR)(UK). PIVOTALboost: RT Planning and delivery guidelines. Version 1. ISRCTN80146950. 2017.

64. Institute of Cancer Research (ICR)(UK). PIVOTALboost: RT Planning and delivery guidelines. Version 1. ISRCTN80146950. 2017.

Chapter 2: Outlining Variation in Radiotherapy

2.1 Sources of error in the radiotherapy pathway

The radiotherapy planning pathway involves a number of steps from patient set up through to treatment delivery (Figure 2.1). Each step carries a potential source of error, which may in turn affect radiotherapy delivery to the patient. Potential issues include incorrect patient set up, errors in target volume delineation and organ/patient motion. In turn, these errors may lead to 'geographical miss' of the tumour i.e. potentially undertreating the intended target therefore increasing the risk of tumour recurrence as well as potentially over-irradiating normal tissue which may cause significant toxicity to the patient.

Several measures have been employed to minimise the risks of errors within the radiotherapy planning pathway. At the patient set up stage, immobilisation strategies are applied to minimise patient movement, stabilise the target position and allow reproducibility from the planning stage through to completion of treatment. With regards to prostate cancer, standard knee and ankle support have been shown to be as effective as more advanced immobilisation techniques in patient set up [1, 2].

Given the influence of rectal and bladder size on prostate position, patients undergoing prostate radiotherapy are usually required to follow bowel and bladder protocol for both the CT planning scan and during treatment. Maggio *et al* [3] found that prostate radiotherapy patients treated with a comfortably full bladder and empty rectum had significantly higher biochemical and clinical disease free survival as well as prostate cancer specific survival at ten years compared with patients who had no bowel or bladder preparation.
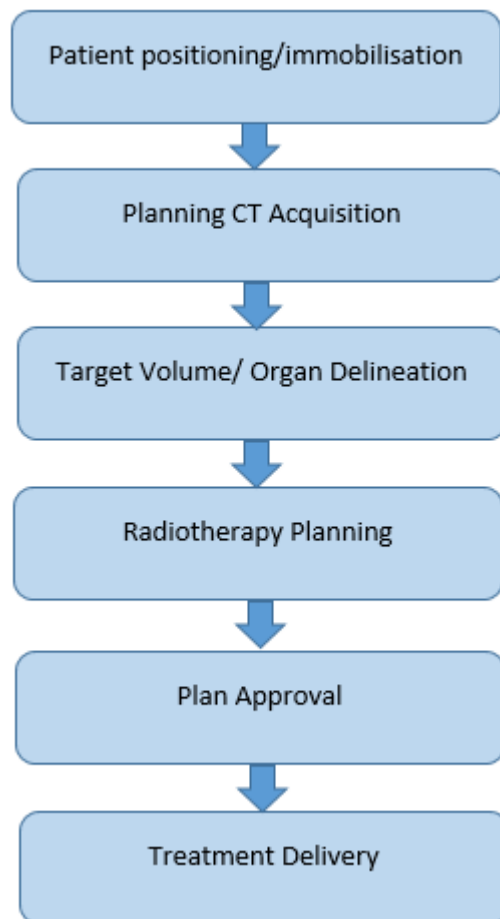
Figure 2.1: Radiotherapy planning pathway

In the era of precision radiotherapy, given the emergence of volumetric arc therapy and hypofractionation, the historic technique of using bony landmarks to assess prostate position during treatment is no longer adequate, due to the prostate's potential to displace more than 10mm in the posterior position [4]. The planning target volume (PTV) allows for organ motion/set up uncertainties. However the larger this margin, the greater the inclusion of normal tissue. Through acquisition of cone-beam CT (CBCT) cross sectional imaging based on soft tissue matching during treatment and/or implantation of prostate fiducial markers, PTV margins of 5-8mm may be used [5]. Furthermore, the use of real time tracking to correct for intrafraction movement during treatment may further reduce PTV margins to 2-3mm [6].

2.2 Target volume outlining variation

Although advances in image-guided RT (IGRT) have enabled greater precision of delivering radiotherapy, the 'weakest link' in the radiotherapy planning pathway relates to the accuracy of the target defined i.e. the target volumes delineated [7]. If the target volumes defined are inaccurate, the

measures described above to correct set up and positional errors during treatment will not compensate for this systematic error in the radiotherapy planning pathway.

2.2.1 Inter-observer variation in prostate radiotherapy outlining

The variability of target volume delineation (TVD) between observers is a major factor in geometric accuracy [8]. Cazzaniga *et al* [9] reviewed the PTVs of three prostate cancer cases delineated by six radiotherapists and found there was considerable variation between the volumes, with the differences between measured and mean volume for each case ranging from -53.64 to +60.48%. Similarly, Dubois *et al* [10] demonstrated a statistically significant difference in the CT-determined mean prostate volume outlined between two trained physicians of 41 prostate cases (-8.5cm³ +/- 9.74 SD).

Although the inter-observer variability (IOV) for prostate radiotherapy outlining may be less than in other tumour sites such as head and neck, cervix, lung and gastrointestinal [11], there are particular aspects within outlining for prostate radiotherapy that incur greater variation between outliners. Villiers *et al* [12] found that the highest level of variation between 13 prostate cases outlined by three radiation oncologists were at the prostate apex followed by the prostate base and seminal vesicles (SV). Another study of ten prostate patients outlined by seven observers showed that the correlation between prostate outlines was much higher than for SV [13].

A number of factors influence the variation amongst outliners including the imaging modality and technique used [8]. Although in their study Dubois *et al* [10] found a statistically significant difference between the two outliners' CT-derived mean prostate volumes, their MR guided prostate volumes were not significantly different. Villiers *et al* [12] also found that using MRI alongside CT compared to CT alone significantly decreased the mean CTV, prostate and SV outliners' volumes by 6.54%, 5.21% and 10.47% respectively. Indeed, in a study of 18 prostate cases outlined by three radiation oncologists, Rasch *et al* [14] concluded that CT-derived prostate volumes were larger than MR-derived volumes, particularly in the prostatic apex and SV and that inter-scan variation was greater than the IOV.

Bhardwaj *et al* [15] studied the estimated clinical impact of IOV in radiotherapy for prostate cancer using radiobiological models. Four observers outlined prostate and SV, bladder and rectum for nine patients. The difference in normal tissue complication probability (NTCP) for the rectum was statistically significant between observers, although not for bladder. There was also a significant difference between observers in the tumour control probability (TCP) of the prostate. Given the trend for CT-guided prostate outlining to produce larger target volumes than MR guided outlines [12, 14],

Steenbakkers *et al* [16] showed that when multiple observers outlined the prostate on 18 patients using CT and MRI separately, the dose to the rectal wall on the CT-outlined plans was on average 5.1Gy higher than on plans using the MR-outlined volumes. The mean dose to the penile bulb was also 11.6Gy lower using the MR-outlined volumes.

Potential dosimetric impact from IOV for prostate cancer is not limited to variation in delineation of the target volume i.e. prostate/SV. Inconsistencies in OAR outlining may lead to dosimetric uncertainties and potential over-irradiation of normal tissue. In a study of ten prostate patients' planning CTs with the penile bulb delineated by 15 observers, Perna *et al* [17] found that seven of the observers systematically over or under estimated the penile bulb volume with deviations from the average volumes ranging between -48% and +34%. This variation translated into a difference in mean dose to the penile bulb ranging between -20% and +20%.

2.2.2 Intra-observer variation in prostate radiotherapy outlining

The variation in delineation of radiotherapy structures by the same outliner i.e. intra-observer variation has also been noted in prostate radiotherapy although to a lesser degree. Gao *et a*l [18] recruited six radiation oncologists to outline the same prostate on CT 20 times, each outlining session occurring a minimum of three days apart. The intra-observer prostate volume variation ranged between 2 to 8% for the 6 observers, whereas the inter-observer variation in outlined volume was much greater ranging from -20% to +25% of the observers' mean volume. This finding was echoed in Fiorino *et al's* [19] study of five expert radiotherapists who contoured the prostate and SV of five patients; intra-observer variability of volumes delineated was on average 5% whereas the average inter-observer volume variation was much higher at 18%.

2.3 Strategies to optimise target volume delineation

A number of strategies have been used to optimise TVD accuracy and are discussed below.

2.3.1 Imaging modality

As discussed above, using MRI to aid outlining of radiotherapy target volumes alone or in combination with CT has been shown to improve IOV and reduce volume treated [10, 12, 14]. In standard clinical practice, radiotherapy treatment planning and dose calculation requires a planning CT, the images from which offer three-dimensional spatial information. From these images, attenuation values of different tissues can be determined (Hounsfield units) which in turn are converted to electron densities to calculate target volume and OAR doses [20]. As MRI offers more accurate soft tissue definition than CT, there has been a move towards an MRI-aided radiotherapy planning for prostate

cancer where patients have both a planning MRI and CT in the set-up/ treatment position. The two modalities are co-registered, enabling outlining of the target volumes on MRI and planning/dosimetric analysis on the CT component. However, Roberson *et al* [21] found that the minimum axial MR to axial CT registration error is approximately 2mm and limited by the data set resolution. Therefore, there is the risk that mis-registration of the CT and MRI will lead to a systematic error throughout treatment. The insertion of fiducial markers into the prostate may aid co-registration of CT and MR images as well as facilitate on-line correction of treatment set up and therefore reduce PTV margins that compensate for these errors [22]. Subsequent studies however, have shown that despite fiducial marker insertion, co-registration errors persist. Persson *et al* [22] recruited four experienced observers to register CT and MR images for 42 prostate cancer patients using gold fiducial markers (GFM) and found that the absolute difference in identification of GFM between observers was up to 3mm. Given the introduction of MR-guided adapted radiotherapy for prostate cancer patients using an MR treatment machine [23], the move towards an MR-only radiotherapy planning workflow is advocated [22], although the challenge of electron density transfer to MRI for dose calculation remains [20].

2.3.2 Implementation of outlining guidelines

Another approach to reduce IOV in radiotherapy outlining is the use of guidelines. In a review of interventions to reduce IOV, Vinod *et al* [24] found that in seven out of nine studies studying the effect of written guidance across tumour sites there was a statistically significant reduction in IOV with their use. There are currently no published studies on the direct effect of guideline implementation on IOV for in-situ prostate cancer radiotherapy. In a study of three post-prostatectomy patients having salvage radiotherapy to the prostate bed, Mitchell *et al* [25] recruited six radiation oncologists to outline the CTV using their routine clinical practice and again using the Radiotherapy and Androgen Deprivation In Combination After Local Surgery (RADICALS) trial contouring protocol [26]. Although following the protocol intervention IOV improved, the mean CTV increased from 40.6cm³ to 53.9cm³. This finding was echoed in Wiltshire *et al's* [27] work when seeking to derive consensus delineation guidelines for prostate bed radiotherapy. This group also created larger CTVs using guidelines they developed compared to those they had delineated in routine clinical practice, attributed to expanding the volume to include sites of surgical clips region that may encompass microscopic disease, suggesting outliners were actually under-estimating the extent of the prostate bed before guidelines were implemented.

There have been mixed outcomes from studies reviewing the impact of an anatomic atlas as part of outlining guidelines. Feng *et al* [28] found that heart and coronary vasculature delineations of seven radiation oncologists significantly improved after provision of an anatomical atlas. However, in a study

of rectal cancer outlining by 14 observers, in the group provided with an anatomical atlas, only one of two CTVs included in the protocol had improved IOV compared to the group without and there was no significant improvement in GTV delineation [29].

In 2018, the European Society for Therapeutic Radiotherapy and Oncology (ESTRO), published consensus guidelines on CT and MRI based target volume delineation for primary radiation therapy of localised prostate cancer [30]. These guidelines, using descriptors and atlas images, were novel in their approach to not only reduce IOV in prostate CTV delineation but to also guide outliners how to outline using CT and/ or MRI, although their impact on IOV has not been published.

There are however, some general pitfalls regarding outlining guidelines. In a survey of 85 radiation oncologists, participants correctly identified existence of outlining guidelines 42% of the time, with main barriers to guideline use including lack of awareness of publication, poor ease of use, poor comprehensibility and limited time [31]. In their systematic review of outlining guidelines in radiation oncology, Lin *et al* [32] found that inconsistent recommendations and a lack of structure regarding guideline content challenged their optimal use with no widely accepted standard for their development. To optimise the quality of consensus guidelines, they recommend involvement of a multidisciplinary panel including a radiologist, a literature review that includes patterns of recurrence and dissemination of a complete reference image set.

2.3.3 Educational interventions

A number of educational interventions have been used to reduce IOV in radiotherapy outlining. In their review of interventions to reduce outlining IOV, Vinod *et al* [24] found teaching interventions reduced IOV in eight out of nine studies reviewed, four of them statistically significant, although the heterogeneity of tumour sites and teaching methods used, makes it difficult to ascertain which is the most effective method.

One teaching method includes an interactive teaching program. Khoo *et al* [33] recruited five oncologists to contour three prostate cancer cases on CT then MRI. This process was repeated two to four weeks later. Observers were then given an hour interactive teaching session once a week for three weeks: two on MRI prostate anatomy and the third a practical session where they received real-time feedback on their contouring. To measure the impact on IOV, the observers completed the same three outlining exercises twice, each session two to four weeks apart. Both inter and intra-observer variation improved following the intervention, with a greater impact on MRI delineations compared to CT, which was attributed to observers' greater familiarity with the latter. To assess longevity of the impact on IOV, four of the five original observers re-outlined the same three cases twelve months

after the initial contouring session [34]. Although there was an improvement in prostate outlining consistency between baseline and twelve months, the improvement from baseline was not significant. A short educational intervention as above however, may not be the optimal teaching intervention for less experienced outliners. Schick *et al* [35] recruited six radiographers with varying radiotherapy outlining experience, to outline the prostate, bladder and rectum on three CT datasets on two occasions two weeks apart. They then each had a 'one to one' two hour interactive teaching session with a radiation oncologist on 'how to contour' the three organs. The six radiographers then repeated the contouring exercises on the same three cases on a further two occasions, two weeks apart. The results showed that improvements in IOV were minimal following the educational intervention and that general consistency with the observers between each other and the 'gold-standard' contours was poor. This study was undertaken in 2010, when the role of the non-medical outliner (NMO) was in its infancy although highlights that less experienced outliners may require a more prolonged, structured teaching program. This finding was supported by an educational intervention conducted by Szumacher *et al* [36]. Thirty-one inexperienced observers (combination of radiation oncology and radiation therapy trainees) were invited to outline the prostate and rectum before and after a teaching session. The participants were split into two groups. The experimental group had a 50-minute interactive lecture with expert outliners regarding prostate and rectum MRI anatomy and its correlation with CT followed by a 30-minute per-student hands-on practice session using MRI and CT. The control group had a 50-minute lecture about generic 3D-CT planning without focus on pelvic anatomy followed by a 30-minute individual practical session using CT only. Participants were then required to outline the same case as before the sessions. There was no significant difference in prostate or rectum outlining congruence between the two groups after the educational intervention However, following a post-workshop survey in the experimental group, 69% of trainees in the experimental group felt they could contour the rectum and prostate more precisely following the educational session compared to less than 10% of those in the control group. The teaching also appeared to motivate the participants, with 81% of the experimental group and 54% of the control group expressing a desire for more interactive teaching.

Interactive, face-to-face workshops however can be time and resource consuming, reaching only a relatively small number of learners at any given time. Since 2010, ESTRO have organised outlining workshops using the Fellowship in Anatomical delineation and CONtouring (FALCON) online radiotherapy contouring platform [37]. The sessions include blended face-to-face and online interactive tumour site- specific teaching sessions in which participants can learn delineation guideline and validate their delineation skills by comparing their outlines with other participants and experts. Similarly, the Royal College of Radiologists (RCR), responsible for clinical oncology training and

professional standards in the UK, hold in-person and online interactive site-specific contouring workshops throughout the year. Through adoption of an online learning platform, national and international access to these educational sessions are possible and provide a more equitable learning opportunity for both established and trainee outliners.

2.4 Radiotherapy trials quality assurance

Establishing radiotherapy treatments as standard of care requires robust design, analysis and reporting of clinical trials assessing their role [38]. It is therefore imperative that vigorous measures are implemented to minimise the impact of potential treatment variations on trial outcome and validate trial results. The European Study Group for Pancreatic Cancer 1 Trial (ESPAC-1) led a Randomised Control Trial (RCT) assessing the role of adjuvant treatment following surgical resection of pancreatic cancer. From their results, they concluded that adjuvant chemoradiotherapy had a deleterious effect on patient survival [39]. Koshy *et al* [40] however argued inferences regarding the role of adjuvant chemoradiotherapy could not be made from the trial given the heterogeneity of radiotherapy doses, archaic radiotherapy techniques and lack of dosimetry specifications in the trial.

Subsequently, many radiotherapy clinical trial groups have mandated radiotherapy quality assurance (RTQA) procedures to validate trial outcomes. Methods used in the RTQA process include participating facility questionnaires and beam output audits, 'dummy run' delineation and planning cases, individual patient case review during trial and complex dosimetry checks [41] although exact procedures have historically varied between trial groups. Given this procedural heterogeneity, in 2010 the Global Clinical Trials RTQA Harmonisation Group (GHG) was established to collate RTQA data from clinical trial groups, provide a platform for prospective RTQA discussion and develop a framework to endorse future and existing RTQA processes [42].

Although RTQA is necessary to validate radiotherapy trial outcomes, deviations from trial protocols may also have a significant clinical impact. In a systematic review of nine prospective radiotherapy trials depicting RTQA violation and impact on patient outcome, *Weber et al* [43] found that non-adherence to protocol-specified radiotherapy parameters is associated with both lower patient survival and local tumour control rates.

2.4.1 TVD in radiotherapy clinical trials

Given the extent of IOV in TVD, there is the risk within radiotherapy clinical trials that significant variations at the outlining stage are carried through the treatment pathway as a systematic error, although the clinical impact of TVD protocol deviations specifically has not been established [44]. To

minimise TVD within trials, TVD assessment of participating centres may be performed both before and during the trial.

2.4.2 Pre-trial TVD assessment

Gwynne *et al* [45] identified four steps in pre-trial outlining assessment:

1. Radiotherapy protocol and atlas development and workshops
2. Assessment of RT protocol adherence i.e. development of pre-accrual benchmark outlining exercise
3. Definition of reference volume(s) for the benchmark case(s)
4. Assessment of investigator outlines of the benchmark case(s)

Radiotherapy outlining and planning protocols form the backbone of RTQA, providing comprehensive guidelines for participating trial teams to follow. The merits of radiotherapy outlining protocols on reducing IOV are discussed above. Consensus from a number of experts in the development of radiotherapy protocols is more likely to reduce ambiguity and inconsistencies before the trial starts. To standardise contouring of upper gastrointestinal organs for Radiation Therapy Oncology Group (RTOG) led trials, twelve experts delineated on CT the oesophagus and oesophageal junction, stomach, common bile duct, liver and duodenum [46]. Their outlines were imported into MATLAB, a programming and numeric computing platform developed by MathWorks, which calculated the binomial distribution to generate 95% group consensus contours and reviewed by the panel. Consistency amongst the outlines had been high and an outlining protocol and atlas developed from this method.

Radiotherapy trial outlining protocols, not only serve purpose to streamline RTQA but can also shape clinical practice. From the triad of UK oesophageal 'SCOPE' trials group (SCOPE1, NeoSCOPE, SCOPE2), participating centres were sent a questionnaire to establish the role the trials played in their routine clinical practice [47]. Of the 27 centres that responded, 100% stated their local TVD protocols were based on the relevant SCOPE trial. Moreover, more advanced radiotherapy techniques (4DCT) were adopted by 71% of respondents following the NeoSCOPE and SCOPE2 trials due to provision of a comprehensive radiotherapy protocol, compared to 42% before.

As discussed above, not only can educational workshops improve IOV, but in the context of pre-trial RTQA, provides an opportunity for presentation of the outlining protocol, discussion regarding common site-specific TVD errors and protocol ambiguities to be addressed prior to the trial [45]. The trials management group (TMG) for ARISTOTLE, a phase III UK trial of chemoradiotherapy for rectal cancer, held seven workshops across five radiotherapy centres prior to the trial to minimise TVD IOV

during the trial [48]. Furthermore, the PersonaLising Anal cancer radioTherapy dOse (PLATO) trial used their pre-trial workshop to evaluate contouring variation identified in the pre-trial benchmark case for educational purposes and to refine their RT trial protocol accordingly [49].

The pre-trial benchmark case is a 'dummy run' for participating centres to outline and plan a case using the trial protocol prior to recruiting any 'real' patients. The selected case is usually representative of a 'typical' patient on trial, albeit with one or more critical features to assess protocol compliance [50]. Although there will be a slight variation between trials, potential investigators are typically sent the trial contouring and planning protocol, patient clinical vignette and a CT data set for contouring and planning. In the 22991 European Organisation for the Research and Treatment of Cancer (EORTC) Radiation Oncology Group (ROG) trial in localised prostate cancer 'dummy run', anonymised DICOM images of a planning CT scan were sent to participating centres with a clinical description of the case and trial protocol [51]. Investigators were asked to outline the prostate CTV, PTV and OARs as stipulated in the protocol and return the volumetric data of the target volumes. Centres successful at first dummy run attempt in this trial, were significantly more likely to deliver protocol-compliant RT on trial [50].

To assess an investigator's TVD performance, submitted volumes are usually compared against a pre-determined 'gold-standard' or 'reference volume'. Any discordance between the investigator and reference volume can be identified and its significance determined i.e. is it minor enough to 'pass' the investigator or will they be required to re-submit another attempt. Any consistent discordance between investigators and the reference volume provides the RTQA team opportunity to review their protocol for any misunderstandings or ambiguities and identify common TVD errors that may be the focus of discussion at associated pre-trial contouring workshops. [45]

The reference volume is usually determined by a panel of experts, often the TMG, and may be determined manually i.e. the panel agreeing a consensus contour together [45, 52]. A Simultaneous Truth and Performance Level Estimate (STAPLE) contour [53] may also be determined from an algorithm that creates a single contour from multiple (expert) contours using statistical methods, which can be modified if required, to create a consensus volume. The SCOPE1 RTQA trial data showed that by using a TMG STAPLE derived contour, significantly more investigators achieved 'excellent' conformity to the reference volume compared to using a single clinician/radiologist defined volume [54].

Participation in a pre-trial 'dummy run' can have positive effects on future pre-trial dummy run performances. In a review of two decades of pre-trial RTQA data from the EORTC ROG, radiotherapy centres which had previously participated in a pre-trial outlining exercises were significantly more

likely to pass future trial pre-accrual benchmark case attempts first time. [50]. Subsequently, some trials have 'streamlined' the RTQA pre-trial process; if a centre has passed a pre-trial RTQA program for a previous trial with similar outlining requirements, they are eligible to participate in the new trial without submitting another pre-trial exercise. The SCOPE TMG took this approach for SCOPE2 for participating centres that had already undertaken NeoSCOPE pre-trial RTQA [47].

2.5 Methods to measure TVD conformity

It is impossible to eliminate outliner inter- and intra-observer variation fully. Therefore, parameters to assess TVD conformity are required to determine whether volumes are 'acceptable' in the case of RTQA or, for example, improved in the context of assessing the impact of an intervention. Both qualitative and quantitative measures have been used to assess TVD conformity and are described below.

2.5.1 Qualitative assessment of TVD conformity

In the context of RTQA, acceptability of outline assessment maybe required in both the pre-trial and on-trial settings. In the pre-trial setting, techniques to assess the investigators' performance in this exercise include visual inspection of outlines either by an expert(s) against a reference volume using visual measures such as inappropriate inclusion of a particular structure or the length of investigator's volume [55].

During the trial however, there is no reference volume to assess TVD conformity. Therefore, radiotherapy trials are increasingly defining acceptable and unacceptable outlining protocol deviations that can assess participants' conformity to TVD protocol both before and during the trial. NeoSCOPE, a phase II RCT of two oesophageal cancer neoadjuvant chemoradiotherapy regimens, undertook a comprehensive RTQA programme with both pre-accrual and prospective on-trial outlining quality assurance (QA) of every patient. Each participating centre was provided with detailed feedback regarding their outlines against pre-determined criteria for their two pre-trial test cases against a consensus reference volume. Subsequently, all patients recruited to the trial had their outlines reviewed by the TMG, against the outlining acceptability criteria. Those with unacceptable deviations were required to re-submit their outlines following detailed feedback and no delays to the patients' treatment were encountered due to the QA process. [56]

2.5.2 Quantitative assessment of TVD conformity

Qualitative measures of TVD conformity that rely upon visual assessment however, can be prone to errors, subjection, bias (particularly in RTQA if the chief investigator the assessor, for example) and

are time-consuming [55]. Therefore, a number of quantitative metrics have been used to assess inter- and intra-observer variation and conformity to a reference volume.

Simple quantitative measures include the measured volume of the delineated structure. In pre-trial RTQA, an acceptable range of volumes against the reference may be stipulated e.g. investigator volumes should be within +/- 25% of the reference volume. To compare intra- or inter-observer variation of TVD volumes, the distribution of the volumes may be presented as an assessment of investigator performance against the reference volume, indicating whether investigator volumes may be too generous or too tight. Another simple quantitative metric includes superior and inferior extent of the outlined structure.

2.5.3 Conformity indices

Although simple quantitative measurements have commonly been used to describe TVD variation in the literature [56], this does not provide any spatial conformity of a structure. Conformity Indices on the other hand, are quantitative measurements of the common volume included between two volumes or comparison of a consensus volume with each of its constituent volumes [56, 57]. The spatial relationship between two volumes, i.e. investigator outline and reference volume outline can be analysed according to the conformity index used.

The most commonly used conformity index metrics to assess investigator volume conformity to a reference volume are those that evaluate the ratio of the volume of overlap of two structures over union volume of the two structures, known as the concordance index [45]. These include van't Riet Index, Jaccard Conformity Index (JCI) and DICE Coefficient (DICE). JCI is the ratio of intersection of two volumes compared to the union of the two comparative volumes (Figure 2.1)

AUB
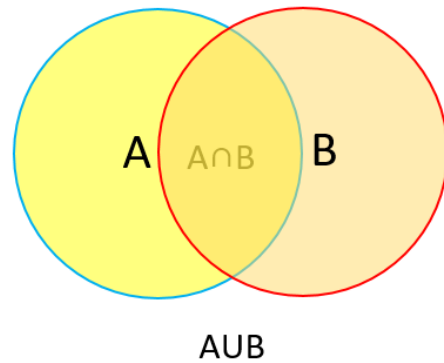
Figure 2.1: Pictorial representation of overlap of investigator and reference volumes that form the basis of overlap conformity indices

$$\text{JCI (Concordance Index)} \quad = \quad \frac{A \cap B}{A \cup B}$$

DICE, will give slightly different values to JCI for the same volumes compared and tend to be slightly higher than JCI except for relatively small volumes entirely included in a larger volume [56]. The DICE similarity coefficient is defined as:

$$\text{DICE Coefficient} \quad = \quad \frac{2(A \cap B)}{A + B}$$

Similarly, the van't Riet formula [58] is defined as:

$$\text{Van't Riet Index} \quad = \quad \frac{A \cap B}{A} \times \frac{A \cap B}{B}$$

For JCI, DICE and van't Riet, perfect concordance is represented by a value of 1 and complete discordance a value of 0. The advantage of these conformity indices are their use of a single metric to describe investigator delineation performance and they are widely used in the literature [45]. This has led to benchmark levels for poor concordance to be described e.g. a JCI of <0.5 in breast cancer radiotherapy delineation [59]. Disadvantages include the whole-volume metrics potentially missing

areas of variation with the volume, poor correlation with length and failure to detect small but potentially clinically significant errors such as inclusion of an OAR in the volume [45].

Some conformity indices more specifically relate to either over- or under- outlining of one volume compared to another. The Discordance Index (DI) [45] is a metric that describes how 'over-outlined' a volume is (compared to another) and is defined as:

$$\text{Discordance Index (DI)} = 1 - \frac{(A \cap B)}{A}$$

Conversely, the Geographical Miss Index (GMI) [45] describes the extent of 'under-outlining' and is defined as:

$$\text{Geographical Miss Index} = \frac{B - (A \cap B)}{B}$$

For both DI and GMI, perfect concordance is represented by a value 0 and complete discordance 1 (converse to van't Riet's, JCI and DICE).

The above conformity indices are all measures of overlap but other groups of evaluation metrics have been described including metrics which are statistical measures of agreement: an example are the kappa (κ) statistics [60]. Fleiss' kappa is a measure of magnitude of agreement between multiple outlines whereas Cohen's kappa is a measurement of magnitude of agreement between two outlines [45]. For the former, no reference volume is required for comparison whereas for Cohen's kappa, a reference volume may be one of the outlines used as comparison. For both κ statistics, the value of perfect concordance would be 1.

Given the limitations of overlap conformity indices and measures of agreement not providing information regarding the differences in shape between two volumes, Jena *et al* [59] devised a morphometric statistic known as the 'mean distance to conformity' (MDC) as a quantitative measure of target volume delineation conformity. For an investigator volume under comparison against a reference volume, MDC represents, in mm, the average distance that all outlying points in the investigator volume must be moved in order to achieve perfect concordance with the reference volume. The MDC value is a single scoring statistic that is representative of the overall conformity of the two volumes being compared but is usually presented with additional statistics that provide

information on whether non-conformity is due to over- (MDC over-contouring) or under-contouring (MDC under-contouring). [59]

A summary of quantitative metrics and their respective advantages and disadvantages are shown in Table 2.1 [45].

Metric	Description of metric	Value of perfect concordance	Advantage	Disadvantage
JCI/DICE/Van't Riet	Ratio of the volume of overlap of two structures over union volume of the two structure	1	• Widely used in the literature across tumour sites • Benchmark levels defined for poor concordance	• Whole-volume metric may miss areas of variation within the volume • Concordance increases with larger volumes • Correlates poorly with length • Failure to detect small but potentially clinically significant errors e.g. inclusion of OARs in volume • No information on the direction of error
DI	Calculates the amount of over-outlining	0	• Well correlated with volume • Calculates the amount of over-outlining	• No benchmark for comparison, tumour site- and case-dependent
GMI	Calculates the amount of under-outlining	0	• Well correlated with volume • Calculates the amount of under-outlining	• No benchmark for comparison, tumour site- and case-dependent

Metric	Description of metric	Value of perfect concordance	Advantage	Disadvantage
K statistic (Fleiss)	Measurement of magnitude of agreement between multiple volumes	1	• No reference volume required for calculation • Objective benchmark values to assess agreement	• Only valid for multiple investigator volumes • Value dependent on investigators and not a reference volume • Decision required regarding acceptable level of agreement • No information on the direction of error
K statistic (Cohen)	Measure of magnitude of agreement between two outlines	1	• Can be used to compare two volumes e.g. the investigator and reference volume • Objective benchmark values to assess agreement	• No benchmark for comparison, tumour site- and case-dependent • Decision required regarding acceptable level of agreement • No information on the direction of error

Metric	Description of metric	Value of perfect concordance	Advantage	Disadvantage
MDC	Shape-based statistic that measures the mean displacement needed to transpose every voxel of the investigator volume onto the reference volume	0mm	• Gives measurements of variation (in mm) • Has an over- and under-outlining component • Independent of size and volumes under comparison	• No benchmark for comparison, tumour site- and case-dependent • Use of under- and over-outlining MDC results in two metrics offsetting the advantage of a single metric • No information on the direction of error • Correlates poorly with length and volume

Table 2.1: Advantages and disadvantages of quantitative metrics to describe outlining variation and their respective values of perfect concordance [45]

2.6 RTQA for PIVOTALboost

To minimise the impact of radiotherapy outlining and planning on trial outcome, the national Radiotherapy Trials Quality Assurance (RTTQA) group devised a quality assurance programme for PIVOTALboost. Expert clinical oncologists on the TMG led this process, alongside physicists, radiographers and HDR brachytherapists. Centres that had satisfactorily completed RTQA in PIVOTAL, an earlier phase II prostate and pelvic nodal IMRT trial in the UK [61], were eligible for a more streamlined QA process. All centres wishing to offer a boost volume however, were required to complete the DIL outlining aspect of RTQA given the relatively new concept of DIL boosting.

PIVOTALboost RTQA is comprised of two main components: a pre-accrual component and on-treatment component. The pre-accrual component included development of a radiotherapy delivery and planning protocol, and both pelvic nodal and boost contouring guidelines. Centres wishing to participate within the trial had to complete a pre- accrual survey, benchmark outlining and planning cases, attend an outlining workshop or webinar and complete a facility questionnaire.

The on-treatment RTQA component includes prospective and/or retrospective case reviews, review of HDR implant parameters, collection of staging MRI imaging and a potential dosimetry site visit (subject to previous RTQA dosimetry accreditation).

2.6.1 Pre-trial RTQA: protocol

The trial's Chief Investigator (IS) and members of the TMG (JS, AT, AH) developed the PIVOTALboost RT guidance with physicist input. The guidance is comprised of three main documents: PIVOTALboost RT planning and delivery guidelines [62], PIVOTALboost boost contouring atlas [63] and PIVOTALboost pelvic node atlas [64].

The planning and delivery guidelines detail the trial schema and treatment arms, pre-RT procedures, planning scan requirements, TVD and OAR definitions, EBRT planning guidelines, HDRB guidelines, on treatment verification methods, treatment scheduling and RTQA processes.

2.6.2 Pre-trial RTQA: pre-accrual benchmark cases

Each clinician (investigator) wishing to participate in PIVOTALboost was required to outline two anonymised cases: case 1 and case 2. Investigators in centres unable to offer a boost were required to outline case 2 only. For each centre wishing to participate in PIVOTALboost, the respective physics department were also required to plan one case for evaluation. Potential investigators participating in the trial were expected to have completed specialist training in clinical oncology and sub-specialise

at consultant level in prostate radiotherapy. The pre-trial outlining exercise will be discussed in more detail in Chapter 4.

2.6.3 On-trial RTQA

To minimise the risk that variations or errors identified during the pre-accrual process would continue into the on-treatment phase, centres were expected to submit each recruited patient's completed radiotherapy plan and accompanying set of TVDs along with the diagnostic mpMRI and planning MRI if performed. For review of outlining, the first two cases with an IMRT boost volume required prospective review i.e. outlines and plans reviewed by the trial's team prior to the patient starting radiotherapy. For nodal outlining, the first case could be reviewed retrospectively i.e. once the patient has started treatment. Unacceptable outlining variations were fed back to the submitting centre with recommended changes to be implemented by the investigator. Once the RTQA group were satisfied with the standard of outlining across all arms, no further on-treatment reviews for that centre were required.

2.7 Thesis aims

Radiotherapy outlining variation is an area within radiotherapy planning that is of particular interest to me. From 2017-2020, I was the Clinical Fellow on the ARENA (Assurances in Radiotherapy through Education and Assessment) project, aimed to standardise high-quality TVD training for clinical oncology trainees through development of educational packages based on principles of RTQA [65]. Part of this project was development of FIELD[RT], an open-source software that offers qualitative and quantitative feedback to outliner submitted volumes [66]. For the basis of this thesis, I wanted to study outlining variation within a clinical context and given my particular interest in prostate radiotherapy, was keen to explore outlining variation of the DIL, given this is a relatively new technique within the field.

This thesis aims to explore the extent of DIL outlining variation identified amongst UK prostate oncologists using pre-trial data from PIVOTALboost RTQA program. Following on from this, I shall explore the role of quantitative metrics to measure DIL outlining variation and methods to improve DIL outlining consistency.

To achieve this, in Chapter 3, I will review the feedback proformas completed by the PIVOTALboost Chief Investigator (IS) of investigator submissions of the pre-accrual benchmark cases to establish whether a high rate of resubmission was required. If attributed to unacceptable DIL outlining variation, using the completed Chief Investigator proformas, I will aim to determine what the causes of unacceptable DIL outlining are.

Following on from the work in Chapter 3, in Chapter 4, I will use all available PIVOTALboost pre-accrual DIL volume submissions to explore whether there is a particular conformity index (or indices) and associated threshold that can be used as a semi-automated tool to identify acceptable DIL outlines.

Chapter 5 will then focus on the impact on inter-observer TVD performance following a national outlining workshop for UK clinical oncologists through assessment of DIL outlining both before, during and one month after the DIL focused educational workshop.

Finally, in Chapter 6, I will review DIL volume submissions of a case outlined by a group of local prostate outliners (medical and non-medical outliners) to establish which step within the DIL outlining process most variation occurs i.e. mpMRI interpretation and/or cognitive transfer. The local outliners will be required to delineate the same case on the planning CT using cognitive transfer as per PIVOTALboost, then again having delineated the case on the diagnostic mpMRI T2W and ADC sequences and finally re-outline on the planning CT having been provided with pre-outlined mpMRI T2W images.

2.8 References

1. Nutting C, Khoo V, Walker V, McNair H, Beardmore C, Norman, A, Dearnaley D. A randomized study of the use of a customized immobilization system in the treatment of prostate cancer with conformal radiotherapy. Radiother Oncol. 2000 Jan; 54(1): 1-9.

2. Rosewall T, Chung P, Bayley A, Lockwood G, Alasti H, Bristow, R, et al. A randomized comparison of interfraction and intrafraction prostate motion with and without abdominal compression. Radiother Oncol. 2008 Jul; 88(1): 88–94.

3. Maggio A, Gabriele D, Garibaldi E, Bresciani S, Delmastro E, Di Dia A, et al. Impact of a rectal and bladder preparation protocol on prostate cancer outcome in patients treated with external beam radiotherapy. Strahlenther Onkol. 2017 Sep; 193(9): 722-732.

4. Crook JM, Raymond Y, Salhani D, Yang H, Esche B. Prostate motion during standard radiotherapy as assessed by fiducial markers. Radiother Oncol. 1995 Oct; 37(1): 35-42.

5. Yartsev S, Bauman G. Target margins in radiotherapy of prostate cancer. Br J Radiol. 2016 Nov; 89(1067): 20160312.

6. Litzenberg D, Balter J, Hadley S, Hamstra D, Willoughby T, Kupelian P, et al. Prostate intrafraction translation margins for real-time monitoring and correction strategies. Prostate Cancer. 2012; 2012:130579.

7. Njeh C. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys. 2008 Oct; 33(4): 136-40.

8. Weiss E, Hess CF. The impact of gross tumour volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. Strahlenther Onkol. 2003 Jan; 179(1) : 21-30.

9. Cazzaniga L, Marinoni A, Bossi A, Bianchi E, Cagna E, Cosentino D, et al. Interphysician variability in defining the planning target volume in the irradiation of prostate and seminal vesicles. Radiother Oncol. 1998 Jun; 47(3): 293-6.

10. Dubois D, Prestidge B, Hotchkiss L, Prete J, Bice W Jr. Intraobserver and interobserver variability of MR imaging- and CT-derived prostate volumes after transperineal interstitial permanent prostate brachytherapy. Radiology. 1998 Jun; 207(3): 785-9.

11. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? Radiol Oncol. 2016 May 9; 50(3): 254-62.

12. Villeirs G, Van Vaerenbergh K, Vakaet L, Bral S, Claus F, De Neve W. Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer. Strahlenther Onkol. 2005 Jul; 181(7): 424-30.

13. Valicenti R, Sweet J, Hauck W, Hudes R, Lee T, Dicker A, et al. Variation of clinical target volume definition in three-dimensional conformal radiation therapy for prostate cancer. Int J Radiat Oncol Biol Phys. 1999 Jul 1; 44(4): 931-5.

14. Rasch C, Barillot I, Remeijer P, Touw A, van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. Int J Radiat Oncol Biol Phys. 1999 Jan 1; 43(1): 57-66.

15. Bhardwaj A, Kehwar T, Chakarvarti S, Sastri G, Oinam A, Pradeep G, et al. Variations in inter-observer contouring and its impact on dosimetric and radiobiological parameters for intensity-modulated radiotherapy planning in treatment of localised prostate cancer. J Radiother Pract. 2008; 7(2):77–88.

16. Steenbakkers R, Deurloo K, Nowak P, Lebesque J, van Herk , Rasch CR. Reduction of dose delivered to the rectum and bulb of the penis using MRI delineation for radiotherapy of the prostate. Int J Radiat Oncol Biol Phys. 2003 Dec 1; 57(5): 1269-79.

17. Perna L, Cozzarini C, Maggiulli E, Fellin G, Rancati T, Valdagni R, et al. Inter-observer variability in contouring the penile bulb on CT images for prostate cancer treatment planning. Radiat Oncol. 2011 Sep 24; 6: 123.

18. Gao Z, Wilkins D, Eapen L, Morash C, Wassef Y, Gerig L. A study of prostate delineation referenced against a gold standard created from the visible human data. Radiother Oncol. 2007 Nov; 85(2): 239-46.

19. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. Radiother Oncol. 1998 Jun; 47(3): 285-92.

20. Chandarana H, Wang H, Tijssen R, Das I. Emerging role of MRI in radiation therapy. J Magn Reson Imaging. 2018 Dec; 48(6): 1468-1478.

21. Roberson P, McLaughlin P, Narayana V, Troyer S, Hixson G, Kessler M. Use and uncertainties of mutual information for computed tomography/ magnetic resonance (CT/MR) registration post permanent implant of the prostate. Med Phys. 2005 Feb; 32(2): 473-82.

22. Parker CC, Damyanovich A, Haycocks T, Haider M, Bayley A, Catton CN. Magnetic resonance imaging in the radiation treatment planning of localized prostate cancer using intra-prostatic fiducial markers for computed tomography co-registration. Radiother Oncol. 2003 Feb; 66(2): 217-24.

23. Dunlop A, Mitchell A, Tree A, Barnes H, Bower L, Chick J, et al. Daily adaptive radiotherapy for patients with prostate cancer using a high field MR-linac: Initial clinical experiences and assessment of delivered doses compared to a C-arm linac. Clin Transl Radiat Oncol. 2020 Apr 29; 23: 35-42.

24. Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol. 2016 Jun; 60(3): 393-406.

25. Mitchell D, Perry L, Smith S, Elliott T, Wylie J, Cowan R, et al. Assessing the effect of a contouring protocol on post prostatectomy radiotherapy clinical target volumes and interphysician variation. Int J Radiat Oncol Biol Phys. 2009 Nov 15; 75(4): 990-3.

26. National Cancer Research Institute (NCRI)(UK). RADICALS: Radiotherapy and Androgen Deprivation In Combination After Local Surgery Protocol. Version 4.0. ISRCTN40814031. 2011.

27. Wiltshire K, Brock K, Haider M, Zwahlen D, Kong V, Chan E, et al. Anatomic boundaries of the clinical target volume (prostate bed) after radical prostatectomy. Int J Radiat Oncol Biol Phys. 2007 Nov 15; 69(4): 1090-9.

28. Feng M, Moran J, Koelling T, Chughtai A, Chan J, Freedman L, et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. Int J Radiat Oncol Biol Phys. 2011 Jan 1; 79(1): 10-8.

29. Fuller C, Nijkamp J, Duppen J, Rasch C, Thomas C Jr, Wang S, et al: Radiation Oncology Committee of the Southwest Oncology Group. Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. Int J Radiat Oncol Biol Phys. 2011 Feb 1; 79(2): 481-9.

30. Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiother Oncol. 2018 Apr; 127(1): 49-61.

31. Sherer M, Bryant A, Wu A, Barry P, Lally B, Yashar C, et al. Assessment of contouring resource use and awareness of contouring guidelines among radiation oncologists. J Radiat Oncol. 2018 Feb 8; 7: 103–109.

32. Lin D, Lapen K, Sherer M, Kantor J, Zhang Z, Boyce L, et al. A Systematic Review of Contouring Guidelines in Radiation Oncology: Analysis of Frequency, Methodology, and Delivery of Consensus Recommendations. Int J Radiat Oncol Biol Phys. 2020 Jul 15; 107(4): 827-835.

33. Khoo E, Schick K, Plank A, Poulsen M, Wong W, Middleton, et al. Prostate contouring variation: can it be fixed? Int J Radiat Oncol Biol Phys. 2012 Apr 1; 82(5): 1923-9.

34. Nicholls L, Gorayski P, Poulsen M, Plank A, Schick K et al. Maintaining prostate contouring consistency following an educational intervention. J Med Radiat Sci. 2016 Sep; 63(3): 155-60.

35. Schick K, Sisson T, Frantzis J, Khoo E, Middleton M. An assessment of OAR delineation by the radiation therapist. Radiography. 2011; 17(3): 183-187.

36. Szumacher E, Harnett N, Warner S, Kelly V, Danjoux C, Barker R, et al. Effectiveness of educational intervention on the congruence of prostate and rectal contouring as compared with a gold standard in three-dimensional radiotherapy for prostate. Int J Radiat Oncol Biol Phys. 2010 Feb 1; 76(2): 379-85.

37. European Society of Radiotherapy and Oncology. Fellowship in Anatomic delineation and CONtouring (FALCON) [Internet] 2019 [cited 2023 Feb 03]. Available from https://estro.educase.com

38. Bentzen S. Towards evidence based radiation oncology: improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. Radiother Oncol. 1998 Jan; 46(1): 15-18.

39. Neoptolemos J, Stocken D, Friess H, Bassi C, Dunn J, Hickey H et al; European Study Group for Pancreatic Cancer. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. N Engl J Med. 2004 Mar 18; 350(12): 1200-10.

40. Koshy M, Landry J, Cavanaugh S, Fuller C, Willett C, Abrams R, et al. A challenge to the therapeutic nihilism of ESPAC-1. Int J Radiat Oncol Biol Phys. 2005 Mar 15; 61(4): 965-6.

41. European Organisation for Research and Treatment of Cancer: Radiotherapy Quality Assurance [Internet][2023 Feb 03]. Available from: https://www.eortc.org/quality-assurance/rtqa

42. Melidis C, Bosch W, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials--Global Harmonisation Group. Radiother Oncol. 2014 Jun; 111(3) :327-9.

43. Weber D, Tomsej M, Melidis C, Hurkmans C. QA makes a clinical trial stronger: evidence-based medicine in radiation therapy. Radiother Oncol. 2012 Oct; 105(1): 4-8.

44. Cox S, Cleves A, Clementel E, Miles E, Staffurth J, Gwynne S. Impact of deviations in target volume delineation - Time for a new RTQA approach? Radiother Oncol. 2019 Aug; 137: 1-8

45. Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, Conibear J, Staffurth J; NCRI RTTQA OUTLINING AND IMAGING SUBGROUP. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. Br Radiol. 2013 Apr; 86: 20120398.

46. Jabbour S, Hashem S, Bosch W, Kim T, Finkelstein S, Anderson B, et al. Upper abdominal normal organ contouring guidelines and atlas: A Radiation Therapy Oncology Group consensus. Pract Radiat Oncol. 2014 Mar-Apr; 4(2): 82-89.

47. Gwynne S, Higgins E, Poon King A, Radhakrishna G, Wills L, Mukherjee S, et al. Driving developments in UK oesophageal radiotherapy through the SCOPE trials. Radiat Oncol. 2019 Feb 4; 14(1): 26.

48. Glynne-Jones R, Harrison M, Gollins S, Harte R, Maggs R, Ghuman S et al. PO-0792 The development of a conformal radiotherapy protocol for the phase III ARISTOTLE rectal cancer trial. Radiother Oncol. 2012 May; 103.

49. Robinson M, Christophides D, Abbott N, Bell S, Langley A, Harrison M, et al. A novel pretrial contouring exercise for the UK led PersonaLising Anal cancer radioTherapy dOse (PLATO) trial. Int J Radiat Oncol Biol Phys. 2017 Oct; 99(2): E181-E182.

50. Fairchild A, Collette L, Hurkmans CW, Baumert B, Weber DC, Gulyban A, Poortmans P. Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials? Eur J Cancer. 2012 Nov; 48(17): 3232-9.

51. Matzinger O, Poortmans, Giraud J, Maingon, P, Budiharto T, van den Bergh A, et al. Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: Dummy run and individual case review. Radiother Oncol. 2009 Mar; 90(3): 285-290.

52. Chang A, Tan LT, Duke S, Ng W. Challenges for Quality Assurance of Target Volume Delineation in Clinical Trials. Front Oncol. 2017 Sep 25; 7: 221.

53. Warfield S, Zou K, Wells W. Similtaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004 Jul; 23(7): 903-921

54. Gwynne S, Spezi S, Hurt C, Falk S, Gollins S, Joseph G, et al. Importance of the reference volume in assessing outlining performance for the purpose of training and revalidation. NCRI Cancer Conference. Liverpool, UK: NCRI, 2012.

55. Wills L, Maggs R, Lewis G, Jones G, Nixon L, Staffurth J, et al; SCOPE 1 trial management group and collaborators. Quality assurance of the SCOPE 1 trial in oesophageal radiotherapy. Radiat Oncol. 2017 Nov 15; 12(1): 179.

56. Hanna G, Hounsell A, O'Sullivan J. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin Oncol (R Coll Radiol). 2010 Sep; 22(7): 515-525.

57. Gwynne S, Spezi E, Wills L, Nixon L, Hurt C, Joseph G, et al. Toward semi-automated assessment of target volume delineation in radiotherapy trials: the SCOPE 1 pretrial test case. Int J Radiat Oncol Biol Phys. 2012 Nov 15; 84(4): 1037-42.

58. van't Riet A, Mak A, Moerland M, Elders L, van der Zee W. A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate. Int J Radiat Oncol Biol Phys. 1997 Feb 1; 37(3): 731-6.

59. Jena R, Kirkby NF, Burton KE, Hoole AC, Tan LT, Burnet NG. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. Br J Radiol. 2010 Jan; 83(985): 44-51.

60. Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol. 2012 Feb; 188(2): 160-7.

61. Institute of Cancer Research (ICR) (UK): A randomised phase II trial of Prostate and pelvIs Versus prOsTate Alone treatment for Locally advanced prostate cancer (PIVOTAL) (ISRCTN48709247). [Internet] 2011.[cited 2023 Feb 03]. Available from: https://www.isrctn.com/ISRCTN48709247

62. Institute of Cancer Research (ICR)(UK). PIVOTALboost: RT Planning and delivery guidelines. Version 1. ISRCTN80146950. 2017.

63. Institute of Cancer Research (ICR)(UK). PIVOTALboost: Trial Boost Contouring Instructions & Atlas; Version 1. ISRCTN80146950. 2017.

64. Institute of Cancer Research (ICR)(UK). PIVOTALboost: Pelvic node contouring Atlas; Version 1. ISRCTN80146950. 2017.

65. Evans, E, Radhakrishna, G, Gilson, D, Hoskin P, Miles E, Yuille F, et al. Target volume delineation training for clinical oncology trainees: the Role of ARENA and COPP. Clin Oncol (R Coll Radiol) 2019 Jun; 31(6): 341–343.

66. Piazzese C, Evans E, Thomas B, Staffurth J, Gwynne S, Spezi E. FIELD[RT]: an open-source platform for the assessment of target volume delineation in radiation therapy. Br J Radiol. 2021 Oct 1; 94: 20210356.

Chapter 3: Qualitative Assessment of PIVOTALboost Pre-trial Case Submissions to Evaluate Presence of DIL Boost Outlining Variation

3.1 Introduction

As discussed in chapters one and two, PIVOTALboost is a multicentre randomized control phase III clinical trial, which aims to evaluate the role of dose-escalation to the DIL in prostate radiotherapy with or without pelvic radiotherapy versus standard hypofractionated radiotherapy (with or without pelvic radiotherapy) [1]. As part of the pre-trial RTQA programme, any potential investigators wishing to recruit patients to the boost arms of the trial were required to participate in a pre-accrual benchmark outlining exercise of two cases. Those centres not offering a boost, were required to participate in case two only, unless they had already satisfactorily completed the pre-trial delineation exercise for the PIVOTAL trial [2].

Case 1 included an intermediate-risk prostate cancer with a right central zone PIRADS-5 DIL. The investigator was required to outline the boost volume only: GTVpb. The GTVpb reference volume was created as a consensus volume by three members of the TMG (IS, JS and AT) all of whom had had experience of mpMRI interpretation for several years and DIL outlining within clinical trials (BIOPROP/BIOPROP20 and DELINEATE) [3,4,5]. They individually outlined GTVpb and after review of each other's outlines agreed a final reference volume. CTVp was pre-outlined by the Chief Investigator (IS) and provided for investigators.

Case 2 included a high-risk, locally advanced prostate cancer with a PI-RADS 5 DIL in the right peripheral zone (PZ) suitable for a boost. This case required the investigator to delineate five volumes: GTVpb, CTVp, CTVpsv, vessels and CTVn. Again, for each volume, a consensus reference volume was determined by the TMG clinicians (IS, AT, JS). Centres not participating in the boost arms were not required to outline GTVpb for case 2.

Definitions of the target volumes to be outlined were included in the trial protocol planning and delivery guidelines [6] and summarised in Table 3.1.

Target volume	Structures included in target volume	Definition of target volume
CTVp	Prostate	Prostate + proximal 1cm of seminal vesicles (defined using MRI planning scan if performed) + any extraprostatic extension
PTVp_6000	Prostate	CTVp+3mm isotropic margin
CTVpsv	Prostate, seminal vesicles	CTVp + any remaining seminal vesicles
PTVpsv_4700	Prostate, seminal vesicles	CTVpsv + 6mm isotropic margin
Vessel	Pelvic vessels	Left and right common iliac, external iliac, internal iliac and obturator vessels
CTVn	Pelvic lymph nodes	Elective nodal volume
PTVn_4700	Pelvic lymph nodes	CTVn + 5mm isotropic margin
GTVpb	Boost volume	Intraprostatic lesion defined on staging mpMRI
CTVpb	Boost volume	GTVpb+3mm isotropic margin, not extending outside CTVpsv

Table 3.1: Definition and description of target volumes with the PIVOTALboost planning and delivery guidelines

Other than the description of structures to include in CTVp and CTVpsv (as described in table 3.1), no detailed outlining guidance was provided for CTVp and CTVpsv, as it was similar to another recent UK trial, PACE C [7]. Vessel and CTVn delineation guidance was provided in a supplementary document, 'PIVOTALboost pelvic node atlas', detailing step-by step instructions and worked examples for pelvic node contouring[8].

Focal boost (GTVpb) outlining guidance was provided in a supplementary document 'PIVOTALboost boost contouring atlas' [9]. The first part of the boost contouring atlas focused on prostate anatomy and interpretation of mpMRI using the PI-RADS v2 scoring system (discussed in Chapter 1). Boost volumes were defined as PI-RADS 4 or 5 lesions and interpretation of mpMRI to define the volume depended on anatomical location of the lesion.

Having defined the boost volume on mpMRI, the guidance recommended using the diagnostic mpMRI and histopathology report to aid outlining. The guidance recommends initially defining the hypo-intense region on the ADC map using the T2W sequences to aid definition. Once defined on mpMRI, a similar volume should be outlined on the corresponding CT planning scan using the MRI to guide position and size in a process termed 'cognitive fusion'. The guidance recommended to not fuse the CT planning scan and diagnostic mpMRI given the potential impact of ADT on prostate and boost volume. Following completion of outlining on CT in the axial plane, the DIL volume should be viewed in the sagittal and coronal planes to ensure consistency of outlines. Any additional DILs should be outlined using the same technique. Once complete, CTVp and CTVpsv should be outlined, ensuring

the boost volume does not extend beyond CTVp. Two worked examples of peripheral zone DILs outlined on their respective mpMRI and CT planning scans were provided.

Referring back to the delivery and planning guidance [5], the DIL should be labelled GTVpb (GTVpb1, GTVpb2 etc if multiple lesions) and a 3mm isotropic margin applied to this volume to create CTVpb. CTVpb should not extend beyond CTVp (Figure 3.1).
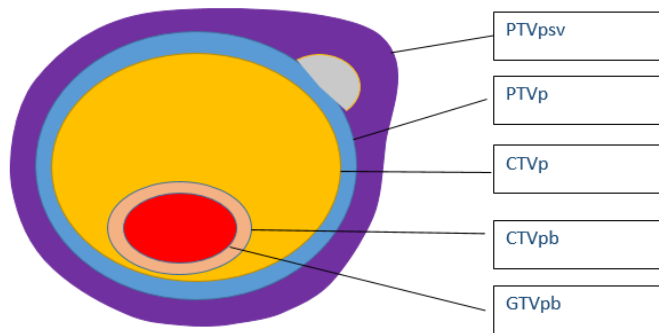
Figure 3.1: Diagram of target volume delineations for a prostate boost

For each pre-accrual benchmark case, the participating centre was sent the respective CT planning scans as a secured DICOM file. An accompanying clinical vignette detailing the patient's PSA result, MRI, bone scan and histopathology report were also provided. Screenshots of the T2W and ADC map of the diagnostic mpMRI were provided to aid delineation.

Participating centres were required to upload the CT planning scan into their own treatment planning software (TPS) for outlining, therefore avoiding potential contouring errors due to use of unfamiliar contouring software. Each target volume required outlining using the nomenclature and outlining guidance described in the protocol and supplementary outlining documents detailed above. Completed structure sets for case 1 and case 2 were transferred together from participating centres as a secure DICOM file to the RTQA team based in Clatterbridge Cancer Centre (CCC). Submitted structures were subsequently uploaded into the CT planning scan in Aria (Eclipse) TPS at CCC for review.

Submitted target volume delineations were directly compared against the pre-determined reference volumes for each structure by IS; investigator volumes were superimposed onto the reference volumes allowing direct visual comparison between investigator volume and reference volume.

The TMG pre-determined limits for outlining variations for each target volume to assess whether the investigator's structures should be considered acceptable or unacceptable (Table 3.2). The reviewer (IS) completed a proforma for each investigator's attempts of the cases. For the purpose of feedback,

given the similarity in creating the volumes, CTVp and CTVpsv were considered one structure (CTVp/CTVpsv), as are vessels and CTVn (vessel/CTVn). IS classified each investigator structure (GTVpb case 1, GTVpb case 2, CTVp/CTVpsv, vessel/CTVn) as having an 'unacceptable' or 'acceptable' variation. Screen shots for each investigator volume overlaying the respective reference volume were provided to complement written feedback detailing outlining deviations in accordance with the criteria in Table 3.2.

Target Volume	Unacceptable Variations
GTVpb	• Extends beyond CTVp outline • Incorrect DIL outlined (or in incorrect region of prostate) • Outline extends >3mm beyond superior and/or inferior extent of reference volume • Total DIL volume +/- 25% of reference volume
CTVp/psv	• Outline extends >3mm beyond apex or base of reference volume (too large) • Outline is >3mm above apex or below base of reference volume (too small) • Extension into anal canal at apex of CTVp • Not including central SV into CTVp for 9mm above prostate level • Not including obvious extracapsular extension in CTVp • Extending CTVp into rectum or bladder
CTVn/vessels	• Outline extends >6mm beyond superior or inferior extent of reference volume (too large) • Incorrect inclusion of bowel or bladder structure in CTVn • Incorrect inclusion of 'bowel +3mm' in CTVn • Not excluding pre-sacral muscles from CTVn if there were other minor deviations from the contouring protocol • Incomplete CTVn (nodal regions missed) • Incorrect vessels outlined/ Ureters outlined as vessels/ Vessels outside CTVn

Table 3.2 Unacceptable variation criteria for PIVOTALboost pre-accrual benchmark cases

Any outlined structure that had a variation from the reference volume not specified in Table 3.2, was highlighted during feedback and noted as having 'acceptable' variation(s). For each investigator, any structure considered having an 'unacceptable' variation required resubmission. Any structure considered 'acceptable' or had an 'acceptable' variation did not require resubmission. Once all target volume submissions were satisfactorily achieved, the outliner was deemed eligible to outline TVDs for patients they subsequently recruited into the trial.

3.2 Aims and Objectives

3.2.1 Aims

The aim of this chapter is to explore the extent of prostate radiotherapy TVD variation amongst UK oncologists using data from the PIVOTALboost pre-accrual case submissions.

1. To establish the incidence of unacceptable DIL outlining variation in Case 1 and Case 2 of the PIVOTALboost pre-accrual cases as defined by the Chief Investigator

2. To identify/ categorise the causes of unacceptable DIL outlining variation as defined by the Chief Investigator

3. To identify the incidence and cause of prostate and pelvic nodal outlining variation in Case 2 of the PIVOTALboost pre-accrual cases as defined by the Chief Investigator

4. To assess outlining performance in the re-submitted PIVOTALboost pre-accrual cases

3.3 Methods

All 32 investigators who had submitted PIVOTALboost pre-accrual case submissions up until September 2018 were sent a completed proforma by the Chief Investigator (IS) with written and pictorial feedback regarding their target volume outlining. These proformas were kept as secure word documents at CCC. HM (RTTQA member) at CCC anonymised all proformas, including re-submission reports, to OWW (RTTQA member) at Velindre Cancer Centre (VCC) via a secure file share. Documents were uploaded onto the secure VCC RTTQA network drive.

For case 1, I reviewed each proforma completed by IS to assess whether GTVpb submissions were marked (by IS) as having 'unacceptable' variations using the pre-determined 'unacceptable variation' criteria in Table 3.2 and the nature of the variation(s). Investigators could have more than one acceptable variation and not require resubmission. I also reviewed all proformas for cases that were re-submitted for GTVpb re-outlining.

For case 2, I reviewed each completed proforma by IS to assess whether GTVpb, CTVp/CTVpsv, vessel/CTVn had been allocated 'unacceptable' variations (Table 3.2) and assessed the nature of outlining variation. Investigators could have more than one type of 'acceptable' variation for each target volume outlined, however only structures assessed as having 'unacceptable' variations by IS required re-submission. I also reviewed all proformas for re-submitted outlining attempts.

3.4. Results

3.4.1 First submissions of PIVOTALboost pre-accrual benchmark case 1 and case 2

Thirty-two investigators completed the PIVOTALboost pre-trial benchmark cases. Two investigators were from centres not offering a DIL boost so therefore did not complete case 1 or outline GTVpb for case 2.

Of the 32 investigators, 26 (81%) were required to resubmit at least one case due to unacceptable variations from the outlining protocol. 16 (50%) investigators were required to re-submit both cases. GTVpb was the structure that had the most unacceptable variations: 22/30 (73%) for case 2 GTVpb and 17/30 (57%) for case 1 (Table 3.3). Fewer unacceptable variations were seen for CTVp/psv (12/32, 38%) and CTVn (8/32, 25%).

	Case 1 (n=30)	Case 2 (n=32)		
	GTVpb	GTVpb (n=30)	CTVp/psv	CTVn
Acceptable/Acceptable Variation(s)	13	8	20	24
Unacceptable Variation(s)	17	22	12	8

Table 3.3 Number of outliners with acceptable outlines/acceptable variations vs unacceptable variations in case 1 and case 2

3.4.2 PIVOTALboost pre-accrual benchmark case 1

The causes of unacceptable variations for Case 1 are shown in Figure 3.2.
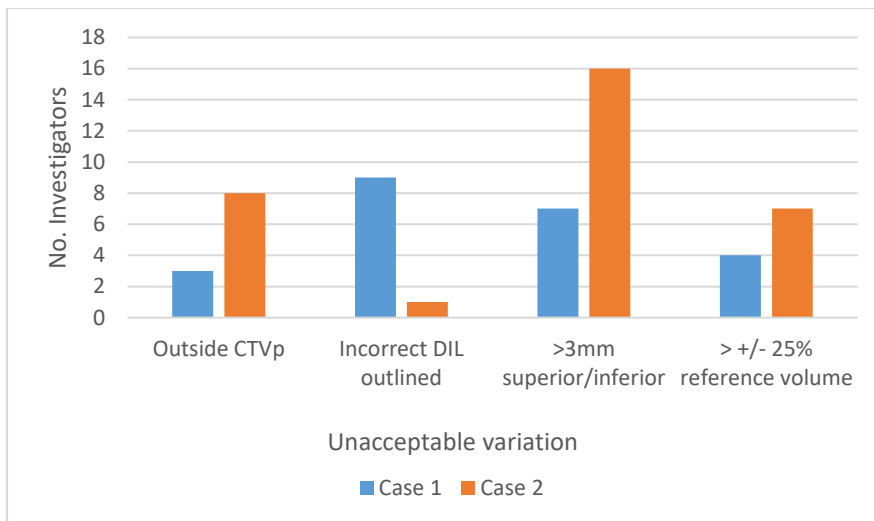
Figure 3.2 Causes of GTVpb unacceptable outlining variation in PIVOTALboost pre-accrual case 1 and case 2

The most common unacceptable variation in GTVpb outlining in case 1 was delineation of an lesion in a different region of the prostate, which was only 3mm in diameter and therefore did not meet criteria as a boost volume. Of the seven GTVpb submissions that were >3mm (i.e. >1 slice in either superior and/or inferior direction) than the reference volume, the additional slices were in the inferior direction for six of them. No GTVpb submissions had unacceptable variations in both superior and inferior

extent. The volume of GTVpb delineated was unacceptable by four outliners (13%), predominantly due to volumes ≥25%, with only one outliner (3%) having a volume ≤25% of the total reference DIL volume.

3.4.3 PIVOTALboost pre-accrual benchmark case 2

3.4.3.1 GTVpb

Causes of unacceptable variation of GTVpb outlines are shown in Figure 3.2. The most common was superior/inferior extent of GTVpb with seven outliners outlining the inferior extent of the DIL unacceptably and four in the superior extent. Five outliners had unacceptable variations in both inferior and superior extent. Investigators more commonly submitted smaller GTVpb volume in case two, with four outliners delineating GTVpb ≤25% of the reference volume and three outliners ≥25% of the reference volume. Only one outliner delineated a GTVpb that did not regionally correspond to the reference volume GTVpb and two outliners (6%) delineated two lesions instead of one.

3.4.3.2 CTVp/CTVpsv

With regards to CTVp/CTVpsv outlining, the most common unacceptable variation was larger volume outlining at the prostate apex followed by excess bladder included within CTVp/CTVpsv and the CTVp/CTVpsv not being delineated at all (Table 3.4).

Target Volume	Unacceptable Variations	No. outliners (%)
CTVp/psv	CTVp/psv not outlined	2(6)
	Rectum included in CTVp/psv	1(3)
	Excess bladder included in CTVp/psv	2(6)
	2 or more CT slices beyond base of reference volume	1(3)
	2 or more CT slices beyond apex of reference volume	4(12)
	1 CT slice above apex **and** below base of reference volume	1(3)
	Extracapsular extension not included in CTVp	1(3)

Table 3.4 Causes of unacceptable outlining variation identified for CTVp/psv in PIVOTALboost pre-accrual case 2

3.4.3.3 CTVn/vessels

Eight outliners (25%) were required to resubmit CTVn due to unacceptable variation. This was predominantly due to incorrect vessel outlining (12%) although 9% of investigators incorrectly failed to include CTVn in their submission. Two outliners (6%) incompletely delineated CTVn and one outliner (3%) incorrectly excluded the pre-sacral region in conjunction with other minor CTVn outlining deviations (Table 3.5).

Target Volume	Unacceptable Variations	No. outliers (%)
CTVn/vessels	Incorrect vessel delineation (total)	4(12)
	No CTVn/ vessel volumes delineated	3(9)
	Incomplete CTVn (nodal regions missed)	2(6)
	Pre-sacral muscles not excluded from CTVn	1(3)

Table 3.5 Causes of unacceptable outlining variation identified for CTVn/vessels in PIVOTALboost pre-accrual case 2

3.4.4 Resubmissions of PIVOTALboost pre-accrual benchmark case 1 and case 2

Of the 26 outliners required to resubmit at least one case, 23 had acceptable outlines and were subsequently approved to recruit to PIVOTALboost. Of the three outliners who needed to submit delineations for a third time, two outliners were required to submit structures for both cases. All three resubmissions had unacceptable deviations relating to GTVpb outlining and two resubmissions had unacceptable CTVp outlining deviations.

Two outliners had identical variations in their first and second submissions. For the first attempt of case 1, both outliners delineated the incorrect left posterior DIL. For the case 2 GTVpb, both outliners were initially two slices too short both superiorly and inferiorly. On resubmission of case 1, both outliners were too short superiorly and DIL volumes were < 25% of the reference volume. For case two, both outliners delineated GTVpb volumes that breached CTVp. The third outliner, was required to resubmit GTVpb for case 2 only, having delineated a GTVpb volume that was too small and too long on their first attempt and then breached CTVp on their second attempt. Following their third submissions, all three outliners were approved to recruit to the trial due to acceptable contour submission.

3.5 Discussion

The proportion of outliners having to re-submit at least one case for the pre-trial RTQA for PIVOTALboost was high at 81%. There have been several previous UK multi-centre prostate cancer radiotherapy trials [2,3,10,11] that have had outlining assessed since the formation of the NCRI RTTQA group, but this is one of the first Phase III trials to include DIL outlining.   Although there is no direct DIL or in-situ prostate radiotherapy pre-trial outlining performance data to compare our results to, Fenton et al [12] on behalf of the post-operative prostate bed radiotherapy EORTC 22043-30041 trial, had 18% of pre-trial 'dummy case' submissions initially rejected due to unacceptable target volume delineation errors. NeoSCOPE, the phase II randomized control trial of two neoadjuvant chemoradiotherapy regimens in oesophageal cancer, had only a 15% resubmission rate for their pre-trial RT outlining test case [13] which is most likely explained by familiarity of investigators with

oesophageal radiotherapy delineation given the standardised protocol implemented across the UK of the earlier SCOPE 1 trial [14].

The relatively high resubmission rate for PIVOTALboost is predominantly due to unacceptable GTVpb outlining. DIL boosting is a new technique for most UK prostate clinical oncologists. The phase II UK trial BIOPROP20, which explored hypofractionated RT with intraprostatic boosts, used direct co-registration of the diagnostic choline PET-CT with the planning CT to define the DIL [15]. This trial however was only open in two centres, CCC and VCC, with all cases delineated by IS and JS, two of the three PIVOTALboost TMG, therefore keeping the familiarity of this technique to a small number of experts. Cognitive transfer of the DIL from diagnostic mpMRI to planning CT for boost volume outlining therefore is a relatively new concept; PIVOTALboost is the first phase-III trial in the UK to define DIL outlining this way and is likely to have underpinned the variation in GTVpb outlining. Variable familiarity with interpreting mpMRI images is also likely to have impacted DIL outlining. As of 2018, 47% of UK centres were not offering mpMRI before biopsy in their diagnostic pathway and 13% of centres did not have access to mpMRI at all [16] suggesting variable mpMRI knowledge amongst UK oncologists and those without access more likely to find DIL outlining more difficult.

For case 1, the most common GTVpb outlining error was delineation of the incorrect intraprostatic lesion. This was due to another DWI restricting lesion in the left posterior aspect of the prostate evident on the pre-diagnostic mpMRI which investigators outlined instead of or in addition to the intended DIL. Although the trial permits boosting more than one lesion in a single patient, the incorrectly delineated left posterior lesion was smaller than the 5mm required to qualify as a suitable DIL to dose escalate. This variation may be explained by general unfamiliarity with DIL outlining, the inclusion of DIL boosting for the first time into UK prostate radiotherapy within PIVOTALboost and lack of clarity of or lack of adherence to the trial's protocol.

The second most common GTVpb unacceptable variation in case two was CTVp breach by the GTVpb volume. This error was seen more commonly in case 2 than case 1. Investigators having to delineate the CTVp in case two, therefore introducing another contouring variable, is likely have exacerbated this outlining variation as the CTVp contour was provided for investigators in case one. This is however the scenario that most matches the on-trial outlining situation.

Across the two cases, the most common GTVpb variation was determining the inferior extent of the DIL. A number of factors preclude the trial from permitting direct co-registration of the diagnostic mpMRI and planning CT. Firstly, diagnostic MRI scans performed across different centres use soft couches and no designated bladder or bowel preparation unlike planning MRI scans which use flat, harder couches and a designated bladder and bowel protocol. Fusion therefore of the diagnostic MRI

introduces differences in patient set up which means direct organ fusion is not always possible, whereas the planning MR mimics the planning CT set up. However, the planning MRI does not have the DWI sequences available to identify the DIL appropriately. Additionally, patients have had a minimum of three months ADT, after which not only can the prostate shrink, rendering fusion of the diagnostic MRI even less precise at the time of planning, but the tumour commonly 'responds' to the ADT and is not always visible at the time of planning scans. Therefore, reliance upon cognitive transfer of diagnostic imaging to the planning CT is used in PIVOTALboost. In the PIVOTALboost 'boost' contouring protocol [9], outlining initially on the MRI first is suggested, however only screenshots of the diagnostic mpMRI sequences were provided in the pre-accrual case submissions, therefore this could not have been possible. However, given the impact of ADT on prostate size, outliners are still required to cognitively transfer a final boost volume onto the planning CT that is proportional to the change in prostate size following ADT. In addition, prostate apex definition i.e. defining the most inferior aspect of the prostate, is recognized as the most common outlining error in prostate gland contouring [17], as reflected in the CTVp/CTVpsv outlining deviations in this data. GTVpb volume was either too large or too small by 25% of the reference volume in 11/32 (34%) submissions and of these 8 also had inferior extension errors, which is likely to have increased the total DIL volume. In 2 out of these 11 cases, both inferior and superior extent of the lesion was assessed as being excessive, possibly representing an oncologist preference to not miss a cancer (over-contouring). Circumferential extent of the lesion was not independently assessed in the outlining criteria, but might also have affected the overall DIL volume.

Two outliners required to submit the case for a second resubmission were both from the same centre and had the same unacceptable variations for their outlines. This suggests that the outliners did the pre-accrual benchmark exercises together and were unfamiliar with the protocol. Both outliners carried the same error of GTVpb length forward to the second resubmission albeit on a different case. CTVp breach and ineligible lesion outlining may both be considered protocol 'rules' that once appreciated are less likely to be carried forward. DIL length and volume outlining however, relate to cognitive transfer of DIL from mpMRI to CT planning scan following ADT, which is more skill-based and therefore it is less surprising this error was carried forward by the two outliners into their second resubmission.

CTVn outlining was associated with the fewest unacceptable errors. Although the role of pelvic nodal irradiation in prostate radiotherapy is yet to be established, pelvic nodal irradiation is routinely included in 13% and 17% of prostate radiotherapy plans in England and Wales respectively [18]. UK consensus guidelines for delineation of an acceptable lymph node CTV were developed for the UK phase II PIVOTAL trial, evaluating prostate and pelvic nodes vs prostate only IMRT [2]. PIVOTALboost

nodal delineation guidance does vary from the international RTOG pelvic nodal delineation guidelines [19] slightly, for example superior extent of the volume, but generally represents a technique familiar to UK prostate clinical oncologists. The most common CTVn outlining error was vessel delineation. This requires anatomical interpretation and some experienced clinicians may not delineate vessels in non-trial patients, preferring to create CTVn as a direct volume, particularly as this can be time consuming, therefore being less familiar with direct vessel contouring.

3.6 Conclusion

The majority of PIVOTALboost pre-accrual benchmark case submissions had 'unacceptable' outlining variations requiring resubmission of at least one case. This was predominantly due to DIL 'boost' outlining, which is a relatively new skill for most prostate clinical oncologists in the UK and is not currently standard of care. This suggests not only is robust on-treatment RTQA required in PIVOTALboost to identify further outlining 'errors' to minimize impact on trial outcome, but further work is also required to improve DIL outlining if the technique is to become future standard of care in prostate radiotherapy. Chapter 5 will explore the impact of an outlining workshop on UK prostate oncologists DIL outlining performance with a view to improving consistency amongst outliners. However, before this, following on from the work in this chapter, I will explore the role of conformity indices as a semi-automated tool in the assessment of investigator GTVpb volumes in the PIVOTALboost pre-trial benchmark cases in Chapter 4.

3.7 References

1. Institute of Cancer Research (ICR) (UK). PIVOTALboost: A phase III randomised controlled trial of prostate and pelvis versus prostate alone radiotherapy with or without prostate boost. Protocol; Version 5. ISRCTN80146950. 2019.

2. Institute of Cancer Research (ICR) (UK): A randomised phase II trial of Prostate and pelvIs Versus prOsTate Alone treatment for Locally advanced prostate cancer (PIVOTAL) (ISRCTN48709247). [Internet] 2011.[cited 2023 Feb 03]. Available from: https://www.isrctn.com/ISRCTN48709247

3. Syndikus I, Uzan J, Mayles P, Nahum, A. BIOPROP: Biologically Optimised Prostate Cancer Radiotherapy or Dose Painting. NCRI Cancer Conference. UK: NCRI, 2010.

4. Syndikus I, Chan J, Rowntree T, Howard L, Staffurth J. Hypofractionated dose painting IMRT using 20 fractions for intermediate to high-risk localised prostate cancer: Two-year outcome data (BIOPROP 20, NCT02125175). J Clin Oncol. 2019 Mar 1; 37(suppl 7): 59-59.

5. Murray J, Tree A, Alexander E, Sohaib A, Hazell S, Thomas K, et al. Standard and Hypofractionated Dose Escalation to Intraprostatic Tumor Nodules in Localized Prostate Cancer: Efficacy and Toxicity in the DELINEATE Trial. Int J Radiat Oncol Biol Phys. 2020 Mar 15; 106(4): 715-724.

6. Institute of Cancer Research (ICR)(UK). PIVOTALboost: RT Planning and delivery guidelines. Version 1. ISRCTN80146950. 2017.

7. Institute of Cancer Research (ICR) (UK). The PACE Trial (Prostate Advances in Comparative Evidence) International randomised study of prostatectomy vs stereotactic body radiotherapy (SBRT) and conventional radiotherapy vs SBRT for organ-confined prostate cancer (ISRCTN17627211).[Internet] 2012 [cited 2023 Feb 04]. Available from: https://www.icr.ac.uk/our-research/centres-and-collaborations/centres-at-the-icr/clinical-trials-and-statistics-unit/our-research/clinical-trials/pace

8. Institute of Cancer Research (ICR)(UK). PIVOTALboost: Pelvic node contouring Atlas; Version 1. ISRCTN80146950. 2017.

9. Institute of Cancer Research (ICR)(UK). PIVOTALboost: Trial Boost Contouring Instructions & Atlas; Version 1. ISRCTN80146950. 2017.

10. Dearnaley D, Jovic G, Syndikus I, Khoo V, Cowan R, Graham J, 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 Apr; 15(4): 464-73.

11. Dearnaley D, Syndikus I, Mossop H, Khoo V, Birtle A, Bloomfield D, et al; CHHiP Investigators. 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 Aug; 17(8): 1047-1060.

12. Fenton, P.A., Hurkmans, C., Gulyban, A., van der Leer, J., Matzinger, O., Poortmans, et al. Quality assurance of the EORTC 22043-30041 trial in post-operative radiotherapy in prostate cancer: results of the Dummy Run procedure. Radiother Oncol. 2013 Jun; 107(3): 346-51.

13. Evans E, Jones, G, Rackley T, Maggs R, Radhakrishna G, Mukherjee S et al. PO-0769: NeoSCOPE RTTQA: pre-accrual and on-treatment review of all patients in a UK oesophageal RT trial. Radiother Oncol. 2018; 127.

14. Gwynne S, Higgins E, Poon King A, Radhakrishna G, Wills L, Mukherjee S, et al. Driving developments in UK oesophageal radiotherapy through the SCOPE trials. Radiat Oncol. 2019 Feb 4; 14(1): 26

15. The Clatterbridge Cancer Centre (CCC). BIOPROP20: Radiotherapy Planning and Treatment Document. Version 1. LCTU-CTSU Protocol Number RDD547. 2016

16. Prostate Cancer UK :''Access to mpMRI before biopsy increasing but parts of UK are still appalling''. [Internet] 2018. [cited 2022 Nov 27]. Available from: https://prostatecanceruk.org/about-us/news-and-views/2018/6/access-to-mpmri

17. McLaughlin P, Evan, C, Feng M, Narayana V. Radiographic and Anatomic Basis for Prostate Contouring Errors and Methods to Improve Prostate Contouring Accuracy. Int J Radiat Oncol Biol Phys. 2010; 76(2): 369-378

18. National Prostate Cancer Audit: Annual Report 2019 [Internet]. 2020 [cited 2023 Feb 04] Available from: https://www.npca.org.uk/content/uploads/2020/01/NPCA-Annual-Report-2019_090120.pdf

19. Harris V, Staffurth J, Naismith O, Esmail A, Gulliford S, Khoo V, et al. Consensus Guidelines and Contouring Atlas for Pelvic Node Delineation in Prostate and Pelvic Node Intensity Modulated Radiation Therapy Int J Radiat Oncol Biol Phys. 2015 Jul 15. 92(4): 874-83.

Chapter 4: Evaluation of Conformity Indices for DIL Outlining in PIVOTALboost Pre-trial Data

4.1 Introduction

Having reviewed the PIVOTALboost pre-trial test case submissions that were deemed acceptable and unacceptable by the trial Chief Investigator (IS) in chapter three, I was particularly interested in the high rates of re-submissions relating to the DIL boost volume (GTVpb). Given the infancy of DIL boosting, and therefore general lack of familiarity of this technique amongst the wider oncology community, this result was unsurprising to me. However, during my analysis of the completed proformas, I was aware that there was inconsistency between whether investigator GTVpb (iGTVpb) submissions had been deemed 'acceptable' or 'unacceptable' by IS, against the four pre-defined criteria for 'unacceptable' variations:

- GTVpb breaches CTVp
- GTVpb outside correct region of prostate
- GTVpb exceeds +/- 1 slice (3mm) superiorly and/or inferiorly
- GTVpb investigator volume exceeds +/- 25% of the reference volume

The criteria developed at the start of the trial recommended that if an iGTVpb demonstrates any of the above unacceptable four criteria, the volume requires re-submission. However, some submissions, despite meeting one or more of the unacceptable criteria, were being considered 'acceptable' by IS. This highlighted a number of points. Firstly, despite pre-defined 'unacceptability' criteria, it appeared that in some cases, clinical judgement was applied additionally by IS in determining whether a volume was 'acceptable' or 'unacceptable'. Secondly, reliance upon one expert clinician to assess all the pre-trial submissions can be problematic. It is likely to be time-consuming for that individual, risk introduction of assessor-bias and introduce intra-observer variation that is unidentified by lack of additional assessors. It has been postulated by Gwynne et al [1], that conformity indices may allow some form of automated assessment of investigator delineation conformity against a reference volume. For GTVpb, however conformal the volume, any slight breach of CTVp would not be reflected in a quantitative measure of conformity against a GTVpb reference volume so a conformity index cannot act solely as an automated assessment tool for GTVpb and some form of clinical assessment would still be required.

Therefore, I set out to explore the role of conformity indices as a semi-automated assessment tool for the PIVOTALboost pre-trial boost volume, that could be applied in conjunction with clinical assessment

by an assessor (in addition to, or in place of the Chief Investigator) and also to minimise assessor subjectivity.

4.2 Aims and Objectives

4.2.1 Aim

The aim of this chapter is to explore the role of conformity indices as a semi-automated tool in the assessment of investigator GTVpb volumes in the PIVOTALboost pre-accrual benchmark cases.

4.2.2 Objectives

1. To establish whether there is a difference between the Chief Investigator's (IS) assessment of the PIVOTALboost pre-trial iGTVpb volumes and my assessment using the four qualitative 'unacceptable' variations criteria.

2. To establish whether there is a conformity index that discriminates between the 'acceptable or 'unacceptable' iGTVpb volumes (i.e whether they were a 'pass' or 'fail') for PIVOTALboost pre-accrual benchmark cases.

3. Having identified a conformity index applicable to the PIVOTALboost pre-accrual benchmark cases GTVpb volumes, establish a threshold above/below which acceptable volumes are deemed 'acceptable' or 'unacceptable'.

4.3 Methods

4.3.1 Detailed qualitative analysis of PIVOTALboost pre-trial investigator GTVpb volumes

Sixty-four GTVpb submissions for the PIVOTALboost pre-accrual benchmark case 1(Figure 4.1) and 74 for case 2, which included TMG attempts and those without IS RTQA review, were anonymised and sent in DICOM format from CCC to VCC along with the respective planning CT images. Using the open-source MATLAB based software platform Computational Environment for Radiological Research (CERR) [2], I imported and re-labelled all case 1 and case 2 iGTVpb volumes to ensure consistency in nomenclature. I then assessed all investigator iGTVpb volumes against their respective reference GTVpb (rGTVpb) volumes, using each of the four parameters of unacceptable deviations, allocating them a 'pass' or 'fail'. For structures which had been previously reviewed by IS, I compared whether she had deemed them requiring resubmission (i.e. a fail) or not (i.e. a pass) against my (EE) assessment.
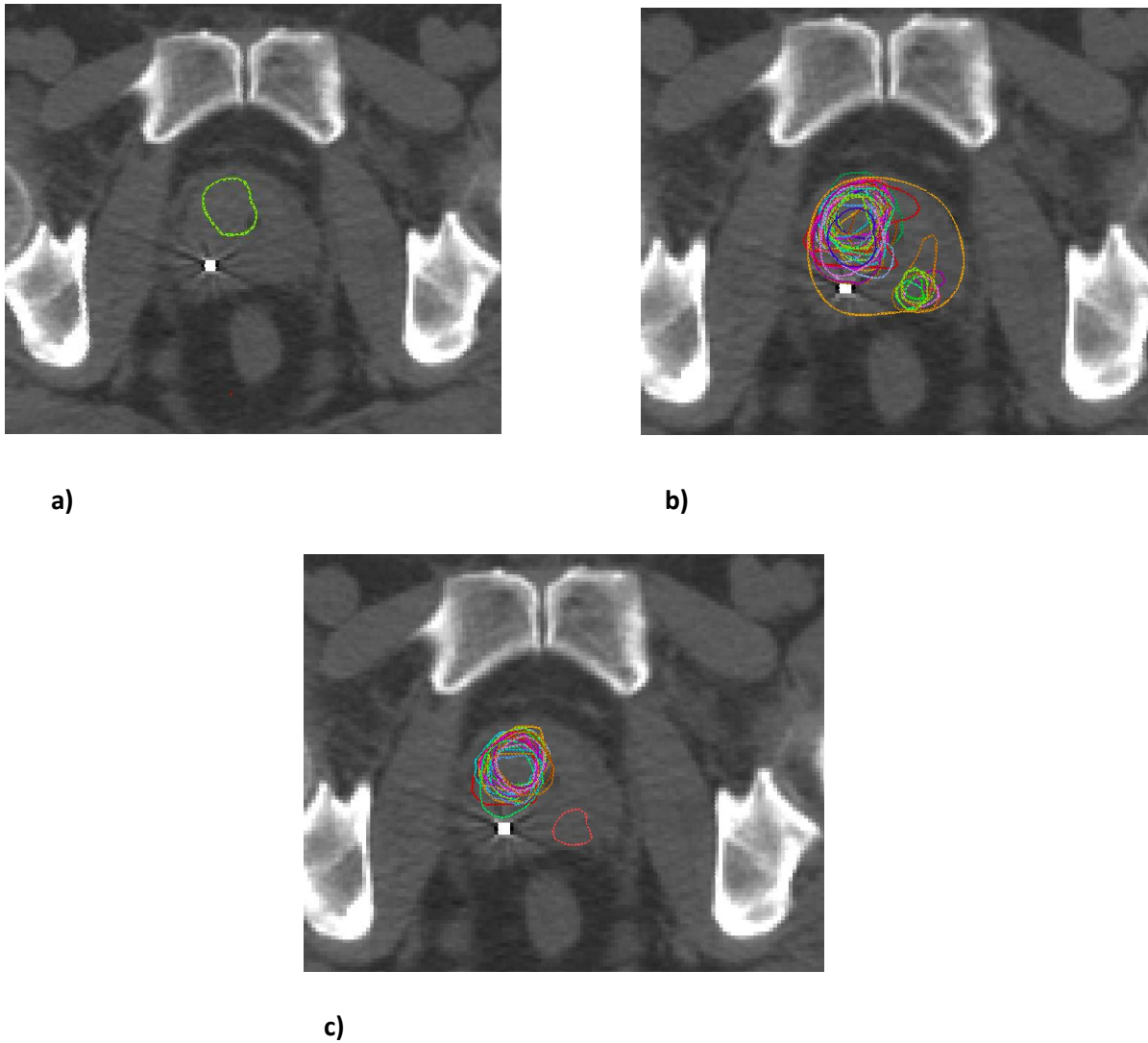
a)

b)

c)

Figure 4.1 PIVOTALboost pre-trial case one volumes: a) reference volume (green outline) b) reference volume with all iGTVpb first submissions and c) subsequent resubmissions following Chief Investigator feedback

4.3.2 Conformity Index Analysis of PIVOTALboost pre-trial benchmark iGTVpb volumes

For the conformity index analysis, I used my own decision on 'pass'/'fail' analysis rather than IS outcomes, having taken a more objective assessment by strictly applying the four qualitative 'unacceptable' criteria. Any pre-trial iGTVpb structures that had 'failed' on my assessment due to CTVp breach or outlining within the wrong prostate segment were excluded from further analysis. The conformity index calculation will not detect this specific variation as it is not able to take into account additional structures that are needed, in this case the CTVp. An initial visual check should also be able to recognise any lesion which has incorrectly been delineated as a GTVpb (i.e. a lesion which does not qualify as a boost lesion because it is in different region of the prostate to the rGTVpb). If two GTVpb volumes had been delineated instead of one for a particular case, the lesion in the 'incorrect' prostate

segment was deleted within CERR so that the DIL in the 'correct' segment could be included in the analysis and I re-scored as a 'pass'/'fail' based on the superior/inferior extent and volumetric parameters.

As 'training data', I initially selected the first submission for case 1, so that any conformity indices that were statistically significant between the 'pass' and 'fail' groups could be validated on the resubmission group and again on the case 2 group to ensure the conformity indices were not case specific.

Case 1 first-submission iGTVpb volumes were imported into CERR by a post-doctorate software analyst (CP) and run through an in-house script that calculated the following conformity indices for each structure: van't Riet's, Jaccard's, DICE, DI, GMI and MDC (total, -over and –under) (see Chapter 2 for description of each conformity index). Following guidance from statistician JW, I used the statistical software SPSS® to establish whether the conformity index data (continuous) was normally distributed. As the data was non-parametric and sample size less than 30, I applied the Mann-Whitney U test in SPSS® for each conformity index, to determine whether the conformity index values between the 'pass' and 'fail' groups were significantly different.

4.3.3 Determining a conformity index 'Pass'/'Fail' threshold for PIVOTALboost GTVpb volumes

To establish a conformity index threshold which could discriminate between the 'pass' and 'fail' iGTVpb submissions, I ranked the relevant conformity index values in numerical order to establish the conformity index value (i.e. the threshold) above (for van Riet/JCI/DICE/MDC) or below (GMI/DI/MDC -under or –over) which all 'true passes' (as determined by my qualitative analysis) passed so that no 'true fails' passed.

4.4 Results

4.4.1 EE vs IS analysis of PIVOTALboost pre-trial investigator GTVpb volumes

Of all the iGTVpb structures received, IS reports were available for 37 case 1 submissions (29 first submission, 8 resubmission) and 31 case 2 submissions (20 first submissions, 11 resubmissions). The Pass/fail scores I allocated the case 1and case 2 iGTVpb volumes compared to those allocated by the PIVOTALboost Chief Investigator (IS) are shown in figure 4.2.
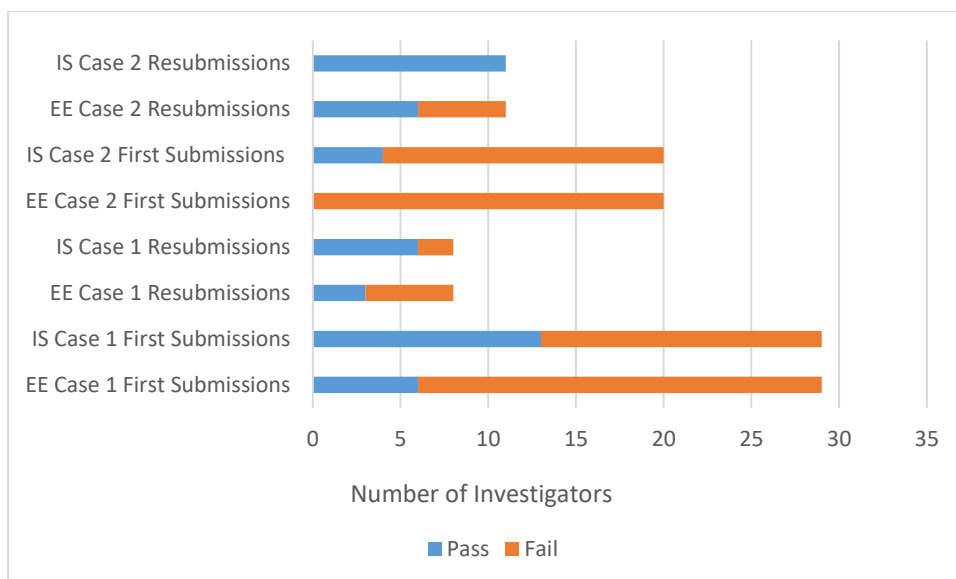
Figure 4.2: Investigator GTVpb Outlines Allocated Pass/Fail for Case 1 and Case 2 by EE vs the PIVOTALboost Chief Investigator (IS)

For both case 1 and case 2, I consistently failed more investigator GTVpb volumes. Proportionately, both IS and I passed more investigators on their resubmissions. For case 2 resubmissions, IS passed all resubmission attempts, whereas I failed nearly half of them. Reviewing the proforma feedback, IS passed investigator GTVpb volumes (for both first and resubmissions) that were technically 'unacceptable' for at least one of the four parameters.

4.4.2 Conformity index analysis of PIVOTALboost pre-trial benchmark iGTVpb volumes

Twenty-four iGTVpb first submissions and 23 re-submissions were eligible to include for quantitative analysis for case 1, 29 first submissions and 24 resubmissions were available for case 2.

The volume of the case 1 GTVpb reference volume was 3.4cm³. The mean investigator volume for case 1 first submissions and resubmissions was 4.9cm³ (SD 2.1) and 3.5cm³ (SD 1.3) respectively.

The GTVpb volume of the case 2 reference volume was 3.2cm³. The mean investigator volumes for case 2 first and resubmissions were 5.2cm³ (SD 3.4) and 4.1cm³ (SD 1.7) respectively.

4.4.2.1 Conformity index analysis of case 1 GTVpb structures (first submissions)

The descriptive statistics for each conformity index for the 24 case 1 first submissions are shown in Table 4.1. Individual investigator results are included in Appendix 1.

Conformity Index (CI)	Minimum CI Value (across all investigators)	Maximum CI Value	Mean CI Value	Standard Deviation	Value in Perfect Concordance	Mann-Whitney U-test P value
van't Riet	0.12	0.64	0.37	0.11	1	0.53
Jaccard	0.21	0.65	0.42	0.10	1	0.42
DICE	0.34	0.78	0.59	0.10	1	0.42
GMI	0.08	0.60	0.31	0.18	0	0.03
DI	0.03	0.69	0.46	0.15	0	0.01
MDC	0.18	0.76	0.47	0.10	0mm	0.64
MDC (over-contouring)	0.18	0.42	0.29	0.07	0mm	0.02
MDC (under-contouring)	0.07	0.40	0.18	0.08	0mm	0.26

Table 4.1: Descriptive statistics for each CI for case 1 GTVpb first submissions

The range of individual conformity index values for case 1 iGTVpb first submissions, according to whether I had marked them as a 'pass' or 'fail' is shown in Figures 4.3-4.4.

Figure 4.3: Box plots showing range of conformity index values for PIVOTALboost case 1 first submissions depending on whether they were a 'pass' or 'fail':

van't Riet (top left),

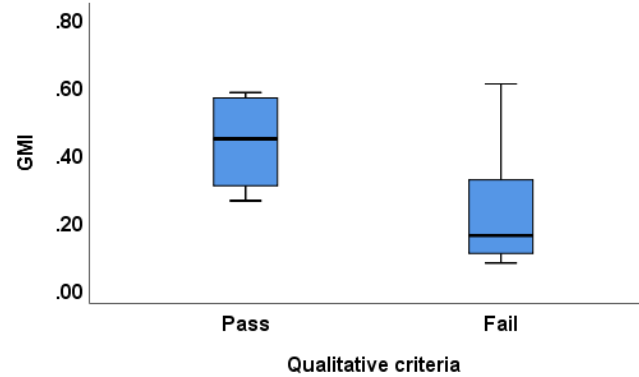*JCI (top right),*

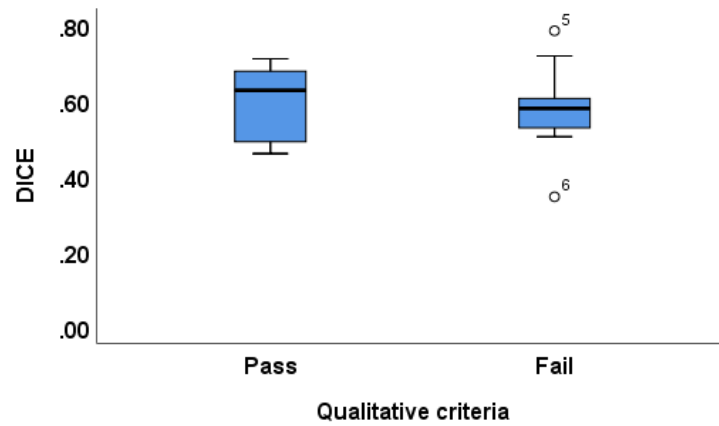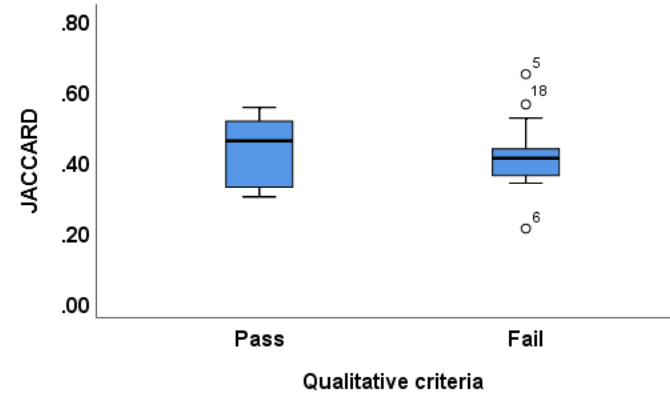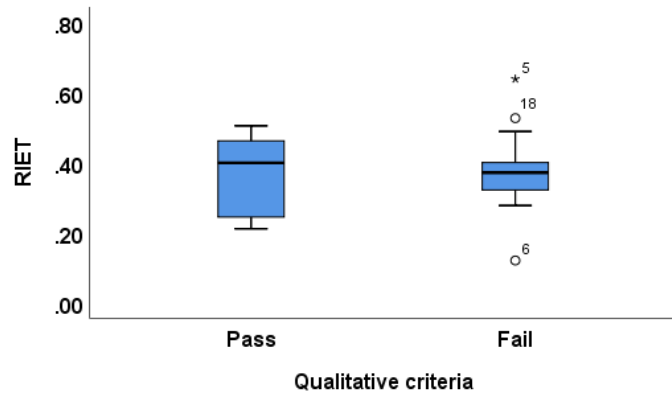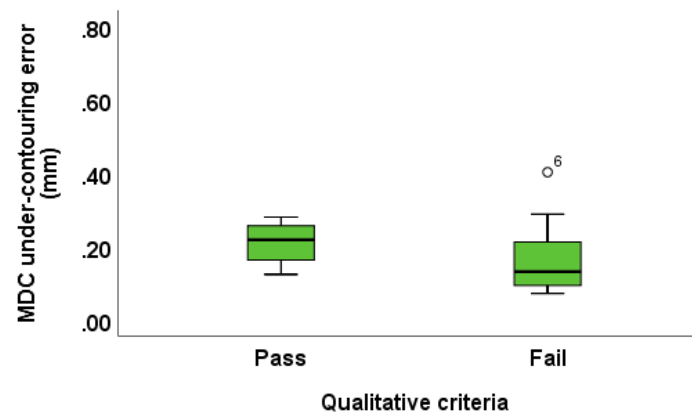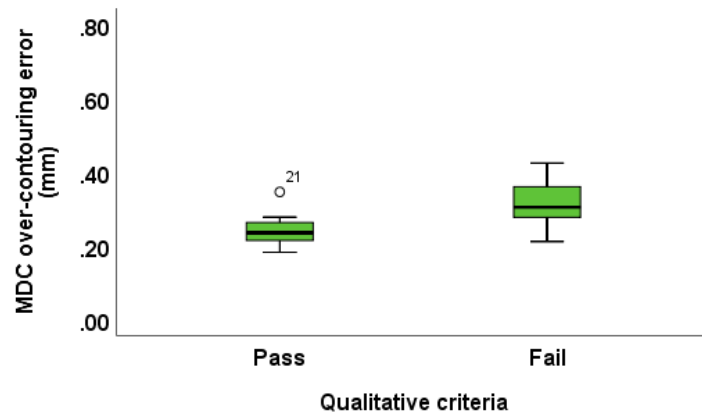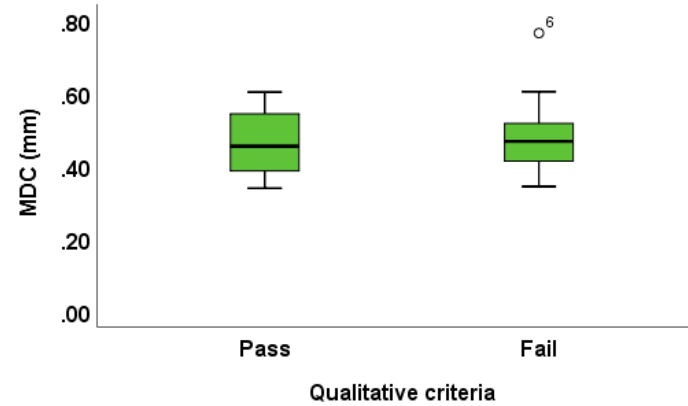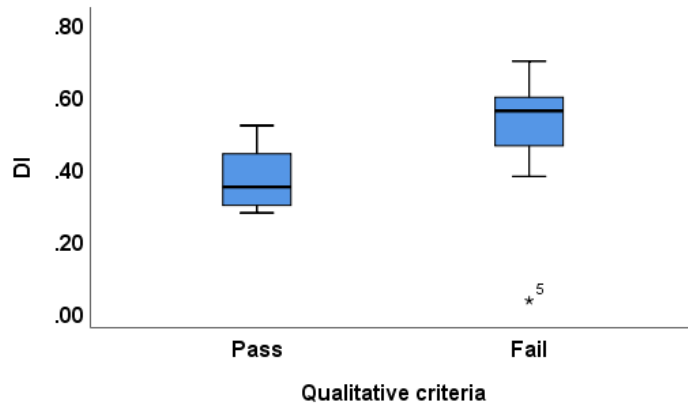*DICE (bottom left) and GMI (bottom right)*

Figure 4.4: Box plots showing range of conformity index values for PIVOTALboost case 1 first submissions depending on whether they were a 'pass' or 'fail':

DI (top left),

MDC (top right),

MDC over contouring (bottom left) and

MDC under contouring (bottom right)

From the results above, GMI, DI and MDC over-contouring showed a statistically significant difference between the pass and fail groups for the iGTVpb case one first attempts.

4.4.3   Determining a conformity index 'pass'/'fail' threshold for PIVOTALboost GTVpb volumes

To determine a threshold conformity index value for GMI, DI and MDC over-contouring to identify the 'true passes', I ranked the investigator GMI, DI and MDC over-contouring values in numerical order from lowest to highest to see whether there was a value below which all 'true passes' (determined by EE) passed (Figure 4.5). For GMI, DI and MDC over-contouring, the lower the value the more concordant the outline (i.e. less under- contouring for GMI, less over- contouring for DI/MDC over-contouring).

iGTVpb DI value	Pass/fail	iGTVpb GMI value	Pass/fail	iGTVpb MDC over- value	Pass/fail
0.03	fail	0.08	fail	0.18	pass
0.27	pass	0.08	fail	0.20	pass
0.28	pass	0.08	fail	0.21	fail
0.29	pass	0.1	fail	0.21	pass
0.35	pass	0.11	fail	0.22	fail
0.35	pass	0.13	fail	0.22	pass
0.38	fail	0.14	fail	0.23	fail
0.38	pass	0.16	fail	0.23	pass
0.43	fail	0.21	fail	0.25	fail
0.44	fail	0.26	pass	0.26	pass
0.44	fail	0.28	fail	0.26	pass
0.48	pass	0.3	fail	0.28	pass
0.48	fail	0.3	pass	0.30	fail
0.5	fail	0.3	pass	0.30	fail
0.51	fail	0.34	fail	0.30	fail
0.52	pass	0.38	pass	0.30	fail
0.56	fail	0.4	fail	0.31	fail
0.56	fail	0.44	pass	0.32	fail
0.58	fail	0.44	fail	0.35	pass
0.59	fail	0.48	pass	0.36	fail
0.6	fail	0.56	pass	0.36	fail
0.63	fail	0.56	fail	0.38	fail
0.65	fail	0.58	pass	0.40	fail
0.69	fail	0.6	fail	0.42	fail

More Concordant ↑

Less Concordant ↓

Figure 4.5: Investigator DI, GMI and MDC over-contouring values (mm) ranked from lowest to highest and corresponding pass/fail outcomes determined by EE

From my data, lower GMI values, which are usually associated with *better* concordance, appear to be associated with 'true fails', whereas for DI and MDC over-contouring, there are more 'true passes' associated with lower values as expected. The converse trend for GMI is most likely due to investigators tending to over-outline GTVpb, as reflected in the mean iGTVpb volumes and DI, MDC over-contouring values. As GMI is a measure of under-outlining, investigators who have 'over-outlined' are more likely to have 'better' i.e. lower GMI values because they are not 'under-outlining'.

For DI I was unable to determine a clinically acceptable threshold which the 'true passes' would be likely to pass because the lowest DI value was a 'true fail' so there was no value below this that could be used as the 'true pass' threshold.

With respect to MDC over-outlining, a value of 0.2mm was the lowest value at which only 'true passes' passed. However, at this threshold only two investigators would have passed out of 24, meaning 22 investigator volumes would still require full review making this conformity index and corresponding threshold of little practical use.

As I was unable to identify an appropriate conformity index and corresponding pass/fail threshold that would be of clinical benefit, I did not proceed with using resubmissions or case 2 as 'testing data' to validate the results.

4.5 Discussion

The aim of this chapter was explore the role of conformity indices and determine whether there was one or more that could be used as a semi -automated tool to determine 'pass'/'fail' based on superior/inferior extent and volume, for PIVOTALboost pre-accrual benchmark case investigator submissions. From my analysis, I was unable to identify a conformity index and corresponding 'pass'/'fail' threshold that could be used in this role. To date, no radiotherapy trial has used a semi-automated tool to determine acceptable or unacceptable investigator volumes as part of pre-trial RTQA assessment. In an RTQA context, conformity indices are usually used in the pre-trial setting for retrospective assessment of investigator benchmark case performance against the reference volume [3]. More recently in an on-trial setting, conformity indices have been used to review whether investigator adherence to the protocol is maintained [4], although this requires a central review team re-contouring trial patients' target volumes and is therefore time-consuming and labour intensive.

Prior to evaluating the use of conformity indices as a 'pass'/'fail' for the PIVOTALboost pre-trial data, GTVpb outlines that breached the CTVp were excluded from my evaluation because this parameter did not relate specifically to concordance between investigator and reference volume and therefore a conformity index could not be used in place of all four qualitative parameters. I had also excluded

GTVpb outlines in the incorrect region of the prostate although some conformity indices could indicate whether there was no concordance with the reference volume. For example, the degree of overlap in this scenario would be nil and therefore JCI/DICE/van't Riet's would be 0 and DI/ GMI would be 1 (if the investigator had delineated an ineligible DIL in a different part of the prostate). However, the same conformity index value would not discriminate between a GTVpb lesion delineated in the correct region adjacent to but not overlapping with the reference volume. Therefore, even before starting data analysis for this chapter, any conformity index identified that could distinguish 'pass'/fail' investigator volumes, could only be used as an adjunct to expert review of the investigator volumes and 'screen' investigators who passed based on volumetric and superior/inferior extent.

Terparia *et al* [5] investigated the feasibility of building an automated tool, based on conformity indices and supervised machine learning to evaluate contouring conformity using 393 investigator contours from 253 Stereotactic Ablative Body Radiotherapy (SABR) pre-trial benchmark outlining exercises. They found this was achievable for specific structures and machine learning models, but could not identify a single conformity index or machine learning model to apply to the whole dataset. They concluded that an automated method cannot replace clinical review, but, if individualised to specific TVDs, could be used to pre-assess investigator contours prior to clinical review, to highlight any gross outlining errors.

The conformity indices measuring degree of investigator and reference volume overlap i.e. DICE, JCI and van't Riet's, did not demonstrate a statistically significant difference between the 'pass' and 'fail' investigator volumes in this study yet are the most commonly applied conformity indices in the literature [1]. Reviewing the application of conformity indices for assessment of pre-trial benchmark cases using the SCOPE 1 [6], SCALOP [7] and ARISTOTLE [8] trials, Gwynne *et al* [1] identified the advantages and disadvantages relating to each conformity index and concluded a site-specific approach is required, with different conformity indices suited to different tumour sites.

From my analysis, DI and MDC over-contouring were statistically significant in discriminating between pass and fail volumes for the pre-trial case 1 first submission cohort. Given the mean volume of investigator volumes was consistently higher than the reference volumes across my data, this could be expected. Holyoake *et al* [9], in their analysis of investigator against reference volumes for pancreatic cancer, also found investigator delineations were larger than their respective reference volumes and subsequently had higher DI scores. This was also reflected in Gwynne *et al's* [10] analysis of the inter-observer variation seen in the SCOPE1 oesophageal cancer pre-trial test case, with the mean investigator GTV volume also larger than the reference volume and DI values higher than GMI. This trend for investigators to over-outline suggests a tendency to err on the side of caution, to ensure

that disease is included rather than excluded from the target volumes to be irradiated. However, in their analysis, Holyoake et al [9], found through NTCP modelling, investigator over-contouring was associated with toxicity risk, in their case the risk of duodenal bleed, thus emphasising the need for a robust on-trial QA programme.

The statistical significance of GMI was initially misleading. As GMI is a measure of under-outlining, given the 'significance' of over-contouring conformity indices, one would expect GMI to be 'insignificant' in my data set. However, as demonstrated in figure 4.5, investigators with high DI scores typically had low GMI scores; if they over-outlined GTVpb, they were less likely to under-outline the same volume resulting in a lower GMI. Therefore, the relationship between 'fail' iGTVpb volumes and high DI scores translated into 'fail' iGTVpb volumes and low GMI scores, resulting in exclusion of GMI as an appropriate conformity index for this data set.

I was unable to establish a DI threshold at which I could discriminate pass or fail investigators. Although there is no definitive method to determine the threshold value, for clinical significance, it should pass the 'true passes' at very least on the test data set. The iGTVpb with the lowest DI however, was a true fail, therefore I could not use this data to determine a DI value below which all iGTVpb volumes had passed. In their evaluation of investigator contours of the pre-trial test case for the EMBRACE II trial in cervical cancer, Duke *et al* [11], using a JCI 'pass' threshold of 0.7, found that only 45% of all contours passed by the expert review ('true -positives') would have passed this threshold. Thirteen percent of contours, who failed expert review, would have also passed ('false-positives'). By raising the JCI threshold to 0.75, all but one of the failing contours would have been identified yet the false-negative rate (i.e. those who failed the threshold but passed expert review) was 74%, concluding JCI was not a reliable alternative assessment to expert review.

Whereas I had endeavoured to identify a suitable DI threshold using the PIVOTALboost pre-trial data, Duke *et al* [11] initially chose a JCI value of 0.7 as a ''commonly used threshold for clinically adequate delineation in studies''. Gwynne *et al* [1] assessed the SCOPE 1 pre-trial case data against a JCI threshold of 0.8 based on Jena *et al's* [12] analysis of glioblastoma TVD, yet found no investigator passed this threshold despite passing expert review. This further supports that even if appropriate conformity indices can be identified as a semi-automated assessment tool, a pass/fail threshold would likely be tumour-site specific and possibly even case-dependent [1].

With respect to reviewing the qualitative 'pass'/'fail' parameters, I consistently failed more investigator GTVpb volumes than IS. I adhered rigidly to the four qualitative parameters therefore objectively assessing the volumes in accordance with the protocol, identifying iGTVpb volumes passed by the conformity indices that did not fulfil all four outlining criteria. There are likely to be a number

of reasons for this. Firstly, in a newer technique such as DIL boosting, there are few trials to adopt acceptable outlining criteria from, and those that are therefore identifying appropriate qualitative parameters in PIVOTALboost is novel and likely to be an iterative process. Reviewing subsequent iGTVpb submissions may inform the Chief Investigator, who is an expert in DIL boosting, that the criteria may be 'too strict' and that clinical judgement is required to determine whether a volume is acceptable or not. Secondly, there will be an element of bias in Chief Investigator assessment of investigator volumes because clinical trial teams will be keen to recruit participating centres and if the recruitment process is too prohibitive, they run the risk of deterring potential investigators.

Setting volumetric parameters (in this case iGTVpb should be within +/- 25% of the rGTVpb), aims to identify volumes that are at risk of under-dosing the tumour or conversely, over-irradiating normal tissue. DILs however, are relatively small target volumes, and from this data, +/- $0.1cm^3$ of volume included in the iGTVpb, could translate into as much as 4% deviation from the rGTVpb. From review of individual investigator volumes, Investigator 5 (Appendix 1) had the highest DICE and JCI scores and the lowest DI score, suggesting excellent conformity. However, they 'failed' objective qualitative assessment because their GTVpb volume was 29% smaller than the rGTVpb. Visually, the iGTVpb conformed well to the rGTVpb, which is unsurprising as the investigator was part of the TMG and therefore contributed to creation of the rGTVpb volumes. The rGTVpb volumes for case 1 and case 2 were $3.4cm^3$ and $3.2cm^3$ respectively yet DILs 5mm or greater in diameter qualify as boost volumes, therefore potentially much smaller volume lesions could be boosted. In this scenario, applying the +/- 25% volumetric criteria would inevitably fail more investigators, as an $0.1cm^3$ increase or decrease in a smaller volume, would translate into a much greater percentage change in volume. The rigidity of this parameter therefore may have affected the significance of the 'overlapping' conformity indices such as DICE or JCI, and their role in assessment of DIL outlining cannot be determined from this data. Following feedback from IS regarding their first submissions, the resubmission iGTVpb mean volumes for both case 1 and case 2 were smaller and closer to the respective rGTVpb volumes suggesting feedback helps improve iGTVpb outlining although as seen in other RT trials, this may not be maintained through the trial and a robust on-trial RTQA process is required [3].

As discussed above, both from this data and other radiotherapy pre-trial data, the role of using a conformity indices and an appropriate threshold as a semi-automated assessment tool for investigator TVDs is yet to be established. Using the ARISTOTLE pre-trial benchmark test case, Sweeney et al [13], used MDC-OVER-UNDER analysis to identify volumes considered 'unacceptable'. This required creation of a maximally accepted reference volume and minimally accepted reference volume by expert clinicians, supported by the STAPLE algorithm. MDC assessment related to investigator breach of the maximum and minimum reference volume. For DIL boost volumes, this assessment method may

be an alternative semi-automated tool to consider because, the maximum acceptable volume could take into consideration the CTVp volume and an MDC over-contouring value could potentially pick this up. It would also be case-specific, therefore the maximum and minimum acceptable volume for each rGTVpb could be determined and avoid the more generic volumetric parameter of iGTVpb being within -/+ 25% of the rGTVpb. Using a minimum and maximum acceptable volume against which to compare outlines has been used to provide semi-automated feedback for TVD training [14] but has not yet been implemented within RTQA. The difficulty with this method however is that it requires a panel of experts to determine the STAPLE algorithm and from this a clinically acceptable minimum and maximum to be determined which can be time- consuming, particularly when a range of target volumes require assessment.

There are limitations associated with the work in this chapter. The data was collected by another centre (CCC), therefore when I received the data, the nomenclature of structures was inconsistent therefore not only did all the structures require relabelling, it was not always possible to identify the same investigator's submissions and resubmissions and who had a completed feedback proforma by IS. I had used CERR to review the iGTVpb volumes but IS had used a different platform therefore volumetric data did not quite marry between the two systems, so I recalculated all the volume data for both the rGTVpb and iGTVpb volumes ensuring that the volumes did not pass/fail depending on which system used. In addition, the number of investigator GTVpb first submissions for case 1 upon which to determine an appropriate conformity index was relatively small at 24. Despite having a number of other investigator submissions (i.e. case 1 resubmissions and case 2 submissions), I had hoped to identify an appropriate conformity index and threshold on the initial data set and then 'test' conformity index on the remaining data. As I was unable to identify a suitable threshold, the remaining submissions were not tested because for a conformity index and threshold to be useful in this setting, it had to apply to all data.

4.6 Conclusion

For the PIVOTALboost pre-accrual benchmark cases, I was unable to determine a conformity index and corresponding threshold that could be used as a semi-automated tool to pass or fail submitted iGTVpb volumes. The qualitative parameters used to assess the iGTVpb volumes, particularly the volumetric parameter, require adjunct expert clinical assessment to determine whether the iGTVpb is clinically acceptable or not. Conformity indices however, have been commonly used to compare the performance of outliners before and after an educational intervention [15]. In view of the results from chapter 3, suggesting UK prostate oncologists require further training with respect to DIL outlining, in the next chapter, I will assess the impact of an outlining workshop on DIL outlining performance using both qualitative and quantitative metrics.

4.7 References

1. Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, Conibear J, Staffurth J; NCRI RTTQA OUTLINING AND IMAGING SUBGROUP. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. Br J Radiol. 2013 Apr; 86(1024): 20120398.

2. Deasy J, Blanco A, Clark V. CERR: a computational environment for radiotherapy research. Med Phys. 2003 May; 30(5): 979-85.

3. Chang A, Tan LT, Duke S, Ng W. Challenges for Quality Assurance of Target Volume Delineation in Clinical Trials. Front Oncol. 2017 Sep 25; 7:221.

4. Fokas E, Spezi E, Patel N, Hurt C, Nixon L, Chu, K, et al. P-222: Analysis of tumour contours and radiotherapy planning of "on-trial" patients undergoing chemoradiotherapy (CRT) in SCALOP trial: does pre-trial Radiotherapy Quality Assurance (RTQA) improve the quality of "on-trial" radiotherapy?. Ann Oncol. 2016; 27(suppl 2): ii64-ii65.

5. Terparia S, Mir R, Tsang Y, Clark CH, Patel R. Automatic evaluation of contours in radiotherapy planning utilising conformity indices and machine learning. Phys Imaging Radiat Oncol. 2020 Dec 1; 16: 149-155.

6. Gwynne S, Hurt C, Spezi E, Crosby T, Staffurth, J. PO-0664: Mean distance to conformity as a tool fro assessing outlining in the UK SCOPE 1 oesophageal chemoradiotherapy trial. Radiother Oncol. 2012 May; 103: S259

7. Mukherjee S, Hale C, Spezi E, Joseph G, Branagan M, Hurt C, et al. Comparison of investigator-delineated GTV in chemoradiotherapy (CRT) for locally advanced non-metastatic pancreatic cancer (LANPC): Analysis of the pre-trial test case for the SCALOP trial. J Clin Oncol. 2013 Feb 1; 31(suppl 4): 230-230.

8. Cole N, Gwynne S, Spezi E, Maggs R, Sebag-Montifiore D, Adams. Quality assurance of target volume definition in th ARISTOTLE Phase III rectal cancer trial- initial assessment. Radiother Oncol. 2012 May 1; 103.

9. Holyoake DL, Robinson M, Grose D, et al. Conformity analysis to demonstrate reproducibility of target volumes for Margin-Intense Stereotactic Radiotherapy for borderline-resectable pancreatic cancer. Radiother Oncol. 2016 Oct; 121(1): 86-91.

10. Gwynne S, Spezi E, Staffurth J, Diez P, Nixon, L, Al-Mokhtar R, et al. Inter-observer variation in target volume definition for the SCOPE 1 pre-trial test case. NCRI Cancer Conference. UK: NCRI, 2011.

11. Duke S, Tan LT, Jensen N, Rumpold T, De Leeuw A, Kirisits C, et al. Implementing an online radiotherapy quality assurance programme with supporting continuous medical education -

report from the EMBRACE-II evaluation of cervix cancer IMRT contouring. Radiother Oncol. 2020 Jun; 147: 22-29.

12. Jena R, Kirkby N, Burton E, Hoole A, Tan L, Burnet N. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. Br J Radiol. 2010 Jan; 83 (985): 44-51.

13. Sweeney L, Spezi E, Cole N, Sebag-Montifiore D, Adams R. PO-0949: Automated approval of a pre trial benchmark RTTQA case. The ARISTOTLE experience. Radiother Oncol. 2016 Apr 1; 119.

14. Piazzese C, Evans E, Thomas B, Staffurth J, Gwynne S, Spezi E. FIELD[RT]: an open-source platform for the assessment of target volume delineation in radiation therapy. Br J Radiol. 2021 Oct 1; 94(1126); 20210356.

15. Cacicedo J, Navarro-Martin A, Gonzalez-Larragan S, De Bari B, Salem A, Dahele M. Systematic review of educational interventions to improve contouring in radiotherapy. Radiother Oncol. 2020 Mar; 144: 86-92.

Chapter 5: Evaluation of the Impact of an Educational Workshop on DIL Target Volume Outlining Variation

5.1 Introduction

In Chapter 3, I established that there was a high rate of resubmission of the pre-accrual PIVOTALboost test cases, predominantly due to unacceptable dominant intraprostatic lesion (DIL) outlining variation and its relative infancy as a technique amongst UK prostate oncologists. In chapter 2, I discussed the role of educational interventions as one method used to reduce inter-observer outlining variation. In their review of interventions to reduce inter-observer variation, *Vinod et al* [1] discovered teaching interventions reduced inter-observer variation in eight out of nine studies reviewed. Although the most effective teaching method could not be ascertained from this review, one method commonly used, is an outlining workshop. With a view to improving DIL outlining consistency, on this chapter I will explore the impact an outlining workshop, using a combination of didactic and interactive teaching methods, has on DIL outlining variation.

5.2 Aims and Objectives

5.2.1 Aim

To assess the impact of an outlining workshop on the outlining variation of the dominant intraprostatic lesion (DIL).

5.2.2. Objectives

1. To assess the immediate impact of an outlining workshop on inter-observer DIL outlining variation using qualitative and quantitative metrics.
2. To assess short and long term retention of knowledge following an educational workshop on inter-observer DIL outlining variation using qualitative and quantitative metrics.
3. To assess the impact of an educational workshop on delegates perceived knowledge regarding DIL outlining.

5.3 Methods

5.3.1 Workshop overview

The outlining workshop was held at the Royal College of Radiologists (RCR) annual conference (RCR18) in Liverpool, UK on the 10th September 2018. This workshop was the first time the RCR had used the

web-based outlining platform Share Place (now Onco Place) by AQUILAB [1] as a real-time interactive outlining platform within an outlining workshop.

The outlining workshop was a two-hour session included at the end of a prostate cancer themed day aimed at clinical oncology consultants, fellows and trainees with variable experience in prostate radiotherapy.

The first hour was dedicated to DIL outlining and the second, prostate bed outlining. The same two cases, a DIL and prostate bed respectively, were to be outlined by delegates before, during and immediately after the workshop (see below) and again one year following the workshop. During the first hour of the workshop, delegates were given a fifteen minute didactic lecture on 'how to outline' the DIL and interpretation of mpMRI sequences to aid delineation, followed by a five minute discussion of the pre-workshop outlines with discussion of common DIL outlining 'pitfalls'.

 To assess the immediate impact of the educational intervention, delegates were required to re-outline the same case in real-time within the workshop over a thirty minute session. The second half of the workshop on prostate bed radiotherapy followed the same format as the DIL session. Prior to finishing the workshop, delegates were asked to complete a questionnaire detailing their previous DIL outlining experience and the impact the workshop had on the confidence of their DIL and prostate bed outlining skills.

I had originally planned to re-distribute the DIL case to delegates one year after the workshop to assess the long- term impact of an educational workshop. Unfortunately, following inevitable delays in the RCR and AQUILAB being able to release the case, the eventual release date coincided with the emergence of the COVID-19 pandemic and therefore this aspect of the project was abandoned. A summary of the DIL steps of the workshop exercise is shown in Figure 5.1.

Before workshop

Delegates to complete DIL outlining exercise 2 weeks before workshop

Workshop

- 15 minute presentation on DIL mpMRI interpretation and 'how to outline'
- 5 minute discussion and review of delegates anonymised contours submitted before the workshop
- Delegates re-attempt the same DIL outlining exercise during a 30 minute time slot
- At end of workshop, delegates complete questionnaire regarding their confidence in DIL outlining

After workshop

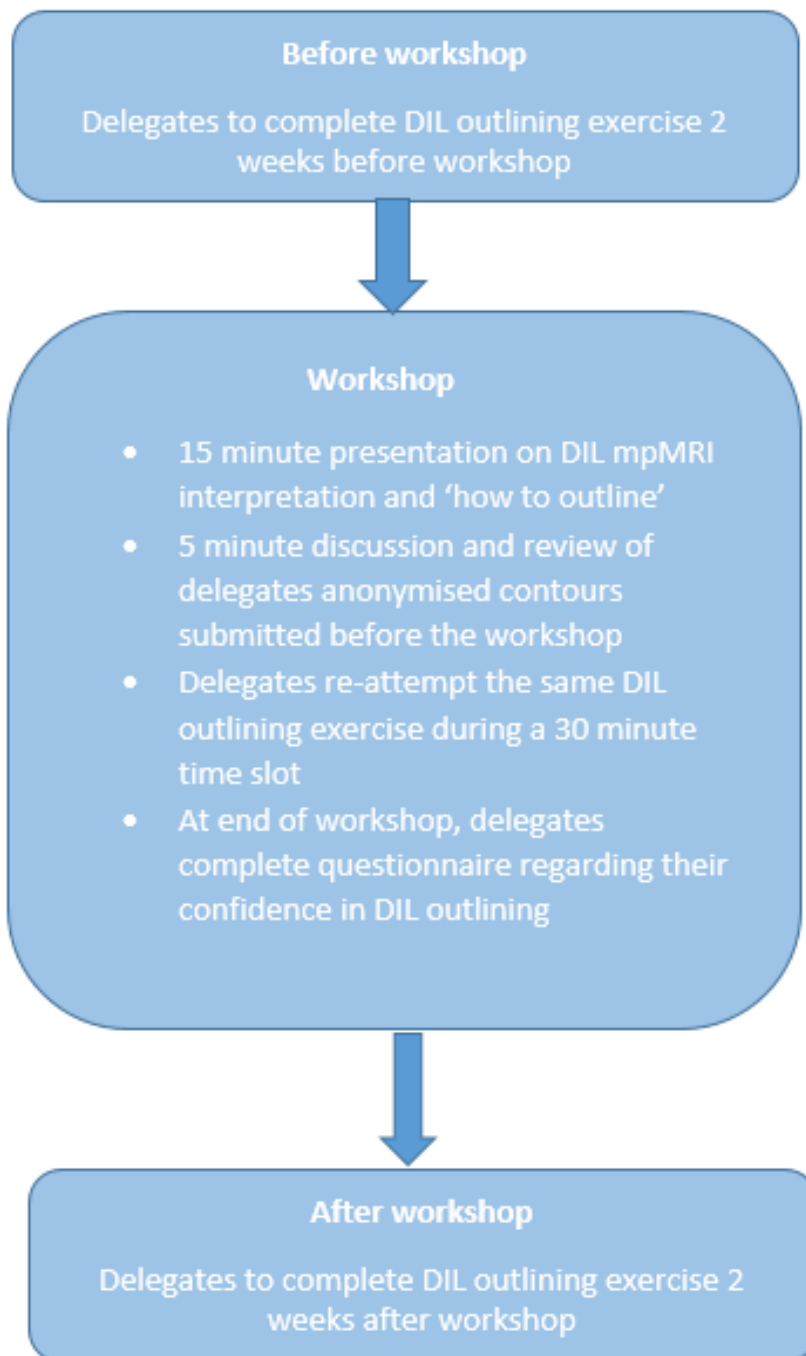Delegates to complete DIL outlining exercise 2 weeks after workshop

Figure 5.1: Summary of steps before, during and after outlining workshop

5.3.2 Case development

An anonymised on-trial PIVOTALboost case with a DIL boost was used as the outlining case. AT (PIVOTALboost TMG) delineated the case 'on-trial' and contours were amended and approved by the PIVOTALboost Chief Investigator (IS) in the context of RTQA. AT was granted permission from the patient, to use his case for the outlining workshop. The diagnostic MRI and planning MRI were sent in anonymised format from the Royal Marsden Hospital NHS Trust (RMH) to Velindre Cancer Centre (VCC) via the Picture Archive and Communication System (PACS). The CT planning scan and delineated structures were sent in an anonymised format as a DICOM file from RMH to VCC and imported into PROSOMA for EE and JS to review suitability to use as reference volumes.

Three target volumes were identified; a right-sided DIL, a left-sided DIL and the prostate gland. Target volume nomenclature was determined by the RCR approved prostate target volume matrix: GTVp (prostate and extra-prostatic extension) and GTVpb (prostate tumour volume on staging mpMRI). Left and right were used to differentiate between the two DILs. Upon review of the outlines (EE, JS) no further amendments were required and three reference volumes were finalised:

- GTVp
- GTVpb_right- Right transitional zone tumour; 4 slices in length (see Figure 5.2)
- GTVpb_left - Left peripheral zone tumour; 3 slices in length (See Figure 5.2)

The CT planning scan and reference volumes were uploaded directly onto Share Place by AQUILAB [1]. The T2 weighted sequence of the planning MR and the ADC map and T2 weighted sequences of the diagnostic MRI were sent in DICOM format via a secure fileshare to CB, AQUILAB Oncology Project Manager, who uploaded the MRI images onto the outlining platform. CB then registered the four imaging sequences (CT planning scan, T2 diagnostic MRI, ADC diagnostic MRI and T2 planning MRI) in alignment, allowing all four sequences to correlate as the user scrolls through the case (Figure 5.3). The reference volumes were 'disabled' within Share Place so that delegates could not see them prior to the workshop.
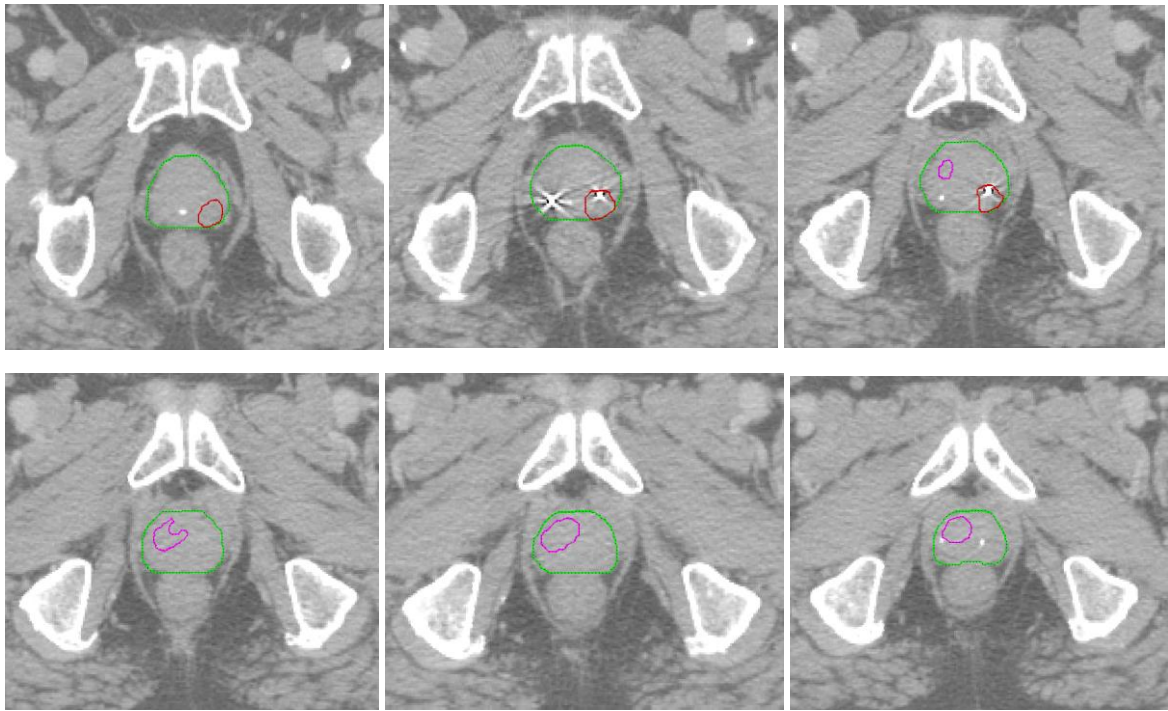
Figure 5.2: GTVpb_left (red) and GTVpb_right reference volumes shown on axial plane of planning CT
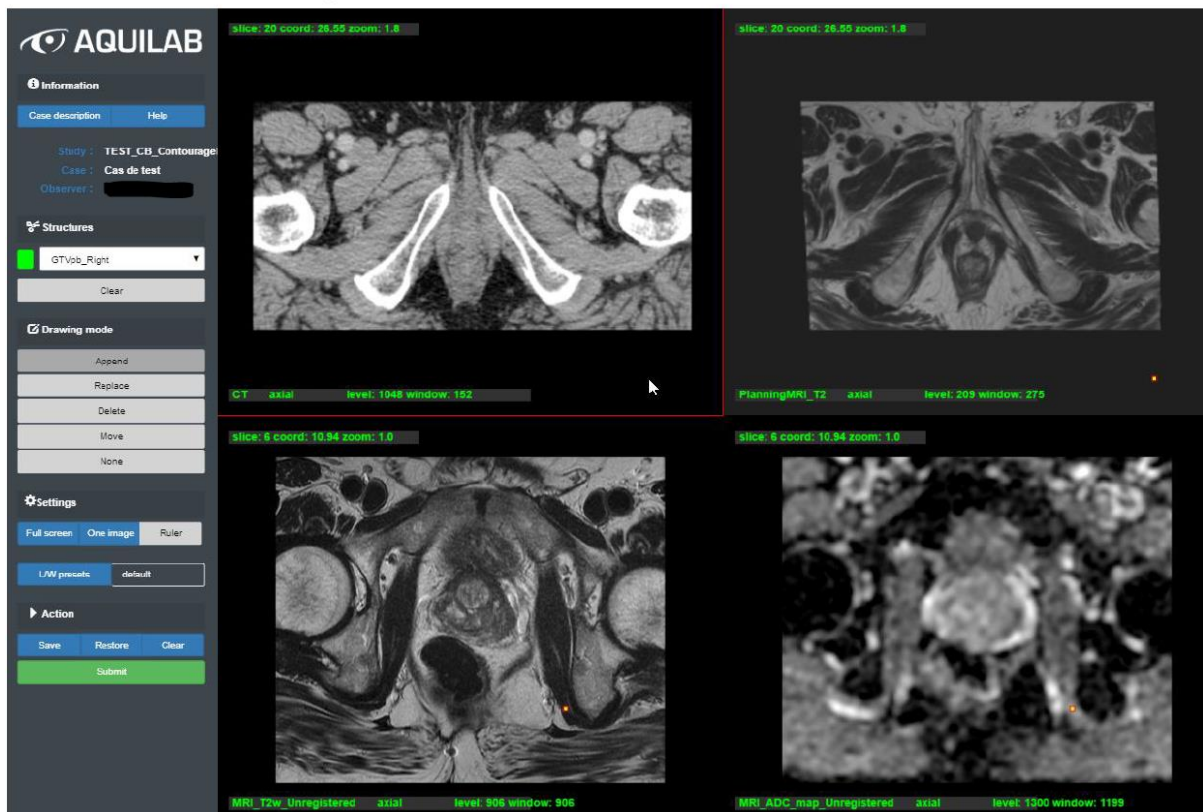

Figure 5.3: AQUILAB Share Place Online Outlining Platform: DIL outlining case used in RCR18 Outlining Workshop showing CT planning scan (top left), MRI planning scan (top right), diagnostic T2W mpMRI sequence (bottom left) and corresponding mpMRI ADC sequence. Note: in this figure, the T2W and ADC sequences are not registered to the planning CT and MRI.

5.3.3 Pre-workshop exercise

Two weeks prior to the workshop, AQUILAB registered delegates were sent a username and password to access the case via Share Place and an accompanying outlining software user manual. Delegates were also sent PIVOTALboost outlining guidelines (with permission of the PIVOTALboost Trials Unit) and instructions detailing the three target volumes to delineate and standardised nomenclature to use (Appendix 2). A clinical vignette (written by EE) detailing the patient's clinical history, histopathology results and MRI findings was also provided (Appendix 3).

Contour submission was open from 28th August to 7th September 2018, to allow time for submitted contour review using the radiotherapy outlining analysis platform, Artiview (AQUILAB) by the workshop leads (EE, JS, JM, IS, AT). It was intended that submitted contours would be reviewed to identify common themes of error by the delegates, to form part of the discussion at the educational workshop. Technical difficulties accessing the Artiview software by workshop leads however, meant delegate contours could not be adequately reviewed prior to the workshop.

5.3.4 Educational outlining workshop

52 delegates attended the outlining workshop. They were required to bring their personal laptops and a mouse with a wheel to the workshop to enable them to re-contour the same case during the workshop. The first hour of the workshop was dedicated to DIL outlining (EE, IS, JS, AT) and the second hour to prostate bed outlining (JM, AT). At the start of the session, a fifteen minute PowerPoint presentation (Appendix 4) was delivered to the delegates (EE) which outlined DIL definition, how to use MRI to define the DIL, DIL outlining steps with pictorial examples and common DIL outlining 'pitfalls'. Delineation guidance of the prostate was also included given the range of experience of delegates. Following the presentation, the delegates' contours were displayed anonymously against the reference volumes and overt sources of outlining 'errors' discussed for five minutes (IS). The delegates were then given approximately thirty minutes to re-contour the three target volumes of the same case via Share-Place. In the interest of time, delegates were asked to prioritise outlining of GTVpb_right and GTVpb_left (not GTVp) and to re-submit their contours at the end of the hour session.

5.3.5 Workshop questionnaire

Delegates were asked to complete a questionnaire prior to leaving the workshop. The first part of the questionnaire included questions regarding pre-workshop DIL outlining experience so that they could be correlated with delegate outlining performance (Appendix 5). Unfortunately, the user identification section referred to the delegates' individual user login for the Share Place platform that was different

to the candidate number assigned by AQUILAB to their contours, so I was unable to correlate the delegate's feedback with their respective contours. The second part of the questionnaire included questions regarding the impact of the outlining workshop on delegates' future practice, confidence in DIL and prostate bed outlining and experience of using the AQUILAB Share Place platform.

5.3.6 Post-workshop exercise

Following the workshop, the same DIL case was re-opened on AQUILAB Share Place platform for two weeks between 24th September 2018 and 10th October 2018. Delegates were encouraged to re-attempt the case and submit their contours within a month following the educational workshop.

5.3.7 Initial review of delegate contours

Thirty-two delegates submitted contours for the DIL case before the workshop, 29 during and 5 afterwards. Submissions that had nonsensical or incomplete contours were excluded from analysis. Of those who submitted contours GTVpb_left contours, 30 'before', 27 'during' and 5 'after' workshop contours were eligible for analysis. For GTVpb_right, 29, 28 and 5 eligible contours were available for the before, during and after workshop cohorts respectively. Only 5 candidates submitted contours after the workshop that were eligible for analysis; 3 delegates submitted GTVpb_left volumes for all three time points and only 2 submitted GTVpb_right volumes for all three time points. Therefore, I was unable to assess the impact the workshop had between the three sittings. Eighteen and 17 delegates submitted GTVpb_left and GTVpb_right volumes respectively both before and during the workshop so I focused on these cohorts to assess whether the workshop had an impact on DIL outlining performance immediately after the educational talk. A summary of the eligible delegate outlines is shown in Figure 5.4.
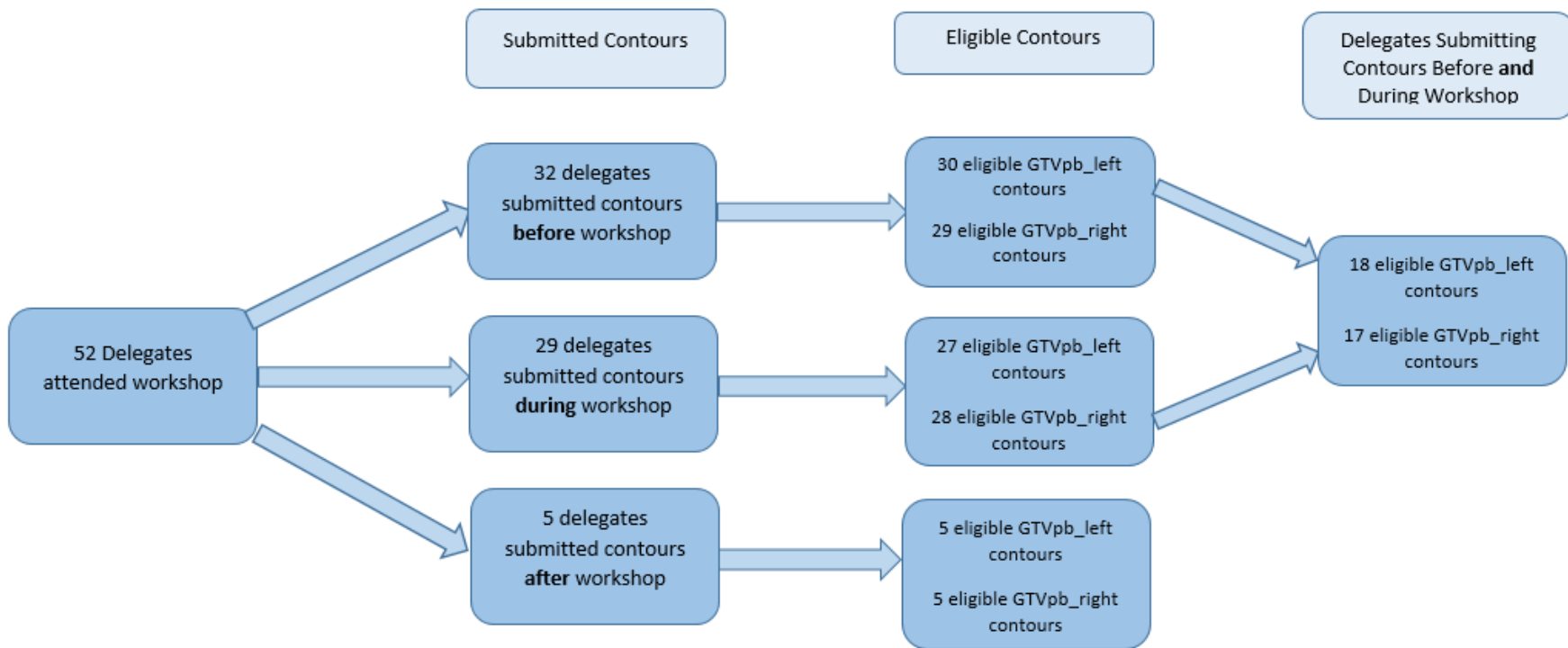
Figure 5.4: Summary of delegates who attended the workshop, submitted contours before, during and after the workshop and of those contours submitted, the number eligible for analysis

5.3.8 Qualitative analysis of delegate contours

All delegate contours were sent as three separate files (before, during and after workshop) with contours designated a four-digit candidate number by CB from AQUILAB to VCC as secure DICOM files. The three structure sets were imported into CERR and nonsensical/ incomplete volumes were excluded from further analysis as above. I re-allocated eligible candidate numbers in numerical order (1 upwards) corresponding to the numerical order of their four digit candidate number. Within CERR, I calculated the volume of each eligible GTVpb_left and GTVpb_right (in cc). I then qualitatively assessed the delegates' GTVpb structures against the 'unacceptable' variations described by the PIVOTALboost pre-accrual benchmark case outlining criteria:

- GTVpb breaches CTVp
- GTVpb outside correct region of prostate
- GTVpb exceeds +/- 1 slice superiorly and/or inferiorly
- GTVpb investigator volume exceeds +/- 25% of the reference volume

I assessed the performance of delegates between each time point i.e. before, during and after the workshop, as cohorts. Where there was no delegate GTVp outline, I used the reference GTVp structure to assess GTVpb breach of the prostate. If delegates had no unacceptable variations, their GTVpb volumes were considered 'acceptable'. If their volumes had one or more 'unacceptable' variations, then the GTVpb volume was considered 'unacceptable'.

5.3.9 Quantitative analysis of delegate contours

All three eligible structure sets (before, during and after workshop) were sent as DICOM files to ES, Professor of Healthcare Engineering, Cardiff University, who ran the delegates GTVpb_left and GTVpb_right structures within CERR against an in-house script which generated RTOG, van't Riet, Jaccard and DICE conformity indices values for each delegate structure. In Chapter 4, I was unable to determine a specific conformity index that could be used to pass/fail PIVOTALboost pre-accrual GTVpb attempts. However, conformity indices have been frequently used as a comparative quantitative metric assessing the impact of educational interventions on radiotherapy delineation, with DICE similarity coefficient being the most common [2]. Therefore, I used DICE as a metric to compare the performance of delegate outlines at the three separate time points as a measure of inter-observer variation. As discussed in chapter 2, perfect conformity of the delegate outline with the reference volume would have a DICE score of 1, whereas outlines with no conformity to the reference volume would have a DICE score of 0. Mean/SD GTVpb_left/ GTVpb_right DICE scores and volumes for eligible delegate contours at each sitting were calculated in excel.

As above, I was unable to assess the impact the workshop had on delegate outlining performance between the three attempts and focused on whether there was a statistical comparison between delegate GTVpb DICE and volume data, before and during the workshop. Eighteen and seventeen delegates submitted GTVpb_left and GTVpb_right volumes respectively both before and during the workshop. Given the sample size for each group was less than thirty, I used the Wilcoxon signed rank test to compare the delegate DICE and volume data between the before and during contours as this was paired data.

5.4 Results

The number of delegates attending the conference and those included for analysis is summarised in Figure 5.4. Screenshots of delegate contours are shown in Figure 5.5.
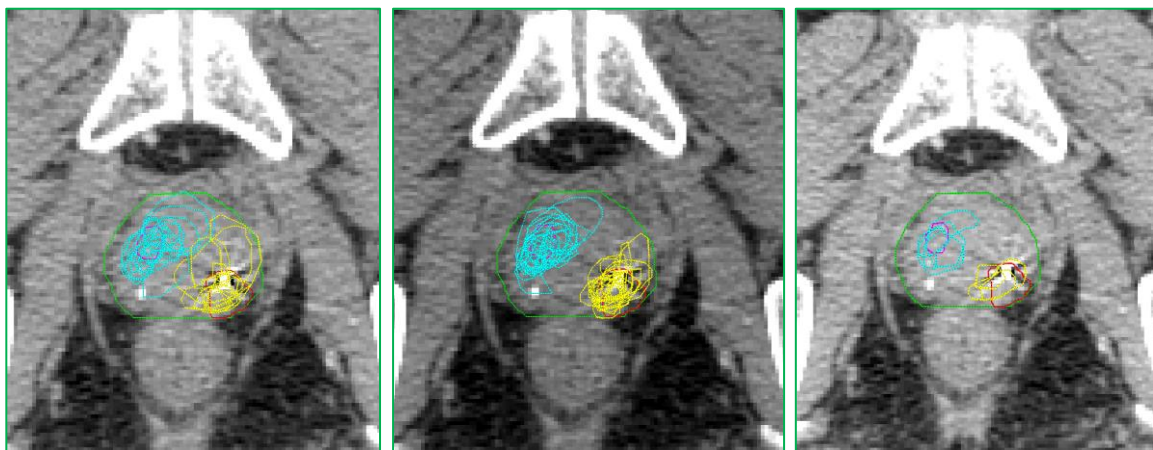

Figure 5.5: Axial cross section of CT planning scan showing delegate contours of GTVp (green), GTVpb_right (yellow) and GTVpb_left (red), before (left image), during (middle image) and after (right image) the workshop

For GTVpb_right, 6/30 (20%) candidates had acceptable volumes before, 4/27 (14%) during and 1/5 (20%) after the workshop. For GTVpb_left, no delegates had acceptable volumes before and after the workshop and only 3 (11%) delegates had acceptable volumes during the workshop (Figure 5.6).
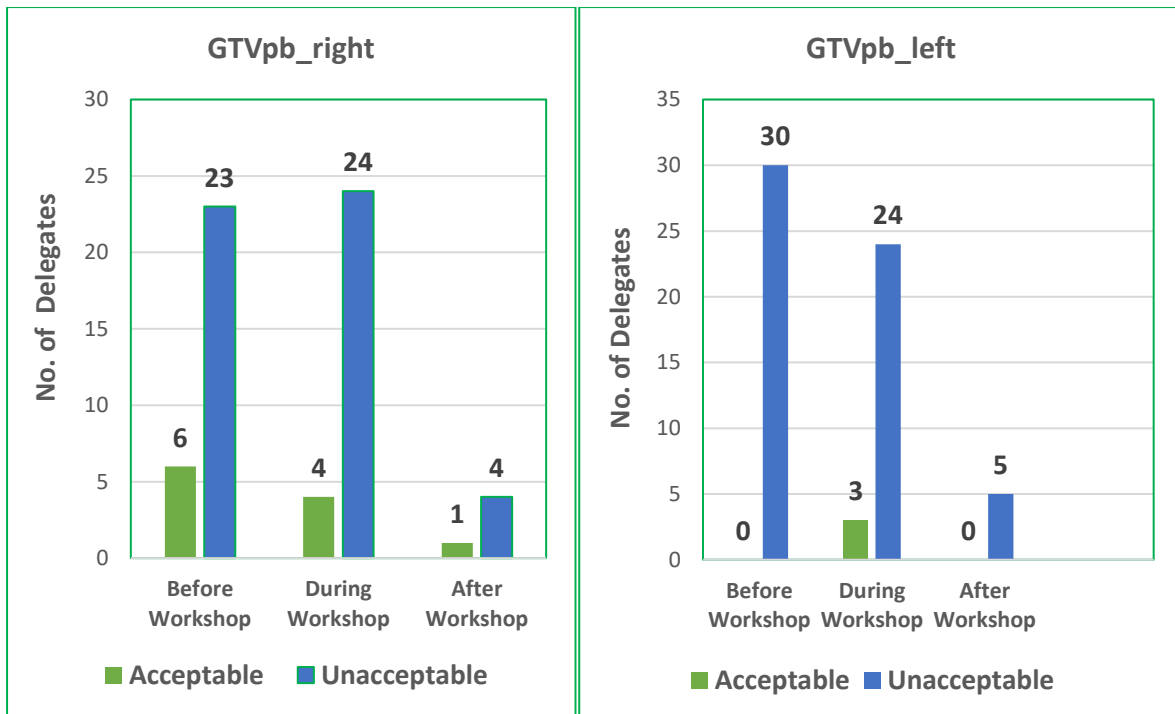
Figure 5.6: Bar chart showing the number of delegates who had 'acceptable' versus 'unacceptable' GTVpb_right and GTVpb_left volumes before, during and after the workshop.

The most common unacceptable GTVpb_right variation before, during and after the workshop was delegate volume being >+/- 25% of the reference volume. The most common unacceptable GTVpb_left variation before and during the workshop was delegate volume being >+/- 25% of the reference volume although more candidates also had unacceptable superior/inferior extension compared to the GTVpb_right and this was the most common unacceptable variation following the workshop (Table 5.1).

GTVpb Unacceptable Parameters	GTVpb_right			GTVpb_left		
	Before Workshop (n=29 Delegates)	During Workshop (n=28 Delegates)	After Workshop (n=5 Delegates)	Before Workshop (n=30 Delegates)	During Workshop (n=27 Delegates)	After Workshop (n=5 Delegates)
Outside CTVp	0	0	0	0	0	0
Incorrect DIL Outlined	1	1	0	1	1	0
>3mm Superior/Inferior	4	4	1	23	11	3
>+/- 25% Reference Vol.	21	23	3	25	19	1

Table 5.1: Table showing causes of unacceptable variation in GTVpb_left and GTVpb_right outlining

5.4.2 Quantitative assessment: volume

The volume of GTVpb_right and GTVpb_left reference volumes were 1.42cc and 1.17cc respectively. The volumetric descriptive statistics for all delegates who submitted eligible contours before, during and after workshop cohorts are shown in Table 5.2. For both GTVpb_right and GTVpb_left, average delegate volumes decreased across the three consecutive time points and were consistently below the absolute reference volume.

	GTVpb_right			GTVpb_left		
	Before Workshop	During Workshop	After Workshop	Before Workshop	During Workshop	After Workshop
GTVpb Volume(cc)	Reference volume 1.42cc			Reference volume 1.17cc		
Mean ± SD	1.15 ± 0.15	1.00 ± 0.66	0.86 ± 0.25	0.96 ± 1.06	0.78 ± 0.37	0.73 ± 0.37
Minimum	0.36	0.45	0.53	0.19	0.07	0.19
Max	4.24	2.93	1.11	5.49	1.58	1.17

Table 5.2: Table showing GTVpb_right and GTVpb_left volumetric statistics for all delegates before, during and after the workshop

The volumetric statistics for those delegates who submitted contours both before and during the workshop and results of statistical comparisons are shown in Table 5.3. There was no significant difference in GTVpb_left volumes between the before and during cohorts but there was a significant decrease in delegate GTVpb_right volumes. However, for GTVpb_left, the standard deviation decreased between attempts suggesting a reduction in inter-observer variation.

	GTVpb_right		GTVpb_left	
	Before Workshop	During Workshop	Before Workshop	During Workshop
GTVpb Volume (cc)	Reference volume 1.42cc		Reference volume 1.17cc	
Mean ± SD	1.26 ± 0.92	1.03 ± 0.62	0.89 ± 0.69	0.87 ± 0.37
Minimum	0.43	0.45	0.19	0.39
Max	4.24	2.9	2.51	1.58
Wilcoxon Signed Rank Test p-value	0.04		0.71	

Table 5.3: Table showing GTVpb_right and GTVpb_left volumetric data and the comparative statistics for delegates who submitted contours both before and during workshop.

The distribution of delegate volumes are shown in Figures 5.7 and 5.8. They also illustrate that for both GTVpb_right and GTVpb_left range of delegate volumes reduce immediately following the educational talk on DIL outlining (i.e. during the workshop) compared to before the outlining workshop.
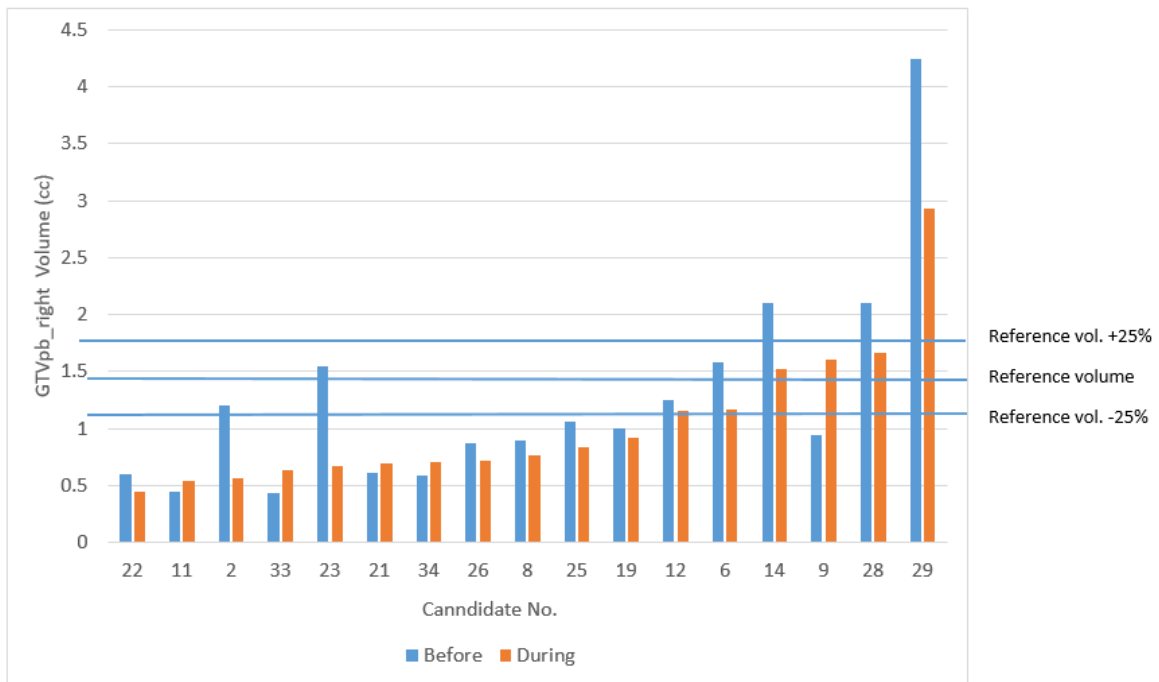

Figure 5.7: Bar chart showing individual delegate GTVpb_right volumes before and after the workshop in ascending order as per the 'during' volumes with reference to the +/-25% reference volume qualitative criteria
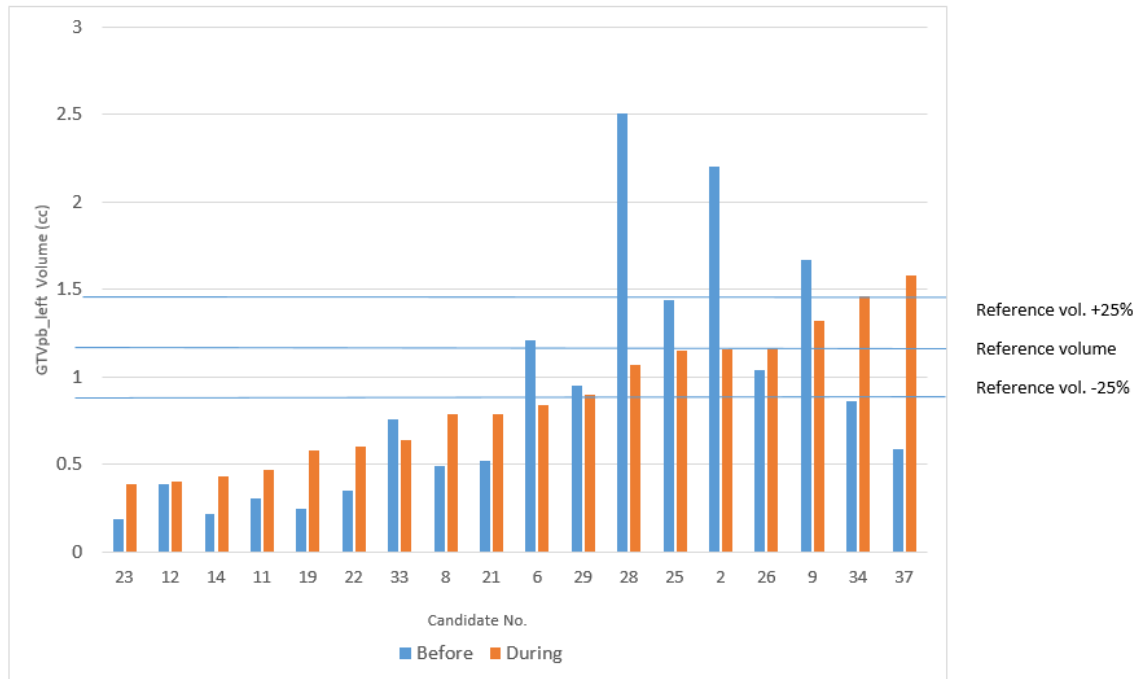
Figure 5.8: Bar chart showing individual delegate GTVpb_left volumes before and after the workshop in ascending order as per the 'during' volumes with reference to the +/-25% reference volume qualitative criteria

5.4.3 Quantitative assessment: DICE scores

The DICE score descriptive statistics for all delegates who submitted eligible contours before, during and after workshop cohorts are shown in Table 5.4. GTVpb_right DICE scores were higher across the three time points than for GTVpb_left, but where GTVpb_right scores stayed consistent, GTVpb_left DICE scores rose incrementally with consecutive attempts.

	GTVpb_right			GTVpb_left		
	Before Workshop	**During Workshop**	**After Workshop**	**Before Workshop**	**During Workshop**	**After Workshop**
DICE Score						
Mean ± SD	0.43 ± 0.14	0.38 ± 0.17	0.41 ± 0.13	0.17 ± 0.13	0.27 ± 0.18	0.35 ± 0.20
Minimum	0.02	0	0.19	0.00	0.00	0.00
Max	0.63	0.71	0.49	0.50	0.61	0.49

Table 5.4: Table showing GTVpb_right and GTVpb_left DICE score statistics for all delegates before, during and after the workshop

The quantitative descriptive statistics for those delegates who submitted contours both before and during the workshop and results of statistical comparisons are shown in Table 5.5. There was no

statistical significance in DICE score for GTVpb_right delineation between the sittings, but there was a statistically significant improvement in GTVpb_left DICE scores. The distribution of delegate DICE scores (Figures 5.9 -5.10) further illustrate these trends.

	GTVpb_right		GTVpb_left	
	Before Workshop	During Workshop	Before Workshop	During Workshop
DICE Score				
Mean ± SD	0.44 ± 0.15	0.43 ± 0.13	0.17 ± 0.16	0.32 ± 0.16
Minimum	0.12	0.14	0.00	0.05
Max	0.63	0.71	0.50	0.61
Wilcoxon Signed Rank Test p-value	0.46		0.01	

Table 5.5: Table showing GTVpb_right and GTVpb_left DICE sore metrics and the respective comparative statistics for delegates who submitted contours both before and during workshop
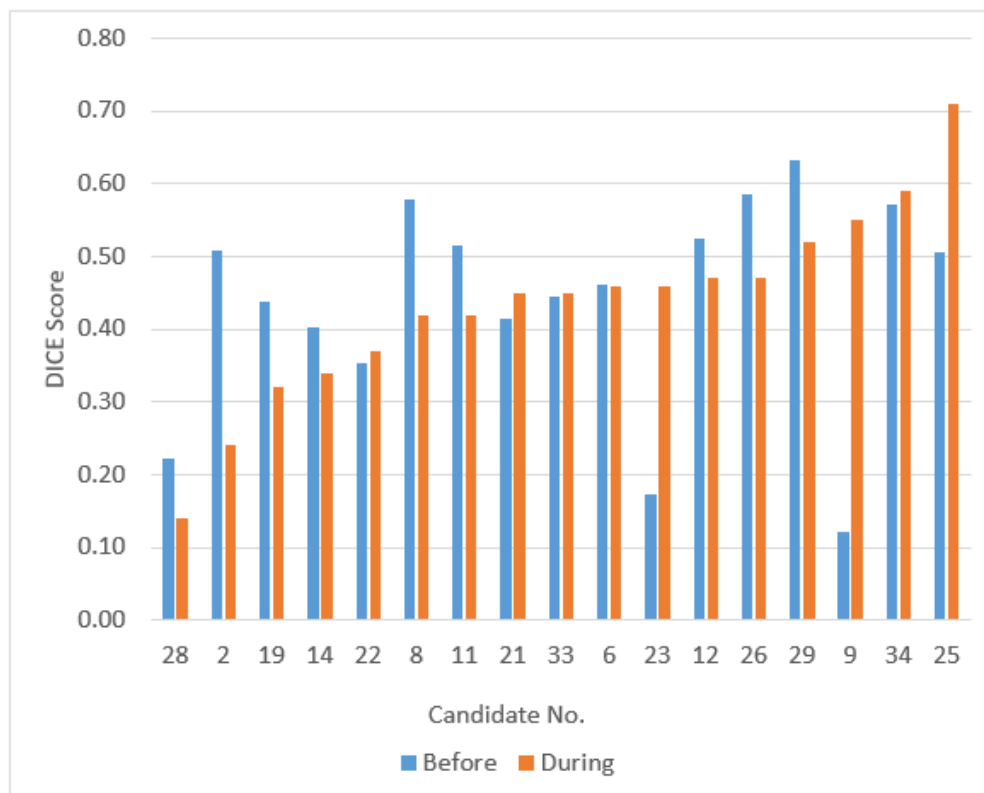

Figure 5.9: Bar chart showing individual delegate GTVpb_right DICE scores before and during the workshop in ascending order as per the 'during' scores
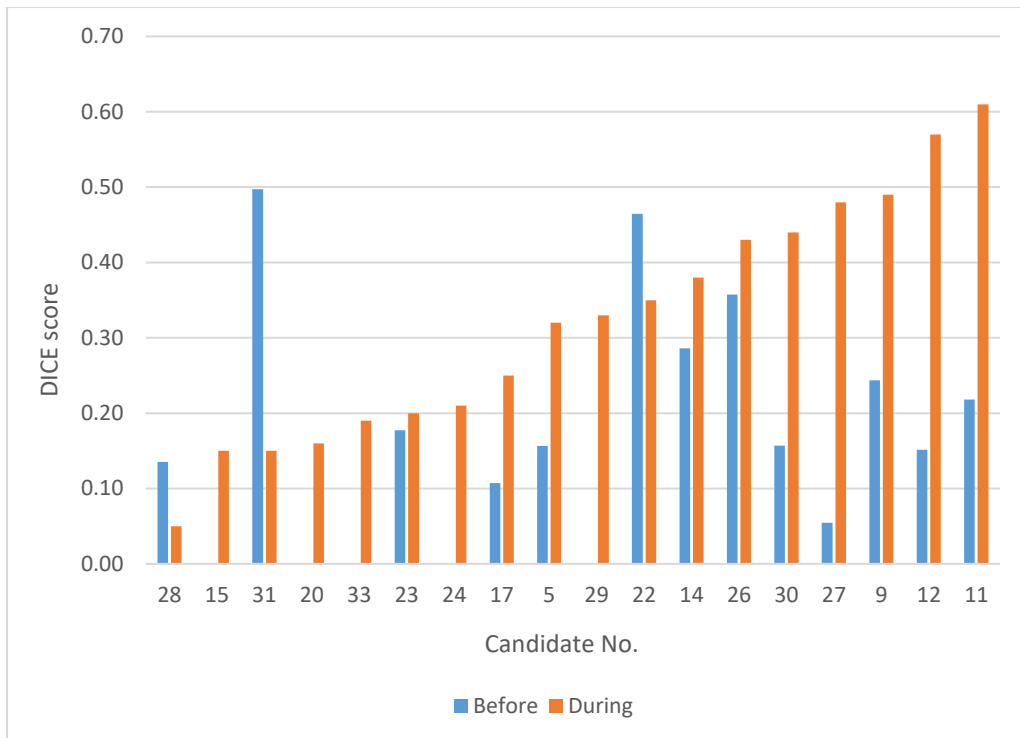
Figure 5.10: Bar chart showing individual delegate GTVpb_left DICE scores before and during the workshop in ascending order as per the 'during' scores

5.4.4 Questionnaire

Thirty-one delegates completed the 'pre-workshop' questionnaires and 29 completed 'post-workshop' questionnaires at the end of the workshop. Results are shown in Figures 5.11-5.13. Most of the attendees were consultants with three or more years prostate radiotherapy experience, although only 3/31 delegates (10%) reported having delineated a DIL prior to the outlining exercise. However, 8/31 (26%) delegates had attempted the PIVOTALboost pre-accrual cases prior to the workshop, but only 3/8 (38%) were approved to treat patients within the trial and only 1/8 (13%) attended the PIVOTALboost pre-trial workshop. There was an improvement in delegates' theoretical DIL outlining confidence immediately after the workshop than beforehand.
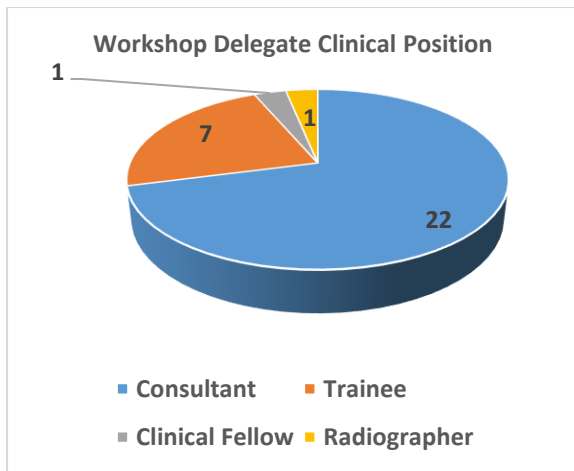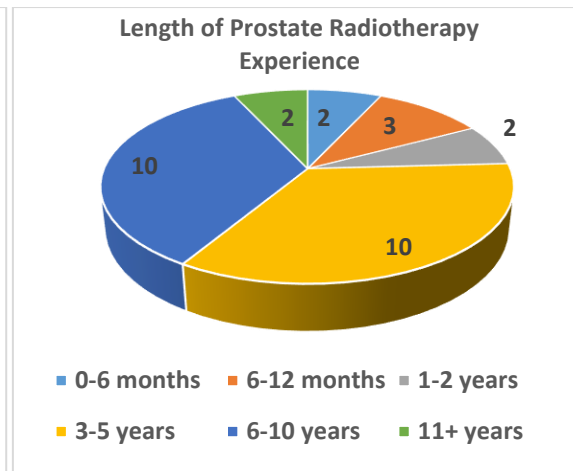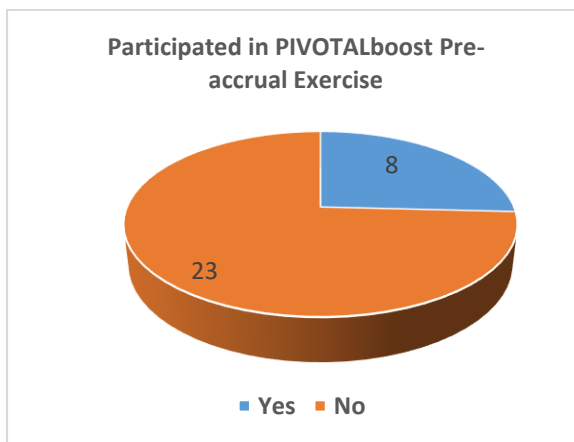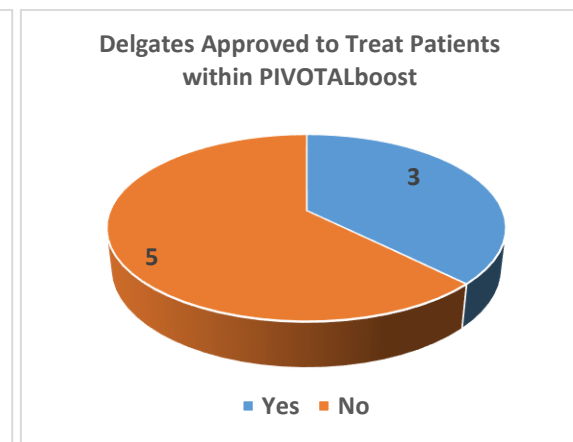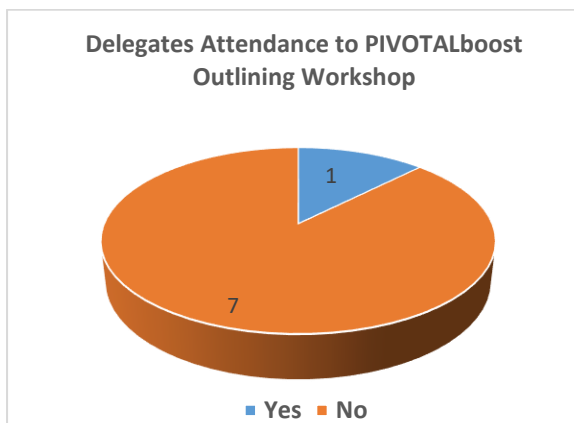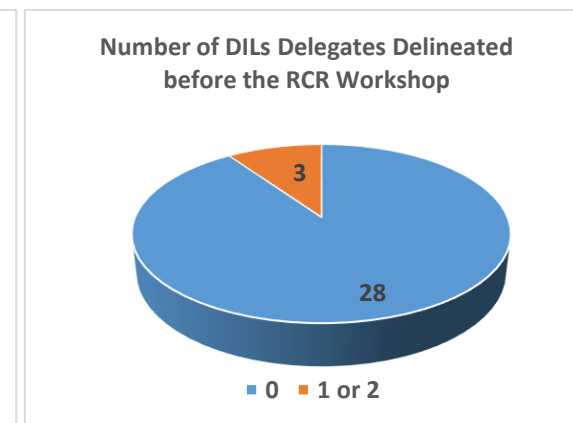
Figure 5.11 (1-7): Pie charts showing the results for each question in the feedback form returned by delegates who attended the workshop

Number of Delegates who had Radiologist Assistance for DIL Delineation
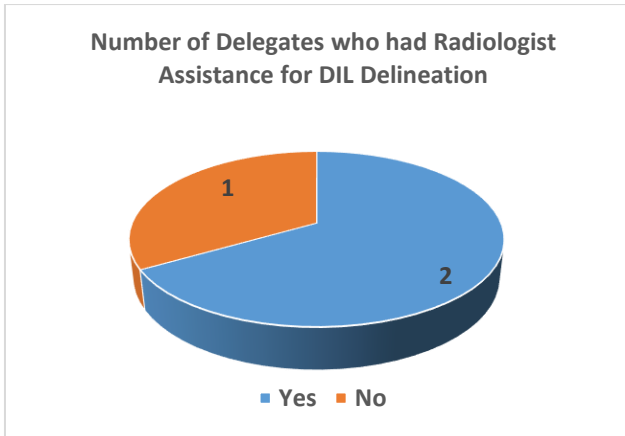
7
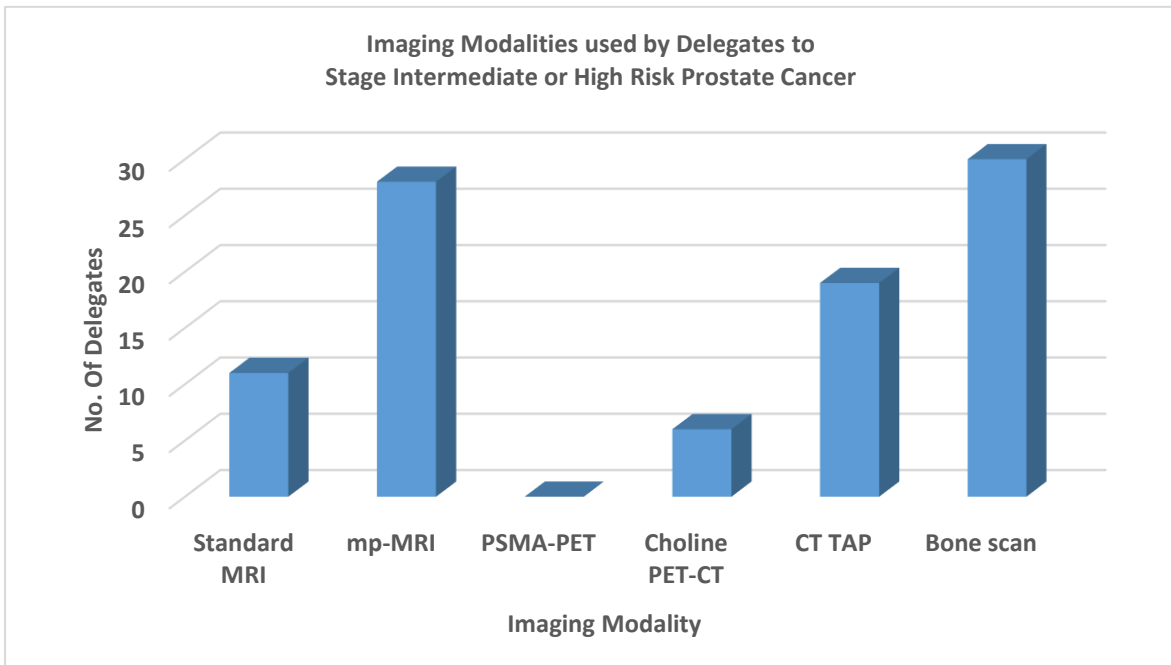
Figure 5.12: Bar chart showing types of imaging modalities delegates use routinely in diagnostic work up of intermediate- or high-risk prostate cancer patients (could choose more than one)

Delegate Confidence in DIL Delineation Pre- and Post-Workshop

	Not a clue	Know in theory, not confident	Know in theory, can perform some parts	Know in theory, confident	Competent. needs minimal support
Pre-Workshop	12	8	7	1	1
Post-Workshop	1	6	15	8	1

Figure 5.13: Bar chart showing delegate confidence in DIL delineation before and immediately after the outlining workshop

5.5 Discussion

The outlining workshop was well attended and was the first time the RCR had used the Share Place contouring platform (AQUILAB) for an outlining workshop. Despite delegates having used the platform for the first time, most delegates were able to submit at least one contour for review.
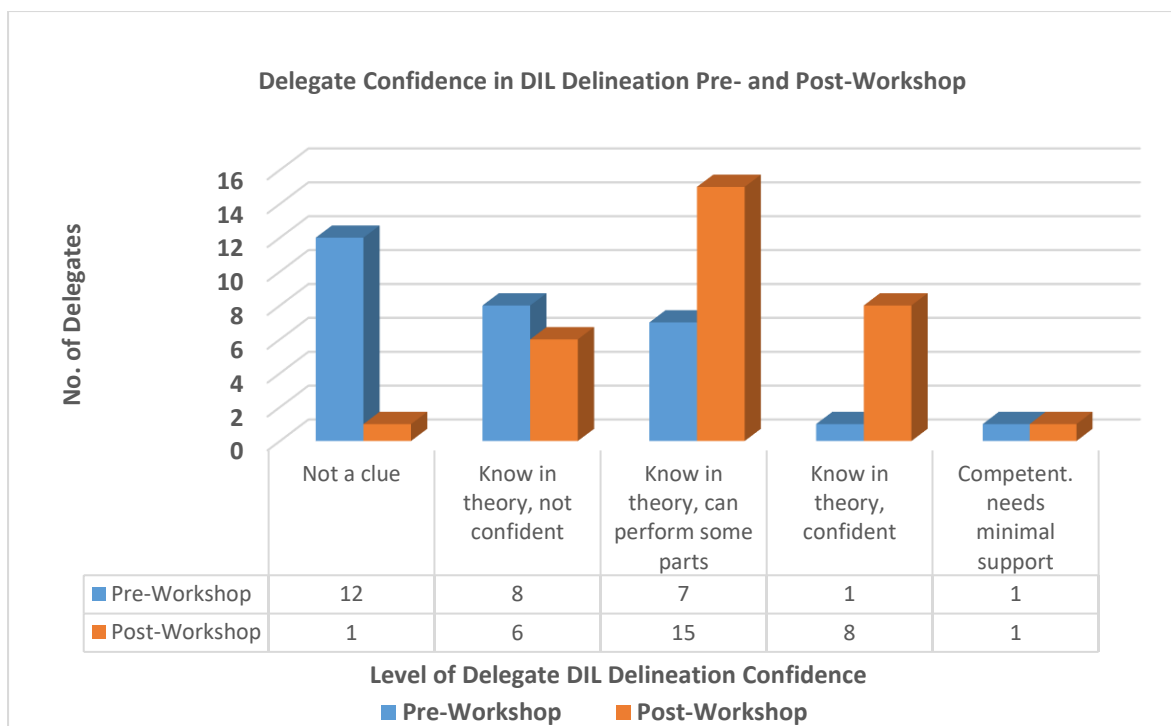
With respect to qualitative assessment, there was a noticeable difference between the GTVpb_right and GTVpb_left outlining performances of delegates. Although delegates were asked to outline a single case, there were two DILS to contour, both of which seemingly posed different challenges. Despite high numbers of +/- 25% of reference volume 'unacceptable' variation for both DILs, GTVpb_left also had relatively a high number of superior/inferior extent 'unacceptable' variation. The GTVpb_left reference volume was only 3 CT slices long compared to the GTVpb_right volume which was 4 slices and initially I thought delegates may have found it more difficult to cognitively transpose a shorter volume from mpMRI onto CT. However, the Share Place outlining platform was set up so that the four viewing windows showing the planning CT, planning MR, mpMRI T2W and mpMRI ADC map were registered meaning all sequences scrolled together in alignment and delegates could see the corresponding MRI sequences whilst delineating on CT, a positive feature not afforded to standard radiotherapy planning systems. This suggests that mpMRI interpretation of the DIL may have been more of a challenge for GTVpb_left rather than GTVpb_right.

More delegates had acceptable volumes across the three time points for GTVpb_right then left. Not only, as discussed above, did delegates struggle with GTVpb_left length, it was also smaller in volume than GTVpb_right. As discussed in chapter 4, smaller volume lesions are more likely to be 'unacceptable' on qualitative assessment due to the +/- 25% of reference volume parameter being applicable to lesions of all sizes.

Interestingly, for both DILs, no delegates across any of the three sittings had GTVpb volumes that breached GTVp, yet this was seen in both case 1 and 2 of the PIVOTALboost pre-accrual cases (Chapter 3). This could be because workshop delegates had to delineate their own GTVp (although not prioritised during the workshop). It could also be due to both DILs' location in the mid-prostate and not extending into the apex or base of the prostate, where prostate outlining is most inconsistent [3]. Also GTVpb_right was a fairly central lesion and therefore away from the prostate capsule (i.e. prostate edge) making it easier to avoid breaching GTVp. Although GTVpb_left was a peripheral zone lesion and therefore adjacent to the capsule, there was a clear demarcation between the prostate and mesorectum, so delegates were less likely to extend outside the prostate, particularly given the senior prostate radiotherapy experience of the audience.

Only one delegate before the workshop and one after delineated either GTVpb_right or GTVpb_left in the incorrect region of the prostate. This is likely to be due to the presence of two eligible DILs to delineate in the same prostate and therefore delegates identifying the two most 'obvious' DILs and not an 'ineligible' one. Also, delegates were aware from the instructions they were expecting to identify a 'left' and a 'right' DIL, which is likely to have also helped them to correctly identify the two lesions.

With respect to DICE scores, a different trend was seen between the two different GTVpb volumes across the three sittings. The GTVpb_right DICE scores stayed fairly consistent between the three time points and were higher than the corresponding GTVpb_left DICE scores, the latter incrementally improving across the three time points. The mean DICE score for the first submissions of PIVOTALboost pre-accrual case 1, was higher (0.59 ± 0.1 SD) than any of the workshop cohorts. This may be explained by the difference in populations participating in the trial exercise versus the workshops. Although most of the workshop delegates were senior clinicians, as per the pre-workshop questionnaire, nearly a third reported they had 'no clue' about DIL delineation and by attending an outlining workshop are more likely to want to improve their radiotherapy skills. Also, there was a time limit of two weeks to complete the outlining exercises before and after the workshop, which can be difficult to accommodate amongst other clinical commitments and there was a much greater time pressure during the workshop itself which may have impacted delegate performance. Additionally, the

workshop exercise was for clinicians to improve their outlining skills, with no immediate consequence if they performed poorly. However, for the PIVOTALboost pre-accrual exercise, until submission of acceptable volumes, they were unable to recruit any of their patients into the trial.

In terms of volume of GTVpb, for both DILs, but particularly GTVpb_left, the trend across the three time points was a reduction in mean delegate volume, in keeping with other studies that have evaluated the impact of a workshop on radiotherapy outlining [4], suggesting outliners become less cautious after an educational intervention. However, the mean delegate volume for both DILs at each time point was smaller than their respective reference volume. Therefore, although delegate volumes got smaller following the educational intervention, the mean values deviated further away from the reference volume, therefore actually introducing greater risk of geographical miss. This is in stark contrast to the pre-accrual PIVOTALboost case 1 data (Chapter 4) where delegates tended to over-outline the DIL and therefore the most appropriate conformity indices to apply to that data set were measures of over-outlining (DI and MDC over-outlining). Delegates from the workshop however consistently under-outlined the DIL suggesting that DI and MDC over-outlining would not have been appropriate measures of conformity for this data set.

One of the main differences between the PIVOTALboost pre-trial cases and the RCR workshop case was the software used to outline. For the pre-trial cases, outliners would have had to use their in-house planning software , which although may allow you to fuse the diagnostic images with the CT planning scan (not recommended in the PIVOTALboost protocol), does not have the functionality for outliners to observe four different imaging sequences at the same time. For the Share Place platform, all four sequences could be visualised at the same time whilst outlining and scrolled through in alignment. It could be, therefore, that cognitive transfer of DIL lesions from T2W/ADC sequences required in the PIVOTALboost outlining process may add an element of uncertainty that cause outliners to delineate volumes which are larger, to ensure the lesion is encompassed.

The mean delegate DICE and volume values for those candidates who completed both before and during GTVpb contours were in keeping with the corresponding whole group values. Although there was no difference in the mean GTVpb_right DICE scores or volume between the cohort who submitted contours both before and during the workshop, there was a significant improvement in conformity of delegate GTVpb_left contours immediately following an educational talk on DIL outlining compared to before. The extremity of non-conformity in the GTVpb_left 'before' workshop cohort (five delegates in the 'before' comparative cohort had a DICE score of 0 i.e. no overlap with the reference volume), lent itself to room for improvement following the educational intervention and probably why an improvement in outlining can be seen.

Conversely, there was no significant difference in the mean volume of delegate GTVpb_left contours before and during the workshop but there was for GTVpb_right (i.e. they became significant smaller during the workshop). Therefore, even though both average DIL volumes were even smaller than their respective reference volumes after the interventions, delegate conformity to the reference volume as a measure of volume overlap, either stayed the same (GTVpb_right) or improved (GTVpb_left). These quantitative metrics in part (particularly the DICE scores) mirror the findings of the qualitative assessment. For GTVpb_right, more delegates had unacceptable variations during the workshop that in the before session, so a lack of improvement in DICE score is unsurprising. However, for GTVpb_left, less than half the number of delegates had superior/inferior extent unacceptable variation, during the workshop than they had beforehand and this was reflected in an improvement in DICE scores.

Whilst systematic reviews have demonstrated an improvement in TVD following an educational intervention [2, 5], some studies show educational interventions have a mixed impact on TVD performance. In a study evaluating the impact of teaching on outlining variation during a virtual stereotactic ablative radiotherapy (SABR) outlining workshop, Slevin *et al* [6] identified a significant improvement in DICE scores following the workshop of only three out of twelve structures evaluated. The authors identified a number of limitations in their study which may have impacted the efficacy of their intervention including heterogeneity in experience of participants, time pressures of the workshop and the same case used pre-and post-educational workshop, meaning delegates are more likely to submit a volume following an intervention based on their initial attempt. This latter point may explain another finding from my data. Although most delegates under-contoured GTVpb_left and GTVpb_right, as individuals they appeared to stay fairly consistent in their volume size between the two sittings i.e. those who outlined smaller volumes before the workshop generally outlined smaller volumes during the workshop and those who outlined large volumes before the workshop, also had a tendency to outline larger volumes during the workshop (Figures 5.8 and 5.9). D'Souza *et al* [7] assessed the impact of a teaching programme on head and neck outlining and found no improvement in post-educational contours, attributing this to the lack of integrated outlining teaching during the seminar. Although we did offer a practical outlining session as well as a didactic lecture in the RCR workshop, the session was very time pressured and some delegates experienced technical difficulties with the software, putting further pressure on submission of optimal GTVpb volumes.

From the pre-workshop questionnaire, most of the delegates were senior clinicians with a number of years' experience of prostate radiotherapy. However, most delegates did not have DIL outlining experience. Of those eight delegates that did participate in the PIVOTALboost pre-trial exercise, only one attended the workshop and three approved (at the time of the workshop) to treat patients within the trial. Unfortunately, due to the different user identification used on the survey, I was unable to

correlate those delegates who had DIL outlining experience and approved to treat patients within PIVOTALboost, with their GTVpb contours submissions, to see if their structures were more conformal with the reference volume and whether they had fewer or no unacceptable variations at any time point. What this survey has highlighted however, is that DIL outlining is a relatively new skill and even those who have years of prostate radiotherapy experience are unfamiliar with this technique.

Most delegates use mpMRI as standard in their diagnostic work up of intermediate- and high-risk patients, although what is unclear from this study is whether mpMRI DIL interpretation is lacking among prostate oncologists and hence relatively poor DIL outlining skills, or whether finalising the CT volume is causing delineation difficulty. As the workshop took place in 2018, it is unsurprising that no delegates reported the routine use of PSMA-PET CT in their routine practice; in Wales, it was commissioned by the Welsh Health Specialised Services Committee (WHSCC) for routine use in high risk patients in May 2021 [8], although recommendation as per the PIVOTALboost trial is to use mpMRI for outlining

The post-workshop questionnaire showed that before the workshop, most candidates reported they 'had no clue' about DIL delineation, but following the workshop, most candidates reported they 'knew in theory and were confident in some parts' suggesting the workshop had a positive impact on delegates DIL outlining confidence. However, this did not directly translate into an overall improvement in outlining ability for the GTVpb_right structure. In their study, D'Souza *et al* [7] found that participants in their head and neck outlining workshop demonstrated significant improvements theoretical knowledge and had positively rated participant satisfaction, yet there was only a significant improvement in three out of 20 structures per- and post-workshop. The authors attributed this to a lack of practical component to their workshop, and although our workshop did have a practical session, this highlights the need for oncologists to have opportunity to practice their outlining skills out of a clinical capacity. A lack of access to TVD resources with feedback in an educational setting has been identified, prompting the RCR to facilitate online TVD training with practice cases to outline through the ARENA project [9].

As alluded to above, there were a number of limitations associated with this study. Firstly, the workshop format was not ideal. The workshop was the last session following a full day of didactic lectures on prostate cancer and delegates fed back that the workshop would have been better at another time of day (e.g. lunchtime as a standalone session), suggesting fatigue may have impacted delegate engagement or performance. Candidates were also unfamiliar with AQUILAB Share Place, as it was the first time the RCR has used this platform in their outlining workshops. Some delegates did require technical support before and during the session, and some of those with Mac software had

technical difficulty viewing the four sequences together, which may have impacted some delegates' ability to submit contours. The workshop also focused on two sub-sites of prostate radiotherapy; as the DIL session was first, some delegates may have felt rushed to complete their contours. It was also in the interest of time, why we advised delegates to not delineate GTVp, which may have impacted delegates' ability to orientate where the DILs were located within the prostate. Only five delegates' submitted volumes following the workshop which meant although we tested delegates' DIL outlining immediately after the didactic teaching session, short term retention could not be adequately assessed. This was also experienced by Slevin *et al* [10] following their RCR pancreatic SABR outlining workshop who could only reliably identify seven participants who produce pre- and during workshop contours and of these only three who submitted contours following the workshop. Although the authors do not stipulate the reason why they could not 'reliably identify' and pair all participants contours, they had used same course organisers and software as I had.

For my session, a different candidate number was allocated to delegates' contours compared to platform their user identification which meant I was unable to correlate the pre-workshop feedback regarding their prostate radiotherapy and DIL outlining experience with their outlining performance. As part of their recommendations for future workshops, Slevin et al [10] suggest clearer identification of paired participants contours and greater encouragement of participants to complete post-workshop contours, possibly by provision of individual feedback. I did not provide individual feedback on delegates' outlining performance other than experts providing direct support during the workshop. This was predominantly due to time constraints. Qualitative feedback is likely to be of more translatable clinical benefit to delegates than quantitative metrics [10], but is time-consuming, labour intensive and likely to be unsustainable for larger workshops unless a pool of experts are recruited. Longer term retention of knowledge could not be assessed in my study. This was initially due to delays in the RCR/AQULIAB being able to re-open the case for delegates, but the emergence of the COVID-19 pandemic subsequently thwarted any further attempts to re-send the case due to unpredictability of clinicians' commitment to non-essential clinical work.

5.6 Conclusion

Following a teaching session on DIL outlining, there was some improvement in DIL outlining skills although short and long-term retention of skills could not be assessed. Delegates perceived DIL outlining knowledge as reflected in their self-reported theoretical confidence, improved following the outlining workshop. However, DIL outlining skills generally remain poor amongst prostate oncologists with few submitting 'acceptable' contours in this study, also reflected in the PIVOTALboost pre-accrual case submission (Chapter 3). The exact cause of why experienced prostate radiotherapy outliners

struggle with DIL outlining remains unclear. In the following chapter, I will aim to establish whether this may be attributed to mpMRI interpretation skills or ability to transfer mpMRI images cognitively to the planning CT from the diagnostic mpMRI.

5.7 References

1. AQUILAB. [Internet] [Cited 2023 Feb 04]. Available from: https://www.aquilab.com/

2. Cacicedo J, Navarro-Martin A, Gonzalez-Larragan S, De Bari B, Salem A, Dahele M. Systematic review of educational interventions to improve contouring in radiotherapy. Radiother Oncol. 2020 Mar; 144: 86-92.

3. Villeirs G, Van Vaerenbergh K, Vakaet L, Bral S, Claus F, De Neve WJ, et al. Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer. Strahlenther Onkol. 2005 Jul; 181(7): 424-30.

4. Kumar S, Holloway L, Roach D, Pogson E, Veera J, Batumalai V, et al. The impact of a radiologist-led workshop on MRI target volume delineation for radiotherapy. J Med Radiat Sci. 2018 Dec; 65(4): 300-310.

5. Vinod S, Min M, Jameson M, Holloway L. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol. 2016 Jun; 60(3): 393-406.

6. Slevin F, Rieu R, Beasley M, Speight R, Aitkin K, Good J, et al. Evaluation of the impact of teaching on delineation variation during a virtual stereotactic ablative radiotherapy outlining workshop. J Radiother Pract. 2023; 22; e19

7. NHS Wales Welsh Health Specialised Services Committee: Specialised Services Commissioning Policy: CP50a. Version 7.0 May 2021.

8. D'Souza L, Jaswal J, Chan F, Johnson M, Tay KY, Fung K, Palma D. Evaluating the impact of an integrated multidisciplinary head & neck competency-based anatomy & radiology teaching approach in radiation oncology: a prospective cohort study. BMC Med Educ. 2014 Jun 26; 14: 124.

9. Evans, E, Radhakrishna, G, Gilson, D, Hoskin P, Miles E, Yuille F, et al. Target volume delineation training for clinical oncology trainees: the Role of ARENA and COPP. Clin Oncol (R Coll Radiol) 2019 Jun; 31(6): 341–343.

10. Slevin F, Beasley M, Speight R, Lilley J, Murray L, Hawkins M, et al. Evaluation of Clinician Contouring for Pancreatic Stereotactic Ablative Radiotherapy During a Contouring Workshop Organised by the Royal College of Radiologists. Clin Oncol (R Coll Radiol). 2021 Mar; 33(3) :e196-e197.

Chapter 6: Investigation into the Causes of Variations During the DIL Outlining Process

6.1 Introduction

As suggested in Chapters 3 and 5, DIL outlining skills amongst experienced prostate outliners seem to be generally lacking. As per the PIVOTALboost contouring guidelines [1], DIL outlining requires mpMRI interpretation and cognitive transfer of the defined lesion on the diagnostic mpMRI onto the planning scan, usually a CT scan. From the workshop data in Chapter 5, access to mpMRI as part of routine diagnostic work-up is relatively high amongst UK prostate cancer centres. However, there are no studies evaluating oncologists and non-medical outliners (NMOs) in mpMRI interpretation and it is not clear whether inadequate mpMRI interpretation skills or inability to cognitively transfer the DIL as determined on mpMRI onto planning CT, is the predominant reason for the errors in outlining. I therefore set out to explore which step(s) in the DIL outlining process i.e. mpMRI interpretation or cognitive transfer, led to the greatest variation against the reference volumes among experienced prostate outliners.

6.2 Aims and Objectives

6.2.1 Aim

The aim of this chapter is to determine whether mpMRI interpretation and/ or cognitive transfer onto planning CT leads to the observed DIL outlining variation amongst prostate radiotherapy outliners.

6.2.2 Objectives

1. To assess prostate radiotherapy outliners (clinicians and NMOs) DIL outlining skills using the PIVOTALboost contouring guidelines (as a one-step process) to outline a DIL using qualitative and quantitative metrics.

2. To assess prostate radiotherapy outliner mpMRI interpretation and cognitive transfer skills for DIL outlining as a deconstructed process through delineation of mpMRI (ADC and T2W) sequences followed by cognitive transfer of combined ADC/T2W DIL delineation onto the CT planning scan using qualitative and quantitative metrics.

3. To assess prostate radiotherapy outliner cognitive transfer only skills in the DIL outlining process through cognitive transfer of a pre-delineated mpMRI DIL onto the CT planning scan using qualitative and quantitative metrics.

4. To assess whether there is a difference in DIL outlining performance between prostate radiotherapy outliner cohorts i.e. clinicians or NMOs

6.3 Methods

6.3.1 Outlining exercise overview

Thirteen prostate outliners from Velindre Cancer Centre (VCC) and the South West Wales Cancer Centre (SWWCC) (eight clinicians -seven consultants, one clinical fellow and five NMOs) were invited to participate in a multi-step DIL outlining process (summarised in Figure 6.1). Step A entailed outlining a DIL directly on planning CT using the PIVOTALboost DIL contouring guidelines [1] to assess DIL outlining performance as a 'one-step' process, reflecting standard practice. Outliners were provided with a clinical vignette (Appendix 6) regarding the patient including the mpMRI report, PIVOTALboost boost contouring guidelines [1] and I met with each outliner individually to go through instructions on how to access the diagnostic mpMRI, CT planning scan within the treatment planning software (TPS) and how to save the final DIL volume. After a minimum period of three weeks after completing step A (to minimise case recall), outliners were then invited to participate in step B. This step aimed to evaluate outliners' mpMRI interpretation and cognitive transfer skills separately, by deconstructing the DIL outlining steps. Outliners were sent detailed instructions on how to outline the same DIL used in Step A, firstly on the ADC mpMRI sequence (B1), then on the T2W mpMRI sequence (B2) (Appendix 7), which were available within the treatment planning software (TPS). Following this, outliners were instructed to outline a composite volume of B1 and B2 onto T2W i.e. using both sequences to create a final mpMRI volume (B3). Outliners were then required to recreate their B3 volume onto the planning CT (B4) through cognitive transfer. After a further period of at least three weeks, to assess cognitive transfer skills only, outliners were required to complete step C. In this final step of the exercise, outliners were provided with screenshots of the DIL used in steps A and B, delineated on the T2W mpMRI sequence (by EE/JS) as a composite volume of ADC and T2W and given instructions to recreate the DIL on CT by cognitively transferring the pre-delineated DIL onto the planning CT. As the COVID-19 pandemic emerged during the early part of this exercise, due to suspension of non-clinical work, outliners had a gap of up to two years between step B and step C, with the exception of two clinicians, recruited later in the exercise.
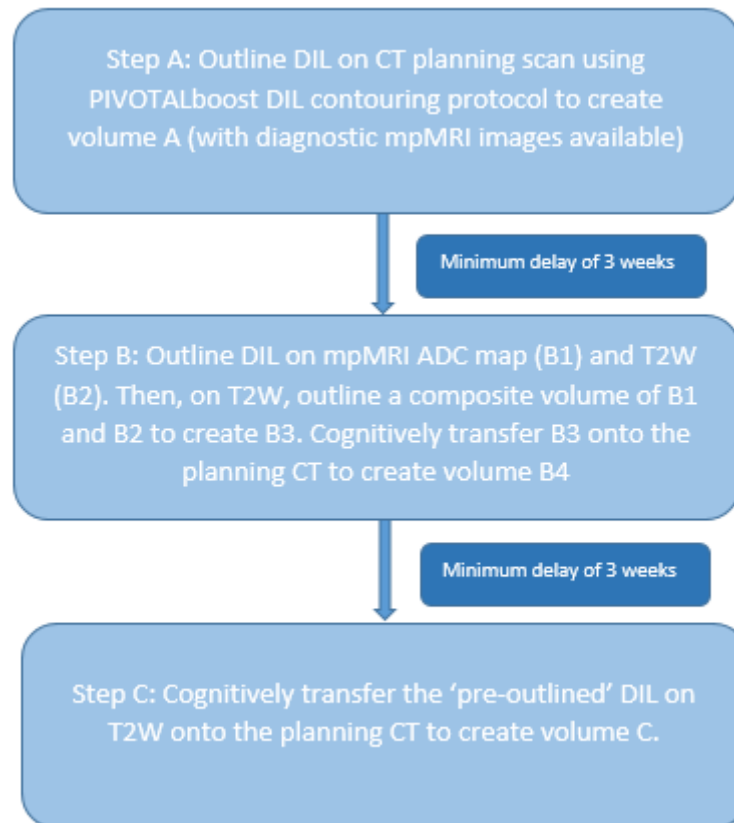
Figure 6.1: Summary of steps of DIL outlining exercise to establish cause of DIL outlining variation

6.3.2 Test case selection and reference volume definition

I identified a patient treated at VCC with an intermediate-risk localised prostate cancer patient whose mpMRI imaging showed two left sided DILs: one in the left peripheral zone and one adjacently in the left transitional zone. JS (PIVOTALboost TMG) and I (EE) delineated the DIL separately on the planning CT and ADC map and T2W sequences of the mpMRI. Via email, I sent anonymised screenshots of both my and JS's outlines, along with the radiology report, to two other members of the PIVOTALboost TMG (AT and IS) regarding which volumes they thought was 'most correct'. Following feedback from AT and IS, given the proximity of the DILs to each other, it was agreed that the DILs should be incorporated into one volume. JS and I subsequently finalised five reference volumes under an anonymised pseudonym within the VCC TPS (ProSoma®):

- Volume A: DIL outlined directly onto the planning CT using mpMRI images as a guide
- Volume B1: DIL outlined directly onto ADC map of mpMRI
- Volume B2: DIL outlined directly onto T2W image of mpMRI
- Volume B3: Composite DIL volume of B1 and B2 outlined onto T2W Image

- Volume B4: Cognitive transfer of B3 outlined onto planning CT

For step C, outliners would be required to cognitively transfer the B3 reference volume, provided as screenshots on PowerPoint onto CT. Therefore, given that the B4 reference volume was created through cognitive transfer of the B3 reference volume onto CT, the B4 volume would also act as the reference C volume used to evaluate outliners' volume C outlines. DIL reference volumes are shown in Figures 6.2-6.4. The prostate volume (CTVp) was outlined as a consensus volume on planning CT for qualitative assessment of DIL outline submissions.
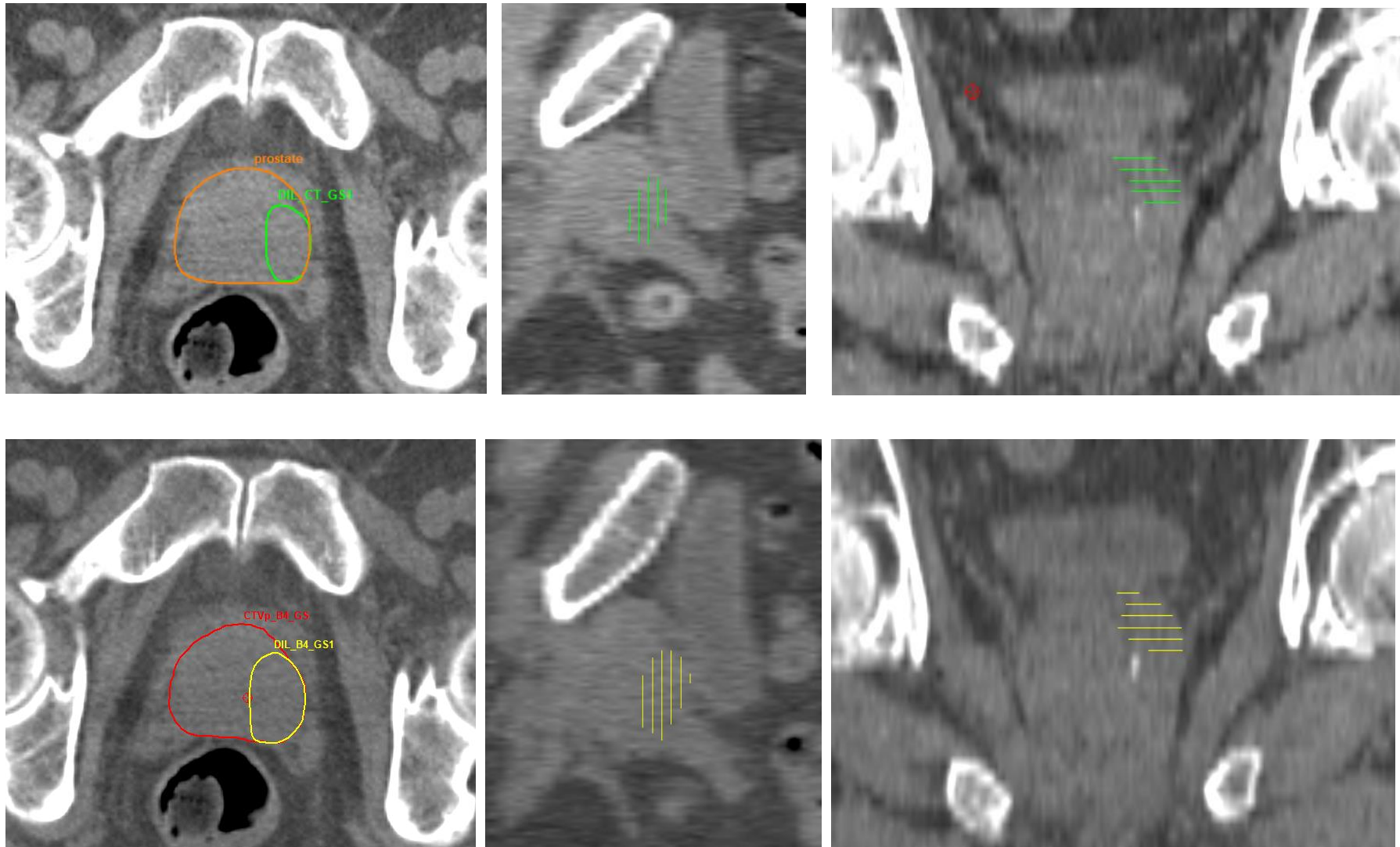
Figure 6.2: Planning CT showing reference volumes for Step A (top row) and B4 (bottom row) in axial, sagittal and coronal plane.

Figure 6.3: ADC map mpMRI images showing the reference volume for Step B1 in axial, sagittal and coronal plane.

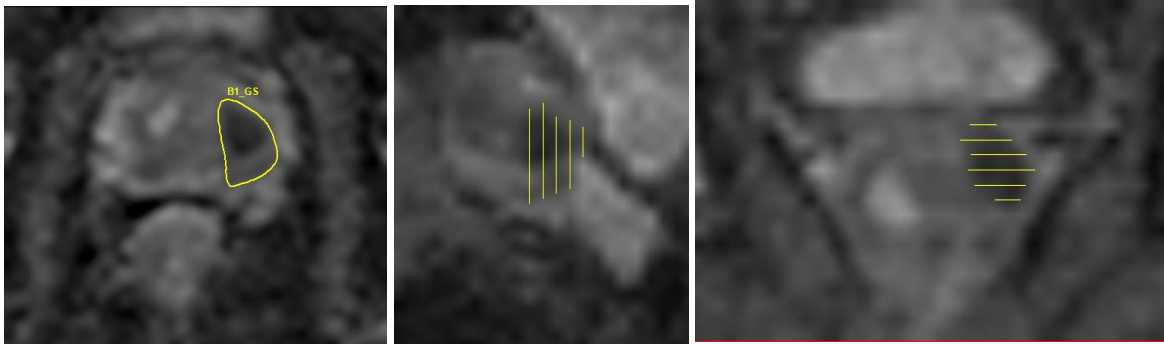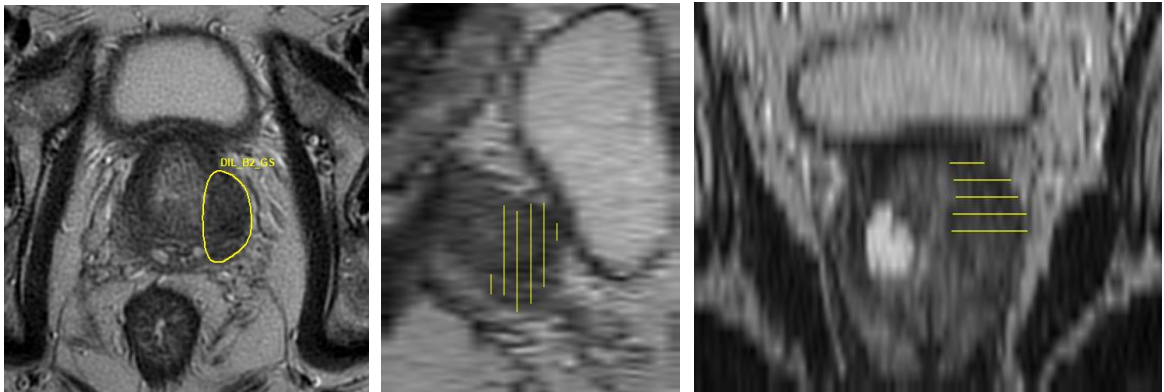*Figure 6.4: T2W mpMRI images showing the reference volume for Step B2 in axial, sagittal and coronal plane.*
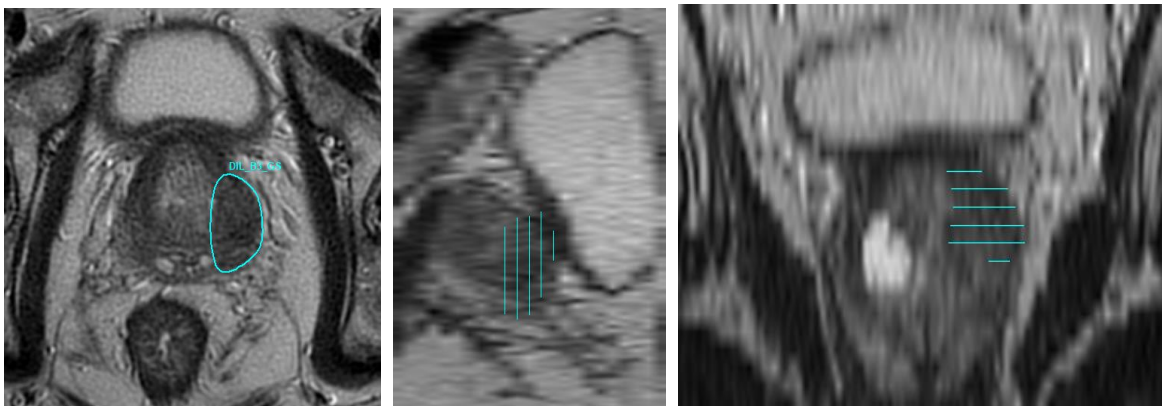
Figure 6.5: T2W mpMRI images showing the reference volume for Step B3 in axial, sagittal and coronal plane.

6.3.3 Case and data transfer

For outliners participating in the outlining exercise within VCC, the CT planning scan and the mpMRI sequences on which to delineate volumes A, B1-B4 and C, were all available under an anonymised pseudonym on ProSoma®. Outliners had been given instructions on how to name their saved volumes to allow me to identify them. Although in theory outliners could look at other saved volumes within the TPS, the instructions requested that outliners avoid reviewing any other volumes. The patient's anonymised diagnostic imaging was also available for outliners to access in the centre's PACS (Picture Archiving and Communication System) software. For those participating in SWWCC, the anonymised planning CT was sent from VCC via DICOM format to SWWCC and imported by a local physicist (JW) into their ProSoma® software. The anonymised diagnostic images were transferred by the VCC PACS team to the Swansea Bay University Health Board (SBUHB) PACS team for the SWWCC outliners to access locally. Completed outliner volumes from SWWCC were packaged and saved in DICOM format by JW and saved within a secure shared network drive between VCC and SWWCC.

6.3.4 Quantitative analysis

The reference volumes and outliner structures for the CT volumes (A, B4 and C) were imported in DICOM format into the CERR (Computational Environment for Radiotherapy Research) [2] platform to perform both quantitative and qualitative analysis. Unfortunately, the mpMRI sequence structures could not be imported into CERR without deformation of the outliner structures. Attempts to import the structures outlined on MRI sequences into other radiotherapy software that may support quantitative analysis (Raystation® and Velocity™) were also unsuccessful due to volume distortion. Therefore, within CERR, I calculated the volumes of each of the outliners' A, B4 and C structures and the volumes of their respective reference volumes. Via DICOM format, I sent the CT and the structure sets to ES (Professor of Healthcare Engineering) who ran them against an in-house conformity indices script within CERR to compute outliner DICE scores against their respective reference volumes. I was unable to compute DICE scores for the structure outlined on MRI sequences (B1, B2 and B3) within Prosoma®, but I was able to determine the volumes of these structures. I calculated the descriptive statistics (mean, range and standard deviation) for all steps within Excel.

6.3.5 Qualitative analysis

Within CERR, I was able to review the outliner CT structures (A, B4 and C) superimposed onto their respective reference volumes to determine whether they had unacceptable variations against the PIVOTALboost pre-accrual benchmark case DIL criteria:

- GTVpb breaches CTVp

- GTVpb outside correct segment of prostate

- GTVpb exceeds +/- 1 slice superiorly and/or inferiorly

- GTVpb investigator volume exceeds +/- 25% of the reference volume

For the structures delineated on MRI, I was able to determine CTVp breach, region of prostate DIL delineated and volumetric data within ProSoma®. As I was unable to directly overlay the outliner structures with their respective reference volumes, I created a PowerPoint of the reference volumes and used these to compare against the outliner structures to assess superior/inferior extent.

6.4 Results

The number of outliners who completed each step of the exercise is shown in Table 6.1.

| Outlining Step	A	B1 (ADC)	B2 (T2W)	B3	B4	C
No. Outliners	13	11	12	12	11	10

Table 6.1: Number of outliners who performed each outlining step on CT; A (mpMRI interpretation and cognitive transfer as one-step only), B4 (composite volume (B3) of ADC (B1) and T2W (B2) outlines transferred onto CT), C (cognitive transfer only of B3 reference volume).

6.4.1 CT quantitative data: volume and DICE

The volume and DICE summary statistics for outliners CT delineations for steps A, B4 and C are shown in Table 6.2. Bar charts showing individual outliner volumes and DICE scores for planning CT outlines (steps A, B4 and C) are shown in Figures 6.3 & 6.4.

	A	B4	C
Reference Volume (cc)	3.11	3.98	3.98
Outliners DIL Volume (cc)			
Mean ± SD	3.42 ± 1.84	4.07 ± 2.40	5.25 ± 1.25
Minimum	1.15	1.52	3.56
Max	6.50	9.29	7.22
DICE Scores			
Mean ± SD	0.53 ± 0.15	0.49 ± 0.15	0.67 ± 0.10
Minimum	0.19	0.07	0.47
Max	0.73	0.63	0.76

Table 6.2: Table showing quantitative metrics and their respective descriptive statistics of outliners for the CT steps of the outlining exercise: A (mpMRI interpretation and cognitive transfer as one-step only), B4 (composite volume (B3) of ADC (B1) and T2W (B2) outlines transferred onto CT), C (cognitive transfer only of B3 reference volume).
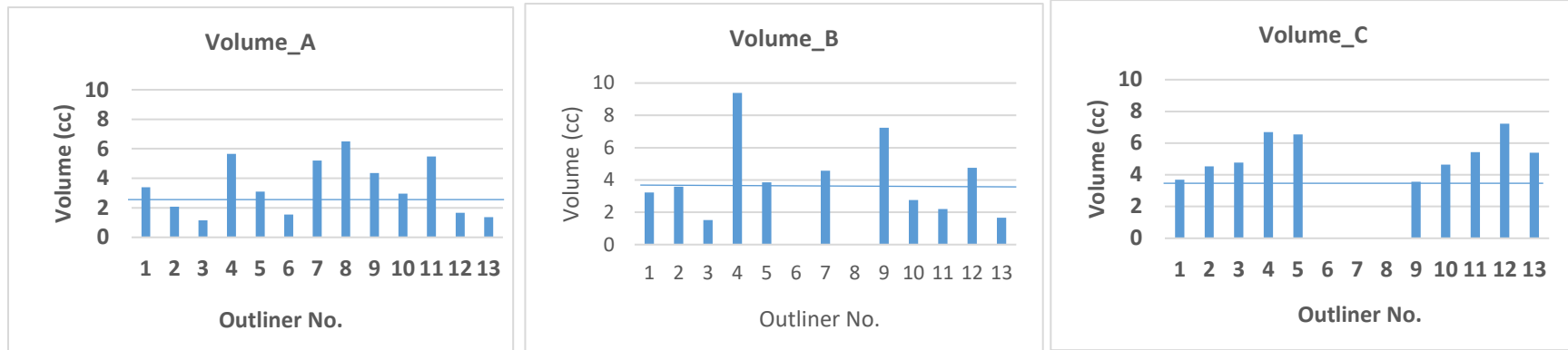
(a)   (b)   (c)

Figure 6.3: Bar chart showing individual outliner DIL volumes (cc) outlined on planning CT in steps A (a), B4 (b) and C (c). N.B. No outlines were submitted by outliner 6 or 8 for B4 or C and, no outline was submitted by outliner 7 for C. Respective reference volumes shown as a horizontal blue line.
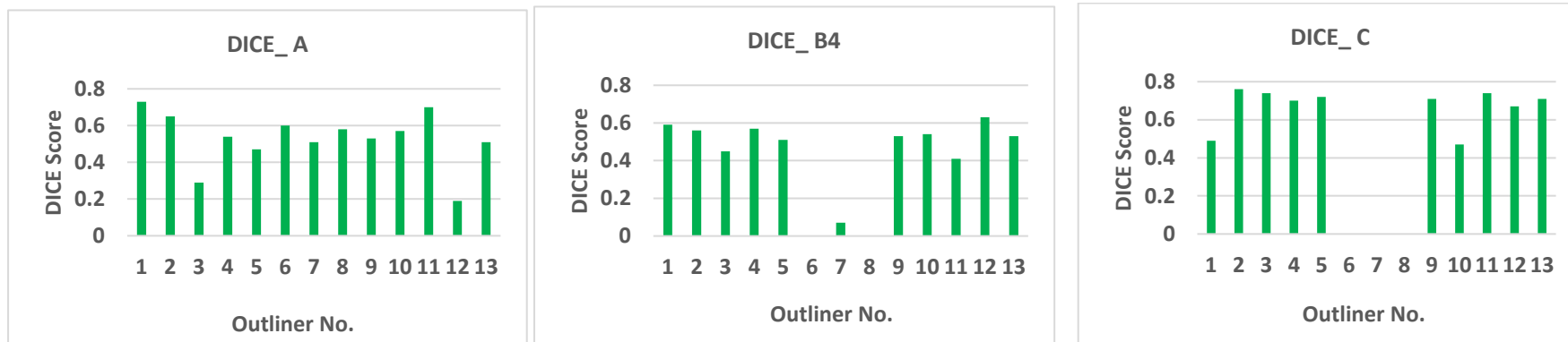
(a)   (b)   (c)

Figure 6.4: Bar chart showing individual outliner DICE scores (compared to reference volume) outlined on planning CT in steps A (a), B4 (b) and C (c). N.B No outlines were submitted by outliner 6 or 8 for B4 or C and, no outline was submitted by outliner 7 for C.

The results show that in terms of volumetric data, the mean outliner volume was similar for A and B4 to their respective volumes, but on average, outliners' delineated bigger volumes on CT (compared to the reference volume) for step C when they had to transfer a pre-outlined mpMRI volume only. However, as shown by the standard deviation of the mean, there were more 'outlier' volumes seen in B4 where candidates had to outline each deconstructed step on mpMRI in the DIL outlining process and transfer a composite volume onto CT, compared to the whole outlining process as one step (A) or cognitive transfer of a pre-outlined mpMRI volume (C).

With respect to the DICE scores, outliners produced the most 'conformal' outlines on CT during step C with a mean DICE of 0.67 compared to 0.53 and 0.49 for step A and B4, respectively. Outliner 7 had a particularly low DICE score for B4 (0.07), whereas the lowest DICE seen in step A was 0.19. I performed a sensitivity analysis on the DICE score for step B4 with this value removed and the mean was still 0.53 (range 0.41 to 0.63) but the standard deviation was lower at 0.06; thus the highest DICE scores remained Step C.

6.4.2 mpMRI quantitative data: volume

The volume and DICE summary statistics for outliners' mpMRI delineations for steps B1, B2 and B3 are shown in Table 6.3. Bar charts showing individual outliner volumes for mpMRI sequences (steps B1, B2 and B3) are shown in Figure 6.5.

	B1	B2	B3
Reference Volume (cc)	3.9	5.2	5.7
Outliners DIL Volume (cc)			
Mean ± SD	2.22 ± 0.94	3.53 ± 2.10	3.47 ± 2.07
Minimum	1.30	1.30	1.30
Max	4.40	8.00	8.40

Table 6.3: Table showing outliners' quantitative metrics and their respective descriptive statistics for the mpMRI sequence steps of the outlining exercise: B1 (ADC alone), B2 (T2W alone) and B3 (combined ADC/T2W volume delineated on T2W)
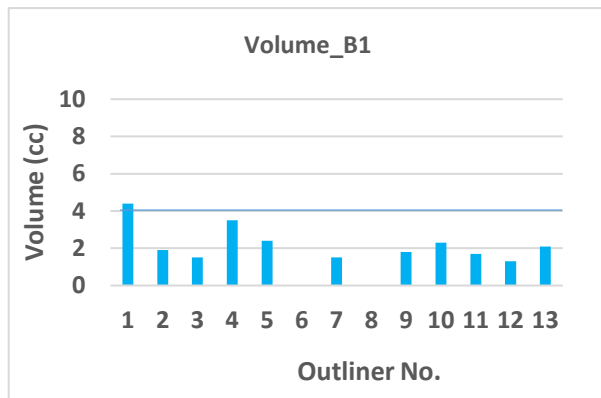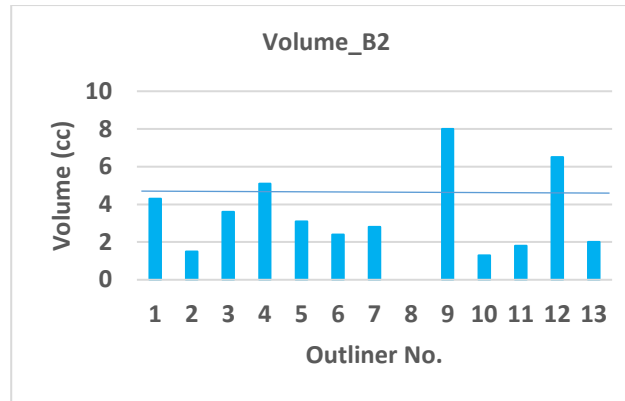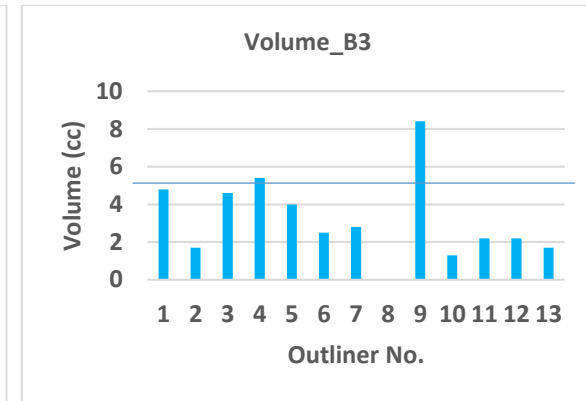
Figure 6.5: Bar chart showing individual outliner DIL volumes (cc) outlined on mpMRI in steps B1 (ADC only) (a), B4 (T2W only) (b) and C (ADC/T2W combined volume on T2W) (c). N.B: Outliner 8 submitted no outlines for B1, B2 or B3 and outliner 6 submitted no eligible outline for B1. Respective reference volume shown as a horizontal blue line.

For the three consecutive volumes, B1, B2 and B3, the reference volume increased in size between the three steps, which I expected as the lesion appeared more demarcated on ADC than T2W (and there is a tendency to outline a larger region when there is uncertainty). As B3 was a composite volume of ADC and T2W, I had expected this to be larger than B1 and B2 individually. Average outliner volumes were larger for B2 than for B1 but surprisingly, given B3 was a composite volume, the mean outliner volume for B3 was actually smaller than for B2. As shown in Figure 6.5, most outliners' delineated volumes were either similar in volume or bigger for B3 compared to B2 with the exception of outliner 12, who delineated a much smaller volume for B3 than B2 (2.2cc compared to 6.5cc). Of note, B3 to B4 reference volumes decrease from 5.7cc to 3.98cc respectively, but the mean outliners' volumes increase from 3.47cc to 4.07cc. Interestingly, B1, B2 and B3 outliner volumes were, on average smaller than their respective reference volumes.

6.4.3 CT and mpMRI: qualitative analysis

Outliner DIL delineation performance against the four PIVOTALboost pre-accrual DIL qualitative 'unacceptable variation' criteria are shown in Figure 6.6. The total number of outlines submitted by outliners across all six steps was 69. The most common 'unacceptable' variation was in total volume +/- 25% (48/69 outlines). The least common 'unacceptable' variation was outlining of a DIL in the incorrect region of the prostate (2/9 outlines). CTVp breach most commonly occurred on the CT delineation steps (A, B4, C) with all outliners having this unacceptable variation in step C (10/10 outliners). The fact that CTVp was not outlined by the delegates means that this is unlikely to be a true reflection of this parameter. Unacceptable superior/inferior extent of DIL was most commonly seen on the CT outlining steps A and B4 (7/13 and 7/11 respectively) but fewer (2/10) outliners had this unacceptable variation in step C i.e. cognitively transferring pre-outlined mpMRI volumes onto CT.
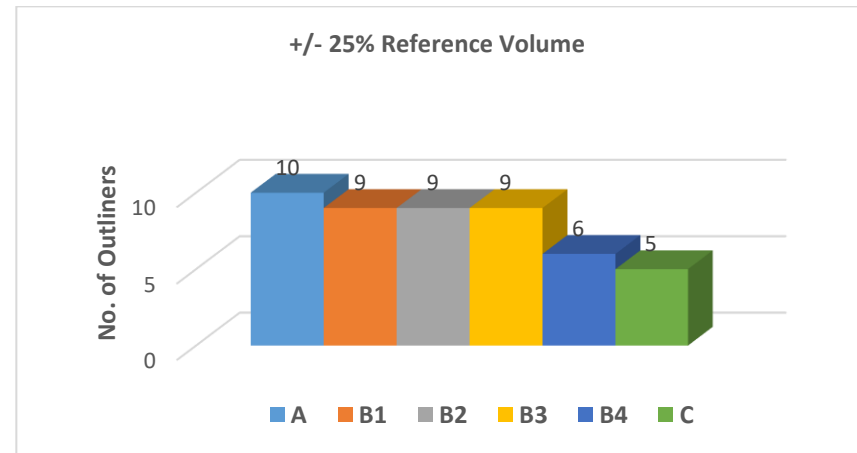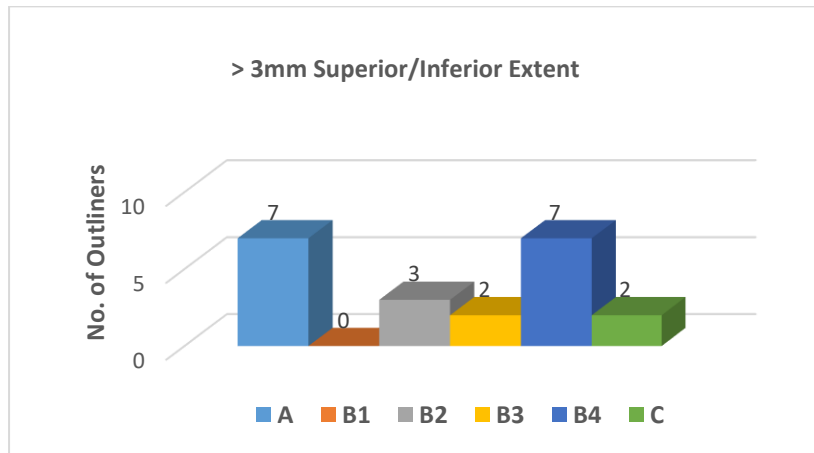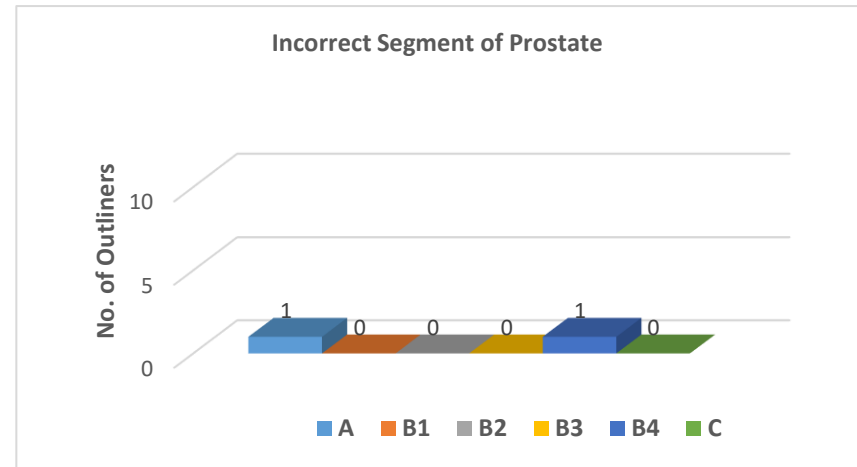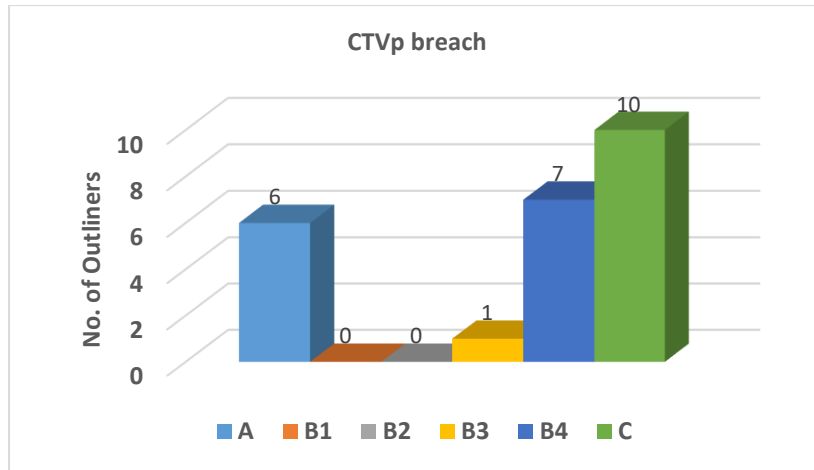
Figure 6.6: Bar charts showing number of outliners for delineation steps A, B1, B2, B3, B4 and C that had PIVOTALboost criteria unacceptable DIL outlining variations of CTVp breach (top left), incorrect segment of prostate (top right), >3mm superior/ inferior extent (bottom left) and volume +/- 25% reference volume (bottom right)

6.4.4 Clinician vs NMO quantitative metrics

The quantitative metrics, volume and DICE, for the outlining steps on CT (A, B4 and C) are shown in Table 6.4, split by two groups: oncologists ('clinicians') and NMOs (therapy radiographers or clinical technologists). NMOs delineated larger volumes on average, than both the clinicians and reference volumes for Steps A and B4, but were very similar to clinicians for step C i.e. cognitive transferral of pre-delineated mpMRI images only. NMO DICE scores however were higher on average than clinicians for both A and B4, although as for volume, were very similar for step C.

	A_Clinician	A_NMO	B4_Clinician	B4_NMO	C_Clinician	C_NMO
Reference Volume (cc)	3.11	3.11	3.98	3.98	3.98	3.98
Outliners DIL Volume (cc)						
Mean ± SD	3.08 ± 2.06	4.08 ± 1.26	3.23 ± 1.39	5.09 ± 3.10	5.38 ± 1.5	5.37 ± 1.32
Minimum	1.15	2.95	1.52	2.20	3.69	3.56
Max	6.50	5.65	4.76	9.39	7.22	6.69
DICE Scores						
Mean ± SD	0.51 ± 0.2	0.56 ± 0.09	0.46 ± 0.23	0.51 ± 0.06	0.67 ± 0.12	0.65 ± 0.12
Minimum	0.19	0.47	0.07	0.41	0.49	0.47
Max	0.73	0.70	0.63	0.57	0.76	0.72

Table 6.4: Table showing clinician versus non-medical outliner quantitative metrics and their respective descriptive statistics for the CT steps of the outlining exercise: A (mpMRI interpretation and cognitive transfer as one-step only), B4 (composite volume (B3) of ADC (B1) and T2W (B2) outlines transferred onto CT), C (cognitive transfer only of B3 reference volume).

Volumetric data for mpMRI sequence outlining steps (B1, B2, B3) is shown in Table 6.5. Again, NMOs outlined all three mpMRI steps on average, larger than the clinicians, but their mean volumes were closer to their respective reference volumes than clinicians' volumes. However, for B2 and B3, NMOs demonstrated a greater range in volumes delineated with both very small and very large outliers compared to the clinicians.

	B1_Clinician	B1_NMO	B2_Clinician	B2_NMO	B3_Clinician	B3_NMO
Reference Volume (cc)	3.9	3.9	5.2	5.2	5.7	5.7
Outliners DIL Volume (cc)						
Mean ± SD	2.12 ± 1.16	2.34 ± 0.84	3.52 ± 1.75	3.86 ± 2.74	2.90 ± 1.29	4.26 ± 2.81
Minimum	1.30	1.70	1.50	1.30	1.70	1.30
Max	4.40	3.50	6.50	8.00	4.80	8.40

Table 6.5: Table showing clinician versus non-medical outliner quantitative metrics and their respective descriptive statistics for the mpMRI sequence steps of the outlining exercise: B1 (ADC alone), B2 (T2W alone) and B3 (combined ADC/T2W volume delineated on T2W)

6.5 Discussion

The volumetric data showed that throughout the CT outlining steps (A, B4 and C) outliners' outlined progressively larger volumes. I had anticipated this change between volume size for step A and step B4 for a number of reasons. Firstly, although both A and B4 reference volumes were outlined on the same planning CT, the reference volume for A was smaller than B4. This is because provisional reference volumes for A, B1 and B2 were consensus outlines agreed by three experts (and me), whereas the final reference volumes for A and B1-B4 were determined by only two of us (JS and EE). Although this did create a slightly bigger B4/C volume than A and therefore questions which volume is the 'ground truth', we followed the same steps as the outliners to create the volumes and demonstrated the same growth in volume between A and B4/C, therefore resulting in a bigger A reference volume than B4/C reference volume. As discussed in chapter 2, determining a consensus reference volume ideally requires a panel of at least four experts [3] and some clinical trials determine a single consensus volume from a group of experts using the STAPLE algorithm [4], yet due to practicality and time constraints compounded by the COVID-19 pandemic, this was not possible for this study.

Secondly, B4 was the result of cognitive transfer of B3, which was a composite volume of B1 (ADC) and B2 (T2W) which is likely to explain the difference in volume between A and B4 for both the reference volume and outliner volumes. However, the standard deviation of the mean outliner volume was greater for B4 than A or C reflecting a greater variation in volume size. This is most likely due to B4 being the product of a deconstructed, multi-step process (delineation of B1-B3) which allows greater scope to introduce variation and/or error within each step outlined. For example, outliner 13 outlined a 'small' DIL on B1, a 'big' DIL on B2 and then a 'very small' composite B3, which was even smaller than their B1 volume. Interestingly, outliner 13 then outlined a volume which was much bigger onto CT. The B4 CT volume should ideally be smaller than the B3 mpMRI T2W composite volume. This is because the prostate volume shrinks after a period of androgen deprivation therapy, meaning the CT prostate volume is smaller than the diagnostic MRI prostate volume; 35.8cc vs 44.6cc i.e. 20%

reduction in this study. The outliner should therefore consider this volume reduction when cognitively transferring their final mpMRI DIL volume onto CT. This may also explain why the outliner average volume for C was larger than the reference volume. When the only step in the DIL process was to cognitively transfer a pre-outlined mpMRI volume onto the CT scans, most outliners did not proportionately reduce the DIL volume onto CT, and this was also reflected in the high numbers of CTVp breach for C, as most outliners overtly breached the prostate capsule, where they hadn't before. This may be explained by the long interval between each delineation step and outliners 'forgetting the concept' of cognitive transfer. It may also be that when the preliminary steps in DIL outlining are omitted and outliners don't have to create the composite DIL volume to transfer, less consideration is applied to the final volume i.e. much of the 'thinking' has been taken away from the outliner. Also, outliners did not have to delineate CTVp, which, if performed by outliners, may have mitigated the CTVp breach error.

With respect to outliner DICE scores, the lowest mean score was for step B4. As above, this volume was a cumulatively constructed volume with a number of steps during which variations could occur, resulting in a final CT volume, that on average, was less conformal with its respective reference volume than for A and C. Therefore, the accumulation of variation throughout the steps within the DIL outlining process is likely to be one of the main causes of DIL outlining variation. For step A, outliners were given the PIVOTALboost boost contouring protocol [1] and instructed to create volume A on CT as per the protocol. Within the protocol, it recommends outlining the DIL on the diagnostic mpMRI sequences prior to cognitive transfer to CT, but despite this, for step A, all outliners only delineated a volume on the planning CT. When creating the reference volume for A, given our collective DIL outlining experience, we also outlined directly onto CT first without delineating on mpMRI. To outline on the mpMRI, the diagnostic imaging has to be imported by the outliner into their TPS, which is not routine practice for all radiotherapy outliners and can be time-consuming. Time spent on radiotherapy outlining is dependent on the outliner's experience [5], and 'short cuts' may be adopted. For example, despite PIVOTALboost pelvic nodal contouring guidelines [6] recommending outlining the vessels before creating the CTVn volume, I was aware during my clinical oncology training, experienced clinical supervisors would outline CTVn directly either freehand or using tool within the TPS. With the exception of two outliners, who had been prostate oncologists for less than a year, the remainder of outliners delineated prostates as part of their routine clinical practice for over a year and therefore may feel more comfortable to 'skip' a suggested step. In addition, this exercise was in addition to outliners' clinical workload and therefore may have been less of a priority to complete precisely.

The mean DICE scores for A, B4 and C (0.53, 0.49 and 0.67 respectively) are either comparable or relatively high for DIL outlining compared to other datasets evaluated in this thesis. In chapter 1, the

PIVOTALboost pre-trial case 1 first submissions had a mean DICE score of 0.59 yet in the RCR workshop outlining exercise, the mean DICE scores for GTVpb_right ranged from 0.38-0.43 across the three outlining attempts and from 0.17-0.35 for GTVpb_left across the three attempts. The comparatively low DICE scores seen in the workshop are likely to be due to the small DIL volume size in the workshop case as GTVpb_right was 1.42cc and GTVpb_left 1.17cc. The PIVOTALboost case 1 DIL was much bigger at 3.42cc and the reference volumes for A and B4/C in this exercise were 3.11cc and 3.98cc respectively and as concordance increases with size [3], comparatively higher DICE scores may be expected. Another possible cause for the lower DICE scores in the workshop compared to the PIVOTALboost pre-accrual exercise is the relative inexperience of workshop delegates compared to trial Principal Investigators (PIs), who tend to be more experienced consultants.

Given the relatively high DICE score for step C (0.67) which required cognitive transfer alone, mpMRI interpretation could be considered another cause of DIL outlining non-conformity. Indeed, mpMRI interpretation even amongst radiologists, the specialty most trained to interpret these images, can be variable [7]. Although there are no studies evaluating oncologists mpMRI interpretation skills, urologists are another cohort that use mpMRI in their clinical practice to obtain information regarding prostate biopsies and surgical techniques. Yet in a study of 73 urologists reviewing 12 mpMRI scans, mpMRI interpretation skills were considered 'far from proficient' [8]. The mean outliner volumes for B1, B2 and B3 were lower than their respective reference volumes, suggesting outliners were unable to identify the full extent of the DIL on mpMRI sequences. B1 (ADC) had less variation in volumes outlined as shown by a relatively low standard deviation of the mean. For B2 (T2W) however, a much bigger variation in volume was seen compared to B1, with outliner volumes ranging from 1.30-8.00cc in relation to the reference volume of 5.2cc. The DIL was comprised of both a peripheral zone and transitional zone lesion and therefore as per the PI-RADS v2.1 assessment tool [9], require both sequences of the mpMRI for interpretation, although both lesions were well visualised on the ADC map which may explain the greater volumetric variation seen in B2 (T2W).

Regarding qualitative assessment however, outlining performance was better for the mpMRI steps (B1-B3) than for the CT steps (A, B4 and C) for three out of the four parameters. No outliners on any mpMRI step delineated the DIL in the incorrect prostate region and only one outliner (on B3) breached CTVp, compared to 6/13 for A, 7/12 for B4 and 10/10 outliners for step C. No outliners on B1 had superior/inferior extent issues for B1 (ADC), suggesting as above, that the extent of the lesion was best visualised on the ADC map. The most common 'unacceptable' variation on B1, B2 and B3 was the outliner volume exceeding ±25% of the reference volume. As suggested by the relatively small outliner DIL volumes across the exercise, this was predominantly due to outliners' under-contouring and not appreciating the full extent of lesion on mpMRI. As alluded to in Chapter 4 however, this qualitative

parameter does not necessarily reflect conformity to a reference volume and its application is limited by DILs being relatively small structures.

As discussed above, CTVp breach was most commonly seen in step C, where outliners more readily transferred the pre-outlined mpMRI verbatim onto CT without consideration of proportionate reduction of the DIL onto CT. Instructing the outliners to delineate CTVp as well as the DIL, may have mitigated this in part, particularly as this was an 'unacceptable' variation also seen in the other CT outlining steps (6/13 outliners for A and 7/12 for B4). However, I had anticipated from my workshop experience, that it would be a challenge to get full engagement from outliners throughout the whole outlining exercise due to busy clinical schedules, compounded by the COVID-19 pandemic, and therefore I minimised the number of structures required to outline as much as possible. Also, a consensus outline for CTVp had not been created for the ADC and T2W sequences and relied on my assessment only, although prostate definition was more appreciable on MRI sequences than CT.

'Unacceptable' superior/inferior extent of DIL was most commonly seen on the CT outlining steps A and B4 (7/13 and 7/11 respectively) where outliners had to cognitively transfer their own mpMRI determined volumes onto CT. This variation on CT seemed to be mitigated by a pre-determined final mpMRI volume to transfer, further suggesting as above, that the multi-step process of DIL outlining introduces more source of outlining variation than for either mpMRI interpretation or cognitive transfer alone.

In terms of differences in DIL CT outlining between NMOs and clinicians, with the exception of step C, for which mean volumes between the two groups were very similar, NMOs tended to outline bigger volumes than clinicians, whose mean volume for A and B4 were closer to the reference volumes. The tendency to delineate bigger volumes by NMOs was also reflected through bigger average volumes for the mpMRI steps (B1, B2, B3) and for B2 and B3 in particular, greater variation in volume size. Although, the NMOs are all designated prostate outliners, as non-medical staff they do not routinely attend, unlike the oncologists, a urology multi-disciplinary meeting during which patient mpMRI images are reviewed in real time and findings described by the specialist radiologist. Oncologists therefore are more likely to be familiar with mpMRI sequences and interpretation, which may account for the propensity for NMOs to outline more generous volumes i.e. include areas of uncertainty. Interestingly, the mean DICE scores for CT outlines A and B4 were slightly higher for NMOs than the clinicians and less variation in scores were seen, whereas for C the mean DICE scores were similar for both groups. The seemingly poorer performance by clinicians is likely to be due to poor outlier performance in both A and B4, both bringing the mean scores for the clinician cohorts down, particularly as the highest DICE score for each CT step was by a clinician.

There are a number of limitations associated with this study. During this piece of work, the COVID-19 pandemic struck, meaning non-clinical work in our centre was suspended. This not only disrupted the flow of the study, but given staff shortages once non-clinical work resumed in our centre, outliners having the time to dedicate to the exercise was limited, hence the diminishing numbers of participating outliners as the exercise progressed. This too impacted the number of experts I could recruit to create the reference volumes. Another challenge I encountered was mpMRI analysis. Unfortunately, CERR did not support the import of structures onto MRI without deformation and I was unable to find alternative software that could enable MRI structure analysis. This meant, not only did I have to abandon DICE analysis for the mpMRI structures, but I also had to perform volumetric and qualitative analysis within ProSoma®, making analysis more time consuming for the mpMRI data. Using the same case at three time points may have familiarised the outliners with the case, particularly as B had four delineation components. However, initially due to the pandemic and subsequently my clinical commitments, the exercise spanned a long timeframe which meant outliners were more likely to be unfamiliar with DIL outlining at each time point, particularly as DIL boosting, at present, is not standard of care. I was also unable to 'hide' the reference and outliner volumes within ProSoma® meaning outliners could 'cheat' and review these structures before outlining their own, although given the variability of the structures, this did not appear to be the case.

6.6 Conclusion

From the work completed, in this chapter, DIL outlining variation is most likely to due to the cumulative steps within the DIL outlining process including both mpMRI interpretation and cognitive transfer rather than an individual step alone. More focused educational strategies for both clinicians and NMOs including mpMRI interpretation may help improve the image interpretation aspect of the DIL outlining process. As DIL outlining for IMRT boost is currently undertaken within a trial setting only, clinicians should be encouraged to participate in trials including PIVOTALboost to familiarise themselves with a technique that may become standard of care.

6.7 References

1.   Institute of Cancer Research (ICR)(UK). PIVOTALboost: Trial Boost Contouring Instructions & Atlas; Version 1. ISRCTN80146950. 2017.

2.   Deasy J, Blanco A, Clark V. CERR: a computational environment for radiotherapy research. Med Phys. 2003 May; 30(5): 979-85.

3.   Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, et al; NCRI RTTQA OUTLINING AND IMAGING SUBGROUP. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. Br J Radiol. 2013 Apr; 86(1024): 20120398.

4.   Warfield S, Zou K, Wells W. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging.* 2004 Jul; 23: 903–21.

5.   Walker Z, Bartley G, Hague C, Kelly D, Navarro C, Rogers J, et al. Evaluating the effectiveness of deep learning contouring across multiple radiotherapy centres. Phys Imaging Radiat Oncol. 2022 Nov 8; 24: 121-128.

6.   Institute of Cancer Research (ICR)(UK). PIVOTALboost: Pelvic node contouring Atlas; Version 1. ISRCTN80146950. 2017.

7.   Sonn G, Fan RE, Ghanouni P, Wang N, Brooks J, Loening A, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus. 2019 Jul; 5(4): 592-599.

8.   Mantica G, Suardi N, Smelzo S, Esperto F, Chierigo F, Tappero S, et al. Are Urologists Ready for Interpretation of Multiparametric MRI Findings? A Prospective Multicentric Evaluation. Diagnostics (Basel). 2022 Nov 1; 12(11): 2656.

9.   American College of Radiology: Prostate Imaging- Reporting and Data System Version 2.1. [Internet] 2019. [Updated 2019; cited 2023 Feb 04]. Available from: https://www.acr.org/-/media/ACR/Files/RADS/PI-RADS/PIRADS-V2-1

Chapter 7: Conclusion and Future Work

7.1 Conclusion

This thesis has reviewed radiotherapy outlining variation of the dominant intraprostatic lesion (DIL) including common sources of variation, methods to assess 'acceptable' or 'unacceptable' variations, the impact of an educational workshop on DIL outlining and which steps within the DIL outlining process the variation occurs.

In chapter 1, I discussed the emerging role of DIL dose-escalation in radiotherapy for prostate cancer and the requirement for accurate outlining to ensure both tumour control and minimise toxicity. Interpretation of mpMRI sequences are imperative to localisation of DILs and is a relatively new technique for radiologists who have required specific training programmes to gain skills necessary to report their findings adequately. In this chapter, I also discussed the premise of the PIVOTALboost trial; a phase III RCT currently recruiting and treating patients in 38 centres across the UK, which aims to evaluate both the role of DIL dose-escalation and pelvic nodal irradiation in patients with intermediate- and high-risk localised prostate cancer.

Chapter 2 explored the impactof outlining variation in the radiotherapy pathway and methods used to minimise variation amongst radiotherapy outliners including outlining protocols, educational interventions and use of novel imaging modalities to optimise TVD. I also discussed the strategies adopted within RTQA processes to minimise the impact of outlining variation on trial outcome and the qualitative and quantitative metrics used to measure outlining performance against an expert-determined reference volume.

In chapter 3, I established that there was a high rate of re-submission required of the PIVOTALboost pre-accrual benchmark cases predominantly due to 'unacceptable' variation in DIL outlining; I concluded this predominantly due to DIL outlining being a relatively new technique for most prostate oncologists in the UK and not currently standard of care. Therefore, not only is a robust on-trial RTQA process required to ensure outlining consistency within the trial, but also work to improve DIL outlining would need to be undertaken if the technique is to become standard of care in prostate radiotherapy.

Given the relatively small pool of DIL outlining experts and labour-intensive nature of reviewing and providing constructive feedback for the PIVOTALboost pre-accrual benchmark case submissions, in Chapter 4, I set out to determine an appropriate conformity index and associated threshold to use as a semi-automated assessment tool. I was unable to determine a conformity index that could screen

acceptable cases on the basis of outliner DIL volume or extent and concluded that use of the pre-determined qualitative parameters requires expert clinical review to determine acceptability of pre-accrual case submissions.

In chapter 5, I evaluated the impact of an educational workshop on DIL outlining performance. Delegates did show improvement in outlining (measured by DICE scores) in one out of two DILs assessed immediately following a lecture on DIL outlining although short and long-term retention of knowledge could not be assessed. Delegates did report however that their perceived knowledge, as reflected in their self-reported theoretical confidence, improved following the workshop. Few delegates though submitted 'acceptable' DIL outlines before, during or after the workshop as assessed by the PIVOTALboost pre-accrual benchmark case criteria, reinforcing results from chapters 3 and 4.

Finally, in chapter 6, I investigated which steps within the DIL outlining process were the main source of variability amongst experienced prostate medical and non-medical outliners specifically mpMRI interpretation or cognitive transfer. I identified both 'errors' (misunderstanding of the protocol/ instructions) and 'variability' (when no specific errors were apparent) and concluded that DIL outlining variation is most likely due to the cumulative steps within the DIL outlining process including both mpMRI interpretation and cognitive transfer rather than an individual step alone. Therefore, both targeted educational interventions regarding mpMRI interpretation and outliners having more opportunity to delineate DILs may help to improve consistency in the future.

7.2 Future work and recommendations

During the work of this thesis, the DELINEATE trial [1] reported their efficacy and toxicity data. DELINEATE is a single centre prospective phase II multi-cohort study of 'standard' (cohort A: 74 Gy in 37 fractions), moderately hypofractionated (cohort B: 60 Gy in 20 fractions) and standard plus pelvis (cohort C) prostate image guided intensity modulated radiotherapy in intermediate- and high- risk prostate cancer patients. Patients in all groups received an integrated DIL boost (Cohort A & C 82Gy and Cohort B 67 Gy). Significant late toxicity was rarely seen in either cohort and in keeping with other contemporary series of prostate cancer radiotherapy, urinary symptoms affected patients more than rectal symptoms although low levels of 'bother' were reported for either toxicity. Although patient numbers in the trial were too small to draw any full conclusions regarding efficacy of a boost, control rates were encouraging with 5-year freedom from biochemical or clinical failure 98.2% (cohort A), 96.7% (cohort B) and 95.1% (cohort C) which compares well with the 5-year PSA relapse free survival rate of 88% in the control arm of the multi-centre RCT CHHiP trial [2]. This supports results from the Dutch FLAME study [3], a phase III multi-centre RCT of patients with intermediate- or high-risk

localised prostate cancer investigating the role of a simultaneous integrated boost to the DIL (77 Gy in 35 fractions standard arm with boost up to 95Gy in 35 fractions in focal boost arm). The 5-year biochemical disease free survival was significantly higher in the boost arm (92%) compared to the control arm (85%). Therefore, it is likely DIL dose-escalation will become the future standard of care and highlights the importance of improving DIL outlining techniques amongst prostate outliners. Below are considerations for future work and recommendations to try to improve DIL outlining accuracy for expectant routine practice.

7.2.1 Trial quality assurance

Chapters 3 and 4 highlighted the limitations of having one expert review all the pre-accrual benchmark case submissions for PIVOTALboost. Not only is the process time-consuming and labour-intensive, there is more scope for subjectivity to influence acceptability and unacceptability contour variation. By expanding this role to a panel of expert reviewers, common outlining 'errors' can be detected which may require protocol amendments and highlight the constraints of the assessment criteria upon which they are compared against. Given the small nature of DILs, using the parameter of outliner exceeding +/- 25% of the reference volume, could deem volumes 'unacceptable' that may actually be clinically acceptable after clinical review. Therefore, DIL outlining assessment as part of pre-trial quality assurance, may lend itself towards the approach suggested by Sweeney et al [4], where a panel of experts determine maximum and minimum acceptable contours and assess whether outliners submit volumes that lie within these parameters. This concept considers the inevitable inter-observer variability of outlines and achieving a clinically acceptable volume should be the goal. I have adopted this method in the educational setting to assess target volume outlining, providing the clinical input to develop FIELD^{RT} [5]. This is an open source software that to support TVD for clinical oncology trainees and consultants which compares user outlines to expert determined minimum and maximum acceptable volumes using the STAPLE [6] algorithm, enabling outliners to grasp what is clinically acceptable rather than striving to achieve the 'perfect' outline against a single reference volume.

7.2.2 Education and peer review

As discussed in Chapter 2, TVD variability is not limited to DIL outlining. The RCR has recognised the need to improve radiotherapy skills amongst clinical oncologists and launched the digital radiotherapy planning platform (COPP) project [7]. This project enables clinical oncologists to update their radiotherapy planning skills through access to educational resources and outlining workshops. In an extension of this, the ARENA project [7] is due to be launched imminently, which is an educational package that offers tumour site-specific TVD instructional modules with corresponding cases to outline on and compare their attempts against pre-determined reference volumes and/ or minimal

and maximum acceptability. If DIL dose-escalation is to become standard of care, this offers a platform upon which this skill can be refined in a non-clinical capacity.

Peer-review guidelines commissioned and published by the RCR [8] not only define minimum standards for TVD but also promote shared contouring decision-making amongst peers. For DIL outlining, as shown in chapter 6, defining the final DIL volume on mpMRI to transfer to planning CT can be problematic and encouraging multi-disciplinary peer review (i.e. clinical oncology trainees, consultants and non-medical outliners) may allow such discrepancies to be amended prior to finalisation of treatment volumes. However, my work suggests that once these agreed MRI volumes have been created, non-medical outliners may be well placed to reliably transfer the volumes onto CT for margin growth and planning. MRI planning in the future may make this even more viable.

7.2.3. Auto-contouring solutions

Another area to consider with respect to minimising TVD variation is the role of auto-contouring, which in addition to reducing inter-observer variation may reduce outliner workload and time associated with manual contouring [9]. In a study of the clinical utility of deep learning-based auto-segmentation for MR-based prostate radiotherapy planning, a third of auto-contoured prostate CTVs required ''clinically significant'' editing by physicians, and time spent on final contouring was reduced by thirty percent for physicians compared to historic methods [10]. Another study reviewed automatic segmentation masks of the prostate and four surrounding OARs for MR-based prostate radiotherapy and found there was no statistical significance in prostate, rectum or external urinary sphincter dosimetry parameters derived from the automatic-defined compared to physician-refined segmentation masks [11]. Although yet to be fully integrated within routine clinical care, auto-contouring is a rapidly evolving field within radiotherapy planning. In refining prostate radiotherapy auto-contouring methods, DIL outlining solutions should be also be considered in order to optimise a technique that may become future standard of care.

7.2.4 Tumour control probability and normal tissue complication probability modelling

Methods to optimise TVD consistency are based on the premise that under-outlining a tumour may lead to geographical miss of the tumour and over-outlining may lead to excess tissue irradiation and subsequent toxicity for the patient. However, what is unclear for DIL dose escalation radiotherapy, is *how much* outlining variation of the DIL actually increases the risk of under- or over-irradiating the target volume with significant clinical consequences. As an extension of work in this thesis, I would propose tumour control probability (TCP) and normal tissue complication probability (NTCP) modelling on outliner submissions of a range of DILs located in close proximity to OARs (e.g. rectum, urethra

and bladder). This could help facilitate definition of what may be considered 'acceptable' variation in DIL outlining for future trials and TVD educational packages.

7.3 Summary

In summary, DIL outlining is a new technique amongst prostate radiotherapy outliners and further work is required to improve outlining consistency. Future methods to improve DIL outlining could include educational and peer review methods, robust radiotherapy trial quality assurance and expansion of prostate auto-contouring techniques to include the DIL.

7.4 References

1. Tree, A, Satchwell L, Alexander E, Blaisak-Wal I, deSouza N, Gao A et al. Standard and hypofractionated dose escalation to intraprostatic tumor nodules in localized prostate cancer: efficacy and toxicity in the DELINEATE trial. Int J of Radiat Oncol Biol Phys. 2023; 115(2): 305-316.

2. 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 Aug; 17(8): 1047-1060.

3. Kerkmeijer L, Groen V, Pos F, Haustermans K, Monninkhof E, Smeenk R, et al. Focal boost to the intraprostatic tumor in external beam radiotherapy for patients with localized prostate cancer: results from the FLAME randomized phase III trial. J Clin Oncol. 2021; 39(7): 787-796.

4. Sweeney, L., Spezi, E., Cole, N., Sebag-Montifiore, D., & Adams, R. PO-0949; Automated approval of a pre trial benchmark RTTQA case. The ARISTOTLE experience. Radiother Oncol. 2016; 119.

5. Piazzese C, Evans E, Thomas B, Staffurth J, Gwynne S, & Spezi E. FIELD[RT]: an open-source platform for the assessment of target volume delineation in radiation therapy. Br J Radiol. 2021 Oct 1; 94(1126), 20210356.

6. Warfield S, Zou K & Wells M. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004; 23(7): 903–921.

7. Evans E, Radhaskrishna G, Gilson D, Hoskin P, Miles E, Yuille F, et al. Target volume delineation training for clinical oncology trainees: the role of ARENA and COPP. Clin Oncol (R Coll Radiol). 2019; 31(6): 341-343

8. Royal College of Radiologists. Radiotherapy target volume definition and peer review, second edition-RCR guidance. London: Royal College of Radiologists; 2022

9. Zabel W, Conway J, Gladwish A, Skiliarenko J, Didiodato G, Goorts-Matthews L, et al. Clinical evaluation of deep learning an atlas-based auto-contouring of bladder and rectum for prostate radiation therapy. Pract Radiat Oncol. 2021; 11(1): 80-89

10. Cha E, Elguindi S, Onochie I, Gorovets D, Deasy J, Zelefesky M, et al. Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy. Radiother Oncol. 2021;159:1-7.

11. Sanders JW, Kudchadker RJ, Tang C, Mok H, Venkatesan AM, Thames HD, Frank SJ. Prospective Evaluation of prostate and organs at risk segmentation software for MRI-based prostate radiation therapy. Radiol Artif Intell. 2022 Jan 26; 4(2):e210151.

Appendix 1: Individual investigator conformity indicies for PIVOTALboost pre-accrual benchmark case 1 GTVpb first submissions (Chapter 4)

Investigator No.		RIET	JACCARD	DICE	GMI	DI	MDC	MDC over-contouring error	MDC under-contour	criteria
1	vs 'GTVpb_1GOLDSTD'	0.219693724	0.30376569	0.465982028	0.579374276	0.477697842	0.546393862	0.277336478	0.269057384	pass
2	vs 'GTVpb_1GOLDSTD'	0.387764314	0.433926302	0.605228179	0.208574739	0.510043042	0.518042519	0.300729009	0.21731351	fail
3	vs 'GTVpb_1GOLDSTD'	0.39392375	0.419672131	0.591224018	0.110081112	0.557348703	0.399016294	0.31968195	0.079334344	fail
4	vs 'GTVpb_1GOLDSTD'	0.244871802	0.325259516	0.490861619	0.564310545	0.437967115	0.542718986	0.262584075	0.280134912	pass
5	vs 'GTVpb_1GOLDSTD'	0.639110034	0.645856981	0.784827586	0.340672074	0.030664395	0.426233574	0.21953125	0.206702324	fail
6	vs 'GTVpb_1GOLDSTD'	0.120735162	0.208180708	0.344618494	0.604866744	0.694444444	0.764960855	0.362244716	0.402716139	fail
7	vs 'GTVpb_1GOLDSTD'	0.491245464	0.539497307	0.700874636	0.303592121	0.294600939	0.337809973	0.18196959	0.155840383	pass
8	vs 'GTVpb_1GOLDSTD'	0.406738249	0.426264801	0.597735849	0.082271147	0.556799105	0.376163159	0.304106291	0.072056868	fail
9	vs 'GTVpb_1GOLDSTD'	0.340440867	0.359020852	0.528352235	0.082271147	0.629039813	0.466374542	0.377316305	0.089058237	fail
10	vs 'GTVpb_1GOLDSTD'	0.399536846	0.456790123	0.627118644	0.442641947	0.283159463	0.453180162	0.23493816	0.218242002	pass
11	vs 'GTVpb_1GOLDSTD'	0.376882725	0.434329395	0.605620155	0.275782155	0.479600333	0.46785908	0.302163543	0.165695537	fail
12	vs 'GTVpb_1GOLDSTD'	0.46250378	0.512429832	0.677624602	0.259559676	0.375366569	0.385333489	0.261673015	0.123660474	pass
13	vs 'GTVpb_1GOLDSTD'	0.369703538	0.39224572	0.563471971	0.097334878	0.590431125	0.425388161	0.312060243	0.113327918	fail
14	vs 'GTVpb_1GOLDSTD'	0.489039349	0.521156559	0.685210941	0.14368482	0.428902628	0.376577997	0.253129206	0.123448792	fail
15	vs 'GTVpb_1GOLDSTD'	0.308469503	0.336927224	0.504032258	0.130938586	0.645054425	0.49516281	0.397227115	0.097935695	fail
16	vs 'GTVpb_1GOLDSTD'	0.505484463	0.551376147	0.710821999	0.303592121	0.274154589	0.366988074	0.204092547	0.162895527	pass
17	vs 'GTVpb_1GOLDSTD'	0.371548697	0.389078498	0.56019656	0.075318656	0.598187311	0.433488754	0.35720202	0.076286734	fail
18	vs 'GTVpb_1GOLDSTD'	0.527140079	0.56	0.717948718	0.156431054	0.375107296	0.342489691	0.211187202	0.13130249	fail
19	vs 'GTVpb_1GOLDSTD'	0.341056078	0.408719346	0.580270793	0.478563152	0.345930233	0.458761225	0.213823214	0.24493801	pass
20	vs 'GTVpb_1GOLDSTD'	0.295463816	0.357651246	0.526867628	0.301274623	0.57713885	0.603269571	0.424183325	0.179086247	fail
21	vs 'GTVpb_1GOLDSTD'	0.211414343	0.298023715	0.459196102	0.563151796	0.516046213	0.602361899	0.345694727	0.256667172	pass
22	vs 'GTVpb_1GOLDSTD'	0.277520278	0.356929955	0.526084569	0.444959444	0.5	0.559383336	0.303716586	0.25566675	fail
23	vs 'GTVpb_1GOLDSTD'	0.335558767	0.407407407	0.578947368	0.400926999	0.439869989	0.513825145	0.225793454	0.28803169	fail
24	vs 'GTVpb_1GOLDSTD'	0.408020457	0.469049695	0.638575668	0.376593279	0.345498783	0.415105386	0.222649561	0.192455825	pass

RCR18 Prostate Cancer Outlining Workshop

Thank you for registering for the prostate cancer outlining workshop that will be held on 10 September, 2018 in Liverpool, as part of RCR18.

The workshop will be an interactive workshop that combines a lecture and interactive hands-on practical contouring exercises.

Focus of the workshop

The session will focus on 2 areas of prostate radiotherapy outlining:

- Dominant Intraprostatic Lesion (DIL) outlining
- Post-operative outlining

The learning outcomes for the workshop are as follows:

> Dominant Intraprostatic Lesion Outlining
>
> ● Identify the most common errors made in dominant intraprostatic lesion (DIL) radiotherapy outlining.
>
> ● Select the most appropriate imaging modalities and sequences to aid DIL and prostate delineation.

> Post-operative prostate bed outlining
>
> ● Describe the current consensus recommendations for post-operative radiotherapy for prostate cancer.
>
> ● Define the main differences between the post-operative consensus guidelines for prostate radiotherapy.
>
> ● Identify the most common contouring errors made in post-operative prostate outlining.

Preparation for the workshop and further details

To initiate discussion and enhance learning, we would like all delegates to outline two cases prior to the workshop.

The first case focuses on DIL and prostate outlining and requires delineation of 3 volumes:

- GTVp
- GTVpb_left
- GTVpb_right

i.e. there are 2 intraprostatic lesions to outline within the same case. The GTVp is essentially the prostate only and is referred to as CTVp in the PIVOTALboost outlining guidelines. The left sided intraprostatic lesion to be outlined will be named GTVp_left and the right-sided lesion, GTVpb_right. Seminal vesicles do not need to be delineated.

We recommend you follow the PIVOTALboost outlining guidelines provided on Share Place to aid DIL delineation. The document provides both instructions for outlining and an atlas for reference. Only the first three bullet points in section 4.1 of the guidelines are relevant to the outlining required for this case; margins do not need to be applied to your volumes and OARs do not need to be delineated.

The second case focuses on post-operative radiotherapy for prostate cancer and requires delineation of:

- CTVbed-RTOG or CTVbed-EORTC or CTVbed-RADICALS
- Rectum

As there is no general consensus regarding post-prostatectomy target delineation, we have provided delegates with the option of contouring according to either the RTOG, EORTC or RADICALS guidelines. Three versions of this case, corresponding to the three sets of guidelines, are available on Share Place. Please choose the relevant version and outline the prostate bed within the nomenclature associated with the chosen guideline eg CTVbed-RTOG outlined according to the RTOG protocol. A summary of these three guidelines are provided on Share Place and screenshots of the pre-operative MRI prostate are available. You will also need to outline the rectum for this case.

Please note that **the deadline for submitting contours is 5pm on Friday 7 September 2018**. After this deadline point you will still be able to access these cases, contour them and submit your contours. However, late submission may impact the usefulness of the variation analysis and the subsequent feedback provided, and we cannot guarantee that contours submitted after this deadline will be taken into account at the workshop.

We strongly urge delegates to complete the pre-workshop exercise. Within the workshop, we will be reviewing the submitted volumes to identify common areas of outlining errors for

discussion. **All submitted volumes will be anonymized** and therefore should not be identifiable during review within the workshop. Delegates will be required to undertake contouring exercises within the workshop therefore **please bring your laptop and a mouse with a wheel to the outlining workshop**. Following completion of the workshop, delegates will have the opportunity to complete two further contouring exercises, one DIL and one post-operative case for which additional CPD points can be claimed. The volumes will be assessed against the reference volumes and you will receive individualised feedback on each case. The cases will also be analysed within the context of educational and radiotherapy research, but again delineations will remain anonymized for analysis.

We also request delegates complete a questionnaire which will be provided at the end of the workshop. This is both to improve the running of future RCR outlining workshops and to also aid analysis of the outlining variation identified.

Accessing and using Share Place

For the contouring exercises that you will be asked to perform before, during and after the workshop, we will be using the AQUILAB's online Share Place platform. Participation does not require you to install any software, all you need is an internet enabled Windows computer/MAC (with either Internet Explorer, Firefox, Chrome or Safari installed). In order to get started, please follow the attached user guidance for Share Place and the associated contouring tools, which AQUILAB have provided.

If you need assistance

For any clinical queries regarding the DIL case please contact Dr Elin Evans at xxxxxxxxxxxx

For any clinical queries regarding the post-operative case please contact xxxxxxxxxxxxx at xxxxxxxxxxxx

For any queries regarding the Share Place platform please contact xxxxxxxxxxxx

For any other queries relating to the outlining workshop please contact me.

Best wishes

RCR Clinical Oncology Planning Project Manager

Appendix 3: Clinical vignette of DIL outlining case for outlining workshop (Chapter 5)

<u>RCR Prostate Outlining Workshop: Dominant Intraprostatic Lesion Clinical Case</u>

An 80 year old gentleman presents with lower urinary tract symptoms to his GP and is found to have a PSA of 5.2. He proceeds to an MRI pelvis, bone scan and transperineal template biopsy, the results of which are outlined below.

<u>MRI Pelvis</u>

The prostate measures 3.7 x 5.1 x 4.1 cm with an estimated volume of 40 cc.

There is a dominant lesion seen in the right transitional zone at mid gland level extending approximately from the 9 to 12 o'clock position. The lesion is approximately 16x10mm and shows restricted diffusion (5/5).

There is a second lesion seen in the left peripheral zone a 4-5 o'clock position which also shows restricted diffusion (5/5) is approximately 13mm in diameter. The lesion abuts the capsule but is difficult to exclude early extra-capsular extension.

There is no evidence of seminal vesicle invasion.

There is no pelvic lymphadenopathy.

No abnormal bone marrow signal identified.

Likely T2c N0 MX but cannot exclude early extracapsular extension at right base.

<u>Histopathology</u>

18 cores in total taken; 6 from the left and 6 from the right plus 4 cores from left target lesion and 2 from right target lesion. 5/18 cores positive; 2 cores from the left target lesion, 2 from the right target lesion and 1 from the right posterior lateral region contain adenocarcinoma. Overall Gleason score 4+3=7 (ratio 55:45). Maximum core length is 14.5mm (left target lesion). 11% of the total core length contains tumour. No extraprostatic extension is seen or perineural invasion is seen.

<u>Bone Scan</u>

No evidence of bony metastatic disease

Slide 1

Dominant Intraprostatic Lesion Outlining

RCR Prostate Outlining Workshop 10th September 2018

Slide 2

Dominant Intraprostatic Lesion

- Escalation of radiotherapy dose to whole prostate (70-80 Gy) improves biochemical control but increases bladder and bowel toxicity
- Prostate cancer may be multi-focal but local recurrence usually associated with a primary tumour i.e. dominant intraprostatic lesion (DIL)
- Therefore dose escalation to DIL (80-90 Gy) may improve biochemical control without unacceptable increase in toxicity
- There may be >1 DIL within the prostate and it may be possible to boost several lesions

Slide 3

DIL criteria (PIVOTALBoost)

- PI-RADS (v.2) lesion of score 4 or 5 on MRI
- DIL >5mm dimension
- Total DIL volume estimated to be <50% total prostate volume
(volume of total number of lesions should be summated if >1 DIL)

Slide 4

Multiparametric MRI (mp-MRI)

- Conventional MRI – T1/ T2 weighted sequences
- Multiparametric MRI additionally includes functional imaging;
  -Diffusion Weighted Imaging (DWI) including calculated Apparent Diffusion Coefficient (ADC) map
  -Dynamic Contrast Enhanced sequences
- Additional functional sequences improves sensitivity and specificity of tumour detection

Maurer & Heverhagen, 2017

Slide 5

Prostate Anatomy

- Peripheral zone contains 70-80% of glandular tissue and accounts for 70-75% of prostate cancer origin

- Transitional zone surrounding the proximal urethra contains 5% of glandular tissue and accounts for 20-30% of prostate cancer origin

- Central zone tumours, containing 20% glandular tissue but prostate cancers originating here are rare

PIVOTALBoost Trial Boost Contouring Instructions & Atlas

Slide 6

Peripheral Zone DIL

T2 weighted

PI-RADS Score	Peripheral Zone
1	Uniform hyper-intense signal intensity (normal)
2	Linear or wedge-shaped hypo-intensity or diffuse mild hypo-intensity, usually indistinct margin
3	Heterogeneous signal intensity or non-circumscribed, rounded, moderate hypo-intensity includes others that do not qualify as 2, 4, or 5
4	Circumscribed, homogenous moderate hypo-intense focus/mass confined to prostate and <1.5 cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extra-prostatic extension/invasive behavior

ADC map

PI-RADS Score	Peripheral zone (PZ) or Transition zone (PZ)
1	No abnormality (i.e., normal) on ADC
2	Indistinct hypo-intense on ADC
3	Focal mildly/moderately hypo-intense on ADC
4	Focal markedly hypo-intense on ADC <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm, definite extra prostatic extension/invasive behavior

Overall PI-RADS Score

DWI score	T2W score	Overall PI-RADS score
4	Any	4
5	Any	5

T2 weighted: DIL in left PZ

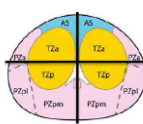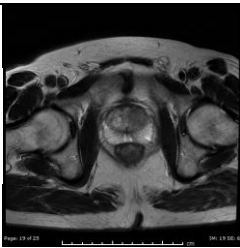**Corresponding ADC map showing restricted diffusion of PZ DIL**

Slide 7
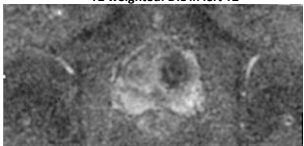
Transitional Zone DIL

T2 weighted

PI-RADS Score	Transition Zone
1	Homogeneous intermediate signal intensity (normal)
2	Circumscribed hypo-intense or heterogeneous encapsulated nodule(s) (BPH)
3	Heterogeneous signal intensity with obscured margins Includes others that do not qualify as 2, 4, or 5
4	Lenticular or non-circumscribed, homogeneous, moderately hypo-intense, and <1.5 cm in greatest dimension
5	Same as 4, but ≥1.5cm in greatest dimension or definite extra-prostatic extension/invasive behavior

ADC map

PI-RADS Score	Peripheral zone (PZ) or Transition zone (PZ)
1	No abnormality (i.e., normal) on ADC
2	Indistinct hypo-intense on ADC
3	Focal mildly/moderately hypo-intense on ADC
4	Focal markedly hypo-intense on ADC <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm, definite extra prostatic extension/invasive behavior

Overall PI-RADS score

T2W score	DWI score	Overall PI-RADS score
3	5	4
4	Any	4
5	Any	5

T2 weighted: DIL in left TZ

Corresponding ADC map showing restricted diffusion of TZ DIL
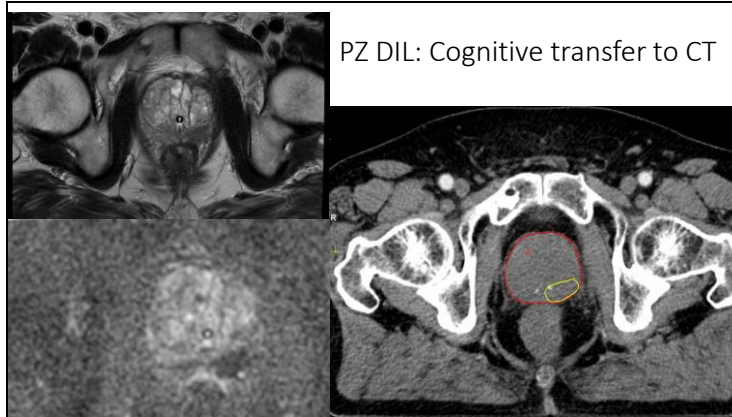

Slide 8

DIL Outlining Steps

- **Outline CTV prostate volume**
- Use T2 weighted MRI imaging to aid delineation

- **Identify DIL on diagnostic MRI**
  - Review all available diagnostic radiology and histopathology to identify DIL to outline
- Peripheral zone tumours: identify hypo-intense lesion on ADC map and use T2 weighted as additional information
- Transitional zone tumours: identify DIL on T2 weighted imaging and use ADC map as additional information
- Ensure DIL volume <50% total prostate volume

- **Outline the DIL on planning CT**
  - Fusion of diagnostic MRI unlikely to be beneficial given prostatic volume change following ADT
- Identify most superior and inferior aspects of DIL on corresponding planning CT
- Re-outline a similar DIL volume in on these slices then delineate remaining DIL volume in between
- Ensure DIL is delineated **within** your delineated prostate volume
- Check DIL volume in sagittal and coronal images to ensure consistency

---

Slide 9

Slide 10

PZ DIL: Cognitive transfer to CT
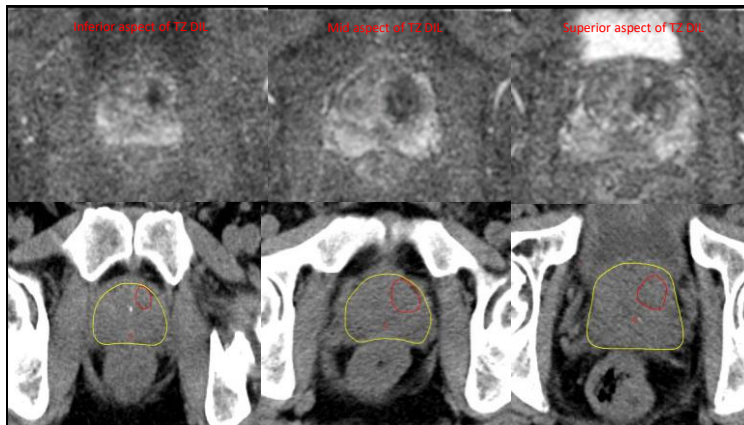
Slide 11

Common DIL outlining errors

- Incorrect inferior/ superior extent of DIL outlined
- Gross over/under contouring of total DIL volume
- Insignificant lesions outlined
- DIL extending outside delineated prostate volume
- Incorrect prostate outlining (usually at apex/ base)
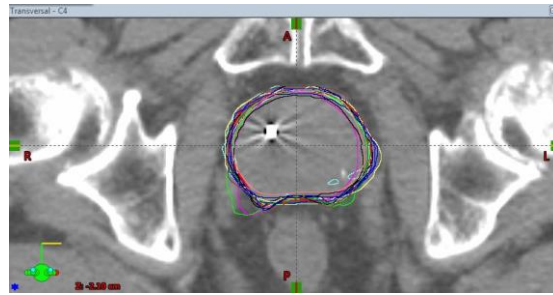

Slide 12

Prostate outlining

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer

Carl Salembier [a], Geert Villeirs [b], Berardino De Bari [c], Peter Hoskin [d], Bradley R. Pieters [e], Marco Van Vulpen [f], Vincent Khoo [g], Ann Henry [h], Alberto Bossi [i], Gert De Meerleer [j], Valérie Fonteyne [k,*]

[a] Department of Radiation Oncology, Europe Hospitals Brussels; [b] Department of Radiology, Ghent University Hospital, Belgium; [c] Department of Radiation Oncology, CHRU Besançon, France; [d] Mount Vernon Cancer Centre, Northwood, United Kingdom; [e] Department of Radiation Oncology, Academic Medical Center/University of Amsterdam; [f] Department of Radiation Oncology, University Medical Center Utrecht, The Netherlands; [g] Department of Clinical Oncology, Royal Marsden Hospital, London, United Kingdom; [h] Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, United Kingdom; [i] Department of Radiation Oncology, Institut Gustave Roussy, Villejuif, France; [j] Department of Radiation Oncology, University Hospital Leuven, Belgium; [k] Department of Radiation Oncology, Ghent University Hospital, Belgium
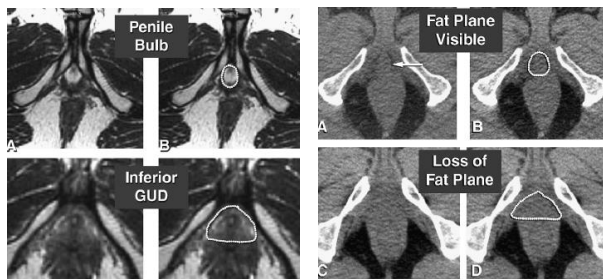
Slide 13

Good consensus usually at mid-prostate...
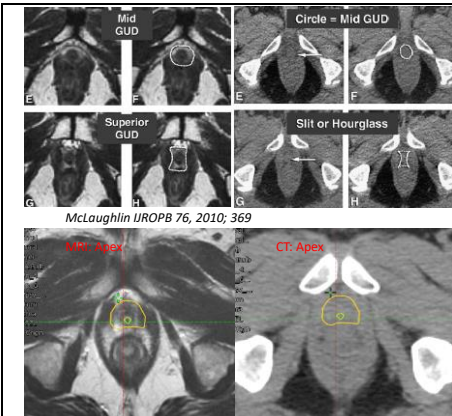
Slide 14

Prostate Outlining: Identifying Apex

McLaughlin IJROPB 76, 2010; 369

- Use T2 weighted MRI sequence to aid delineation (can fuse planning MRI if available)
- Identify penile bulb on MRI and correlate to planning CT (prostate apex approx. 1 cm above penile bulb)
- Genitourinary diaphragm is tissue between penile bulb and prostate apex
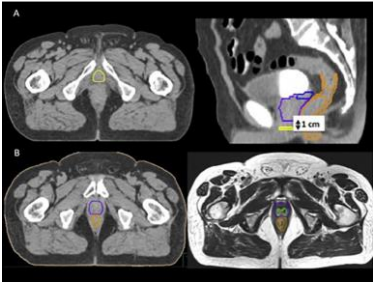
Slide 15

McLaughlin IJROPB 76, 2010; 369

Prostate Outlining: Apex

- Superior GUD may appear as a slit or hourglass region on CT/MRI in approximately 50% of patients
- Prostate apex starts superior to the superior aspect of GUD
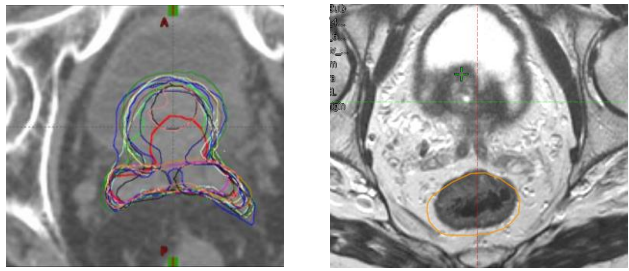
Prostate Outlining: Apex

- If outlining on MRI, can exclude urethra to create a 'butterfly' shape (green); corresponding planning CT slice shows apex delineation including urethra (purple)
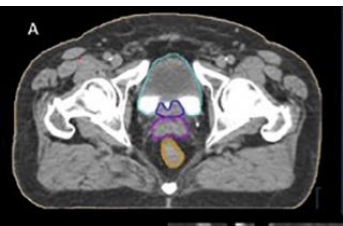
Salembier et al, 2018

Prostate Outlining: Base of Prostate

Prostate Outlining: Base of Prostate

- Base of prostate continuous with bladder mucosa
- As per ESTRO (2018) guidelines, CT planning delineation alone would be aided by use of IV contrast 10mins before planning CT
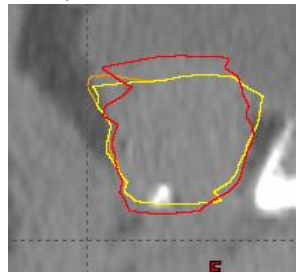- T2-weighted MRI more clearly defines prostate protrusion into bladder

Salembier et al, 2018

Slide 19

Prostate Outlining: Extra Tips

Also remember to include EPE in volume….
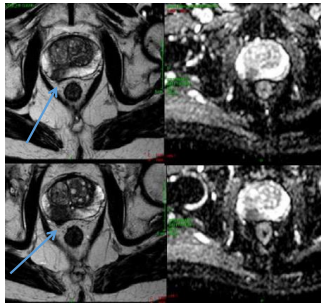
…and check final volumes in sagittal/ coronal plane for inconsistencies

References

- McLaughlin P, Evans C, Feng M, Narayana V. Radiographic and Anatomic Basis for Prostate Contouring Errors and Methods to Improve Prostate Contouring Accuracy. *International Journal of Radiation Oncology, Biology and Physics.* 76(2), 369-378.

- Pathmanathan A, Alexander E, Huddart R, Tree A. (2016) The delineation of intraprostatic boost regions for radiotherapy using multimodality imaging. *Future Oncology,* 12(21), 2495-2511

- RTTQA Group. (2017). PIVOTALBoost Trial Boost Contouring Instructions & Atlas.

- Salembier C, Villiers, G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, Khoo V, Henry A, Bossi A, De Meerleer G, Fonteyne V. (2018) ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. *Radiotherapy and Oncology,* 127(1), 49-61

- Von Eyben F, Kiljunen T, Kangasmaki A, Kairemo K, von Eyben R, Joensuu T. (2016). Radiotherapy Boost for the Dominant Intraprostatic Cancer Lesion- A Systematic Review and Meta-Analysis. Clinical Genitourinary Cancer. 14(3), 189-197.

Prostate cancer outlining workshop: pre-workshop information

User ID:_____

1. Current Position:

Consultant	Trainee (please state level)	Clinical Fellow

2. Length of experience in prostate radiotherapy:

0-6 months	6-12 months	1-2 years	3-5 years	6-10 years	11+ years

3. Have you participated in the PIVOTALboost pre-trial outlining exercise?
a. Yes
b. No

4. Have you participated in the PIVOTALboost pre-trial outlining exercise?
c. Yes
d. No

5. If you answered yes to question 5, have you been approved to treat patients in PIVOTALboost?
a. Yes
b. No

6. Have you attended the PIVOTALboost outlining workshop, London 2017?
a. Yes
b. No

7. Prior to the RCR outlining workshop exercise, how many DIL cases have you outlined?

0	1-2	3-5	6-10	11-15	15+

8. If you have previously done DIL outlining, did a radiologist help you outline the DIL?
a. Yes
b. No
c. If yes, for how many cases (approximately)?_____

9. Which of the following imaging does your centre perform in the diagnostic work-up for intermediate to high-risk prostate cancer (circle all that apply)?
a. Standard MRI
b. mp-MRI
c. PSMA PET-CT
d. choline PET-CT
e. CT-TAP
f. bone scan
g. Other (please specify)_____

10. How confident do you feel about your contouring skills prior to the workshop?

i. For DIL:
a. Not a clue
b. Know in theory but not confident in practice
c. Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
d. Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
e. Know in theory, competent in practice minimal peer support / discussion in highly selected cases (< 1:5 cases)

ii. For Post operative radiotherapy:
a. Not a clue
b. Know in theory but not confident in practice
c. Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
d. Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
e. Know in theory, competent in practice minimal peer support / discussion in highly selected cases (< 1:5 cases).

11. How confident do you feel about your contouring skills prior to the workshop?
iii. For DIL:
f. Not a clue
g. Know in theory but not confident in practice
h. Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
i. Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
j. Know in theory, competent in practice minimal peer support / discussion in highly selected cases (< 1:5 cases)

iv.   For Post operative radiotherapy:
  f.   Not a clue
  g.   Know in theory but not confident in practice
  h.   Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
  i.   Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
  j.   Know in theory, competent in practice minimal peer support / discussion in highly selected cases (<1:5 cases)

Prostate cancer outlining workshop: pre-workshop information

  1.   What impact (if any) will attending this contouring workshop likely to have on your future clinical practice?
  e.   Improved tumour delineation
  f.   No impact (no new information was gained)
  g.   Other (please specify) _____

  2.   How confident do you feel about your contouring skills following the workshop?
i.   For DIL:
  a.   Not a clue
  b.   Know in theory but not confident in practice
  c.   Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
  d.   Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
  e.   Know in theory, competent in practice minimal peer support / discussion in highly selected cases (< 1:5 cases).

ii.   For Post operative radiation therapy:
  a.   Not a clue
  b.   Know in theory but not confident in practice
  c.   Know in theory, can perform some parts in practice independently but need peer discussion to be readily accessible in the majority of cases (> 1: 3 cases)
  d.   Know in theory, confident in practice, need for peer support / discussion in selected cases (<1:3 cases)
  e.   Know in theory, competent in practice minimal peer support / discussion in highly selected cases (< 1:5 cases).

3. Please specify tumour site(s) that you would be interested in attending a contouring workshop on: _____

4. The current contouring workshop format includes pre-workshop homework. Do you think this format is suitable?
a. Yes
b. No
c. Any comments_____

_____

5. Which of the following is the most ideal setting to organise future contouring workshops:
a. As stand-alone live course(s)
b. As part of a tumour-specific course(s)
c. As part of an online webinar
d. Other (please specify)_____

_____

6. Was the Share Place user guidance provided by AQUILAB helpful and clear
a. Yes
b. No
c. Any comments_____

7. Was it easy to access the Share Place platform?
a. Yes
b. No
c. Any comments_____
_____

8. Was it easy to use Share Place and the contouring tools provided?
a. Yes
b. No
c. Any comments_____
_____

9. Did you need to contact AQUILAB support in order to get advice/guidance about using Share Place?
a. Yes
b. No
c. Any comments_____

_____

Appendix 6: Clinical vignette and investigation results for DIL outlining exercise (Chapter 6)

Clinical Vignette

A 70 year-old man presented with a PSA of 9.2. He proceeded to undergo an MRI pelvis, biopsy and a bone scan, the results of which are detailed below:

MRI Pelvis

Prostate volume 51 ml (4.6 x 4.2 x 4.4 cm). There are BPH changes in the central gland. 9mm left base lateral peripheral zone nodule with corresponding restricted diffusion PIRADS 4.There is also a larger restricting lesion in the left transitional zone. Further ill-defined intermediate T2 signal at the right mid-gland posterior peripheral zone with corresponding linear restricted diffusion PIRADS 3. 2 sub-centimetre cysts are seen in the right central zone at mid gland level, these do not appear connected to the urethra. Normal seminal vesical. Scattered sigmoid diverticuli. No lymphadenopathy.

Conclusions: PIRADS 4 lesion left base peripheral zone and left transitional zone. PIRADS 3 lesion right midgland peripheral zone.

Histopathology

10 cores sampled. 2 positive cores from the right (apex, base), 5 from the left (apex, mid-gland, base, horn and medial). Maximum tumour length 9mm (left base). 100% tumour in most involved core. Perineural invasion seen in left base core. Overall Gleason score 3+4-7 (<10% pattern 4), grade group 2.

Bone Scan

No evidence of bone metastas

Appendix 7: Instructions for delegates for steps B and C of the DIL outlining exercise (Chapter 6)

<u>Outlining exercise: B</u>

Step 1. Open synapse and find case Zarena_dil_mri. This is the MRI for the patient to aid DIL delineation.

Step 2. In prosoma, find case ZARENA_DIL1Z

Step 3: Open S1 MR HFS: ADC- please open the images only and not the saved version with structures already delineated!

Step 4: Delineate the boost lesion as you identify it on the ADC map (please name volume ADC_x). Save your attempt as DIL_x_B1

Step 5: Open S3 MR HFS Ax T2 HR Propeller- again, please open the images only and not the saved version with structures already delineated!

Step 6: Delineate the boost lesion as you identify it on the T2 sequence (please name volume T2_x). Save your attempt as DIL_x_B2

Step 7: Modify the T2 DIL outline to incorporate the DIL volume as identified on the ADC map- this may lead to a bigger volume than identified on the T2 images alone. Name this volume T2/ADC_x. Save this attempt as DIL_x_B3

Step 8: Cognitively transfer the final boost volume, saved as DIL_x_B3, from the T2 images to the CT planning images. Name this volume CT/T2/ADC_LT. Save this attempt as DIL_x_B4.

Outlining exercise: C

1.   Open the case **ZARENA_DIL1Z** in PROSOMA

2.   Open the CT planning scan, which is saved as **S6 CT HFS: Pelvis 3.0mm** (please do not be
      tempted to look at other saved delineations!)

3.   In an adjacent window, or ideally on an adjacent screen, open the attached powerpoint titled
      '*DIL COMBINED T2/ADC VOLUME ON T2*'

4.   Create a new volume on the CT planning scan and label it DIL_x_C

5.   Using the MRI images in the powerpoint, delineate the pre-outlined MRI DIL volume on CT
      (under DIL_x_C). In other words, delineate slice by slice, the MRI outlines on the corresponding
      CT planning scan. Please save this attempt as DIL_x_C.