The Role of Automation in Radiotherapy Treatment Planning for Prostate Cancer

A thesis submitted to The University of Manchester for the degree of

Doctor of Clinical Science (D.Clin.Sci)

in the Faculty of Biology, Medicine & Health

Joseph Stephen Wood

School of Medical Sciences

Division of Cancer Sciences

Contents

Word count: 30,000 approx.

List of Figures

List of Tables

List of Common Abbreviations

3D-CRT	3D conformal radiotherapy
AA	Aperture area
AAV	Aperture area variability
AI	Aperture irregularity
ANN	Artificial neural network
AP	Aperture perimeter
ART	Adaptive radiotherapy
BA	Beam area
BI	Beam irregularity
BM	Beam modulation
CNN	Convolution neural network
CTV	Clinical tumour volume
DSC	Dice similarity coefficient
DTA	Distance to agreement
DVH	Dose volume histogram
EBRT	External beam radiotherapy
ESTRO	European Society for Radiotherapy and Oncology
GTV	Gross tumour volume
IMRT	Intensity modulated radiotherapy
KB	Knowledge-base
LSV	Leaf sequence variability
MCS	Modulation complexity score
MI	Mutual information
MLC	Multi-leaf collimator
MPE	Medical physics expert
OAR	Organ at risk
OVH	Overlap volume histogram
PCA	Principle component analysis
PoD	Plan of the day
PTV	Planning target volume
RTP	Radiotherapy treatment planning
SV	Seminal vesicles
TPS	Treatment planning system
VMAT	Volumetric modulated arc therapy

Abstract

Each year in the UK, almost 50, 000 men are diagnosed with prostate cancer and of these approximately 30 % receive external beam radiotherapy (EBRT) as part of their treatment (Cancer Research UK, 2018). Treating prostate cancer patients therefore compromises a significant proportion of a typical EBRT centre's total clinical workload.

It has long been acknowledged that changes to the contents of the bladder, bowel and rectum can cause the prostate to move considerably within the pelvic cavity (Moiseenko *et al.*, 2007; Hosni *et al.*, 2017). To ensure that the malignant disease receives the prescribed EBRT dose under this positional uncertainty, a margin of healthy tissue that includes radiosensitive organs at risk (OAR) surrounding the prostate is generally also treated. Adaptive radiotherapy (ART) can ameliorate this situation by adapting the treatment to account for day-to-day changes in pelvic anatomy thereby allowing the disease to be targeted more accurately and reducing the irradiated volume (Nijkamp *et al.*, 2008; Antico *et al.*, 2019).

Unfortunately, although ART techniques provide dosimetric and potential clinical benefits to patients, clinical adoption of ART is far from widespread. ART brings additional work to the radiotherapy treatment planning process, which is magnified over the large prostate cancer patient population such that it quickly becomes unmanageable in most centres.

Automation therefore has an important role to play in ART workflows for prostate radiotherapy and hence is the focus of this thesis. Chapter 1 outlines and compartmentalises the current non-ART prostate radiotherapy treatment planning pathway and presents a literature review of published attempts to automate it. Subsequent chapters present individual studies that build on these published works and culminate in a fully-automated knowledge-based treatment planning workflow.

Rigorous analyses of treatment plans generated using this fully-automated workflow show that they are generally of at least comparable quality to manually generated clinical treatment plans – although gross inaccuracies in auto-contouring of anatomical structures can be a limitation. Nevertheless, the proposed fully-automated

workflow could provide significant efficiencies for treatment planning departments, which could be exploited to aid with the implementation of ART techniques and provide patients with earlier access to their cancer treatments.

Publication of the fully-automated knowledge-based treatment planning workflow and the data used to drive it would also allow treatment planning information and expertise developed over more than a decade at The Christie NHS Foundation Trust to be shared easily and quickly with other centres worldwide.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the 'Copyright') and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which The University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the 'Intellectual Property') and any reproductions of copyright works in the thesis, for example graphs and tables ('Reproductions'), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses

Acknowledgements

According to legend, Albert Einstein once quipped 'if I knew what I was doing, it wouldn't be called research, would it?!'. A more apt characterisation of my last three years working on the project presented here, I have not been able to find. Fortunately, though, this has not been a solo journey. Marianne Aznar and Phil Whitehurst have provided their guidance and support and I extend my sincere thanks to you both.

Submission of this thesis marks the culmination of 5 years on the HSST programme. Throughout this time, I have regularly been away from my clinical department at The Christie and I am grateful to all of my colleagues who have supported this. Perhaps most notably are my office-mates, past and present: Will Beasley, Rob Chuter, Matthew Lowe, Andrew Pollitt and Dan Welsh. They have put up with rants, given advice and offered consolation through the darkest of HSST times. Thank you all.

Of course, the person who has suffered the most throughout my time on HSST, is my amazing wife Lindsay. Not only has she endured an irritable husband for the past 5 years she also happens to be one of my long suffering colleagues acknowledged above. Thank you for everything.

Preface

This thesis is presented in journal format. Each chapter contains a standalone piece of work with an Introduction that sets out the rationale and aims, the author's contribution and if the work has been submitted for conference presentation or journal publication. For overall consistency within this thesis, some of text presented here differs from that submitted to conferences and journals and Manchester Harvard referencing is used throughout. Each chapter has an individual list of references specific to that chapter and a combined list of all references is presented at the end.

Chapter 1

1 Literature Review

1.1 Prostate Cancer Radiotherapy Treatment Planning

This thesis considers the role of automation in prostate radiotherapy treatment planning. Prostate cancer accounts for almost 50, 000 new cancer diagnoses each year in the UK (Cancer Research UK, 2018) and radiotherapy plays in important role in the treatment of 30 % of prostate cancer patients (National Cancer Registration and Analysis Service, 2018). Current practice throughout the radiotherapy community is for patient specific treatment plans to be generated manually by experienced treatment planners. However, it is widely reported that manual treatment planning is time consuming and introduces variation into the quality and consistency of clinical treatment plans (Chanyavanich *et al.*, 2011). Manual treatment planning is also not feasible for many approaches to adaptive radiotherapy (ART), which are reported to provide dosimetric improvements (Antico *et al.*, 2019), as increased workload quickly becomes prohibitive.

For busy centres, effective automation of the treatment planning process can

generate significant quality improvements and the required workload efficiencies to enable ART. It is therefore extremely desirable for clinical radiotherapy departments.

Approaches to automating the prostate radiotherapy treatment planning process are presented abundantly in the literature but prior to reviewing this literature here, it is first important to introduce the existing prostate treatment planning pathway to provide context to where and how automation techniques can be applied. The standard radiotherapy treatment planning pathway is illustrated in Figure 1.1 and typically from 'Planning CT Scan' to 'Treatment' takes approximately two weeks. Although different radiotherapy centres may implement the specifics of each box in Figure 1.1 differently, the fundamentals of the pathway remain the same. The following subsections briefly outline the essential requirements of each step in the pathway.

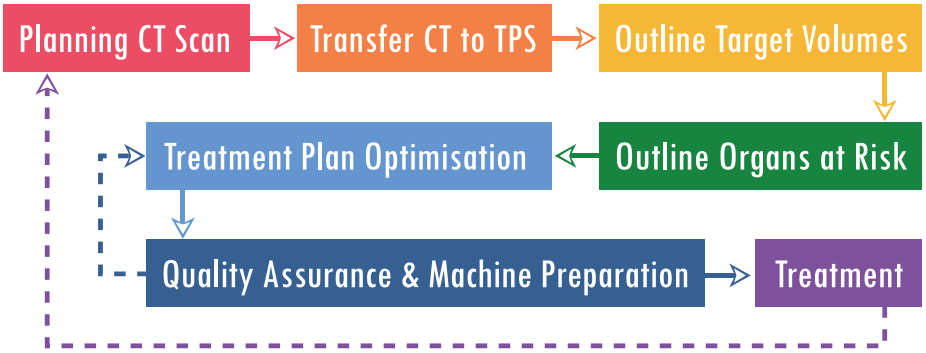

Figure 1.1: Prostate radiotherapy treatment planning pathway.

1.1.1 Target & Organ at Risk Delineation

Following a prostate cancer diagnosis and radiotherapy referral, the patient undergoes a computed tomography (CT) scan in the intended treatment position. This CT scan is transferred from the scanner to the treatment planning system (TPS) where a consultant oncologist outlines the prostate and, depending on disease stage, seminal vesicles (SV). For standard prostate radiotherapy, prostate and SV form the clinical tumour volumes (CTV) (ICRU, 1999). After the prostate and SV have been delineated, a treatment planner outlines the organs at risk (OAR) to which dose must be minimised and reported (ICRU, 1999). OARs for prostate cancer are the rectum, bladder and femoral

heads and possibly the bowel and urethral bulb.

Typically, delineation of these structures is performed manually on each slice of the CT scan and allows the therapeutic radiation dose distribution to be optimised and reported. Manual outlining of anatomical structures arguably constitutes the most significant bottleneck in Figure 1.1 and delays between the outlining stages can account for the majority of time it takes to complete the pathway.

1.1.2 Treatment Plan Optimisation

Once all necessary structures are outlined they are used to drive the optimisation of the patient specific treatment plan. Modern radiotherapy treatment planning follows an 'inverse planning' approach as opposed to tradition 'forward planning' (Hristov *et al.*, 2002). Forward planning is a technique where treatment planners use their expertise to manipulate treatment machine parameters in the TPS (beam energy, gantry and collimator angles, MLC positions etc.) to achieve an optimal dose distribution. For complex intensity modulated radiotherapy (IMRT) and volumetric modulate arc therapy (VMAT), however, there are too many machine parameters and variables to adjust and optimise in this manner. Inverse planning is the alternative and uses an automatic optimisation algorithm in the TPS. The treatment planner defines the desired dose distribution in terms of achievable objectives for the algorithm. The algorithm then automatically optimises the machine parameters to give a treatment plan that best satisfies the objectives defined by the planner.

Manually driven inverse planning for prostate VMAT is an iterative process whereby the treatment plan is increasingly improved with each optimisation until it meets some minimum dosimetric requirements and visually the dose distribution appears acceptable. The doses received by the outlined structures can be assessed using a number of parameters.

Three basic metrics used to characterise radiation dose to a 3D structure are the minimum, maximum and mean doses. Despite the relative simplicity of these metrics, they succinctly capture whether a structure is being grossly under- or over-dosed. With

modern radiotherapy, however, these simple metrics are often insufficient to capture whether the dose to a critical structure is fully optimised. ICRU report 83 (ICRU, 2010) introduced dose volume histogram (DVH) reporting which provides information of how optimally the dose is distributed within a volume.

Figure 1.2 shows a differential DVH on the left for a PTV where most of the volume has received a dose close to the 6000 cGy prescription. This can be converted into the cumulative DVH shown on the right using equation 1.1 (ICRU, 2010).

$$\mathrm{DVH}_{\mathrm{cumulative}}(D) = 1 - \frac{1}{V}\int_0^{D_{\max}} \frac{dV(D)}{dD}\,dD, \qquad 1.1$$

where V is the volume of the structure, $D_{\max}$ is the maximum dose to the structure and the integrand is the rate of change of the differential DVH with respect to dose. Cumulative DVHs make it easy to compare the dose to a structure from different treatment plans. DVH data can also be used to generate quantitative dose statistics and ICRU report 83 defines the nomenclature for these. The minimum dose, D, that is received by the fractional volume of a structure, V, is written as D_V (for example, $D_{95\,\%}$ is a minimum dose that covers 95 % of the structure). Volume may also be specified in absolute terms such as $D_{1\,\mathrm{cc}}$. Similarly, the volume that receives at least some absorbed dose is written as V_D (for example, $V_{6000\,\mathrm{cGy}}$ is the volume of the structure that receives at least 6000 cGy).

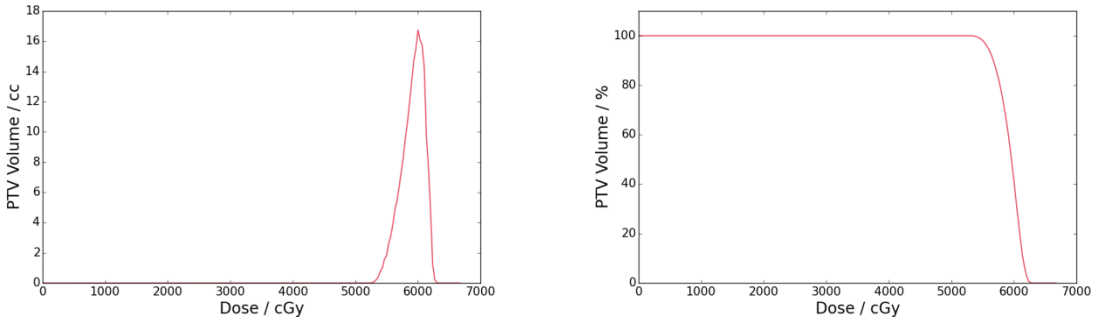

Figure 1.2: Differential (left) and cumulative (right) DVHs for a PTV.

DVH statistics provide a quick and easy means to assess and compare dose distributions. However, when a 3D dose distribution is collapsed into a 2D graphic, information about the spatial distribution of dose is lost and so visual inspection of the dose distribution using isodose lines overlaying the 3D CT scan is required.

1.1.3 Quality Assurance & Machine Preparation

Following the Towards Safer Radiotherapy recommendations (The Royal College of Radiologists *et al.*, 2008), all prostate VMAT treatment plans that are generated by trained treatment planners are independently checked as a means of quality assurance. This well established quality assurance procedure checks that the TPS parameters and settings are correct and appropriate and checks that the dose distribution has been sufficiently optimised. In addition, a consultant clinical oncologist also reviews the plan to ensure the dose is clinically acceptable. Plans that fail either the physics or clinical quality assurance are returned to the treatment planner for correction.

In addition to the treatment plan quality assurance, an independent check of the TPS dose calculation is also performed. Potential inaccuracies with the clinical dose calculation can generally be characterised as systematic and random errors. Systematic errors are caused by insufficiencies in the TPS beam model and dose calculation algorithm such that it does not represent the behaviour of the actual treatment machine. For example, if the TPS beam energy spectrum does not reflect the radiation emitted from the treatment machine, the TPS will systematically get its calculations incorrect. Random errors occur when the TPS does not perform its calculation as expected. These may be the result of, amongst other causes, data corruption or instability of the TPS software or hardware.

TPS dose calculation quality assurance involves comparison of the TPS calculation with an independent measurement and calculation to eliminate the possibility of systematic and random errors. Gamma analysis (Low *et al.*, 1998) is a commonly used means of quantifying the similarity between different dose distributions – i.e. the primary TPS calculation and the independent calculation or measurement. It

considers simultaneously the dose difference between corresponding voxels in the distributions and also the distance to agreement (DTA). Some pass-fail criteria are specified and the percentage of pixels passing the analysis gives a measure of similarity.

Figure 1.3 illustrates the gamma analysis technique in 2D. The x-y plane in Figure 1.3A represents the plane of the two overlaid dose distributions with the origin centred on the current voxel being analysed. The yellow shaded region shows the DTA radius. In Figure 1.3B, the vertical axis gives the pixel-by-pixel percentage difference between the dose distributions and the gamma analysis dose difference criteria is represented by the dark blue arrows. Gamma analysis combines the DTA and dose difference criteria by generating an ellipsoid of acceptance as shown in yellow in Figure 1.3C. If the percentage dose difference surface does not intersect this ellipsoid, the pixel fails the gamma analysis. This process is repeated for every pixel in the reference distribution and the final reported value is the percentage of pixels passing the gamma analysis.

On successful completion of all the quality assurance, the final preparations for clinical treatment delivery are made by treatment radiographers.

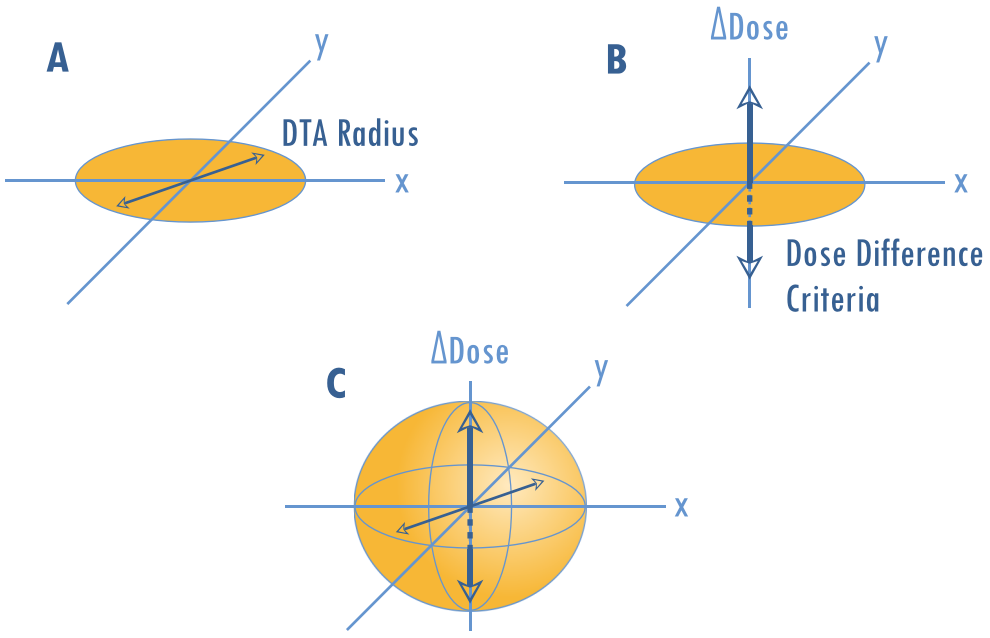

Figure 1.3: Gamma analysis illustration. A) shows the DTA acceptence radius shaded in yellow. B) adds the dose difference acceptance criteria. C) shows the ellipsoid of acceptance as the shaded yellow ellipsoid.

1.1.4 Treatment Delivery & Plan Adaptation

The Royal College of Radiologists' recommends treating intermediate- to high-risk prostate cancer (T1b-T3aN0M0) with radiotherapy prescribed as 6000 cGy in 20 fractions treated daily (excluding weekends) over four weeks (The Royal College of Radiologists, 2016). A regular cone-beam CT (CBCT) online imaging protocol is often followed to review anatomical changes or differences from the CT scan used for radiotherapy treatment planning. When differences are noted by treatment radiographers they are reviewed dosimetrically and clinically by the radiotherapy physics and clinical oncologist teams. Where it is required, a new treatment planning CT scan is acquired and the treatment planning process starts anew to generate a revised treatment plan for the remainder of treatment. This review and re-planning process is represented by the dashed purple line in Figure 1.1

1.2 Auto-Segmentation

Segmentation of a medical image refers to the identification of anatomical structures within it. As discussed in section 1.1.1, manual-segmentation is performed by drawing around each structure on a computer screen on each slice of the CT image. This is time consuming, however, and suffers from significant inter- and intra-observer variation (Han *et al.*, 2008; Collier *et al.*, 2003). To minimise these variations, peer review of clinical structures is increasingly common within hospitals and the Royal College of Radiologists recommend using target volume definition protocols 'across a clinical network and ideally nationally' (Royal College of Radiologists, 2017). Although these measures help, they do not fully remove inter- and intra-observer variations and international protocol consensus certainly does not exist for the majority of treatment sites.

Auto-segmentation of CT images can address these issues by replacing the time consuming manual-segmentation and by providing consistency to the segmented structures. Unfortunately, auto-segmentation presents a number of challenges. First is the poor soft tissue contrast, noise and artefacts observed in CT images (Li *et al.*, 2016).

Poor soft tissue contrast and noise are illustrated in Figure 1.4 where a central anterior-posterior plot of CT number shows little variation other than noise as the profile runs through muscle, adipose tissue, bladder (yellow), prostate (red) and rectum (blue). A larger difference is seen between the soft tissues and bone.

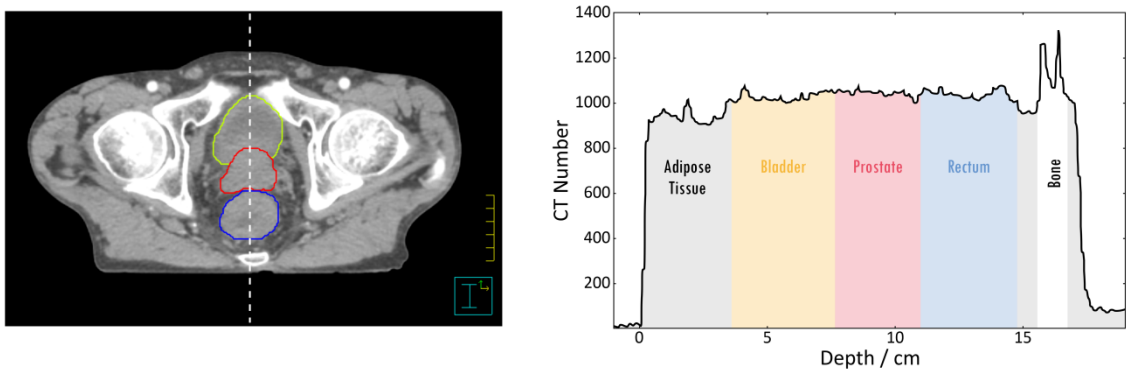

Figure 1.4: A pelvis CT image with an anterior-posterior plot of CT number along the white dashed line. Little variation in the CT number is observed between adipose tissue, bladder, prostate and rectum. A larger difference is observed between these soft tissues and bone.

The second major challenge of automatic segmentation is the variation of organ shape (and to some extent density) between patients (Bzdusek *et al.*, 2012). This is illustrated in Figure 1.5, where six prostate cancer patient CT scans are shown that have been manually segmented and show significant variations in muscle and fat composition and prostate (red), seminal vesicles (blue)[†], rectum (brown) and bladder (yellow) size and shape.

For structures where there is sufficient contrast at boundaries (such as bone, lungs or the whole body), a threshold CT number can be specified and the structure can be semi-automatically outlined based on this value (Özsavas *et al.*, 2014). Alternatively, for structures where there is little variation in the shape between patients (such as the femoral heads) a simple model-based approach can be applied (Seim *et al.*, 2008). In the pelvis, however, where little soft tissue contrast is seen and where there is large variation of the soft tissue structures, more sophisticated approaches need to be employed. These are considered in the following subsections.

[†] Not all of the CT slices in Figure 1.5 display seminal vesicles as slices containing prostate, seminal vesicles, rectum and bladder do not always exist.

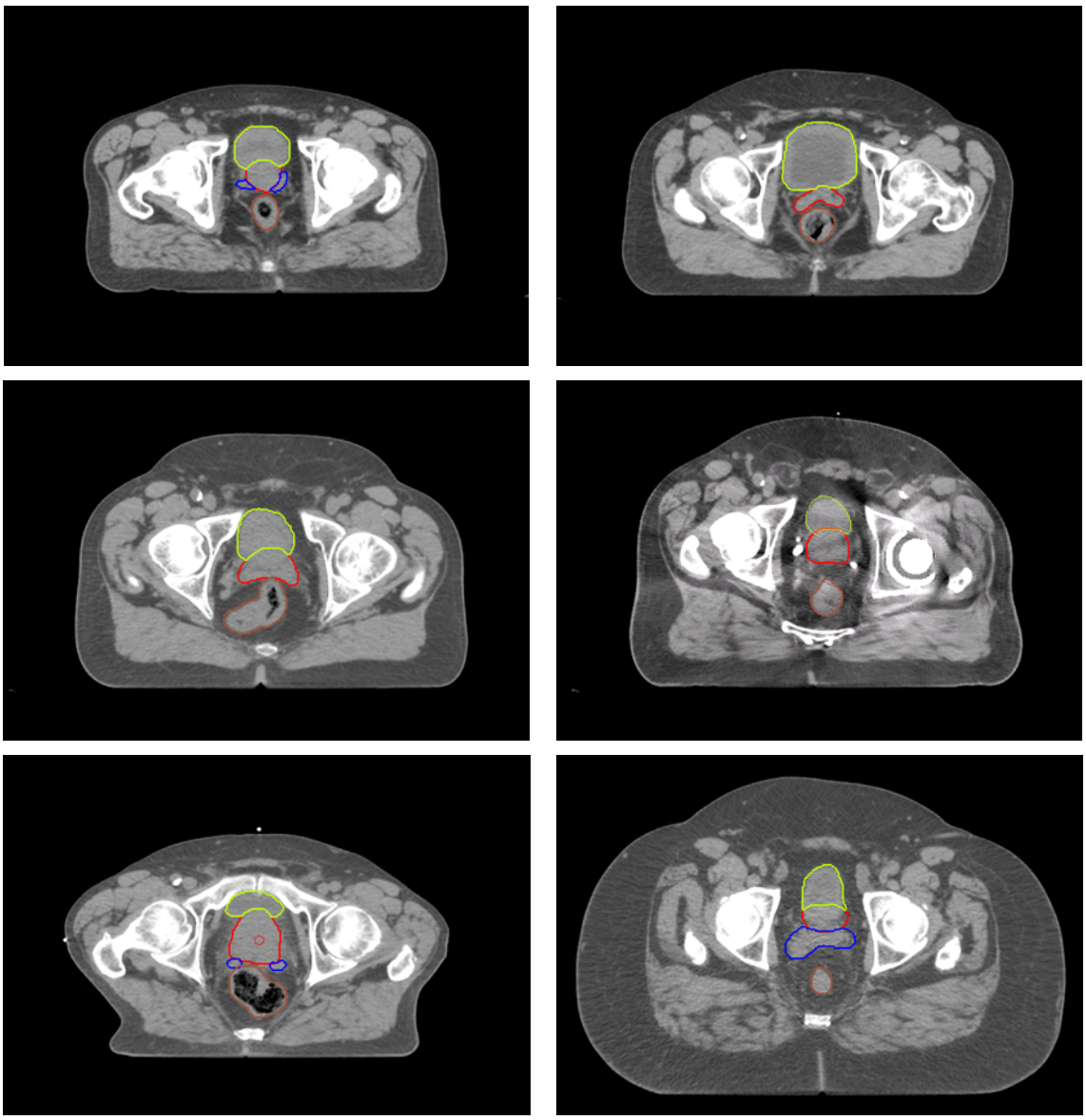

Figure 1.5: CT images demonstrating significant variations in pelvic anatomy where the prostate, SV, bladder and rectum and outlined in red, blue, yellow and brown respectively.

1.2.1 Atlas-Based Segmentation

Atlas-based auto-segmentation is defined by Han *et al.* (2008) as 'the process of performing segmentation on novel data using the knowledge of a prior segmentation – a dataset that has had the structures of interest already labeled (sic)'. Typically, deformable image registration (DIR) is used to map the atlas image to the novel image and produce a deformation vector field (DVF) (Castadot *et al.*, 2008). The DVF can be applied to the atlas structures, which are then transferred to the novel image – thus auto-segmenting the anatomy in the novel image. This process is illustrated in Figure

1.6. The majority of auto-segmentation work reported in the literature in recent years has relied on atlas-based approaches (Aljabar & Gooding, 2017).

Aljabar *et al.* (2009) writes that the performance of atlas-based auto-segmentation is dependent on the accuracy of the DIR, the quality of the atlas segmentation and the similarity of the atlas image to the novel image. Han *et al.* (2008) describes the first of these dependencies as the most critical component of atlas-based auto-segmentation – if the deformed atlas is not well matched to the novel dataset, the propagated contours will not overlie their respective anatomical structures. Collier *et al.* (2003) considers the second dependency and reports significant inter- and intra-observer variations in manually drawn clinical structures which limit the quality of any atlas-based auto-segmentation – inaccuracies and disagreements in the atlas segmentation are transferred to the novel dataset. For the final dependency, Han *et al.* (2008) suggests that atlas-based auto-segmentation accuracy degrades with dissimilarity between the atlas and the novel image. These dissimilarities can be both anatomical and image characteristic such as noise, artefact, etc..

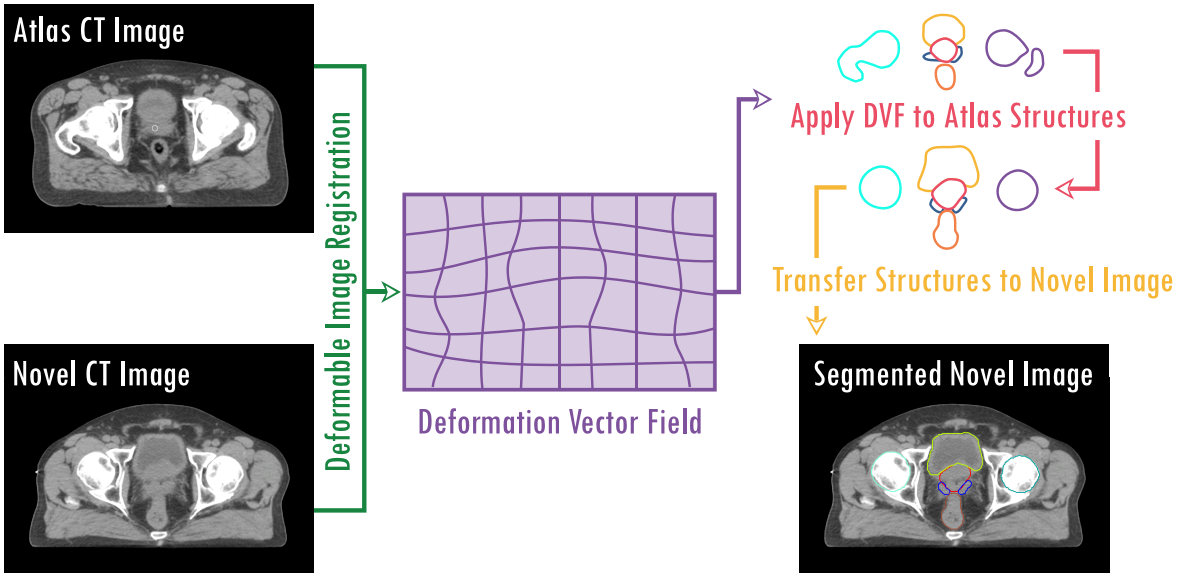

Figure 1.6: Atlas-based auto-segmentation. The atlas image is first mapped to the novel image using DIR. Atlas contours are then deformed with the same deformation vector field and transferred to the novel image.

The use of multiple atlases has been shown to improve the performance of atlas-based auto-segmentation significantly (Acosta *et al.*, 2017; Aljabar *et al.*, 2009). In a multi-atlas approach, either the single atlas that is most similar to the novel image is selected for auto-segmentation from a database of atlases or multiple auto-segmentations are performed and the resulting segmentations are combined in some way – such as with the STAPLE algorithm (Warfield, Zou & Wells, 2004).

A criticism of atlas-based auto-segmentation is that selection of the most suitable atlas from a database is computationally demanding and so in practice this means the number of atlases in a database is typically restricted to 10-20 (Peressutti *et al.*, 2016). When a novel dataset that needs to be segmented deviates significantly from any of the available atlases, the performance of atlas-based approaches is fundamentally limited.

1.2.2 Deep Learning-Based Segmentation

Deep learning is a type of machine learning, where input data are used to drive model parameters with the aim of minimising differences between the model output and observation. A detailed discussion of machine and deep learning is beyond the scope of this thesis, but a brief overview is provided below. A more thorough discussion can be found in Litjens *et al.* (2017).

Machine learning methods generally have an artificial neural network (ANN) as their basis. An ANN consists of layers of software units called neurons where each layer receives an input from the previous layer and performs a process before passing an output to the next layer. Neurons have an activation, a set of weights and a set of biases that describe the behaviour of a particular model parameter. The activation, weights and biases are optimised in a training process using existing data. A simple ANN is illustrated in Figure 1.7, where there are two 'hidden' layers linked by weighted connections. Where the ANN has many hidden layers between the input and output layers, it is referred to as 'deep' and hence can be used for 'deep learning'.

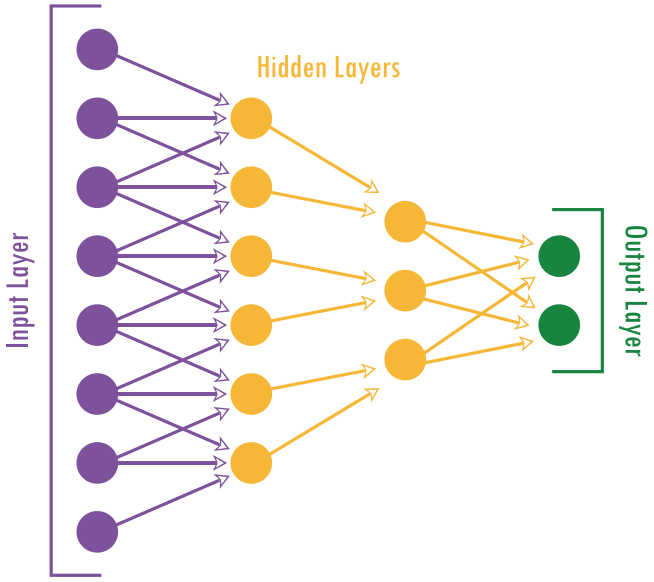

Figure 1.7: A simple artificial neural network (ANN) used for machine learning. The input layer (purple) passes information to the first hidden layer (yellow). This is then passed to the second hidden layer (also yellow) before the final output is realised at the output layer (green). Sets of weights and biases between layers, which are optimised in the training process are represented by left-right arrows. Based on Figure 2 in Litjens *et al.* (2017)

Convolution is a common image processing technique and Aljabar & Gooding (2017) write that incorporation of convolution into deep learning frameworks, with a convolution neural network (CNN), has 'yielded very impressive results' for generating feature maps and hence for automatic segmentation. These results have led to a surge of interest in the field. Indeed, Litjens *et al.* (2017) present a survey on the use of deep learning for medical image analysis where over 300 papers were reviewed mostly from 2016 and 2017 and they state that 'segmentation is the most common subject of papers applying deep learning to medical imaging'.

Deep learning models need to be trained with large clinical datasets to optimise up to 100 M activations, weights and biases. Although this training process is computationally demanding and time consuming, once trained, deep learning models can segment new data rapidly.

1.2.3 Performance Metrics

Evaluations of auto-segmentation techniques that are reported in the literature use a number of parameters to assess geometric accuracy. These range from qualitative

judgements, where clinicians score auto-contours on a pre-defined scale (Stapleford *et al.*, 2010; Hardcastle *et al.*, 2013; Thor *et al.*, 2011) to quantitative metrics that numerically score the similarity of one contour to another. The following subsections describe the two most commonly used quantitative performance metrics in the literature.

1.2.3.1 Volume-based

Volume-based metrics calculate the overlap of two contours as a fraction of their total volume (Sharp *et al.*, 2014). Dice similarity coefficient (DSC) (Dice, 1945) is the most commonly used volume-based metric in the literature. It ranges from 1, where the contours are identical, to 0 where there is no overlap. It is calculated using equation 1.2 where V_a and V_b are the volumes of the contours being compared. A 2D illustration of DSC is shown in Figure 1.8.

$$\mathrm{DSC} = \frac{2(V_a \cap V_b)}{|V_a| + |V_b|} \qquad 1.2$$

Sharp *et al.* (2014) suggests that the strength of DSC lies in its simplicity and ease of understanding. However, they criticise its sensitivity to fine details in large structures. They also suggest it is inappropriate for very small structures where significant overlap cannot be achieved and where small discrepancies in the surfaces can dramatically affect the volume.

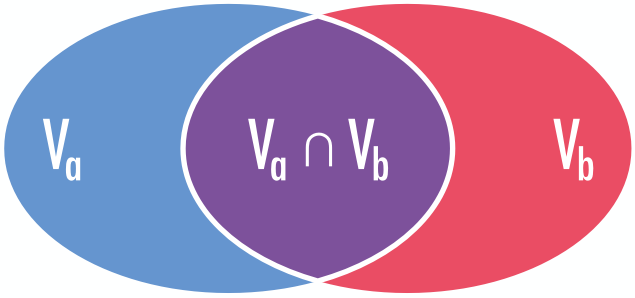

Figure 1.8: DSC is based on the fractional overlap of two contours.

1.2.3.2 Surface-based

Surface-based metrics quantify the distance from the surface of one contour to the surface of another. For each point a on a surface A the minimum distance to agreement (DTA) with point b on surface B is calculated using equation 1.3.

$$\mathrm{DTA}(a, B) = \min_{b \in B} \|a - b\| \tag{1.3}$$

If equation 1.3 is used for all points on surface A, a DTA histogram is produced and from this a number of surface-based metrics can be obtained (Sharp *et al.*, 2014). This is illustrated in Figure 1.9 where a square with side 10 cm is compared with a circle of radius 5 cm.

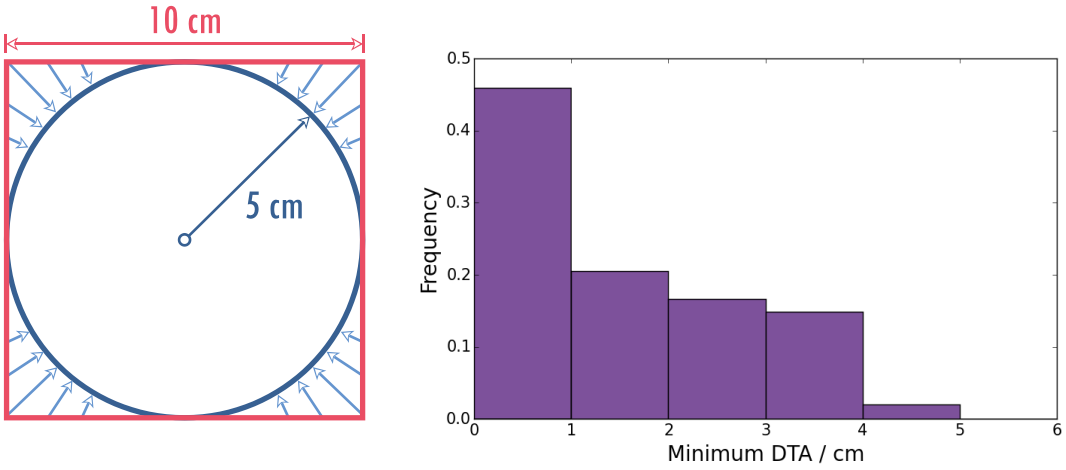

Figure 1.9: Minimum DTA between the two surfaces is calculated and plotted as histogram.

The most commonly used surface-based metrics are the mean and median DTA. DTA is sometimes called the Hausdorff distance (HD) in the literature but, to avoid confusion in this thesis, DTA will be used throughout. Sykes (2014) states that although surface based metrics quantify the similarity of two different contours, interpreting the clinical significance of discrepancies is difficult.

1.2.4 Commercially Available Packages

At the time of writing this literature review, a number of packages that perform atlas-

based auto-segmentation are available commercially and their performances are evaluated in the literature. Auto-segmentation literature tends to focus on studies in the head and neck, arguably because there are more structures to delineate, which makes for greater potential efficiency savings and because the structures tend to suffer less from the variations described in section 1.2.

DLCExport (Mirada Medical, 2021) is the single package that uses a deep learning approach that is available commercially. However, this is relatively new to the market and evaluations of its performance have not been published at the time of writing. A review of three of the atlas-based auto-segmentation packages (SPICE (Bzdusek *et al.*, 2012), Mirada RTx (Mirada Medical, 2017) and ABAS (Elekta, 2013)) is presented below.

1.2.4.1 SPICE

SPICE is an atlas-based auto-segmentation approach that uses 'several deformable image registration algorithms with model-based segmentation and probabilistic refinement' (Bzdusek *et al.*, 2012). It is a fully automated system within the Pinnacle^3 TPS and claims to segment normal and target tissues accurately in head and neck, thorax, prostate and abdominal CT images.

Qazi *et al.* (2011) gives an in-depth description of the SPICE pelvic auto-segmentation pipeline, which is broken into three steps. In the first step a translation, rotation and scaling registration is performed to register trained tissue probability atlases to the patient CT scan. Organs are then positioned as surface mesh models on the CT scan using organ specific probability maps. The final step then adapts the surface meshes to trained image features using model-based segmentation.

The atlases and models used by SPICE cannot be customised by individual centres to match local (or even national if they exist) outlining protocols. Rather, a number of variations are provided from which the planner or clinician selects the most appropriate contour.

Zhu *et al.* (2013) reports the performance of SPICE in the male pelvis where the bladder, bilateral femoral heads, rectum and prostate were auto-segmented on 30

patient CT scans. DSCs for all of these structures were relatively consistent at ~0.8 and mean DTA varied from 0.5 cm for femoral heads to between 0.8 cm and 1.0 cm for the rectum, bladder and prostate. Although Qazi *et al.* (2011) state that a strength of SPICE is its ability to operate in low soft tissue contrast, comparison of the male pelvis results of Zhu *et al.* to similar studies of head and neck structures (Thomson *et al.*, 2014; Zhu *et al.*, 2013) suggest that low soft tissue contrast may be a challenge for SPICE.

1.2.4.2 ABAS

Atlas-Based Auto-Segmentation (ABAS) is a stand-alone software package available from Elekta that is not integrated into a particular TPS (Elekta, 2013). It takes a DICOM CT image input and outputs a DICOM structure file that can be imported into any TPS. Registration atlases are fully customisable by individual centres although ABAS is supplied with an atlas library.

The ABAS auto-segmentation process is divided into three steps (Han *et al.*, 2008). As the algorithm runs through these steps, the degrees of freedom for the registration are increased and each step is used as an initialisation for the next. The first step is a mutual information linear registration (translation, rotation and scaling) of the patient and atlas images to correct for global differences in position, orientation and scaling. The second step takes this registration and uses the structures to be segmented to drive an affine transformation based registration (Arsigny *et al.*, 2006). Large structures are broken down into smaller structures as a single affine transformation is insufficient to account for large inter-patient variation. The final step is a shape constrained dense deformable registration and mapping of atlas structures' boundary points on to the novel image to achieve auto-segmentation of the novel image.

Published literature reporting evaluations of ABAS focusses primarily on head and neck auto-segmentation (Teguh *et al.*, 2011; Speight *et al.*, 2014; Lim & Leech, 2016). Although these papers report that the ABAS auto-contours required manual editing prior to clinical acceptance they also suggest that this modification is quicker than outlining all of the structures *de novo*.

Greenham *et al.* (2014) present the only study in the literature that evaluates the ability of ABAS to auto-segment structures in the male pelvis. Unfortunately, this evaluation is limited to a qualitative assessment of the structures and a comparison of the overall volumes of the auto- and corresponding manual-contours. For the qualitative assessment, clinical oncologists scored the auto-contours on a 7 point scale from 'structure acceptable to treat 'as is'' to 'No resemblance to the clinical structure or > 75 % slices needing edit' and for the volume comparison, Pearson's product-moment correlation coefficient was used to assess the correlation between auto-contours and 'as treated' volumes. They claim that femoral heads, bladder, rectum and prostate auto-contours were clinically acceptable in 90 %, 80 %, 50% and 20 % of cases respectively without manual modification and present strong correlations between auto- and manual-contour volumes. Interestingly, they write that inclusion of more atlases in a multi-atlas approach (ranging between 2 and 10 combined using STAPLE) did not consistently improve the volume correlation and the best results were obtained using the built-in demo atlas. This is in contrast to the work of Acosta *et al.* (2017) and Aljabar *et al.* (2009) where it is claimed that increasing the number of available atlases improves segmentation accuracy.

It is unfortunate that Greenham *et al.* do not report standard geometric accuracy metrics (DSC and DTA as discussed in section 1.2.3) for the ABAS contours and do not consider the utility of the auto-contours for treatment plan optimisation and dose reporting. Voet *et al.* (2011) does conduct such as study (albeit for head and neck rather than prostate cancer patients) and suggest that geometric inaccuracies in auto-contoured targets cause significant under-dosage of the target but similar inaccuracies in OARs do not lead to statistically significant dose differences.

Referring back to Figure 1.1, the outlining process is divided into two: 'outlining target volumes' and 'outlining organs at risk'. Since these two processes are performed separately and, in some cases, up to a week apart, it is important that the utility of auto-segmentation for targets and OARs is considered separately.

1.2.4.3 MIRADA RTx

Mirada Medical Ltd. offers a number of image registration and auto-segmentation products. These products are stand-alone packages and not specific to any TPS. Mirada RTx offers rigid and deformable image registration with CT, MRI and PET images (Mirada Medical, 2017). Manual target contouring can be performed within Mirada RTx on multi-modality datasets simultaneously in any plane through the image. Atlas-based auto-segmentation can also be performed with customisable segmentation atlases.

Workflow Box is an additional product offered by Mirada Medical Ltd. to perform 'Zero-Click Contouring'. In Figure 1.1, Workflow Box would sit between 'Planning CT Scan' and 'Transfer CT to TPS'. Images exported from the scanner to Workflow Box have the structures segmented automatically before being forwarded to the TPS.

As with ABAS described in section 1.2.4.2, the majority of published evaluations of Mirada RTx in the literature focus on auto-segmentation of head and neck structures (Larrue, Kadir & Gooding, 2013; Gugyeras *et al.*, 2017). Although the results of these studies are generally positive and demonstrate similar geometric accuracy results to SPICE and ABAS, they cannot provide any indication of the performance of Mirada RTx in the male pelvis.

1.3 Knowledge-Based Treatment Planning

Knowledge-based (KB) treatment planning automatically incorporates prior experience into the treatment planning optimisation process (Nwankwo *et al.*, 2015). This can make treatment planning more consistent within a radiotherapy centre and can generate workflow efficiencies by removing some of the time consuming human interventions. For small, inexperienced centres, where IMRT plan quality has been shown to be poorer than in larger experienced centres (Chung *et al.*, 2008), KB treatment planning potentially offers a way to disseminate experience indirectly and rapidly.

Nwankwo *et al.* (2014) write that prostate radiotherapy treatment planning is the most common site for KB treatment planning research and the following sections review the literature.

1.3.1 Optimisation Class-Solutions

Perhaps the simplest form of KB treatment planning is the use of a rigid class-solution (Wood *et al.*, 2016). An ideal class-solution consists of a list of optimisation objectives and standard beam parameters that can be used to generate clinically acceptable and dosimetrically optimal treatment plans regardless of the specifics of a patient's anatomy. Typically, class-solutions are developed by generating treatment plans for a representative group of patients and taking the average of the final optimisation objectives. While this is KB treatment planning to a degree, it really only offers a starting point for manual optimisation and so is not considered further here.

1.3.2 Search and Retrieve

Search and retrieve approaches to KB treatment planning generally consist of a number of manually generated treatment plans that ideally cover the full range of anatomical variation that is seen within the patient population. Given a new patient, some quantitative characterisation of the anatomy is performed and the most similar patient from the KB is identified. The treatment planning parameters used to generate the KB treatment plan are then used to optimise the new treatment plan. Selection of the 'most similar patient' can be difficult, however, and the following subsections describe the two most commonly used approaches.

A danger of the search and retrieve approach is that it assumes the single KB treatment plan on which the new treatment plan is based is fully optimised. It is widely accepted in the literature that manual treatment plan optimisation is subjective and often treatment planners accept plans before they are fully dosimetrically optimal (Chanyavanich *et al.*, 2011). Since search and retrieve approaches select a single instance from the KB, the degree to which the new plan is optimal is dependent on the degree to which the KB treatment plan was optimised. This propagation of dosimetrically sub-optimal treatment plans is difficult to prevent. Counters to this are to peer review each treatment plan rigorously prior to inclusion in the KB but this is time consuming and ultimately limits the number of plans that can be included.

1.3.2.1 Overlap Volume Histograms

Wu *et al.* (2009) describe a search and retrieve KB treatment planning methodology that uses overlap volume histograms (OVH). An OVH describes the proximity of the OAR to a target and provides a way to infer the likely DVH based on previously optimised treatment plans. The aim of Wu *et al.* was to quality assure expert-optimised treatment plans rather than to automate the optimisation process. An OVH example is illustrated in Figure 1.10 where the OVH has been calculated using equation 1.4 for the Target and each OAR_i (based on Fig 1 from Wu *et al.*, 2009).

$$\mathrm{OVH}(r) = \frac{|\{p \in \mathrm{OAR}_i | d(p, \mathrm{Target}) \leq r\}|}{|\mathrm{OAR}_i|}, \tag{1.4}$$

where $d(p, \mathrm{Target})$ is the signed distance between a point p within OAR_i and the Target boundary. Although the OARs in Figure 1.10 are the same size and shape they have different geometric relationships to the target and thus markedly different OVHs. Wu *et al.* (2009) write that, for a conformal dose distribution, the OVH of an OAR is directly related to its DVH. Therefore after expert manual optimisation, querying a database of previous OVHs and corresponding DVHs and comparing the retrieved results with the current plan can highlight if the current plan meets existing quality standards.

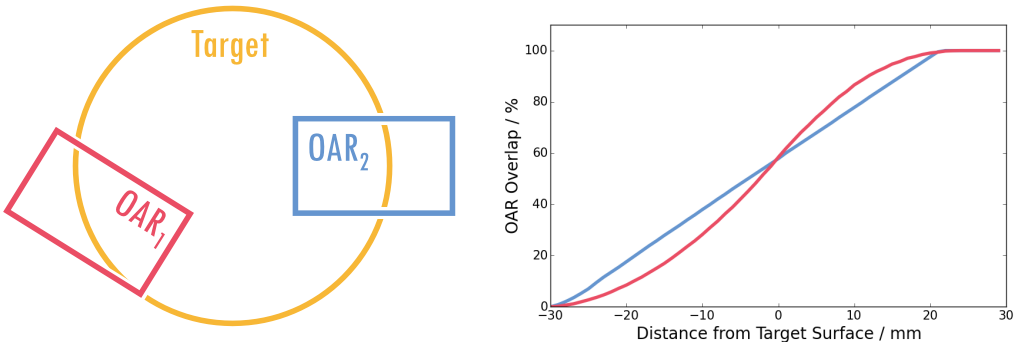

Figure 1.10: Illustration of the OVH concept based on Wu *et al.* (2009).

OVH analysis using a database of previous generated treatment plans is reported to provide a reliable KB quality assurance technique for new treatment plans (Wu *et al.*,

2009; Janssen *et al.*, In press; Wang, Heijmen & Petit, 2017). Petit *et al.* (2012) extended the use of OVHs from retrospective treatment plan quality assurance to prospective prediction of achievable OAR doses to drive treatment plan optimisation automatically. They demonstrated that selection of optimisation objectives from a KB of prior treatment plans can, in the majority of cases, generate superior treatment plans compared to those generated manually.

1.3.2.2 Beam's Eye View Projections

Chanyavanich *et al.* (2011) and Good *et al.* (2013) present search and retrieve approaches where KBs are queried based on 2D beam's eye view (BEV) projections of anatomical structures. In both studies, BEV projections of the new plan are compared with each plan in the KB using mutual information (MI). The treatment planning parameters from the most similar plan are then used directly to drive the optimisation of the new plan. MI is relatively similar to the concept of DSC discussed in section 1.2.3.1 and is calculated using equation 1.5.

$$\mathrm{MI}(O_{ref}; O_{novel}) = \sum_{O_{ref}} \sum_{O_{novel}} P(O_{ref}; O_{novel}) \log_2 \frac{P(O_{ref}; O_{novel})}{P(O_{ref}) P(O_{novel})}, \qquad 1.5$$

where $P(O_{ref})$ and $P(O_{novel})$ are the marginal distributions and $P(O_{ref}; O_{novel})$ the joint probability distribution of projections of the reference and novel organ, O. Figure 1.11 illustrates the MI comparison technique.

Both Chanyavanich *et al.* (2011) and Good *et al.* (2013) present studies of the application of the BEV projection search and retrieve method to prostate IMRT treatment planning. Both report that KB treatment plans were typically at least comparable dosimetrically (and generally superior) to those generated manually. Chanyavanich *et al.* also adds that the approach has the potential to improve treatment planning efficiency. The work of these two studies, however, is based on either step-and-shoot or dynamic IMRT treatment planning and the MI similarity index was computed for each of seven fixed beam angles. It is not clear from either study whether

this approach could be generalised to work for VMAT treatment planning where the subtleties of OAR and target configuration viewed over a full 360° arc might not be captured from a selection of static beam angles.

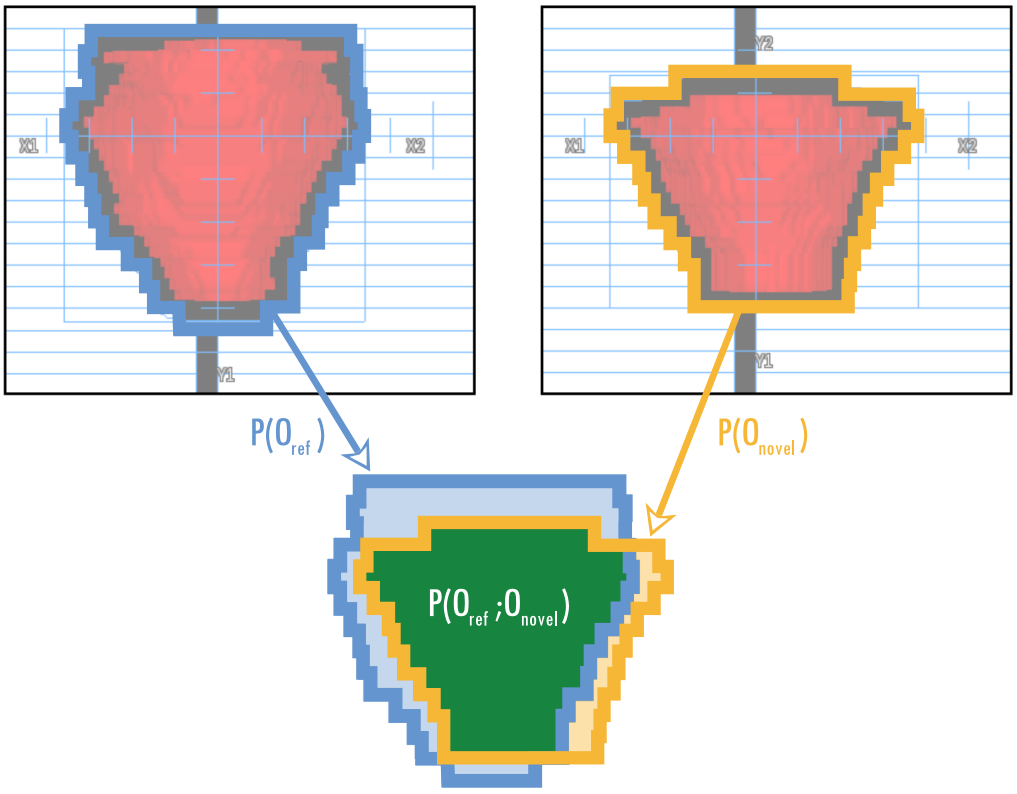

Figure 1.11: Illustration of mutual information (MI) comparison of 2D BEV structure projections.

The test cases used in Good *et al.* (2013) were taken from a small community hospital where prostate IMRT had recently been adopted. Although the results suggest that KB treatment planning can be used to transfer experience from larger, experienced centres to smaller, inexperienced centres, it does not specifically benchmark the KB treatment plans generated for the community hospital against the manually generated treatment plans generated in the larger research centre. Therefore it is not clear whether the improvement in treatment plan quality for the community hospital made them comparable to the manual plans generated at the centre of Good *et al.* (2013).

1.3.3 Machine Learning

The application of machine learning to various aspects of the radiotherapy pathway is becoming increasingly common. Machine learning approaches to KB treatment planning aim to learn patterns from the KB that can be generalised and used to plan or quality assure new treatment plans. Arguably, this offers advantages over the search and retrieve techniques discussed in section 1.3.2 as optimisation is not based solely on one prior treatment plan but on the whole KB population.

Yang *et al.* (2013) present a machine learning KB treatment planning approach where rectum DVH and OVH relationships are evaluated over a population (albeit limited to 21 patients). They suggest relatively strong correlations exist between rectum $D_{15\,\%}$, $D_{20\,\%}$, $D_{25\,\%}$, $D_{35\,\%}$ and $D_{50\,\%}$ and $L_{15\,\%}$, $L_{20\,\%}$, $L_{25\,\%}$, $L_{35\,\%}$ and $L_{50\,\%}$ respectively (where $L_{X\,\%}$ refers to the distance from the PTV surface that encompasses X % of the OAR). Yang *et al.* contend that automatically generated treatment plans optimised based on predictions using these correlations are superior to those generated manually using an iterative trial-and-error approach.

The 21 plans used by Yang *et al.* (2013) to generate the KB may be insufficient to fully encompass the range of patient anatomy seen amongst a population of prostate cancer patients (recall Figure 1.5) and the variation in the degree of optimisation seen in manual treatment planning. Full details of the inclusion- and exclusion-criteria are not provided in the paper and so the reported correlations may be unfairly biased towards a specific group of patients and may not include the noise expected from a larger cohort of clinical prostate treatment plans. However, machine learning approaches benefit from ongoing refinement of the KB. For example, as the variation in optimisation reduces the KB can be influenced by these plans to improve the reliability of the relationships on which the optimisation is informed.

Zhu *et al.* (2011) and Yuan *et al.* (2012) present similar but more complex approaches to machine learning KB treatment planning than Yang *et al.* (2013). They use principal component analysis (PCA) to identify sources of variation in DVH and OVH data. PCA is a statistical technique that takes multi-dimensional data and finds

relationships between linear combinations of the dimensions that account for the most significant sources of variation.

A simple case of PCA is illustrated in Figure 1.12, where the 2D plot in Figure 1.12A shows no correlation between the variables x and y. This lack of direct x-y correlation does not necessarily mean that the variables are not related in a higher dimensional space. Figure 1.12B shows that with the addition of variable z, the 3D data may sit on the plane shaded in yellow. PCA aims to evaluate this plane by creating a new coordinate system with its origin at the centre of the plane. The axes are the principle components (PC1 and PC2) and are shown by the green arrows in Figure 1.12B. Figure 1.12C shows the data plotted in this manipulated coordinate system and it is apparent that the greatest variation in the data is in the direction of PC1 followed by PC2. Here, the 3D distribution of x, y and z has been fully characterised by a lower 2D model in terms of PC1 and PC2 and a 1D approximation of the model could be given solely in terms of PC1.

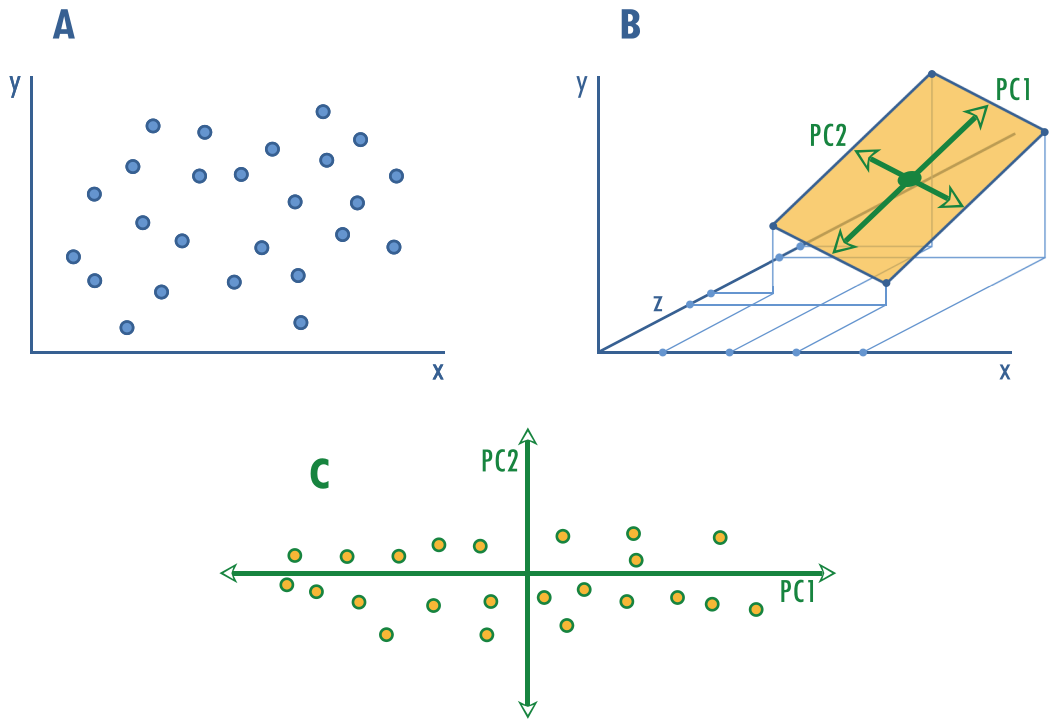

Figure 1.12: Illustration of principal component analysis (PCA).

Zhu *et al.* (2011) and Yuan *et al.* (2012) plot large numbers of bladder and rectum DVHs and OVHs in up to 50 dimensions and claim these can be reduced to either two or three principal components. They then use machine learning algorithms to identify features in the OVHs that allow the corresponding DVH to be predicted. Although the testing of the DVH predictions for treatment plan optimisation were limited to 14 test cases in Zhu *et al.* (2011) and 24 in Yuan *et al.* (2012), both papers report that the PCA method generates accurate DVH predictions.

Nwankwo *et al.* (2014) criticise approaches to KB treatment planning that only predict likely DVHs for OARs and do not describe the specific spatial distribution of dose. They present a method that predicts the dose to each voxel within an OAR and a metric to evaluate the accuracy of the prediction compared to the dose actually achieved by manual treatment planning. Each voxel within the rectum and bladder contours for 95 prostate IMRT treatment plans was characterised in terms of its Cartesian coordinates, reference dose, distance to PTV and slice level (zero on slices containing PTV and greater than zero otherwise). Models of rectum and bladder dosimetry were generated from the voxel characterisation data and used to predict voxel-by-voxel doses of 33 validation cases. They report that on average 18 and 28 of the 33 predictions for the bladder and rectum respectively were within 5 % of the PTV dose from a manual treatment plan and conclude that their KB method is an effective means of quality assuring new treatment plans. A further study by Nwankwo *et al.* (2015) concludes that the algorithm also shows promise as a prospective treatment planning approach.

1.4 Treatment Plan Complexity

The advantage of IMRT or VMAT treatment planning over 3D conformal radiotherapy (3D-CRT) is the degree to which dose can be conformed to an irregularly shaped target whilst simultaneously avoiding healthy OARs. By dividing beams into multiple segments (either as static field IMRT or dynamic arc VMAT), regions of a treatment field can be selectively shielded or exposed to modulate the radiation intensity over the exposed

aperture. As the intensity of the beam becomes increasingly modulated, smaller more irregularly shaped beam segments are required and such segments are typically less accurately modelled in the TPS than larger more open segments (Du *et al.*, 2014).

When generating IMRT treatment plans, a compromise has to be made between segmenting fields to give sufficient high dose conformality and limiting the segmentation so that confidence can be placed in the deliverability. One approach to managing this compromise that is much discussed in the literature is the quantification of treatment plan complexity.

To some extent, complexity can be assessed using simple metrics such as total treatment plan monitor units (MU) on the basis that increased modulation generates increased MLC blocking and hence increased MU to deliver the treatment dose. However, such metrics are described in the literature as 'crude' (Du *et al.*, 2014) as they do not account for natural variation of total MUs between patients and do not indicate the form of the complexity.

More sophisticated complexity metrics are generally divided into two categories: fluence map-based and aperture-based. These are considered in the following sections.

It is worth noting here that a number of authors try to correlate complexity metrics with dose measurements or patient specific quality assurance results (Du *et al.*, 2014; Younge *et al.*, 2012). Although some papers report weak correlations, Du *et al.* writes that a single metric cannot reveal the complete complexity of an IMRT or VMAT treatment plan. Moreover, agreement between the TPS and machine is dependent on the accuracy of the TPS beam model and dose calculation algorithm, which is not necessarily related to aperture or fluence map complexity. For example, a beam model that contains an inaccurate off-axis profile will systematically get off-axis calculations wrong irrespective of aperture complexity. The same beam model may perform on-axis calculations extremely accurately even for highly complex beam apertures.

Complexity metrics are therefore unreliable measures of whether the TPS is likely to have performed its calculation accurately. Correlations between complexity metrics and dose measurement results are therefore not considered in detail in this thesis.

1.4.1 Fluence Map-Based Metrics

Fluence map-based metrics quantify the variation in photon fluence between neighbouring pixels in a beam fluence map. Llacer *et al.* (2001) describe the fluence map complexity (FMC) index, which is calculated for a beam using equation 1.6, where a_j is the fluence of the j^{th} pixel in the beam fluence map, λ_k and a_k are the weighting and fluence of the k^{th} nearest neighbour respectively, and N_j is the set of nearest neighbours being considered in the FMC calculation.

$$\text{FMC} = \left(\sum_j a_j\right)^{-1} \times \sqrt{\sum_j \left(a_j - \lambda_k \sum_{k \in N_j} a_k\right)^2}. \tag{1.6}$$

For a perfectly uniform fluence map, FMC is zero and increases positively as the fluence map becomes increasingly non-uniform – thereby scoring complexity.

Webb (2003) criticises FMC because although it successfully characterises local variation in fluence, it does not relate this to overall beam variation. Overall beam fluence mean and standard deviation are possible ways to do this but Webb also argues that these are insufficient and proposes the modulation index as an alternative. Modulation index is computed by calculating differences between neighbouring pixels, Δp, in the beam fluence map and considering the number of differences, N, greater than some fraction, f, of the overall beam fluence standard deviation, σ. This is then normalised by the total number of pixels, n, as shown in equation 1.7.

$$Z(f) = \frac{1}{(n-1)} N(f;\, \Delta p > f\sigma). \tag{1.7}$$

$Z(f)$ then represents a distribution of differences within a fluence map, which is used to define modulation index using equation 1.8.

$$\text{Modulation Index} = \int_0^{0.5\sigma} Z(f)\, df. \tag{1.8}$$

While fluence map-based metrics can be used to quantify modulation and hence complexity in some settings, they have also been criticised because they only assess the beam as a whole and not the modulation of the individual beam segments (McNiven *et al*, 2010). For example, a beam comprised of many small segments that sum to give a relatively uniform irradiation would have similar fluence map-based complexity scores to a uniform irradiation with a single open field. This criticism is arguably more relevant for VMAT, where the overall beam fluence map of a full 360 ° arc cannot contain the subtleties of individual segment complexity.

1.4.2 Aperture-Based Metrics

Qualitatively, it is relatively straightforward to view a field aperture and comment on its complexity relative to a different segment. For example, Figure 1.13 shows three isocentric field apertures for a 10 cm × 10 cm open field, a typical segment from a step-and-shoot IMRT treatment plan and a VMAT control point and it is immediately apparent that aperture complexity increases from left to right. Reliably quantifying aperture complexity, however, is challenging but a number of metrics have been proposed in the literature.

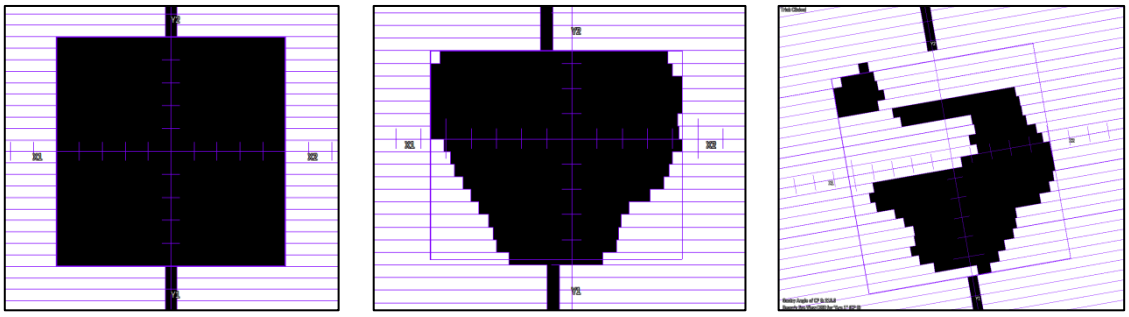

Figure 1.13: Qualitative characterisation of aperture complexity is relatively straightforward: in the examples above complexity increases from left to right.

Aperture-based metrics assess complexity by analysing beam apertures based on MLC and jaw positions and MU weightings. These overcome the limitations of fluence map-based metrics discussed in section 1.4.1 and are arguably more relevant when

considering the complexity of VMAT treatment plans.

McNiven *et al.* (2010) introduced the modulation complexity score (MCS) which is calculated for an individual beam using equation 1.9. It ranges from 1.0 for an unblocked open field with no complexity down to 0.0 for an 'infinitely complex' aperture.

$$\mathrm{MCS}_{beam} = \sum_{i=1}^{I} \mathrm{AAV}_i \times \mathrm{LSV}_i \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{beam}}, \tag{1.9}$$

where MU_i is the monitor units for the i^{th} segment and MU_{beam} is the total monitor units for the beam. LSV and AAV represent the leaf sequence variability and aperture area variability respectively and these are calculated using equations 1.10 and 1.11 based on the K MLCs that define the field from each leaf bank (Sumida *et al.*, 2017).

$$\begin{aligned}\mathrm{LSV}_{segment} = {} & \frac{1}{(K-1) \times \Delta\mathrm{X1}_{\max}} \times \left\langle \sum_{k=1}^{K-1} (\Delta\mathrm{X1}_{\max} - |\mathrm{X1}_k - \mathrm{X1}_{k+1}|) \right\rangle \\ & \times \frac{1}{(K-1) \times \Delta\mathrm{X2}_{\max}} \times \left\langle \sum_{k=1}^{K-1} (\Delta\mathrm{X2}_{\max} - |\mathrm{X2}_k - \mathrm{X2}_{k+1}|) \right\rangle,\end{aligned} \tag{1.10}$$

$$\mathrm{AAV}_{segment} = \frac{1}{K \times (\mathrm{X1}_{\max} + \mathrm{X2}_{\max})} \times \sum_{k=1}^{K} (\mathrm{X1}_k + \mathrm{X2}_k). \tag{1.11}$$

Equations 1.10 and 1.11 were originally presented in McNiven *et al.* but contained an error in the limits of the sum to K, which was corrected in Sumida *et al.*. MLC position nomenclature used here differs from Sumida *et al.* and McNiven *et al.* for consistency within this thesis. Figure 1.14 illustrates the nomenclature definitions.

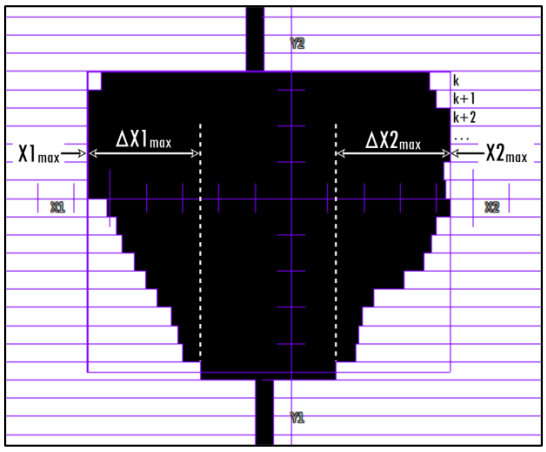

Figure 1.14: MLC position nomenclature definitions used throughout this thesis.

McGarry *et al.* (2011) present a comparison of various complexity metrics (both fluence- and aperture-based) and report that MCS outperforms MI and simple MU based metrics. However, the study was limited to six prostate IMRT patients.

Du *et al.* (2014) write that a single metric cannot reveal the complete complexity of an IMRT or VMAT treatment plan. Instead they present a number of aperture-based metrics that aim to characterise complexity in terms of aperture area (AA), perimeter (AP) and irregularity (AI). For the j^{th} segment of the i^{th} beam, these are calculated using equations 1.12, 1.13 and 1.14.

$$\mathrm{AA}_{ij} = \sum_{k=1}^{K} t_k \times \left(\mathrm{X1}_{ijk} - \mathrm{X2}_{ijk}\right), \tag{1.12}$$

$$\mathrm{AP}_{ij} = \sum_{k=1}^{K} |\mathrm{X1}_k - \mathrm{X1}_{k-1}| + |\mathrm{X2}_k - \mathrm{X2}_{k-1}| + 2t_k, \tag{1.13}$$

$$\mathrm{AI}_{ij} = \frac{\mathrm{AP}_{ij}^2}{4\pi \times \mathrm{AA}_{ij}}, \tag{1.14}$$

where k represents each leaf pair and t_k is the thickness of the k^{th} MLC.

From equations 1.12, 1.13 and 1.14, monitor unit weighted beam aperture area (BA) and irregularity (BI) metrics can be computed using equations 1.15 and 1.16. Beam

modulation (BM) can also be calculated using equation 1.17, where $U(\mathrm{AA}_{ij})$ is the union area of all of the segments of the i^{th} beam.

$$\mathrm{BA}_i = \frac{1}{\mathrm{MU}_i} \times \sum_j (\mathrm{MU}_{ij} \times \mathrm{AA}_{ij}). \tag{1.15}$$

$$\mathrm{BI}_i = \frac{1}{\mathrm{MU}_i} \times \sum_j (\mathrm{MU}_{ij} \times \mathrm{AI}_{ij}). \tag{1.16}$$

$$\mathrm{BM}_i = 1 - \left[\frac{1}{\mathrm{MU}_i \times U(\mathrm{AA}_{ij})} \times \sum_j (\mathrm{MU}_{ij} \times \mathrm{AA}_{ij}) \right]. \tag{1.17}$$

Finally, the complexity metrics at the beam level can be summed over all beams to give the plan complexity metrics plan aperture (PA), irregularity (PI) and modulation (PM). These are calculated using equation 1.18, 1.19 and 1.20.

$$\mathrm{PA} = \frac{1}{\mathrm{MU}_{\mathrm{total}}} \times \sum_i (\mathrm{MU}_i \times \mathrm{BA}_i). \tag{1.18}$$

$$\mathrm{PI} = \frac{1}{\mathrm{MU}_{\mathrm{total}}} \times \sum_i (\mathrm{MU}_i \times \mathrm{BI}_i). \tag{1.19}$$

$$\mathrm{PM} = \frac{1}{\mathrm{MU}_{\mathrm{total}}} \times \sum_i (\mathrm{MU}_i \times \mathrm{BM}_i). \tag{1.20}$$

1.5 Aims & Objectives of this Thesis

Bottlenecks in the treatment planning pathway illustrated in Figure 1.1, can prevent the widespread adoption of modern radiotherapy techniques, such as adaptive radiotherapy, and delay patient access to their cancer treatments. Automation of the pathway therefore has the potential to bring significant benefits to the clinic and the previous sections have reviewed attempts to automate various aspects of the pathway.

Such attempts that are published in the literature have a tendency to focus on

technicalities and lack attention to how they fit into clinical practice. This is the rationale for this thesis: the role of automation in radiotherapy treatment planning for prostate cancer.

Considering the pathway illustrated in Figure 1.1, immediately after the acquisition of the CT scan the image can be transferred to the TPS without delay so automation of this step offers little benefit. At the end of the pathway, 'Quality Assurance & Machine Preparation' tends to happen relatively quickly and given this step could be considered safety critical, automation may offer some benefit but with an associate risk that is unlikely to be accepted by many clinical departments.

This leaves the intermediate stages of 'Outline Target Volumes', 'Outline Organs at Risk' and 'Treatment Plan Optimisation'. These three stages are where each treatment plan spends the majority of time as it moves through the pathway. Most of that time is spent waiting for a clinical oncologist or treatment planner to add their contribution and this is the source of the major bottlenecks in the pathway. Attempts at automation therefore need to focus on these key stages and equally importantly focus on how these stages relate to each other. The following list gives the core aims and objectives of this thesis to address the gaps in the existing knowledge base:

- To establish the geometric accuracy of auto-contours in the male pelvis and their impacts on treatment plan generation.
- To develop a method of automatically generating optimal prostate radiotherapy treatment plans.
- To redesign the treatment planning pathway to integrate automation and maximise its benefit for the clinic and prostate cancer patients.

Chapter 2

2 Geometric Accuracy of Atlas-Based Auto-Segmentation

Joseph Wood[1, 2], Marianne Aznar[2], Philip Whitehurst[1, 2]

[1]The Christie NHS Foundation Trust
[2]The University of Manchester

2.1 Abstract

This chapter presents an evaluation of the geometric accuracy of male pelvic auto-contours generated with a range of atlas-based auto-segmentation packages. Geometric accuracy is quantified in terms of standard metrics, namely Dice similarity coefficient (DSC) and distance to agreement (DTA). All packages performed reasonably consistently although packages with customisable atlases achieved superior agreement with manual-contours. For all OARs (bladder, rectum and femoral heads) the relationships between DSC and DTA were found to be negatively correlated, which indicates both DSC and DTA are reasonably reliable scores of geometric accuracy. The limitations of using geometric accuracy to evaluate the performance of auto-

segmentation packages are also addressed in relation to the inference of clinical utility. From the results presented below it is not possible to determine whether the auto-contours could replace manual-contours directly. However, they do provide an important insight into the performance of atlas-based auto-segmentation in the male pelvis and a useful benchmark for future developments in the field of auto-segmentation.

2.2 Introduction

Manual-segmentation of medical images, where trained experts manually draw around anatomical structures on each slice of the image, is time consuming and has been demonstrated to suffer from inter- and intra-observed variation (Han *et al.*, 2008; Collier *et al.*, 2003). Auto-segmentation is the process of automatically identifying anatomical structures in a medical image and has been much discussed in the literature. Atlas-based auto-segmentation is a common approach and a number of commercially available software packages are available (Han *et al.*, 2008; Aljabar *et al.*, 2009).

Dice similarity coefficient (DSC) (Dice, 1945) and distance to agreement (DTA) (Sharp *et al.*, 2014) are the two commonly used metrics that assess the similarity of two different structures. DSC and DTA are therefore often quoted in research studies that evaluate the geometric accuracy of auto-segmentation algorithms (Zhu *et al.*, 2013; Teguh *et al.*, 2011; Speight *et al.*, 2014; Lim & Leech *et al.*, 2016; Voet *et al.*, 2011; Larrue, Kadir & Gooding, 2013; Gugyeras *et al.*, 2017). The literature is biased towards evaluations of atlas-based auto-segmentation algorithms in the head and neck region and evaluations of their performance in the male pelvis are rare. Zhu *et al.* (2013) is the only quantitative study in the literature and this is limited to an evaluation using SPICE (Qazi *et al.*, 2011) only. There certainly lacks a comparative study where the commercially available packages are benchmarked against each other using the same cohort of patients.

This was the purpose of the work presented in this chapter. From the literature reviewed in section 1.2, it was not possible to predict how geometrically accurate the

auto-contours would be or indeed whether any of the auto-segmentation packages would significantly outperform the others. The hypothesis for this work, therefore, was that all of the auto-contours would differ from corresponding manual-contours and the aim was to evaluate the magnitude of these differences for the packages assessed.

Langmack *et al.* (2014) present an inter-observer agreement study that gives DSCs for multiple manual-contours of male pelvic anatomical structures. Although it was not possible here to establish *a priori* a geometric accuracy threshold above which auto-contours could be considered reliable for radiotherapy treatment planning, the DSCs described by Langmack *et al.* potentially provided a useful benchmark against which auto-contour inaccuracies can be assessed.

The author performed all of the auto-segmentations described in this chapter and used the manual- and auto-contours to create overlap structures from which DSCs were calculated. They also wrote a program in Python to produce DTA histograms and related DTA statistics. All of the data analysis was performed by the author.

An abstract was submitted to the 2019 European Society for Radiotherapy and Oncology (ESTRO) conference based on the work presented in this chapter and Chapter 3. This was accepted for presentation as a poster and a reproduction of the poster is presented in section 3.8.2.

2.3 Method

Eleven prostate radiotherapy treatment planning (RTP) CT scans were randomly selected from a patient population and used to compare the performance of multiple auto-segmentation packages. On each of the 11 CT images an expert clinical oncologist had manually outlined the prostate and seminal vesicles (SV) and an experienced treatment planner had outlined the rectum, bladder and femoral heads according to local clinical protocols. Outlining was performed on RTP CT images only without MR fusion and all contours had been peer-reviewed and approved by a team of clinical oncologists. The model-based auto-segmentation tool that is included in the Pinnacle^3 treatment planning system (TPS) may have been used initially to outline the femoral

heads. These structures, however, generally require manual modification so for this work are considered to have been drawn manually.

Three atlas-based auto-segmentation packages were used for this work based on availability at the author's centre. It was acknowledged that this does not constitute an exhaustive list of available packages but is sufficient for a general evaluation of the atlas-based auto-segmentation approach in the male pelvis.

SPICE, the Philips auto-segmentation package (Qazi *et al.*, 2011), is integrated into Pinnacle[3] and atlases for auto-segmentation cannot be customised. The built-in 'Male Pelvis' atlas was used to auto-segment each of the 11 prostate CT scans. Mirada RTx (Mirada Medical, 2017) and ADMIRE (Elekta, 2013) have customisable atlases so a 'leave-one-out' approach was taken – where for each CT scan, the remaining 10 scans and structure sets were used as an atlas database. ADMIRE is a version of ABAS specific to the MR-Linac, which at the time of writing is not commercially available. However, it is still useful for this work, which just aimed to evaluate the geometric accuracy of the atlas-based auto-segmentation technique in the male pelvis using a range of different algorithms. The final segmentations from Mirada RTx were calculated using a majority vote based on each individual segmentation from the database. With ADMIRE, the STAPLE algorithm was used to combine the multiple atlas segmentations (Warfield, Zou & Wells, 2004). Although different methods to combine the segmentations were used, the specific approaches are the default options in each package and were therefore used to reflect manufacture recommendations.

All of the auto-segmented structure sets were imported into their respective patient treatment plans in Pinnacle[3]. DSCs were calculated from the volumes of each auto-contour, the corresponding manual-contour and the overlap region using equation 1.2. The calculated DSCs for each anatomical auto-contour was compared with a corresponding DSC reported in the inter-observer study of Langmack *et al.* (2014).

DTA histograms were generated from the Pinnacle[3] `plan.roi` file, where for each point in the manual-contour, the shortest distance to the auto-contour in 3D was found. This process was then repeated with the auto-contour as the reference to ensure all

large differences between the manual- and auto-contours were assessed.

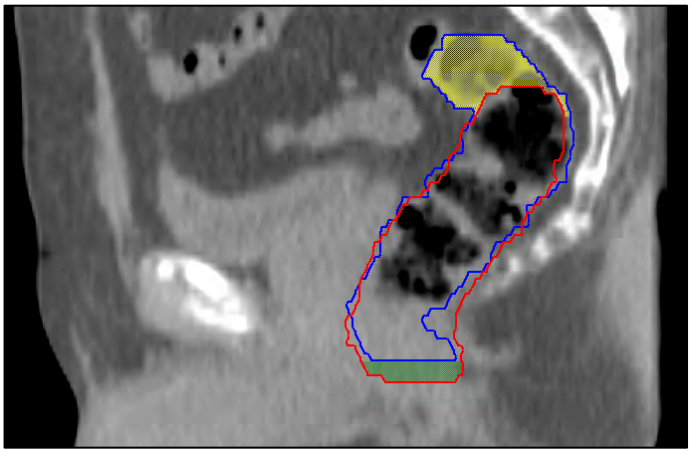

Figure 2.1: Rectum manual- and auto-contours shown as blue and red lines respectively. The shaded yellow and green regions show differences in their superior and inferior extents and the choice of reference contour determines the inclusion or exclusion from the DTA analysis.

Figure 2.1 illustrates the importance of performing these two DTA analyses. The manual- and auto-contours of the rectum are shown as the blue and red solid lines respectively. The shaded yellow area represents the superior region of the manual-contour that is missing from the auto-contour and this is included in the DTA analysis where the manual-contour is the reference. However, the shaded green region shows where the auto-contour extends further inferiorly than the manual-contour. This inferior auto-contour inaccuracy would not be reflected in the DTA analysis where the manual-contour is the reference. When the auto-contour is the reference the reverse is true: the shaded green region is included in the analysis but the shaded yellow region is not. It was therefore important to perform and report both analyses. For each histogram the mean DTA was calculated and the means were averaged to give the average mean DTA.

Following the calculation of DSC and average mean DTA, the reliability of the metrics as measures of geometric accuracy was assessed for each anatomical structure. For two structures that are identical, DSC is 1 and mean DTA is 0 mm (for both of the analyses described above). Any dissimilarities between the structures result in a reduced DSC and an increased mean DTA, where the size of the reduction or increase is dependent on the degree of dissimilarity. If DSC and mean DTA are reliable metrics of

geometric accuracy it is reasonable to expect a negative correlation between them across a range of patients and auto-segmentations. To evaluate this, for each auto-segmented structure, DSC was plotted against average mean DTA and coefficients of determination calculated.

2.4 Results

Figure 2.2 presents box and whisker plots for the distributions of DSC for the auto-segmented bladder, rectum, femoral heads (FHL and FHR), prostate and SV with SPICE, Mirada RTx and ADMIRE. The limits of the boxes represent the lower and upper quartiles of the data, the central horizontal line represents the median, the extent of the whiskers show the 5th and 95th percentiles of the data and outliers are represented by circular data points.

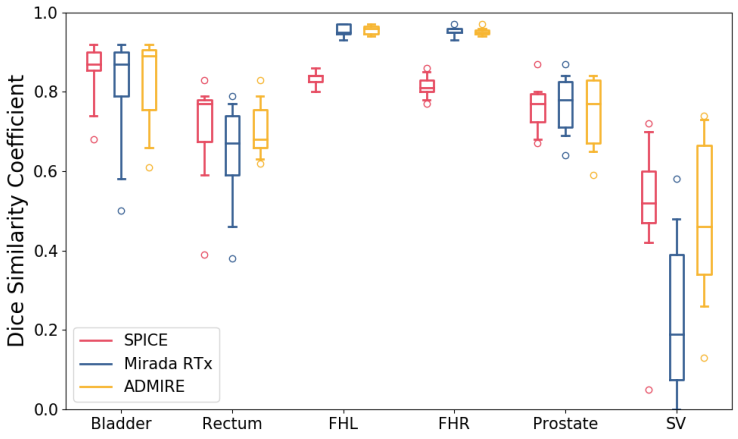

Figure 2.2: DSC box and whisker plots for auto-segmented male pelvic structures. The limits of the boxes represent the lower and upper quartiles of the data, the central horizontal line represents the median and the extent of the whiskers show the 5th and 95th percentiles.

Although box and whisker plots provide a convenient way to visualise variations of DSC between packages, patients and structures, reading detailed quantitative information from them is not possible. Therefore, the range, medians, means and standard deviations on which Figure 2.2 is based are presented in Table 2.1. For reference, the mean DSCs and standard deviations from the inter-observer variation study of Langmack *et al.* (2014) are also presented in Table 2.1 and to aid the

interpretation of the results the mean and standard deviation of the manual-contour volumes for each structure are also included.

Table 2.1: DSC data for auto-contoured structures in the male pelvis. For reference the DSCs from the inter-observer study of Langmack *et al.* (2014) are also provided and to aid interpretation the mean and standard deviation of the manual-contour volumes are included with the structure names.

Structure (Volume / cc)	Package	Range DSC	Median DSC	Mean ± SD DSC	Langmack *et al.* (2014) DSC
Bladder (91.8 ± 43.0)	SPICE	0.68-0.92	0.87	0.85 ± 0.08	0.88 ± 0.05
	Mirada RTx	0.50-0.92	0.87	0.81 ± 0.14	
	ADMIRE	0.61-0.92	0.89	0.83 ± 0.11	
Rectum (63.0 ± 26.0)	SPICE	0.39-0.83	0.77	0.71 ± 0.13	0.76 ± 0.11
	Mirada RTx	0.38-0.79	0.67	0.64 ± 0.13	
	ADMIRE	0.62-0.83	0.68	0.71 ± 0.07	
FHL (168.8 ± 14.8)	SPICE	0.80-0.86	0.84	0.83 ± 0.02	0.89 ± 0.03
	Mirada RTx	0.93-0.97	0.95	0.95 ± 0.02	
	ADMIRE	0.94-0.97	0.96	0.96 ± 0.01	
FHR 168.7± 15.3)	SPICE	0.77-0.86	0.81	0.81 ± 0.03	0.91 ± 0.02
	Mirada RTx	0.93-0.97	0.95	0.95 ± 0.01	
	ADMIRE	0.94-0.97	0.95	0.95 ± 0.01	
Prostate (61.4 ± 31.6)	SPICE	0.67-0.87	0.77	0.76 ± 0.06	0.70 ± 0.08
	Mirada RTx	0.64-0.87	0.78	0.77 ± 0.07	
	ADMIRE	0.59-0.84	0.77	0.75 ± 0.09	
SV (7.6 ± 5.0)	SPICE	0.05-0.72	0.52	0.51 ± 0.18	0.58 ± 0.10
	Mirada RTx	0.00-0.58	0.19	0.24 ± 0.20	
	ADMIRE	0.13-0.74	0.46	0.48 ± 0.20	

In Figure 2.2, it is clear that some anatomical structures generate larger DSCs than others and the variation of DSC between patients varies between structures and auto-segmentation packages. The DSCs for the SPICE contours are consistent with those presented in Zhu *et al.* (2013). Perhaps as expected, the femoral heads, which are the most well defined and regularly shaped anatomical structures considered here, show the smallest DSC variation between patients for all packages and the highest DSCs. Wilcoxon signed-rank tests showed that the differences in the mean DSCs for the Mirada RTx and ADMIRE femoral head auto-contours were not statistically significant using a threshold for significance of $p < 0.05$. However, the SPICE femoral head auto-contour mean DSCs were statistically significantly lower than both Mirada RTx and

ADMIRE. Conversely, again perhaps as expected, the smallest of the outlined structures, the SV, show the largest degree of variation between patients and the lowest DSCs for all packages. Here SPICE and ADMIRE SV mean DSCs were not statistically significantly different, but with Mirada RTx the SV DSCs were statistically significantly lower than with SPICE and ADMIRE.

For small structures, DSC can be unreliable as relatively small absolute volume discrepancies can represent a sizable proportion of the actual structure volume. For example, the volume of the SV manual-contour for one of the patients considered here was 4.2 cc and the corresponding SPICE auto-contour volume was 3.1 cc. This gave a DSC of 0.44 based on an overlap volume of 1.6 cc. In this case the auto-contour missed 2.6 cc of the ground-truth SV and included an erroneous 1.5 cc. The left femoral head for the same patient had a 174.4 cc manual-contour and a 140.5 cc auto-contour. The overlap volume was 135.6 cc, which gave a DSC of 0.86. For the left femoral head, the auto-contour missed 38.8 cc and included an erroneous 4.9 cc. In absolute volume terms, the geometric accuracy of the auto-contoured femoral head is significantly worse than for the SV, yet the femoral head DSC is almost a factor of 2 greater than the SV. For the same structures, the average mean DTAs were 3.1 mm and 1.8 mm for the SV and left femoral head respectively. Although average mean DTA also indicates that the femoral head auto-contour is more geometrically accurate than the SV, the difference is only 1.3 mm, which was approximately the DTA standard deviation for the SV auto-contour (see Table 2.2). This highlights the sensitivity of DSC to the overall structure volume, and hence its limitation when comparing auto-segmentation performance of different anatomical structures or structures with large volume variations.

Wilcoxon signed-rank tests showed that differences in DSC for the rectum, bladder and prostate were not statistically significant between SPICE, Mirada RTx or ADMIRE using a threshold for significance of $p < 0.05$.

It is interesting and important to note here that since the SPICE atlases are not customisable they do not allow for local definitions of anatomical boundaries, which impacts on the geometric accuracy of the auto-segmented structures. With the femoral

heads, for example, Figure 2.2 shows there is a clear offset in the DSCs for SPICE compared to Mirada RTx and ADMIRE – although the variation is similar and small with all three packages. Figure 2.3 shows, for one of the 11 patients, the femoral head manual-contours shaded in orange and the SPICE, Mirada RTx and ADMIRE femoral head auto-contours as red, blue and yellow solid lines respectively. The inferior extent of the manual-contours has been defined following a local protocol to the level of the lesser trochanter. Since the Mirada RTx and ADMIRE atlas databases have been generated from locally drawn manual-contours, they give more locally representative auto-segmentations than the proprietary non-customisable 'Male Pelvis' atlas in SPICE. This is the cause of the systematic offset seen for the SPICE femoral head DSCs and suggests that non-customisable atlases fundamentally limit the geometric accuracy of auto-segmented structures when benchmarked against local contouring practices. This effect may also be present with the other pelvic structures, but the larger variations in DSCs between patients makes this difficult to identify.

Figure 2.3: Femoral head manual-contours (shaded orange) and SPICE (red line), Mirada RTx (blue line) and ADMIRE (yellow line) auto-contours.

From Table 2.1, the auto-contour mean DSCs generally lie within one standard deviation of the inter-observer DCSs reported by Langmack *et al.* (2014). This could imply that, although auto-contours are not consistent with an individual's manual-contours, they are consistent with manual-contours in general. However, care needs to be taken with this inference, because inter-observer variation and auto-segmentation inaccuracy can arise for different reasons. Inter-observer variation can result from differences of expert clinical opinion and the variation may be discrete where an anatomical boundary is ambiguous. As discussed in section 1.2.1, atlas-based auto-

segmentation inaccuracies can result from (amongst other things) the quality of the atlas segmentations and how these are combined to segment the anatomy in a novel image (Aljabar *et al.*, 2009). As a result, differences between auto- and manual-contours can be continuous and so where two observers may disagree on the exact location of an anatomical boundary, they may well completely agree that it is not anywhere between either of their opinions. On this basis, for the specification of clinical utility, it is arguably insufficient to demand that auto-contours lie within the inter-observer variation. This makes determining the acceptability of auto-contours extremely difficult and suggests that geometric accuracy may not be the most appropriate metric for this.

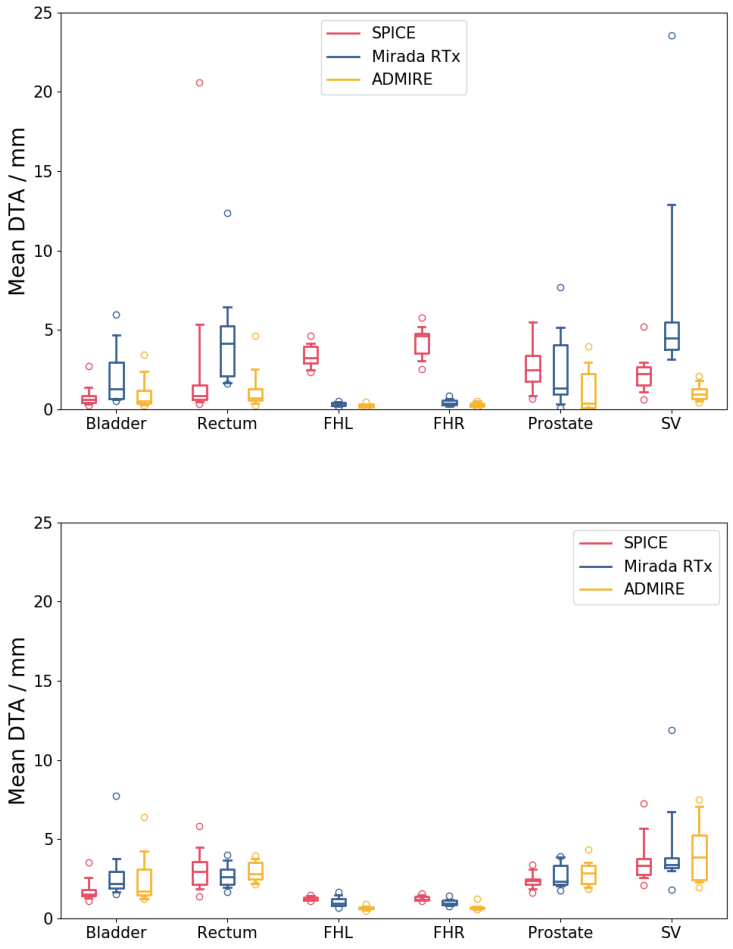

Figure 2.4: Mean DTA box and whisker plots for auto-segmented male pelvic structures with the manual-contour as the reference (top) and the auto-contour as the reference (bottom). The limits of the boxes represent the lower and upper quartiles of the data, the central horizontal line represents the median and the extent of the whiskers show the 5th and 95th percentiles

Figure 2.4 presents box and whisker plots for the distributions of mean DTA for the auto-segmented structures with SPICE, Mirada RTx and ADMIRE. The limits of the boxes represent the lower and upper quartiles of the data, the central horizontal line represents the median, the extent of the whiskers show the 5^{th} and 95^{th} percentiles of the data and outliers are represented by circular data points. The top and bottom plots use the manual- and auto-contour respectively as the reference.

Table 2.2 gives a full breakdown of the mean DTA for each anatomical structure, reference contour and auto-segmentation package. Average mean DTA is also given, which is simply the average mean DTA from the two analyses. Wilcoxon signed-rank tests were performed on the data to determine the statistical significance of differences between auto-segmentation packages. Average mean DTA values presented in bold were statistically significantly higher than values followed by * using a threshold for significance of $p < 0.05$.

Table 2.2: Mean DTA data for auto-contoured structures in the male pelvis.

		Manual Reference		**Auto Reference**		
Structure	**Package**	**Mean DTA / mm**	**σ / mm**	**Mean DTA / mm**	**σ / mm**	**Average Mean DTA / mm**
Bladder	SPICE	0.84	0.70	1.77	0.71	1.30*
	Mirada RTx	2.07	1.96	2.85	1.76	**2.46**
	ADMIRE	1.03	1.02	2.52	1.61	1.78*
Rectum	SPICE	3.05	5.98	3.09	1.28	3.07
	Mirada RTx	4.38	3.13	2.67	0.75	**3.53**
	ADMIRE	1.24	1.30	2.96	0.65	2.10*
FHL	SPICE	3.37	0.72	1.25	0.13	**2.31**
	Mirada RTx	0.34	0.13	1.06	0.32	0.70*
	ADMIRE	0.24	0.11	0.66	0.12	0.45*
FHR	SPICE	4.19	0.99	1.28	0.16	**2.73**
	Mirada RTx	0.42	0.22	1.01	0.20	0.72*
	ADMIRE	0.31	0.14	0.72	0.18	0.51*
Prostate	SPICE	2.76	1.61	2.39	0.51	**2.58**
	Mirada RTx	2.57	2.39	2.72	0.78	2.64
	ADMIRE	1.18	1.41	2.85	0.80	2.02*
SV	SPICE	2.27	1.21	3.70	1.51	2.99*
	Mirada RTx	6.80	6.19	4.34	2.77	**5.57**
	ADMIRE	1.07	0.53	4.12	1.97	2.60*

From Figure 2.4 and Table 2.2, it is clear that the choice of reference contour for the DTA analysis has a considerable impact on the mean DTA. The SPICE rectum and femoral head and Mirada RTx SV auto-contours are particularly extreme examples. Figure 2.1 and Figure 2.3 illustrate why the choice of reference contour can have such a pronounced impact on mean DTA. The average mean DTA for the SPICE contours differ from those reported in Zhu *et al.* (2013) and are generally higher. In Zhu *et al.* the DTA analyses were performed slice-wise whereas those above were done in 3D. It would generally be expected that 3D analysis would generate lower mean DTAs but because of the difference in analysis technique it is not possible to compare the results directly. This highlights the sensitivity of DTA results to the specifics of the analysis and the need to specify exactly how it has been performed.

As noted above, for the rectum, bladder and prostate, the DSC results for all three packages are similar and not statistically significantly different: the bladder tends to give the highest DSC but with considerable variance, the prostate tends to give a lower DSC but a smaller variance and the rectum gives the lowest DSC and largest variance. The performance of the packages in auto-segmenting these structures is also fairly well reflected in the mean DTA results shown in Figure 2.4. A possible reason for this is the variation in the size and shape (and contents for the bladder and rectum) of the structures, which can vary widely between patients. Figure 2.5 shows sagittal CT slices for two of the 11 patients with the bladder, prostate, SV and rectum manual-contours as yellow, red, blue and brown solid lines respectively. The rectum contents in the two images are markedly different: the right CT image contains significantly more gas than the left. Differences in rectum contents between the atlas image and the novel image could limit the performance of the DIR and thus the auto-segmentation accuracy.

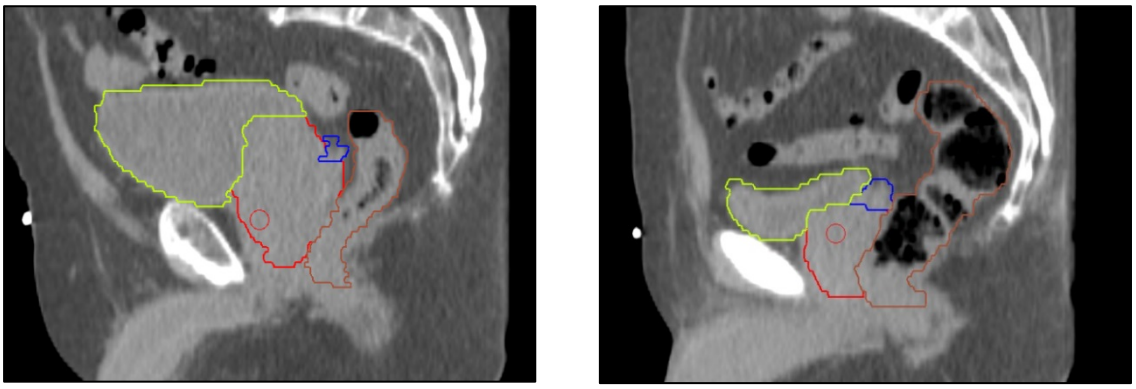

Figure 2.5: Sagittal CT slices showing considerable variation in rectal contents.

To evaluate this, the mean CT number of voxels inside the rectum manual-contour – which should be ~1000 with no gas in the rectum and < 1000 where gas is present – was plotted against the rectum DSC and average mean DTA for all packages. These plots are presented in Figure 2.6 for the Mirada RTx auto-contours and show weakly correlated relationships (coefficients of determination were 0.47 and 0.61 for DSC and average mean DTA respectively). Although more patient data is required to confirm the definitive existence of these relationships, Figure 2.6 suggests that for the rectum, where auto-segmentation has proved least successful in the pelvis, the average CT number inside the contours themselves could potentially be used as a surrogate to quality assure geometric accuracy.

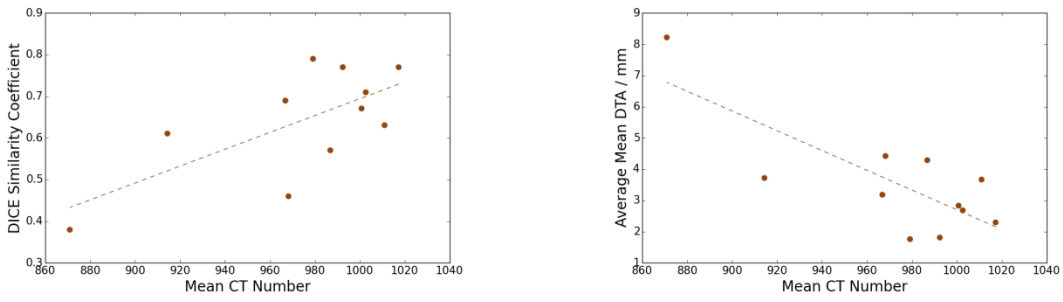

Figure 2.6: Relationships between DSC and average mean DTA and average rectum CT number for auto-contours generated with Mirada RTx.

Figure 2.7 shows plots of average mean DTA against DSC for each of the auto-segmented pelvic structures. Using the average of the mean DTAs ensured all

discrepancies between the manual- and auto-contours were considered and this also gave stronger correlations than using either of the individual mean DTAs. The plots include linear trend lines with coefficients of determination. Although the relationships between DSC and average mean DTA are not strictly linear, linear fits were used for simplicity and the relatively high coefficients of determination (for the OARs at least) suggest that the linear fit approximation is reasonable.

The strongest correlations are observed for the OARs, where all coefficients of determination are greater than 0.7. Average mean DTA and DSC show weaker correlations for the prostate and SV where coefficients of determination are 0.32 and 0.47 respectively.

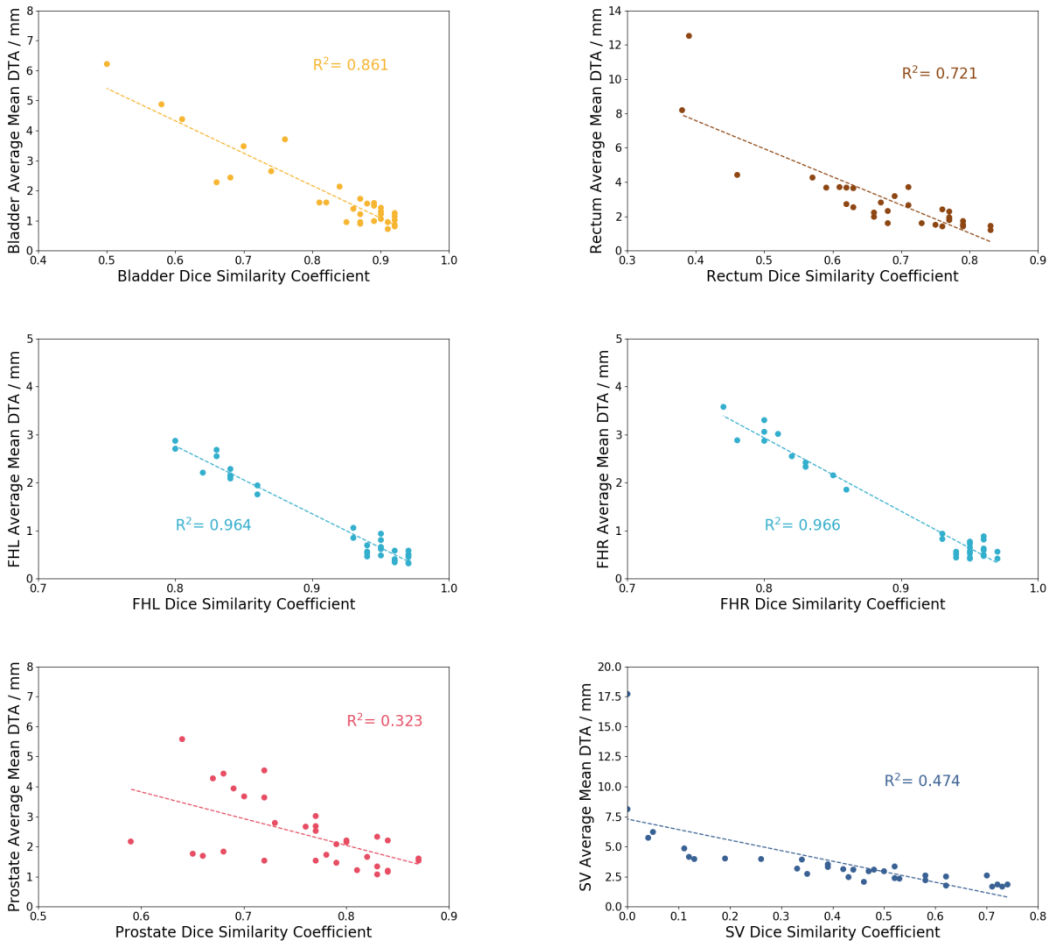

Figure 2.7: Average mean DTA plotted against DSC for auto-segmented male pelvic structures.

Limitations of DSC for comparisons of auto-segmentation performance for different anatomical structures have been discussed above – in particular where the comparison involves structures with significantly different volumes. It has also been discussed that care needs to be taken with DTA analysis to ensure all geometric inaccuracies are considered in the final reported mean DTA value. Figure 2.7 shows reasonably strong negative correlations between DSC and average mean DTA for all of the OARs in the male pelvis. This suggests that for a given anatomical structure either metric can provide a reasonably reliable measure of auto-contour geometric accuracy. If the geometric accuracy of different anatomical structures is to be compared, though, both metrics need to be specified for a reliable comparison to be made.

The prostate and SV plots in Figure 2.7 show weaker correlations than the OARs. It was discussed above that for small structures such as the SV, DSC is particularly sensitive to small inaccuracies and this is the most likely reason for the wide range of SV DSCs plotted in Figure 2.7, which potentially reduces the correlation between DSC and average mean DTA. Despite the prostate and SV having weaker correlations between DSC and average mean DTA than the OARs, the data still show a trend towards larger average mean DTA as DSC reduces. Although this implies that both metrics give an indication of geometric accuracy, specification of both metrics is necessary to interpret the geometric accuracy of the prostate and SV auto-contours reliably.

2.5 Discussion

This work has evaluated the geometric accuracy of male pelvic auto-contours generated using a range of atlas-based auto-segmentation packages in terms of the standard metrics DSC and average mean DTA. It has been shown that the femoral head auto-contours are the most geometrically accurate with the rectum, bladder, prostate and seminal vesicles demonstrating smaller DSCs and larger average mean DTAs. This is perhaps unsurprising given the regularity of femoral head shape and size and the high contrast with surround soft tissues. It was also shown that auto-segmentation packages with customisable atlases can achieve superior geometric agreement with manual-

contours compared to packages with non-customisable proprietary atlases.

Such a comparative study has not been presented in the literature previously and the results presented in this chapter add to the existing knowledge base whilst demonstrating limitations of existing literature. The literature is biased towards evaluations of auto-segmentation performance in terms of geometric accuracy – indeed, this was partly the motivation for the work presented in this chapter. Although it is an intuitive concept when comparing the similarity of two different shapes, the specifics of exactly how geometric accuracy analyses have been performed are often not fully specified and where they differ between studies it can be difficult to compare studies in a meaningful manner. This was demonstrated most clearly here with the specification of which contour has been used as the reference for the DTA analysis.

Despite the limitations of DSC and DTA, it was demonstrated that for any given anatomical structure in the male pelvis, the two metrics do correlate with each other, which indicates that for a given structure either provides a reasonable score of geometric accuracy.

A limitation of this work is that the evaluation was only based on 11 patients. Whilst an effort was made to ensure that this cohort contained a range of patient anatomies that represent the clinical patient population, it is not possible for such a relatively small group to cover every possible clinical situation. Although it would be reasonable and interesting to conduct a larger scale study and possibly to break down the patient cohort by anatomical characteristics, the value of such studies is itself limited for the practical application of auto-segmentation in the radiotherapy treatment planning pathway.

Although it has been demonstrated that DSC and DTA give reliable predictors of geometric accuracy, geometric accuracy does not necessarily provide a reliable indicator of the utility of auto-contours for treatment planning. Some absurd situations can arise where auto-contours with better geometric accuracy scores can be less useful for treatment planning that auto-contours with worse scores. Figure 2.1 can be used to illustrate this: an auto-contour that improves the large geometric inaccuracies at the

superior and inferior extents of the rectum at the expense of a relatively small compromise of geometric accuracy at the cranio-caudal level of the prostate would likely generate better geometric accuracy scores; in practice, however, this would not represent an improved auto-contour for treatment plan optimisation.

In the future, auto-segmentation will be a first step in automated treatment planning workflows and so the utility of auto-contours for treatment planning purposes is of more relevance to clinical practice than pure geometric accuracy. This should be an area for future work to address. If auto-contours are used to generate PTVs and OAR sparing volumes with which a treatment plan can be optimised, small geometric inaccuracies of the order of a few millimetres that have been seen in this study may be of little consequence.

2.6 Conclusion

Eleven CT scans of the male pelvis have been used to evaluate the geometric accuracy of atlas-based auto-segmentation using SPICE, Mirada RTx and ADMIRE. DSC and mean DTA were identified as standard metrics of geometric accuracy in the literature and have been calculated here for each set of auto-contours. The femoral heads and SV consistently showed the highest and lowest DSCs respectively for all 11 patients and prostate, bladder and rectum were generally consistent in terms of DSC across all three packages. Mean DTA was strongly influenced by the choice of reference contour and performing two separate DTA analyses with the manual- and auto-contours as the reference was necessary to ensure all geometric inaccuracies were included in the analysis. Taking the average of the two analyses was found to give a more robust measure of overall geometric accuracy. It is therefore recommended when considering mean DTA results reported in the literature that caution is taken where the exact method of calculation is not thoroughly specified and where multiple analyses have not been performed.

Negative correlations were found between DSC and average mean DTA for all auto-segmented structures although prostate and SV were only weakly correlated. This

suggests that, for a given anatomical structure, DSC and average mean DTA are both reliable measures of geometric accuracy. However, since DSC is sensitive to overall structure volume, to compare the geometric accuracy of different structures both DSC and average mean DTA need to be specified.

Although it has been shown that DSC and average mean DTA can be used to measure the geometric accuracy of auto-segmented structures reasonably reliably, neither provide any indication of the significance of the inaccuracies for radiotherapy treatment planning. The following chapter addresses this issue by using the auto-contours from this chapter to generate radiotherapy treatment plans.

2.7 References

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J. & Rueckert, D. (2009). 'Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy', *NeuroImage*, 46, pp. 726-738.

Collier, D., Burnett, S., Amin, M., Bilton, S., Brooks, C., Ryan, A., Roniger, D., Tran, D. & Starkschall, G. (2003). 'Assessment of consistency in contouring of normal-tissue anatomic structures', *Journal of Applied Clinical Medical Physics*, 4(1), pp. 17-24.

Dice, L. (1945). 'Measures of the amount of ecologic association between species', *Ecology*, 26(3), pp. 297-302.

Elekta (2013). *ABAS Atlas-Based Autosegmentation*. Available at: https://www.elekta.com/dam/jcr:7b262e2d-74c7-412c-bd1d-eec08da8183d/ABAS.pdf, (Accessed: 6 November 2017).

Gugyeras, D., Farkas, A., Gulyban, A., Csima, M., Hadjiev, J. & Lakosi, F. (2017). 'Multi-Atlas based segmentation (Mirada RTx) vs. manual contouring for OARs in the head and neck region: efficiency time analysis', *South East Europe Technology in Radiation Oncology 2017*. Available at: https://www.researchgate.net/profile/Ferenc_Lakosi/project/Time-sparing-effect-of-multi-ABS-Single-vs-Multi-ABS-Atlas-Based-Segmentation/attachment/5a03823eb53d2fed8ad4336f/AS:558547001327616@1510179390631/download/Seetro2017.pdf?context=projectUpdatesLog, (Accessed: 19 February 2018).

Han, X., Hoogeman, M., Levendag, P., Hibbard, L., Teguh, D., Voet, P., Cowen, A., & Wolf, T. (2008). 'Atlas-based auto-segmentation of head and neck CT images', *Medical Image Computing and Computer-Assisted Intervention*, 11(2), pp. 434-441.

Langmack, K., Perry, C., Sinstead, C., Mills, J. & Saunders, D. (2014). 'The utility of atlas-assisted segmentation in the male pelvis is dependent on the interobserver agreement of the structures segmented', *British Journal of Radiology*, 87: 20140299, pp. 1-6.

Larrue, A., Kadir, T. & Gooding, M. (2013). 'SU-E-J-93: Assessment of Contour Fusion Method for Auto-Contouring of Head and Neck CT Images', *Medical Physics*, 40(6), pp. 171.

Lim, J. & Leech, M. (2016). 'Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck', *ACTA Oncologica*, 55, 7, pp. 799-806.

Mirada Medical (2017). *RT Imaging. Agile Workflows.* Available at: http://mirada-medical.com/wp-content/uploads/2017/11/mm6908-1-ROW-RT-Imaging-Brochuretoemail.pdf, (Accessed: 11 December 2017).

Qazi, A., Pekar, V., Kim, J., Xie, J., Breen, S. & Jaffray, D. (2011). 'Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach', *Medical Physics*, 38(11), pp. 6160-6170.

Sharp, G., Fritscher, K., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H. & Yang, J. (2014). 'Vision 20/20: Perspectives on automated image segmentation for radiotherapy', *Medical Physics*, 41(5), pp. (050902)1-13.

Speight, R., Karakaya, E., Prestwich, R., Sen, M., Lindsay, R., Harding, R. & Sykes, J. (2014). 'Evaluation of atlas based auto-segmentation for head and neck target volume delineation in adaptive/replan IMRT', *Journal of Physics*, 489, pp. 1-4.

Teguh, D., Levendag, P., Voet, P., Al-Mamgani, A., Han, X., Wolf, T., Hibbard, L.,Nowak, P., Akhiat, H., Dirkx, M., Hiejmen, B. & Hoogeman, M. (2011). 'Clinical Validation of Atlas-Based Automatic Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck', *International Journal of Radiation Oncology Biology Physics*, 81(4), pp. 950-957.

Voet, P., Dirkx, M., Teguh, D., Hoogeman, M., Levendag, P. & Heijmen, B. (2011). 'Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis', *Radiotherapy & Oncology*, 98, pp. 373-377.

Warfield, S., Zou, K. & Wells, W. (2004). 'Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation', *IEEE Transactions on Medical Imaging*, 23(7), pp. 903-921.

Zhu, M., Bzdusek, K. Brink, C., Eriksen, J., Hansen, O., Jensen, H., Gay, H., Thorstad, W., Widder, J., Brouwer, C., Steenbakkers, R., Vanhauten, H., Cao, J., McBrayne, G., Patel, S., Cannon, D., Hardcastle, N., Tomé, W., Guckenberg, M. & Parikh, P. (2013). 'Multi-institutional Quantitative Evaluation and Clinical Validation of Smart Probabilistic Image Contouring Engine (SPICE) Autosegmentation of Target Structures and Normal Tissues on Computer Tomography Images in the Head and Neck, Thorax, Liver, and Male Pelvis Areas', *International Journal of Radiation Oncology Biology Physics*, 87(4), pp. 809-816.

Chapter 3

3 Utility of Auto-Segmentation for Prostate VMAT Treatment Planning

Joseph Wood[1, 2], Marianne Aznar[2], Philip Whitehurst[1, 2]

[1]The Christie NHS Foundation Trust
[2]The University of Manchester

3.1 Abstract

This chapter presents an investigation into the utility of auto-contours for prostate radiotherapy treatment planning. Auto-contours generated with a range of atlas-based auto-segmentation packages, that contained known geometric inaccuracies, were used to generate treatment plans. For 11 patients, 3 treatment plans were generated: a manual-plan based on target and OAR manual-contours; an auto-plan based on target and OAR auto-contours; and a hybrid- based on target manual-contours and OAR auto-contours. When reported to 'ground-truth' manual-contours, auto-plans were shown to differ considerably from corresponding manual-plans with significant under-dosage of PTVs. Hybrid-plan conversely showed good agreement with corresponding manual-

plans with comparable high dose coverage of target volumes and OAR sparing. This suggests that OAR auto-contours can be used reliably to drive treatment plan optimisation and offers an opportunity to restructure the treatment planning pathway to generate a considerable efficiency saving.

Dosimetric differences between the manual-, auto- and hybrid-plans were not found to correlate with differences in the geometric accuracy of the auto-contours used to generate them. This demonstrates a clear limitation of using geometric accuracy as an evaluation of auto-segmentation performance.

3.2 Introduction

In Chapter 2, the geometric accuracy of male pelvic auto-contours was evaluated using three atlas-based packages: SPICE, (Qazi *et al.*, 2011), Mirada RTx (Mirada Medical, 2017) and ADMIRE (Elekta, 2013). Whilst assessment of geometric accuracy is a common approach to evaluating the performance of auto-segmentation packages (Zhu *et al.*, 2013; Teguh *et al.*, 2011; Speight *et al.*, 2014; Lim & Leech *et al.*, 2016; Voet *et al.*, 2011; Larrue, Kadir & Gooding, 2013; Gugyeras *et al.*, 2017), it is also necessary to consider the utility of the auto-contours for the intended clinical purpose. Auto-contours that contain geometric inaccuracies may still fulfil their clinical requirements and in such cases geometric accuracy serves as a poor performance metric for auto-segmentation. This is often overlooked in the literature where geometric inaccuracies are widely used to infer poor clinical utility. Beasley *et al.* (2016) report that in the head and neck region relationships between auto-contour geometric accuracy and treatment planning utility are weak at best. Such a study of auto-contour utility for prostate radiotherapy treatment planning has not been published in the literature.

At The Christie, standard-of-care radiotherapy for intermediate- to high-risk prostate cancer (T1b-T3aN0M0) is based on the hypofractionated arm of the CHHiP trial (Dearnaley *et al.*, 2016; Wilkins *et al.*, 2015). From initial manual outlines of the prostate, SV and rectum three PTVs are generated according to the expansions in Table 3.1.

Table 3.1: Margin recipe for prostate radiotherapy PTVs.

PTV	Source	Avoid Interior	Avoid Exterior	Expansion
PTV1	Prostate + SV	–	–	1 cm SIRLAP
PTV3	Prostate	Rectum	–	0.5 cm SIRLA
PTV2	PTV3	–	PTV1	1 cm SIRLAP

In accordance with the recommendations of The Royal College of Radiologists (The Royal College of Radiologists, 2016), 6000 cGy in 20 fractions is prescribed as the mean prostate dose, and the treatment plan is optimised in Pinnacle^3 such that PTV3 receives at least 5700 cGy, PTV2 at least 5460 cGy and PTV1 at least 5000 cGy. Patients are scanned supine and head-first into the scanner with no patient specific immobilisation devices other than left, right and anterior Beekley CT-SPOT markers (Beekley Medical, 2021). Treatment plan optimisation is performed using a single 10 MV volumetric modulated arc therapy (VMAT) beam over a full 360° gantry rotation with a 10° collimator twist to avoid interleaf leakage. Although 6 MV is the energy of choice for most pelvic VMAT (Kumar *et al.*, 2015; Shang *et al.*, 2014), 10 MV for prostate VMAT is widely accepted in the literature as a superior energy due to improvements in target coverage, OAR sparing and integral dose (Mattes *et al.*, 2014; Stanley *et al.*, 2015).

Table 3.2: CHHiP and local OAR treatment planning dose limits.

OAR	Dose / cGy	CHHiP Maximum Volume / %	Local Maximum Volume / %
Rectum	2460	80	88
	3240	70	76
	4080	60	60
	4860	50	43
	5280	30	28
	5700	15	7
	6000	3	0
Bladder	4080	50	93 (or 60 if Bladder > 150 cc)
	4860	25	84 (or 46 if Bladder > 150 cc)
	6000	5	5 (or 1 if Bladder > 150 cc)

OAR doses are limited in the treatment plan optimisation based partly on the constraints given by the CHHiP trial but edited to reflect a local bladder emptying protocol that differs from the CHHiP requirements. Local practice at The Christie is for patients to void their bladder immediately prior to the treatment planning CT scan acquisition and every treatment fraction. The CHHiP and local OAR dose limits are given in Table 3.2.

In the prostate radiotherapy treatment planning pathway illustrated in Figure 1.1, outlining of anatomical structures is performed in two stages: outlining of target volumes by clinical oncologists and outlining of OARs by treatment planners. These outlines are first used to drive the treatment plan optimisation and second to report the dosimetry of the final plan. For the reporting of the final plan dosimetry, it is arguably important to have geometrically accurate structures for accurate dose reporting, especially near steep dose gradients[‡]. However, for the optimisation of the treatment plan where anatomical structures (particularly OARs) are used to guide where dose is deposited more generally, small geometric inaccuracies in the structures may have little impact on the final dose distribution.

Treatment plan optimisation cannot proceed without both target and OAR outlines and Figure 1.1 illustrates how this constrains the overall pathway. The outlining constraints cause significant inefficiencies for treatment planning departments and ultimately delay the access of prostate cancer patients to their radiotherapy treatments. The work presented in this chapter aims to determine the utility of auto-contours that contain known geometric inaccuracies for radiotherapy treatment planning. To do this, the author used the auto-contours for the 11 patients described in Chapter 2 to generate auto-plans (optimised with auto-contoured OARs and targets) and hybrid-plans (optimised with auto-contoured OARs and manually-contoured targets) using the same optimisation parameters. They analysed the dosimetry of the treatment plans (always

[‡] Although this is a commonly accepted and intuitive viewpoint, reporting doses to geometrically accurate structures is only accurate if the treatment planning geometry is perfectly replicated on every fraction throughout the course of treatment. This is known not be the case and Orlandini *et al.* (2017) suggest daily image guidance and dose tracking is necessary to report accurate dosimetry.

reported to the manual-contours) and used gamma analysis to compare the global dose distributions. Finally, they investigated relationships between the treatment plans and auto-contour geometric accuracy to establish the relative importance of geometric accuracy in the evaluation of auto-contours.

If clinically acceptable and dosimetrically optimal treatment plans can be generated based on auto-contours (albeit with known geometric inaccuracies) the outlining constraints on the pathway could be released and opportunities to restructure the Figure 1.1 pathway and harness this efficiency could be exploited.

3.3 Method

The clinical prostate VMAT optimisation class-solution used at The Christie was developed in 2009 when VMAT was a relatively new technique to the radiotherapy community (Boylan, Golby & Rowbottom, 2010). Despite having successfully treated more than 3000 prostate cancer patients with this solution since its inauguration, its fine details bear the hallmark of a prototype. To avoid biasing the results of this research with a class-solution that is known to contain conflicting optimisation objectives, a new, simplified solution was developed. The new solution had not been through a rigorous commissioning process that would be typical of a clinical optimisation class-solution and it was not necessarily the intention that it would generate dosimetrically optimal (or even clinically acceptable) treatment plans. Rather, the intention was to generate approximations of clinical treatment plans to evaluate the impacts of using auto-contours for optimisation. Full details of the new optimisation class-solution used for this work are presented in the section 3.8.1.

For each of the 11 patient datasets described in Chapter 2, four sets of PTVs were created based on the expansions in Table 3.1 from the manual- and auto-contours (SPICE, Mirada RTx and ADMIRE). Four sets of OAR sparing volumes were also generated. These consisted of the OAR outside PTV1 with a 0.7 cm SIRLAP expansion. An example of this for the manual-contours is shown in Figure 3.1 where the prostate, SV, rectum, bladder and left and right femoral heads are shown as the red, blue, brown,

yellow, green and pink solid lines respectively, PTV1, PTV2 and PTV3 are the shaded skyblue, maroon and turquoise areas respectively and the OAR sparing volumes are the shaded regions in the same colour as their respective OAR.

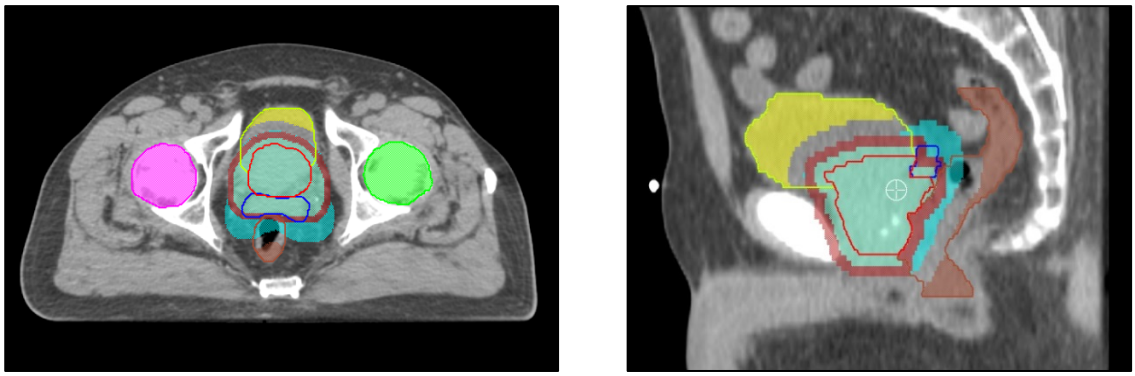

Figure 3.1: Prostate treatment planning PTVs, OARs and sparing volumes.

In-plane OAR structures were also created using the OAR as the source and limiting it to the superior and inferior extents of PTV1. Within the Pinnacle[3] optimiser, maximum DVH and maximum equivalent uniform dose (EUD) objectives can be specified. The extent to which these objectives limit high dose to a partially irradiated structure is dependent on the overall structure volume, which can vary considerably and arbitrarily between patients. In-plane optimisation structures allow dose gradients through OARs to be controlled in a more standardised way.

The simplified optimisation class-solution was created using the volumes described above and this was run using the manual-contours and each set of auto-contours for each of the 11 patients. Since the target volumes perform the precise task of defining the extent of the high dose region but the sparing volumes perform the more general function of steering dose away from the sensitive OARs, it was hypothesised that geometric inaccuracies in the prostate and SV would have a greater impact on the overall dose distribution compared to similar geometric inaccuracies in the OARs. Hybrid-plans were therefore generated using the prostate and SV manual-contours to expand the PTVs (the rectum auto-contour was used to expand PTV3) but the auto-contours for the OAR sparing volumes. Hybrid-plans were generated

using the same optimisation objectives and parameters as the manual- and auto-plans (i.e. those presented in section 3.8.1).

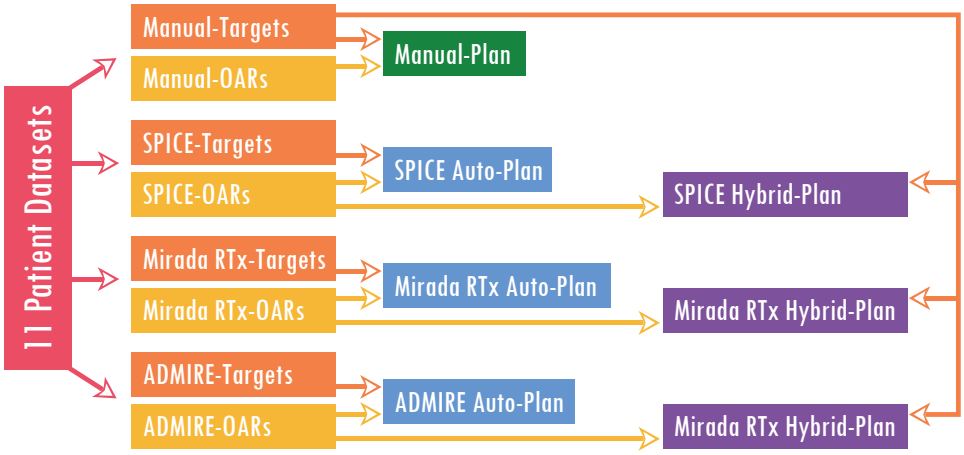

Figure 3.2: Flowchart of the manual-, auto- and hybrid-plan generation.

Figure 3.2 shows a flowchart of the manual-, auto- and hybrid-plan generation. In total, seven treatment plans were generated for each of the 11 patients: one manual- (green), three auto- (blue) and three hybrid-plans (purple). DVH data for the manual-contours were extracted from Pinnacle[3] and analysed for each of the seven treatment plans to investigate the dosimetric impact of performing the optimisation with auto-contours on the ground-truth anatomy. Since the optimisation was always run with exactly the same objectives and parameters, any differences in the dosimetry reported to the manual-contours could be isolated to having been caused by the geometric differences between the manual- and auto-contours – thus allowing the impacts of geometric inaccuracies on auto-contour utility for treatment planning to be investigated.

Although DVH analysis is used clinically in the assessment of treatment plan optimisation, it does not provide information on where within an anatomical structure the high dose is located. Clinically, this would be assessed visually but to compare the impacts of auto-contour geometric inaccuracies on the global dose distribution in this manner is difficult and subjective. To provide a quantitative surrogate for the visual inspection and comparison of the 3D dose distributions, 3D gamma analysis (Low *et al.*,

1998) was performed using VeriSoft v5.0 to compare the dose distributions from the manual-plan with the auto-plans and hybrid-plans.

To compare the manual-plans with auto- and hybrid-plans, six 3D gamma analyses were performed at 3 %/3 mm, 5 %/3 mm and 4 %/4 mm within the 1800 cGy and 3000 cGy isodoses, which approximately represent the lowest doses that are considered on clinical prostate treatment plans. 1800 cGy is 30 % of the 6000 cGy prescription and is the closest threshold available in VeriSoft v5.0 to the 2460 cGy (41 % of 6000 cGy) dose level used for rectum DVH reporting. 3000 cGy is 50 % of the 6000 cGy prescription and was the closest threshold available to the value of the lowest isodose (3155 cGy) displayed on prostate treatment plans at The Christie.

Since gamma analysis is not used routinely to compare dose distributions from different treatment plans in this way the most appropriate criteria were not known so performing multiple analyses with similar criteria to those used for clinical analysis allowed this to be considered retrospectively.

3.4 Results

The data in Table 3.3 show mean and standard deviation, minimum and maximum percentage point differences of the standard prostate radiotherapy DVH parameters between the auto- and manual-plans reported to the manual-contours. Femoral head dosimetry is given as the difference in $D_{1\,cc}$ between the auto- and manual-plans expressed as a percentage of the 6000 cGy prescription. This normalisation to the prescription dose was necessary because generally femoral heads receive a relatively low dose (typically ~3000 cGy) so small absolute differences in the maximum dose can give seemingly large percentage differences. Negative values in Table 3.3 indicate the auto-plan dose statistics were lower than the manual-plan and mean dose differences presented in bold were found to be statistically significant using Wilcoxon signed-rank tests with a threshold for statistical significance of $p < 0.05$.

The results in Table 3.3 show that the SPICE, Mirada RTx and ADMIRE auto-plans give considerably reduced target coverage compared to the corresponding manual-

plans. For SPICE and Mirada RTx the differences were all statistically significant with the exception of the $D_{1\,cc}$ for PTV3. PTV3 $D_{1\,cc}$ is the only target maximum dose that is reported clinically and it was not expected that optimisation using auto-contours would generate regions of dose significantly greater than the prescription dose. The target coverage reduction for SPICE and Mirada RTx would also certainly be considered clinically significant as well as statistically significant. PTV dose agreement between the manual- and auto-plans is best for ADMIRE where the reductions in target coverage were not found to be statistically significant. However, mean under-doses of 11.0 %, 9.4 % and 7.0 % in terms of $D_{99\,\%}$ are observed for PTV1, PTV2 and PTV3 respectively for the ADMIRE auto-plans, which would not be acceptable clinically.

The reduction in target coverage is illustrated in Figure 3.3, which shows cumulative DVHs for the three PTVs for one of the 11 patients. The beginning of the DVH fall-off occurs at a lower dose in the auto-plans compared to the manual-plan.

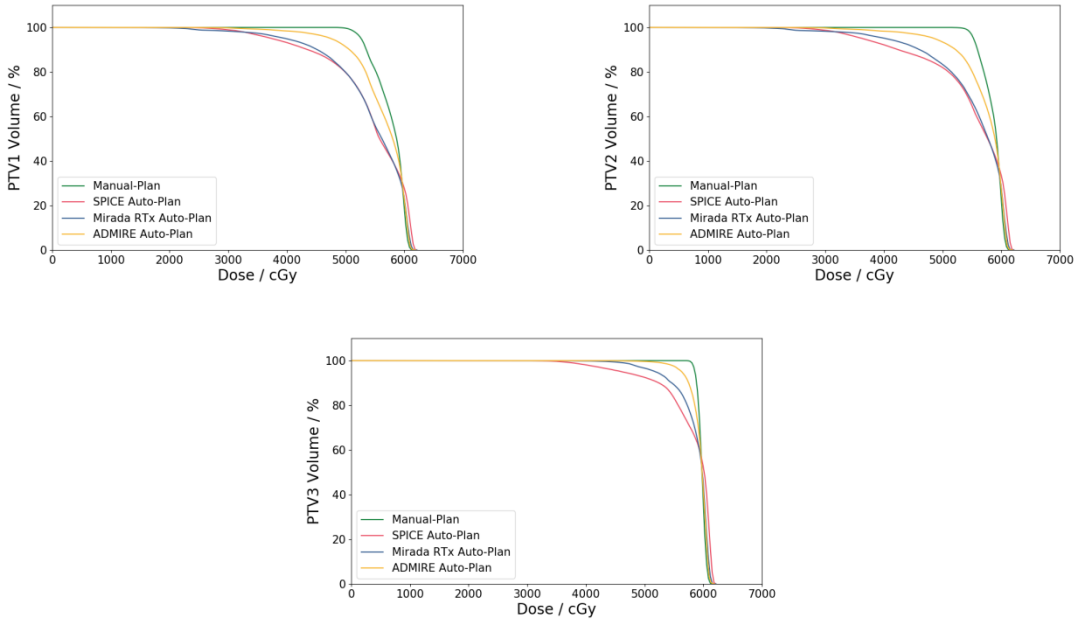

Figure 3.3: Cumulative PTV DVHs for the manual- and auto-plans for one of the 11 patients.

Table 3.3: Average percentage point differences between the auto-plans and manual-plans.

		SPICE / %				Mirada RTx / %				ADMIRE / %			
OAR	**Dose Statistic**	**Mean**	**SD**	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**
PTV1	Min 1 cc	**–27.8**	17.9	–27.8	17.9	**–30.5**	26.5	–75.6	2.9	–13.9	22.4	–69.2	5.6
	Min 99 %	**–22.6**	17.0	–22.6	17.0	**–27.2**	24.5	–68.2	2.9	–11.0	19.6	–62.0	5.6
PTV2	Min 1 cc	**–27.5**	21.6	–27.5	21.6	**–24.0**	26.1	–77.6	6.3	–11.3	25.1	–71.8	11.4
	Min 99 %	**–25.1**	19.9	–25.1	19.9	**–21.7**	24.0	–71.2	5.5	–9.4	22.3	–65.5	10.5
PTV3	Max 1 cc	0.7	1.1	0.7	1.1	0.6	1.8	–0.6	5.7	0.1	0.4	–0.5	0.7
	Min 1 cc	**–15.0**	16.8	–15.0	16.8	**–12.5**	18.5	–63.3	0.9	–7.9	18.1	–61.0	1.5
	Min 99 %	**–14.6**	15.6	–14.6	15.6	**–11.9**	17.5	–60.6	0.9	–7.0	16.3	–55.0	1.6
FHL	Max 1 cc	0.3	5.0	–9.0	8.5	–0.1	3.6	–5.5	4.9	**3.6**	3.6	–1.5	9.0
FHR	Max 1 cc	–1.7	5.2	–10.8	6.5	–1.5	6.1	–11.3	7.6	1.2	5.2	–7.3	10.8
Rectum	$V_{2460\ cGy}$	**8.6**	6.5	8.6	6.5	–0.4	10.0	–21.7	15.1	4.8	7.0	–6.8	17.4
	$V_{3240\ cGy}$	0.9	5.6	0.9	5.6	–1.9	10.8	–23.6	13.9	**5.5**	7.6	–8.7	16.6
	$V_{4080\ cGy}$	–1.8	4.4	–1.8	4.4	–1.3	10.7	–19.5	13.0	5.4	8.1	–10.3	14.3
	$V_{4860\ cGy}$	–2.3	3.3	–2.3	3.3	0.2	9.7	–13.7	16.1	5.0	7.7	–9.7	12.4
	$V_{5280\ cGy}$	–1.2	3.8	–1.2	3.8	2.1	8.4	–8.8	15.8	**4.3**	6.1	–6.3	10.4
	$V_{5700\ cGy}$	1.2	3.1	1.2	3.1	**3.3**	4.7	–2.7	12.4	**3.0**	3.4	–2.8	7.2
	$V_{6000\ cGy}$	0.2	0.3	0.2	0.3	0.2	0.3	0.0	0.8	0.1	0.1	0.0	0.4
Bladder	$V_{4080\ cGy}$	–5.7	14.0	–26.1	12.7	–0.5	15.5	–27.8	26.2	9.6	14.9	–16.4	32.1
	$V_{4860\ cGy}$	–5.0	13.6	–24.4	14.9	0.6	16.6	–23.4	35.5	**11.6**	16.3	–14.0	35.9
	$V_{6000\ cGy}$	0.6	3.0	–3.6	7.5	2.4	6.8	–4.2	21.6	2.6	4.4	–3.3	13.0

Although the results in Table 3.3 and plots in Figure 3.3 show that the auto-plan PTV coverage is significantly reduced compared to the manual-plan, the OARs show closer agreement. Generally, the SPICE and Mirada RTx auto-plans give closer dosimetric agreement for the OARs than ADMIRE. However, care has to be taken with this comparison as ADMIRE auto-plans give superior PTV coverage than SPICE and Mirada RTx and it is not clear from Table 3.3 how OAR auto-plan dosimetry is affected by this. Figure 3.4 shows bladder, rectum and femoral head cumulative DVHs for the same patient as Figure 3.3.

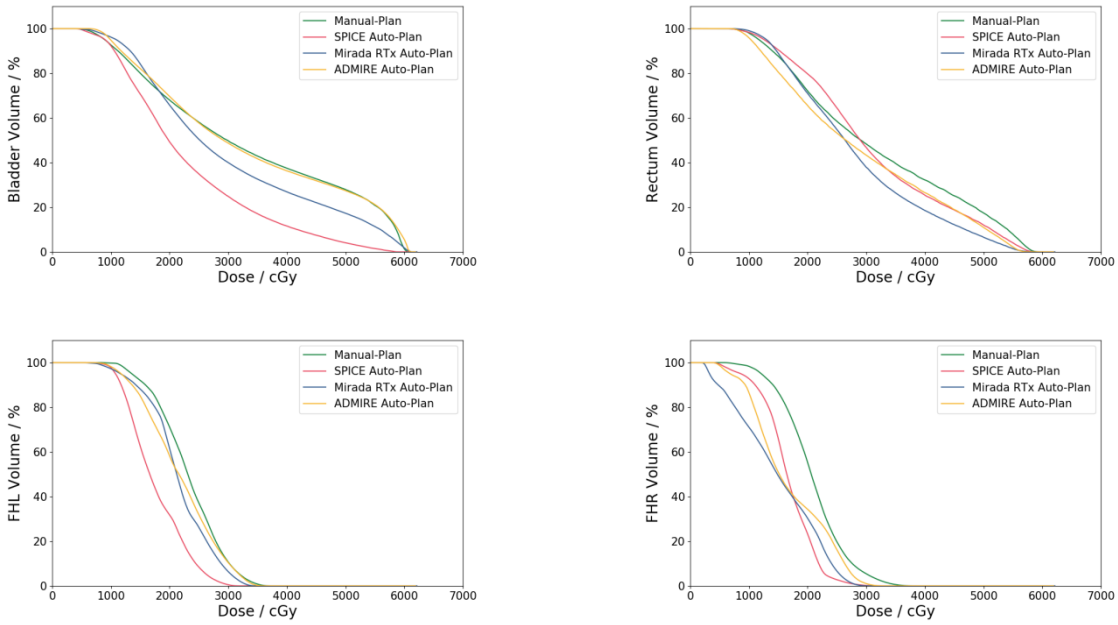

Figure 3.4: Cumulative OAR DVHs for the manual- and auto-plans for one of the 11 patients.

It is worth recalling here that the optimisation objectives and parameters were the same for all manual- and auto-plans. Therefore the dosimetric differences reported in Table 3.3 and illustrated in Figure 3.3 and Figure 3.4 are driven purely by geometric differences in the optimisation volumes. To evaluate this more specifically, the DSCs and average mean DTAs from Chapter 2 were plotted against the dose differences for each dose statistic in Table 3.3 and coefficients of determination were analysed. No significant correlations were found between geometric accuracy scores and any of the reported dose statistics. This implies that whilst DSC and average mean DTA are reasonably reliable predictors of geometric accuracy,

geometric accuracy itself is not necessarily a reliable predictor of auto-contour utility for treatment plan optimisation.

Table 3.4 presents mean and standard deviation, minimum and maximum percentage of pixels passing gamma analysis where the manual-plans have been compared with each of the auto-plans. The gamma analysis results also show poor agreement between the manual- and auto-plan dose distributions. Mean percentage of pixels passing were not statistically significantly different between the auto-segmentation packages across all analyses using Wilcoxon signed-rank tests and a threshold for significance of $p < 0.05$.

It is clear from the data in Table 3.3 and Table 3.4 and the example DVH plots in Figure 3.3 and Figure 3.4 that auto-plans differ significantly from corresponding manual-plans. Even with a fully developed treatment planning optimisation class-solution, in the majority of cases, auto-contours from SPICE, Mirada RTx and ADMIRE would not produce clinically acceptable auto-plans and so their utility for treatment planning is limited.

Table 3.4: Average pixels passing gamma analysis when the auto-plans are compared with the manual-plan.

			Pixels Passing / %			
		Threshold / %	Mean	SD	Min.	Max.
5 % / 3 mm	SPICE	30	57.6	6.2	46.5	66.4
	Mirada RTx		58.8	9.0	41.8	70.8
	ADMIRE		49.4	13.3	19.9	64.6
4 % / 4 mm	SPICE	30	58.0	6.3	45.6	65.3
	Mirada RTx		59.0	9.1	42.9	70.8
	ADMIRE		49.4	13.7	19.5	64.7
3 % / 3 mm	SPICE	30	46.6	5.5	34.9	53.4
	Mirada RTx		47.3	8.5	32.5	58.1
	ADMIRE		39.0	12.3	12.9	54.0
5 % / 3 mm	SPICE	50	69.6	8.4	50.8	80.5
	Mirada RTx		67.7	11.5	48.5	80.9
	ADMIRE		60.9	15.4	29.0	79.2
4 % / 4 mm	SPICE	50	73.2	9.2	51.3	84.5
	Mirada RTx		71.4	12.0	51.5	85.4
	ADMIRE		64.6	15.3	33.6	83.4
3 % / 3 mm	SPICE	50	61.5	9.1	39.7	72.4
	Mirada RTx		59.7	12.3	39.7	74.0
	ADMIRE		53.1	15.5	22.5	71.1

Table 3.5: Average percentage point differences between the hybrid-plans and manual-plans.

		Hybrid SPICE / %				Hybrid Mirada RTx / %				Hybrid ADMIRE / %			
OAR	**Dose Statistic**	**Mean**	**SD**	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**
PTV1	Min 1 cc	0.0	3.2	–6.1	6.6	0.1	3.1	–6.0	6.3	0.1	2.1	–3.6	3.5
	Min 99 %	0.0	2.7	–5.2	5.6	0.1	2.5	–5.1	5.2	0.1	1.8	–2.6	3.0
PTV2	Min 1 cc	1.2	3.1	–1.5	7.9	1.3	3.2	–1.3	7.8	1.7	3.2	–0.7	8.6
	Min 99 %	1.0	2.3	–1.1	4.9	1.0	2.4	–1.0	5.0	1.4	2.4	–0.8	5.5
PTV3	Max 1 cc	0.1	0.3	–0.6	0.7	0.0	0.3	–0.7	0.2	–0.1	0.3	–0.6	0.3
	Min 1 cc	0.2	1.0	–1.4	2.2	0.1	1.0	–1.6	2.1	0.2	1.3	–1.3	3.0
	Min 99 %	0.2	1.0	–1.4	2.2	0.1	1.0	–1.6	2.1	0.2	1.2	–1.1	2.7
FHL	Max 1 cc	–0.2	4.5	–10.5	5.0	0.2	2.8	–3.7	5.8	0.0	4.3	–8.9	6.0
FHR	Max 1 cc	–2.5	3.9	–7.2	3.6	–1.1	3.2	–7.1	3.7	0.2	2.4	–3.5	4.9
Rectum	$V_{2460\ cGy}$	1.6	4.4	–5.5	11.1	4.0	4.3	–3.2	11.7	1.6	3.8	–5.6	6.8
	$V_{3240\ cGy}$	1.3	3.9	–5.2	9.7	2.2	3.5	–2.7	9.4	1.2	2.9	–4.1	6.5
	$V_{4080\ cGy}$	0.8	3.6	–3.5	8.9	1.1	3.2	–3.4	8.2	0.7	2.4	–3.7	4.7
	$V_{4860\ cGy}$	0.2	3.6	–6.2	8.3	0.4	3.4	–5.9	7.9	0.4	2.3	–3.5	3.7
	$V_{5280\ cGy}$	0.1	3.3	–5.9	7.0	0.1	3.1	–6.4	6.2	0.3	2.2	–3.4	3.4
	$V_{5700\ cGy}$	1.0	2.5	–2.4	6.2	0.8	2.5	–2.6	5.9	0.1	1.7	–2.2	3.3
	$V_{6000\ cGy}$	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bladder	$V_{4080\ cGy}$	1.5	4.3	–2.0	12.9	2.1	4.6	–1.8	14.6	1.7	3.7	–1.5	10.8
	$V_{4860\ cGy}$	1.6	4.7	–1.8	14.5	2.1	5.1	–1.5	16.1	1.6	3.7	–1.3	10.5
	$V_{6000\ cGy}$	0.9	3.3	–2.5	9.9	0.2	2.5	–3.1	7.0	0.6	2.6	–2.9	7.2

The data in Table 3.5 show mean and standard deviation, minimum and maximum percentage point differences for the standard prostate radiotherapy DVH parameters between the hybrid- and manual-plans reported to the manual-contours. As with Table 3.3, femoral head dosimetry is given as the difference in $D_{1\,cc}$ expressed as a percentage of the 6000 cGy prescription.

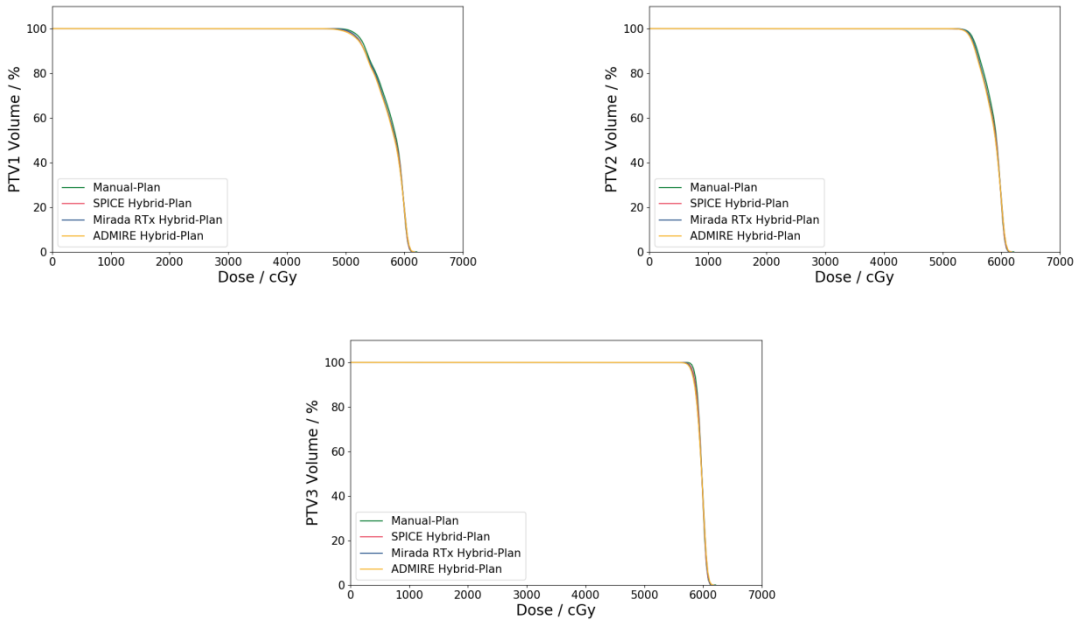

Figure 3.5: Cumulative PTV DVHs for the manual- and hybrid-plans for one of the 11 patients.

The results in Table 3.5 show that hybrid-plan PTV dosimetry is reasonably consistent with manual-plan dosimetry. All three auto-contouring packages generate hybrid-plan PTV doses that are on average within –0.1 % to 1.7 % of the manual-plan for all reported PTV dose statistics. Wilcoxon signed-rank tests were used to assess the mean differences and none was found to be statistically significant with a significance threshold of $p < 0.05$.

Figure 3.5 shows hybrid-plan PTV DVHs for the same patient used in Figure 3.3 and Figure 3.4. It is immediately apparent that the hybrid-plan PTV coverage is improved dramatically compared to the auto-plan and is comparable to the manual-plan.

OAR doses are also in good agreement between the hybrid- and manual-plans.

SPICE and ADMIRE hybrid-plans have rectum and bladder dosimetry that is on average between 0.0 % and 1.7 % different than the manual-plan. A slightly wider range is seen for Mirada RTx, from 0.0 % to 4.0 %. For all three packages, hybrid-plan femoral head doses were similar to the manual-plan. Figure 3.6 shows bladder, rectum and femoral head cumulative DVHs for the same patient as above.

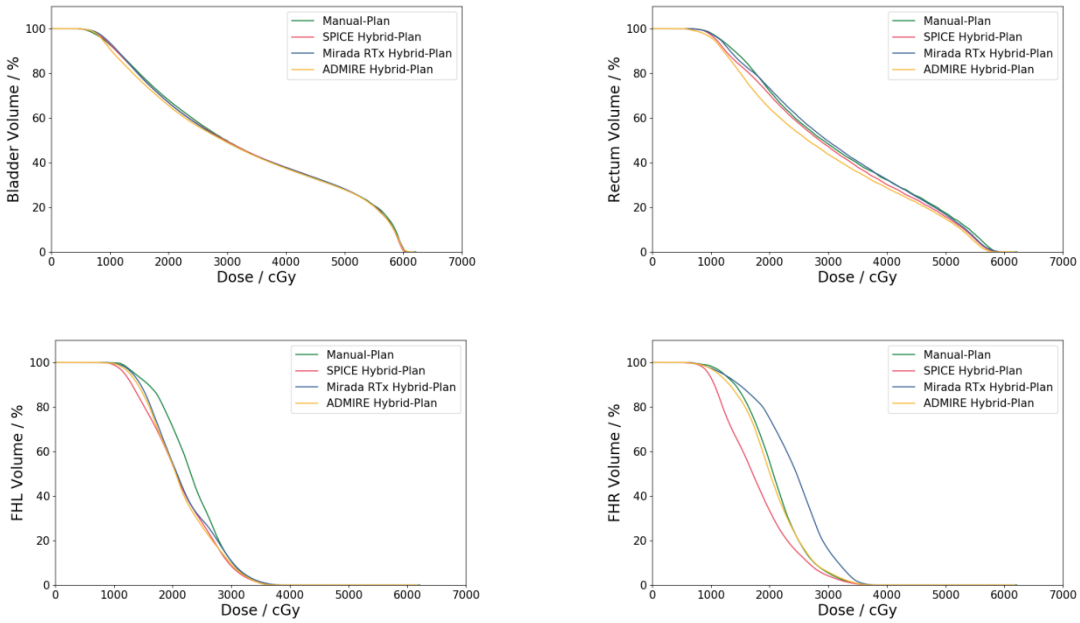

Figure 3.6: Cumulative OAR DVHs for the manual- and hybrid-plans for one of the 11 patients.

Gamma analysis results presented in Table 3.6 show that the hybrid-plan dose distributions are similar to those of the corresponding manual-plans. At 3 % / 3 mm – the strictest gamma analysis criteria that are used for clinical assessments at The Christie – all three auto-segmentation packages generate hybrid-plans with average pass rates greater than 90 % within the 50 % isodose. The pass rate drops by ~20 % when the dose threshold is reduced to 30 %, but since 3155 cGy (i.e. 52.6 % of the prescription) is the lowest isodose displayed on clinical prostate treatment plans, doses lower than this are not considered clinically. The only exception to this is the rectum $V_{2460\,cGy}$, and from Table 3.5 this is the rectum dose statistic that shows the biggest disagreement between the manual- and hybrid-plans – although the average difference is 1.6 %, 4.0 % and 1.6 %

for SPICE, Mirada RTx and ADMIRE respectively, which are differences that could potentially be improved with a more developed optimisation class-solution.

As for the auto-plan gamma analyses, the differences between the pass rates for the hybrid-plans were not statistically significantly different between auto-segmentation packages using Wilcoxon signed-rank tests and a threshold for significance of $p < 0.05$.

The gamma analysis failures were reviewed visually on a patient-by-patient basis and no apparent patterns were observed in where the hybrid- and manual-plans differed. This was expected since, other than the inferior extent of the SPICE femoral head auto-contours (see Figure 2.3), the geometric discrepancies in the OAR auto-contours were not generally considered to be systematic. Therefore it was not surprising that patterns of failures were not observed in the hybrid-plan gamma analyses.

Table 3.6: Average pixels passing gamma analysis when the hybrid-plans are compared with the manual-plan.

			Pixels Passing / %			
		Threshold / %	Mean	SD	Min.	Max.
5 % / 3 mm	SPICE	30	78.2	7.6	65.8	88.1
	Mirada RTx		77.9	7.2	65.4	88.2
	ADMIRE		79.6	8.9	58.4	92.5
4 % / 4 mm	SPICE	30	77.2	7.2	66.5	86.7
	Mirada RTx		76.9	6.9	63.9	86.7
	ADMIRE		78.8	8.3	59.4	91.1
3 % / 3 mm	SPICE	30	67.4	7.7	54.6	76.9
	Mirada RTx		67.5	7.2	55.8	79.4
	ADMIRE		69.4	9.2	50.2	84.2
5 % / 3 mm	SPICE	50	94.4	3.8	85.4	99.3
	Mirada RTx		94.6	3.5	85.7	98.7
	ADMIRE		94.6	4.1	85.7	98.7
4 % / 4 mm	SPICE	50	95.4	3.0	88.7	99.5
	Mirada RTx		95.6	2.8	89.1	99.1
	ADMIRE		95.6	3.1	89.7	98.8
3 % / 3 mm	SPICE	50	90.5	5.5	78.1	98.1
	Mirada RTx		91.0	4.8	79.6	97.0
	ADMIRE		91.0	5.5	81.2	97.1

Despite the dosimetry of the hybrid-plans closely matching the manual-plans, no correlations were found between the differences in any of the OAR dose statistics and either DSC or average mean DTA. This is consistent with the finding noted above for the auto-plans in that where DSC and average mean DTA are reliable predictors of geometric accuracy, they do not necessarily reliably predict the utility of the auto-contours for treatment planning. This may be in part due to the limitations of the standard geometric accuracy metrics, which could be insensitive to the subtle differences in utility between auto-contours from different packages. This is potentially a significant finding because the literature is heavily biased towards evaluations of auto-contours using geometric accuracy scores and this may not translate to utility in the clinical setting.

3.5 Discussion

The work presented in this chapter has shown that auto-plans generated with target and OAR auto-contours suffer from severely compromised target coverage. Hybrid-plans, however, where optimisation is performed with target manual-contours and OAR auto-contours, demonstrated good agreement with manual-plans. This was demonstrated to be the case with three different and fully-independent atlas-based auto-segmentation packages. Furthermore, the degree to which the auto- and hybrid-plans differed from corresponding manual-plans was not found to correlate with geometric accuracy.

This implies that while DSC and DTA are frequently used to assess auto-segmentation performance geometric accuracy itself represents a poor surrogate with which to infer clinical utility. Specifically testing the utility of auto-contours in the intended clinical setting is often overlooked in the literature in favour of geometric accuracy evaluations or increasingly time and motion studies that measure efficiencies of manually correcting auto-contours and drawing them *de novo*. This work has shown that such approaches alone are insufficient to evaluate the performance of auto-segmentation packages fully and a better alternative is to use the auto-contours in the intended clinical setting and then evaluate the impact on the resultant treatment plans.

This work was limited to 11 patients and since the optimisation was performed using a non-clinical optimisation class-solution, the manual-, auto- and hybrid-plans were not necessarily dosimetrically optimal and clinically acceptable. The intention here was not necessarily to produce such optimal and acceptable treatment plans; rather, it was simply to investigate the impact of geometric inaccuracies on treatment plan generation, which has been achieved. It is not particularly useful to comment on how many of the plans were optimal and acceptable (since this was not the purpose of the work) but this is a limitation of the conclusions that can be drawn. Further work should address this issue by developing a clinical optimisation strategy that would allow more widespread testing of the hybrid-plan concept. Automation of this new optimisation strategy would provide an expedient way to produce treatment plans very quickly after clinician outlining of prostate and SV target volumes. This could be used to generate significant efficiencies for treatment planning, which could in turn be used to pursue developments in advanced radiotherapy techniques that are currently unfeasible due to the associated increased workload.

3.6 Conclusion

Prostate VMAT treatment plans generated using SPICE, Mirada RTx and ADMIRE auto-contours have been shown to differ significantly from plans generated using manual-contours. However, if auto-contours are used only to generate OAR sparing volumes and prostate and SV manual-contours are used to generate PTVs, agreement between this hybrid-plan and a manual-plan is on average generally within approximately 1.5 % and not statistically significant. 3D gamma analysis has also shown good agreement between the hybrid- and manual-plan dose distributions.

Despite observing poor dosimetric agreement between auto-plans and manual-plans and good agreement between hybrid-plans and manual-plans, no correlations were found between the compared dose statistics and DSC or average mean DTA. Therefore when considering the utility of auto-contours for radiotherapy treatment planning, it is important not just to consider geometric accuracy but also the intended clinical

purposes of the auto-contours.

Figure 1.1 shows that manual-contouring introduces constraints into the prostate VMAT treatment planning pathway. The work presented here shows that the utility of atlas-based auto-segmentation is severely limited for the prostate and SV and thus for the generation of PTVs. This implies that auto-segmentation cannot remove the first manual-contouring constraint in Figure 1.1. However, the hybrid-plan results suggest that auto-segmentation of OARs could remove the second constraint and allow the treatment plan optimisation stage of the pathway to be brought forward. Even if the auto-contoured OARs need to be manually edited for final dose reporting, which is likely based on the geometric accuracy results presented in Chapter 2, moving and automating the optimisation process such that is could start immediately after the prostate and SV are manually drawn would potentially drive significant efficiencies for clinical treatment planning departments.

3.7 References

Beasley, W., McWilliam, A., Aitkenhead, A., Mackay, R., & Rowbottom, C. (2016). 'The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy', *Journal of Applied Clinical Medical Physics*, 17(2), pp. 41-49.

Beekley Medical (2021). *CT-SPOT 120.* Available at: https://beekley.com/product-details/ct-spot-120, (Accessed: 14 April 2021).

Boylan, C., Golby, C. & Rowbottom, C. (2010). 'A VMAT planning solution for prostate patients using a commercial treatment planning system', *Physics Medicine Biology*, 55(14), pp. 395-404.

Dearnaley, D., Syndikus, I., Mossop, H., Vincent, K., Birtle, A., Bloomfield, D., Graham, J., Kirkbride, P., Logue, J., Malik, Z., Money-Kyrle, J., M O'Sullivan, J., Panades, M., Parker, C., Patterson, H., Scrase, C., Staffurth, J., Stockdale, A., Tremlett, T., Bidmead, M., Mayles, M., Naismith, N., South, C., Gao, A., Cruickshank, C., Hassan, S., Pugh, J., Griffin, C. &Hall, E. (2016). 'Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 17(8), pp. 1047-1060.

Elekta (2013). *ABAS Atlas-Based Autosegmentation.* Available at: https://www.elekta.com/dam/jcr:7b262e2d-74c7-412c-bd1d-eec08da8183d/ABAS.pdf, (Accessed: 6 November 2017).

Gugyeras, D., Farkas, A., Gulyban, A., Csima, M., Hadjiev, J. & Lakosi, F. (2017). 'Multi-Atlas based segmentation (Mirada RTx) vs. manual contouring for OARs in the head and neck region: efficiency time analysis', *South East Europe Technology in Radiation Oncology 2017*. Available at: https://www.researchgate.net/profile/Ferenc_Lakosi/project/Time-sparing-effect-of-multi-ABS-Single-vs-Multi-ABS-Atlas-Based-Segmentation/attachment/5a03823eb53d2fed8ad4336f/AS:558547001327616@1510179390631/download/Seetro2017.pdf?context=projectUpdatesLog, (Accessed: 19 February 2018).

Kumar. L., Yadav, G., Raman, K., Bhushan, M. & Pal, M. (2015). 'The dosimetric impact of different photon beam energy on RapidArc radiotherapy planning for cervix carcinoma', *Journal of Medical Physics*, 40(4), pp. 207-213.

Larrue, A., Kadir, T. & Gooding, M. (2013). 'SU-E-J-93: Assessment of Contour Fusion Method for Auto-Contouring of Head and Neck CT Images', *Medical Physics*, 40(6), pp. 171.

Lim, J. & Leech, M. (2016). 'Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck', *ACTA Oncologica*, 55, 7, pp. 799-806.

Low, D., Harms, W., Mutic, S. & Purdy, J. (1998). 'A technique for the quantitative evaluation of dose distributions', *Medical Physics*, 25(5), pp. 656-661.

Mattes, M., Tai, C., Lee, A., Ashamalla, H. & Ikoro, N. (2014). 'The dosimetric effects of photon energy on the quality of prostate volumetric modulated arc therapy', *Practical Radiation Oncology*, 4(1), pp. e39-e34.

Mirada Medical (2017). *RT Imaging. Agile Workflows.* Available at: http://mirada-medical.com/wp-content/uploads/2017/11/mm6908-1-ROW-RT-Imaging-Brochuretoemail.pdf, (Accessed: 11 December 2017).

Orlandini, L., Coppola, M., Fulcheri, C., Cernusco, L., Wang, P. & Cionini, L. (2017). 'Dose tracking assessment for image-guided radiotherapy of the prostate bed and the impact on clinical workflow', *Radiation Oncology*, 12(78), pp. 1-8.

Qazi, A., Pekar, V., Kim, J., Xie, J., Breen, S. & Jaffray, D. (2011). 'Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach', *Medical Physics*, 38(11), pp. 6160-6170.

Shang, J., Kong, W., Wang, Y., Ding, Z., Yan, G. & Zhe, H. (2014). 'VMAT planning study in rectal cancer patients', *Radiation Oncology*, 9, pp. 1-7.

Speight, R., Karakaya, E., Prestwich, R., Sen, M., Lindsay, R., Harding, R. & Sykes, J. (2014). 'Evaluation of atlas based auto-segmentation for head and neck target volume delineation in adaptive/replan IMRT', *Journal of Physics*, 489, pp. 1-4.

Stanley, D., Popp, T., Ha, C., Swanson, G., Eng, T., Papanikolaou, N. & Gutiérrez, A. (2015). 'Dosimetric effect of photon beam energy on volumetric modulated arc therapy treatment plan quality due to body habitus in advanced prostate cancer', *Practical Radiation Oncology*, 5(6), pp. e625-e633.

Teguh, D., Levendag, P., Voet, P., Al-Mamgani, A., Han, X., Wolf, T., Hibbard, L.,Nowak, P., Akhiat, H., Dirkx, M., Hiejmen, B. & Hoogeman, M. (2011). 'Clinical Validation of Atlas-Based Automatic Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck', *International Journal of Radiation Oncology Biology Physics*, 81(4), pp. 950-957.

The Royal College of Radiologists (2016). *Radiotherapy Dose Fractionation Second Edition.* Available at: https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfco163_dose_fractionation_2nd_ed_march2017.pdf, (Accessed: 30 November 2018).

Voet, P., Dirkx, M., Teguh, D., Hoogeman, M., Levendag, P. & Heijmen, B. (2011). 'Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis', *Radiotherapy & Oncology*, 98, pp. 373-377.

Wilkins, A., Mossop, H., Syndikus, I., Khoo, K., Bloomfield, D., Parker, C., Logue, J., Scrase, C., Patterson, H., Birtle, A., Staffurth, J., Malik, Z., Panades, M., Eswar, C., Graham, J., Russell, M., Kirkbride, P., O'Sullivan, J., Gao, A., Cruickshank, C., Griffin, C., Dearnaley, D. & Hall, E. (2015). 'H ypofractionated radiotherapy versus conventionally fractionated radiotherapy for patients with intermediate-risk localised prostate cancer: 2-year patient-reported outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 16(16), pp. 1605-1616.

Zhu, M., Bzdusek, K. Brink, C., Eriksen, J., Hansen, O., Jensen, H., Gay, H., Thorstad, W., Widder, J., Brouwer, C., Steenbakkers, R., Vanhauten, H., Cao, J., McBrayne, G., Patel, S., Cannon, D., Hardcastle, N., Tomé, W., Guckenberg, M. & Parikh, P. (2013). 'Multi-institutional Quantitative Evaluation and Clinical Validation of Smart Probabilistic Image Contouring Engine (SPICE) Autosegmentation of Target Structures and Normal Tissues on Computer

3.8 Supplementary Material

3.8.1 Treatment Plan Optimisation Parameters

The table below presents the optimisation objectives used in Pinnacle[3] for the new optimisation class-solution used to generate auto- and hybrid-plans.

ROI	Type	Dose / cGy	Volume / %	Weight
PTV3	Min Dose	5950	–	90
PTV3	Max Dose	6100	–	70
PTV3	Uniform Dose	6000	–	10
PTV2-PTV3	Min Dose	5550	–	100
PTV2-PTV3	Max Dose	5950	–	55
PTV2-PTV3	Max DVH	5900	5	70
PTV1-PTV2	Min Dose	5100	–	95
PTV1-PTV2	Max Dose	5500	–	70
PTV1-PTV2	Max DVH	5450	5	70
Rind_1	Max Dose	5250	–	30
Rind_2	Max Dose	4600	–	40
Rind_3	Max Dose	3900	–	50
Background	Max Dose	3250	–	50
Rectum_Sparing_InPlane	Max Dose	3500	–	40
Rectum_Sparing_InPlane	Max EUD	3250	–	30
Bladder_Sparing_InPlane	Max Dose	3500	–	40
Bladder_Sparing_InPlane	Max EUD	3500	–	30
FHR_Sparing	Max Dose	4000	–	15
FHL_Sparing	Max Dose	4000	–	15

3.8.2 ESTRO 2019 Poster

Based on the work presented in Chapters 2 and 3 an abstract was submitted to the 2019 European Society for Radiotherapy & Oncology (ESTRO) conference. The abstract was accepted for a poster exhibition and a representation of the poster is shown on the following page.

A comparative study of male pelvis CT auto-segmentation and its clinical utility

J. Wood[1, 2], M. Aznar[2] & P. Whitehurst[1]
[1]The Christie NHS Foundation Trust, United Kingdom. [2]The University of Manchester, United Kingdom.
joe.wood@christie.nhs.uk : @joewood86

Purpose

This work tested the geometric accuracy of auto-segmented structures within the male pelvis and their utility for prostate VMAT treatment planning.

Method

SPICE (Philips Radiation Oncology), Mirada RTx (Mirada Medical) and ADMIRE (Elekta) were used to segment the bladder, rectum, femoral heads, prostate and SV in 11 pelvic CT scans. For Mirada RTx and ADMIRE, atlases were generated from local manual-structures. Atlases in SPICE cannot be customised and so the proprietary 'Male Pelvis' atlas was used. DICE and mean DTA were used to assess the geometric accuracy of the auto-structures compared to 'ground-truth' manual-structures drawn by expert clinical oncologists and treatment planners. Manual- structures were drawn purely on CT without MR fusion.

PTVs and optimisation volumes were expanded and VMAT treatment plans generated using each set of auto- and manual-structures in the Pinnacle TPS. Hybrid-plans were also created using manual targets (prostate and SV) and auto-OARs. The same optimisation objectives and beam parameters were used for all plans. Auto- and hybrid-plan dose distributions were compared to manual-plans using standard prostate VMAT DVH statistics (reported to the 'ground-truth' manual-structures) and gamma analysis at 3 % / 3 mm inside the 30 % and 50 % isodoses – the two lowest doses reported on clinical treatment plans based on the CHHiP protocol.

Results

In terms of DICE and mean DTA, SPICE and ADMIRE performed best for the bladder, rectum and SV. Mirada RTx and ADMIRE performed best for the femoral heads and all three packages were comparable for the prostate. Figure 1 shows boxplots for each structure and variation across the 11 patients.

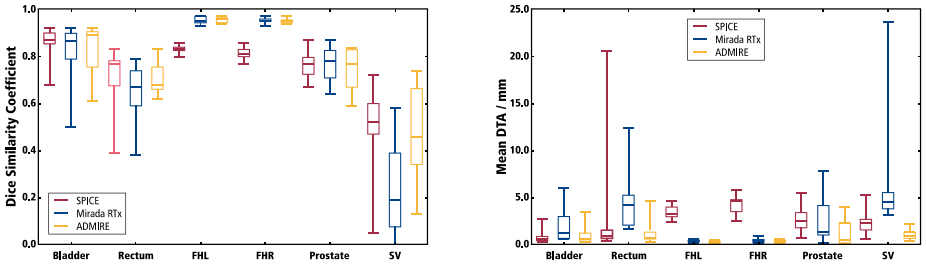

Figure 1: Boxplots showing variation in DICE and mean DTA for auto-structures in the male pelvis.

Gamma analysis pass rates are presented in Table 1 and show that the dose distributions for all hybrid-plans are comparable to the corresponding manual-plans at doses > 50 %. For the hybrid-plans, all reported DVH statistics were within 3 % of those achieved with the manual-structures (mostly within 2 %). No correlation was found between any of the geometric evaluations of the auto-structures and treatment plan doses.

	3 % / 3 mm (30 % threshold) / %	3 % / 3 mm (50 % threshold) / %
SPICE	47.5 ± 7.5	63.9 ± 13.3
Mirada RTx	45.5 ± 8.8	58.8 ± 12.3
ADMIRE	41.1 ± 13.8	53.4 ± 15.5
Hybrid-SPICE	68.1 ± 9.3	90.0 ± 5.2
Hybrid-Mirada RTx	66.4 ± 9.4	90.1 ± 6.4
Hybrid-ADMIRE	70.0 ± 5.3	90.5 ± 4.1

Table 1: Gamma analysis pass rates comparing auto- and hybrid-plans with manual-plans.

Conclusion

Overall, SPICE and ADMIRE gave the best geometric agreement with manual outlines for the 11 patients in this study. The utility of auto-segmented target volumes for treatment planning is severely limited. However, hybrid-plans using manual target volumes and auto-OARs demonstrate good agreement with manual-plans. Further work is required to refine the hybrid treatment planning solution, but this work demonstrates promise for the approach, which could generate significant treatment planning efficiencies.

Chapter 4

4 Knowledge-Based Prostate Treatment Planning

Joseph Wood[1, 2], Marianne Aznar[2], Philip Whitehurst[1, 2]

[1]The Christie NHS Foundation Trust
[2]The University of Manchester

4.1 Abstract

In this chapter a novel approach to knowledge-based (KB) treatment planning is proposed, developed and tested. The KB consists of a model of the ideal prostate radiotherapy treatment plan, which was trained using 562 clinical treatment plans that had been generated manually over a two year period at The Christie NHS Foundation Trust. For the 562 treatment plans used to train the model, optimisation of dose fall-off within OARs from the high dose region around the PTVs was assessed and found to be extremely inconsistent. From the training data, two optimisation class-solutions were developed with the aims of achieving the average dose fall-off to maintain average

treatment plan quality and achieving the 25th percentile dose fall-off to improve average treatment plan quality. The KB class-solutions were tested on a sample of 11 new cases and treatment plans of at least comparable quality were generated in nine of the 11 cases without clinically or statistically significant compromise to PTV coverage.

4.2 Introduction

Standard-of-care radiotherapy for intermediate- to high-risk prostate cancer (T1b-T3aN0M0) delivered at The Christie is based on the hypofractionated arm of the CHHiP trial (Dearnaley *et al.*, 2016; Wilkins *et al.*, 2015). In general, the ideal high dose region in a standard prostate radiotherapy treatment plan can be characterised by the required target dosimetry described in section 3.2. It is therefore relatively simple to convert this characterisation into a short list of objectives for an optimisation algorithm. However, since OARs sit in steep dose gradients and the size and shape of OARs is extremely variable between patients (see section 1.2), predicting ideal OAR dosimetry – and thus converting it to optimisation objectives – is difficult. Typically, therefore, clinical treatment plan optimisation consists of a manual process where OAR doses are iteratively reduced as far as is reasonably practicable.

Such iterative approaches can be subjective and time consuming, however (Djajaputra *et al.*, 2003). Knowledge-based (KB) treatment planning aims to incorporate prior experience into the optimisation of new treatment plans automatically to reduce this subjectivity (Nwankwo *et al.*, 2015). A number of approaches to KB treatment planning have been described in the literature where generally a KB consists of a database of previously optimised treatment plans. When a new treatment plan is to be optimised, it is characterised in some way such that OAR DVH parameters can be predicted using the treatment plan from the KB with the most similar characterisation (Chanyavanich *et al.*, 2011; Wu *et al.*, 2009; Janssen *et al.*, In press; Wang, Heijmen & Petit, 2017; Petit *et al.*, 2017; Good *et al.*, 2013; Yang *et al.*, 2013; Zhu *et al.*, 2011; Yuan *et al.*, 2012).

As noted above, accurate prediction of OAR DVH parameters is difficult due to

variations in the size and shape of OARs relative to the PTVs and these variations have a significant impact on the OAR DVH (Bzdusek *et al.*, 2012). Any uncertainties in the complex process of DVH prediction ultimately become manifest as uncertainties in the degree to which the final treatment plan is dosimetrically optimal. It is perhaps for this reason that where the literature reports that KB treatment planning is used in the clinic, it is often only as a tool to quality assurance manual treatment planning and not a primary method of treatment plan generation (Powis *et al.*, 2017).

A further disadvantage to this approach to KB treatment planning is that the need for computational speed can severely restrict the number of previous treatment plans that can be included in the KB database, which limits the likelihood of finding a suitable match for the new treatment plan.

The work presented in this chapter describes the author's development of a novel approach to KB prostate VMAT treatment planning that aims to overcome the limitations of current approaches described above. The basis of the KB presented here is a generalised model of the ideal prostate radiotherapy treatment plan that applies to all patient anatomies and geometries[§]. The model characterises ideal coverage of PTVs and dose fall-off into the surrounding OARs which are learnt from a large cohort of clinical treatment plans. Since variations in OAR size and shape should not significantly affect the PTV1-OAR dose gradients, using this trained model of the ideal prostate radiotherapy treatment plan to optimise new treatment plans could provide a more reliable and consistent approach to automated KB treatment planning.

An abstract was submitted to the 2020 European Society for Radiotherapy and Oncology (ESTRO) conference based on the work presented in this chapter and Chapter 5. This was accepted for presentation as a poster and a reproduction of the poster is presented in section 5.8.1.

[§] The 'ideal treatment plan' refers to how treatment plans should be generated at The Christie following a local clinical protocol. It is acknowledged that protocols in other centres differ from what is described here.

4.3 Method

A model of the ideal prostate radiotherapy treatment plan was constructed according to a local clinical protocol. This model is illustrated in Figure 4.1: where no part of the anatomy should receive a dose greater than 6300 cGy; the blue 5700 cGy isodose line covers PTV3 (inner turquoise area) but does not extend outside PTV2 (maroon area); the orange 5460 cGy isodose line conforms to PTV2 and the green 5000 cGy isodose line conforms to the skyblue PTV1. The radial dose gradients (i.e. dose falloff per unit distance) from the PTVs into the OARs, (shown as the colour transparency gradients in Figure 4.1) should be as steep as can be practically achieved. For each OAR, the PTV1-OAR dose gradient should be fairly constant amongst a cohort of consistently optimised treatment plans because the practically achievable gradients are determined primarily by machine parameters and are largely independent of patient anatomy.

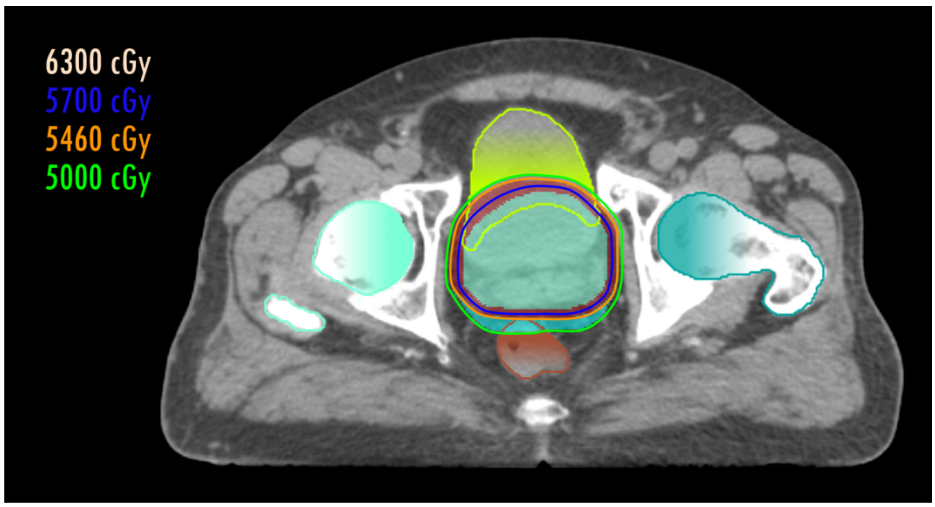

Figure 4.1: The ideal prostate radiotherapy treatment plan. Isodoses show highly conformal target coverage for all three dose levels and the colour gradients within OARs represent dose gradients, which are as steep as machine paramters will reasonably allow.

562 clinical prostate VMAT treatment plans generated in 2017 and 2018 at The Christie were used to train the model illustrated in Figure 4.1. These plans were all generated manually in Pinnacle[3] by trained treatment planners according to a well-established clinical protocol based on manual-contours drawn by clinical oncologists and the planners themselves. Although it has been stated previously that manual

treatment planning is subjective, all 562 treatment plans were accepted by expert treatment planners and clinical oncologists, passed a rigorous quality assurance process and were deliverable on the treatment machines. For the purpose of this work, these 562 plans are taken to be dosimetrically and clinically acceptable if not fully optimal.

To characterise the average PTV1-Rectum, PTV1-Bladder and PTV1-femoral head dose gradients, for each treatment plan, PTV1 was expanded isotropically in 2 mm increments up to 40 mm. OARs were then segmented based on the volume enclosed by pairs of successive PTV1 expansions but limited by the superior and inferior extents of PTV1 as illustrated in Figure 4.2.

The $D_{0.01\,cc}$, $D_{50\,\%}$ and mean dose of each OAR segment were recorded and these were each averaged over all 562 patients. Average PTV1-OAR dose gradients were then determined based on these statistics as functions of radial distance from the PTV1 surface. Limiting the superior and inferior extents of the OAR segmentations was important to avoid biasing the dose statistics by out of plane dose.

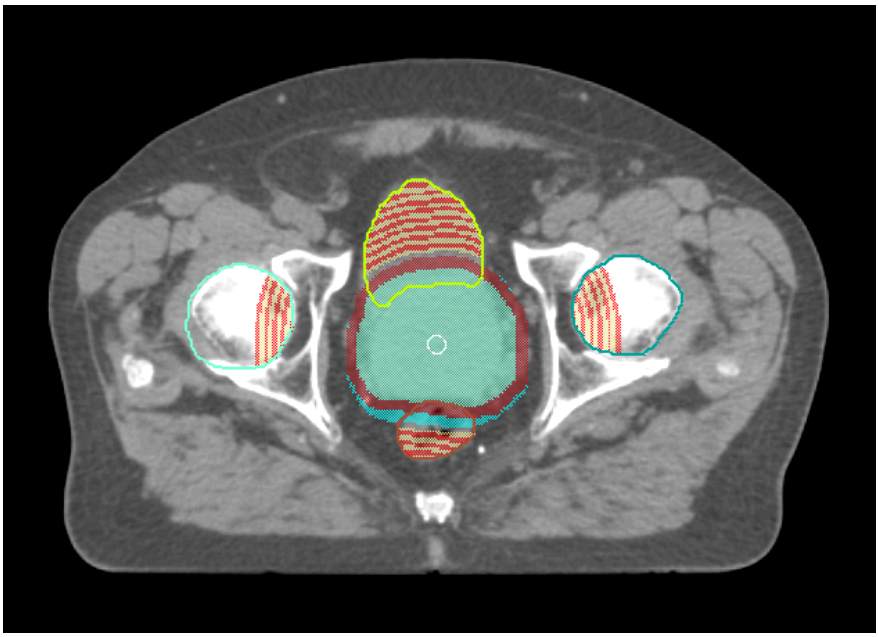

Figure 4.2: OAR segmentation based on 2 mm isotropic expansions of PTV1. OAR segments were also limited by the superior and inferior limits of PTV1 to avoid biasing dose statistics by out of plane dose.

In a busy clinical department, at the point where a treatment plan meets the minimum dosimetric requirements (set out in Table 3.2 and discussed in section 3.2) the treatment planner must trade-off dosimetric improvements generated by further optimisations with time spent running them. It was therefore expected that variations in the PTV1-OAR dose gradients would be observed amongst the 562 patients – a representation of the degree to which treatments plans are not consistently optimised.

To investigate this, histograms of the recorded $D_{0.01\,cc}$, $D_{50\,\%}$ and mean dose for each of the OAR segments were plotted. Strictly, according to the KB model presented in Figure 4.1, if all 562 treatment plans were consistently optimised, the histograms should form delta functions with spikes at decreasing doses as the radial distance from PTV1 is increased. Such an interpretation of the model is an unrealistic expectation and in reality predicting the exact forms of the distributions was not possible. A reasonable hypothesis, however, is that inconsistent optimisation primarily impacts the negative tails of the distributions. Where treatment plans are not optimised consistently, a negative skew would be expected because the upper limits are fixed by the maximum OAR dose criteria and finite target dose but the lower limits are where the variation resides.

For each OAR segment distribution, the skewness was calculated using equation 4.1, where n is the number of patients in the dataset, x_i is the dose to the i^{th} histogram bin and $\bar{x}$ and s are the sample mean and standard deviation respectively.

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^3, \qquad 4.1$$

Once the average PTV1-OAR dose gradients had been characterised they were used to inform the model illustrated in Figure 4.1. Within the TPS, the trained model consisted of two new optimisation class-solutions, which were tested on the 11 patients used in Chapters 2 and 3.

For each of the 11 test cases, OAR manual-contours were segmented into 2 mm

rinds, as per the training data. In the Pinnacle[3] optimiser, maximum dose optimisation objectives were assigned to these OAR segments based on the predictions of the trained model. In the first instance, the optimisation objectives were based on the average of the $D_{0.01\,cc}$ segmentation data from the training cohort with the aim of maintaining average treatment plan quality. Since it was acknowledged that the training cohort contained some treatment plans that were more optimal than others, by maintaining average quality, the treatment plans that were originally superior to the average would necessarily worsen when optimised using the average-KB class-solution. A second class-solution was therefore constructed in the same manner but the 25th percentile of the $D_{0.01\,cc}$ segmentation data from the training cohort were used to set maximum dose optimisation objectives with the aim of improving average treatment plan quality.

For each treatment plan generated with the KB, a cold-start optimisation was performed followed by an additional warm-start optimisation without modification of any optimisation parameters. A single 360 ° 10 MV VMAT arc was used to plan the treatment with a collimator twist of 10 ° to minimise interleaf leakage. Full details of the optimisation parameters used for each KB class-solution are included in section 4.8.

Although the 25th percentile KB class-solution described above can be justified to ensure treatment plan quality does not decrease, care has to be taken as from the dose gradient data generated for this work, it is not clear whether plans with steep PTV1-Rectum dose gradients also have steep PTV1-Bladder dose gradients. Therefore, it may not be possible to reduce bladder, rectum *and* femoral head doses without simultaneously increasing treatment plan complexity, which is difficult to assess definitively and not the intention of the work presented in this chapter. To assess this OAR sparing trade-off amongst the treatment plans in the training cohort, the individual PTV1-OAR dose gradients were calculated for each of the 562 treatment plans by fitting straight lines to the OAR segment $D_{0.01\,cc}$ against distance data over distances where the dose fall-off could reasonably be assumed to be linear. For the 562 patients, the PTV1-Rectum and PTV1-Bladder data were plotted against each other to assess correlation and provide insight into the validity of using either KB class-solution.

Evaluation of the 22 treatment plans generated using the KB class-solutions was performed using standard prostate DVH parameters and visual review by the author – a medical physics expert (MPE) trained and experienced in prostate radiotherapy treatment planning and checking.

4.4 Results

Figure 4.3 shows plots of mean $D_{0.01\,cc}$ as functions of radial distance from the PTV1 surface within the Bladder, Rectum and Femoral Heads for the 562 treatment plans used to train the model of the ideal prostate treatment plan illustrated in Figure 4.1. The 25th and 75th percentiles are also plotted as thinner lines and the 5th and 95th percentiles are plotted as dashed lines.

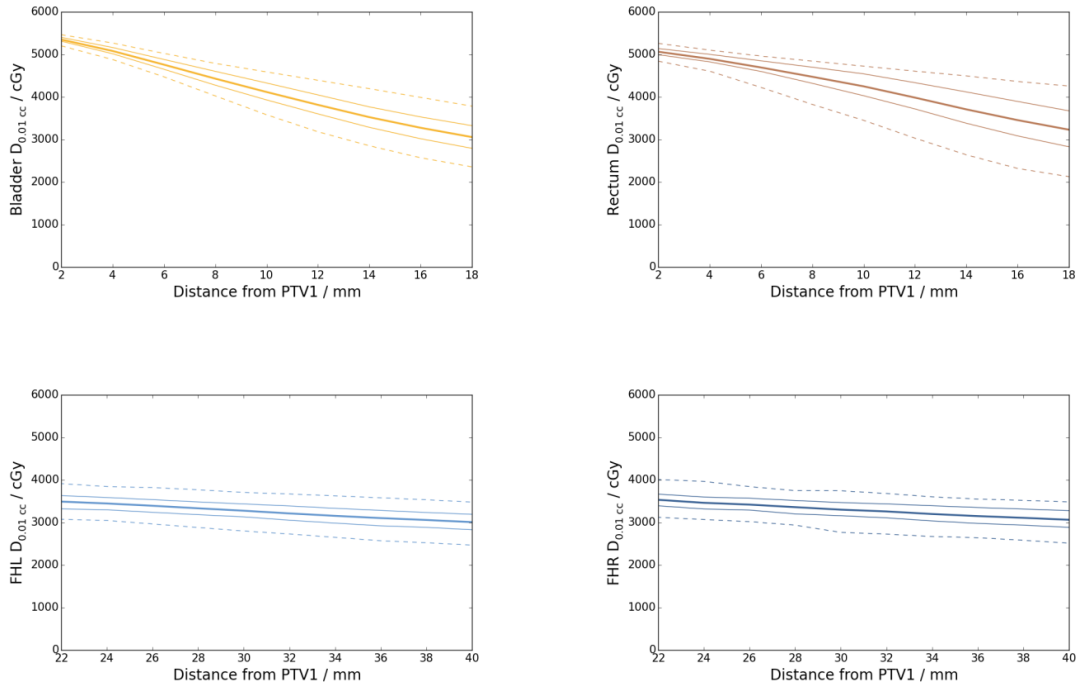

Figure 4.3: Mean dose gradients for the 562 manual treatment plans used to train the KB model within the Bladder, Rectum and Femoral Heads. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

It is worth noting here that the horizontal axes for the bladder and rectum plots are different than those for the femoral heads. Bladder and rectum are plotted from 2 mm to 18 mm, which is the distance range over which the dose fall-off is reasonably

linear – beyond 18 mm the average dose gradients become less steep and similar to those within the femoral heads. The femoral heads are plotted from 22 mm to 40 mm as the majority of the OAR segments closer than 22 mm had zero volume. The numbers of patients on which the bladder and rectum plots in Figure 4.3 are based decrease with increasing distance. This is because small OARs are entirely encompassed within a relatively short radial distance from PTV1 and the 2 mm OAR segments beyond that relatively short distance had zero volume. Similar plots for $D_{50\,\%}$ and mean dose to the OAR segments were generated but for conciseness these are not include here.

Simple straight lines were fit to the mean data to quantify the average PTV1-OAR dose gradients as –147 cGy mm^{-1}, –118 cGy mm^{-1}, –28 cGy mm^{-1} and –26 cGy mm^{-1} for the bladder, rectum, left and right femoral heads respectively. Average PTV1-Rectum and PTV1-Bladder dose gradients were approximately five times steeper than the PTV1-Femoral Head gradients. The relatively modest fall-off of dose throughout the femoral heads is indicative of their sitting in the low dose bath surrounding the target volumes.

For each patient, the PTV1-Rectum and PTV1-Bladder dose gradients was plotted against each other to assess the trade-off of OAR sparing. This plot is shown in Figure 4.4. The coefficient of determination is 0.0005 and it is clear that there is no correlation.

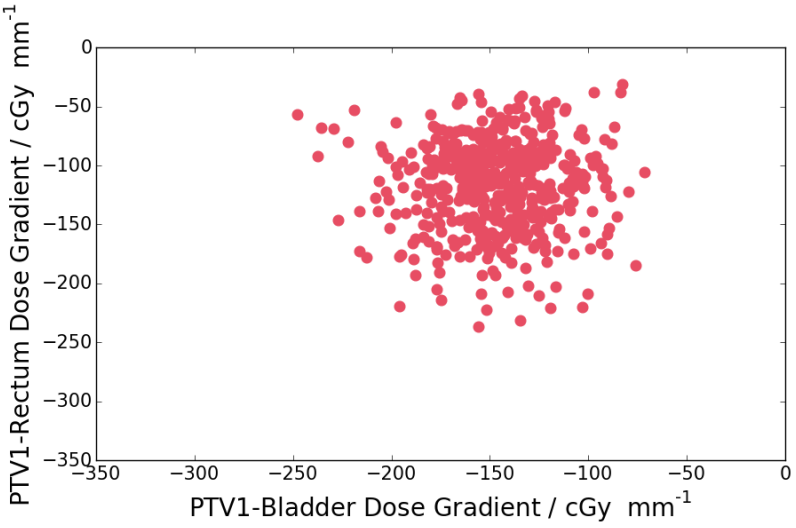

Figure 4.4: PTV1-Rectum dose gradient plotted against PTV1-Bladder dose gradient for the 562 patients used to train the KB model of the ideal prostate treatment plan.

Figure 4.5 shows histograms of $D_{0.01\,cc}$ at a selection of distances from PTV1 for the bladder and rectum. Bladder and rectum data are plotted together and are represented by the yellow and brown histograms respectively. Figure 4.6 shows similar plots for the femoral heads where blue and light blue histograms represent the right and left femoral heads respectively. Visually, it is apparent that the bladder and rectum histograms have a negative skew and Table 4.1 presents the skewness calculations that were performed on the OAR segment $D_{0.01\,cc}$ distributions. These skewness calculations provide a useful baseline for future evaluations of the distributions of $D_{0.01\,cc}$ from treatment plans generated with the KB class-solutions.

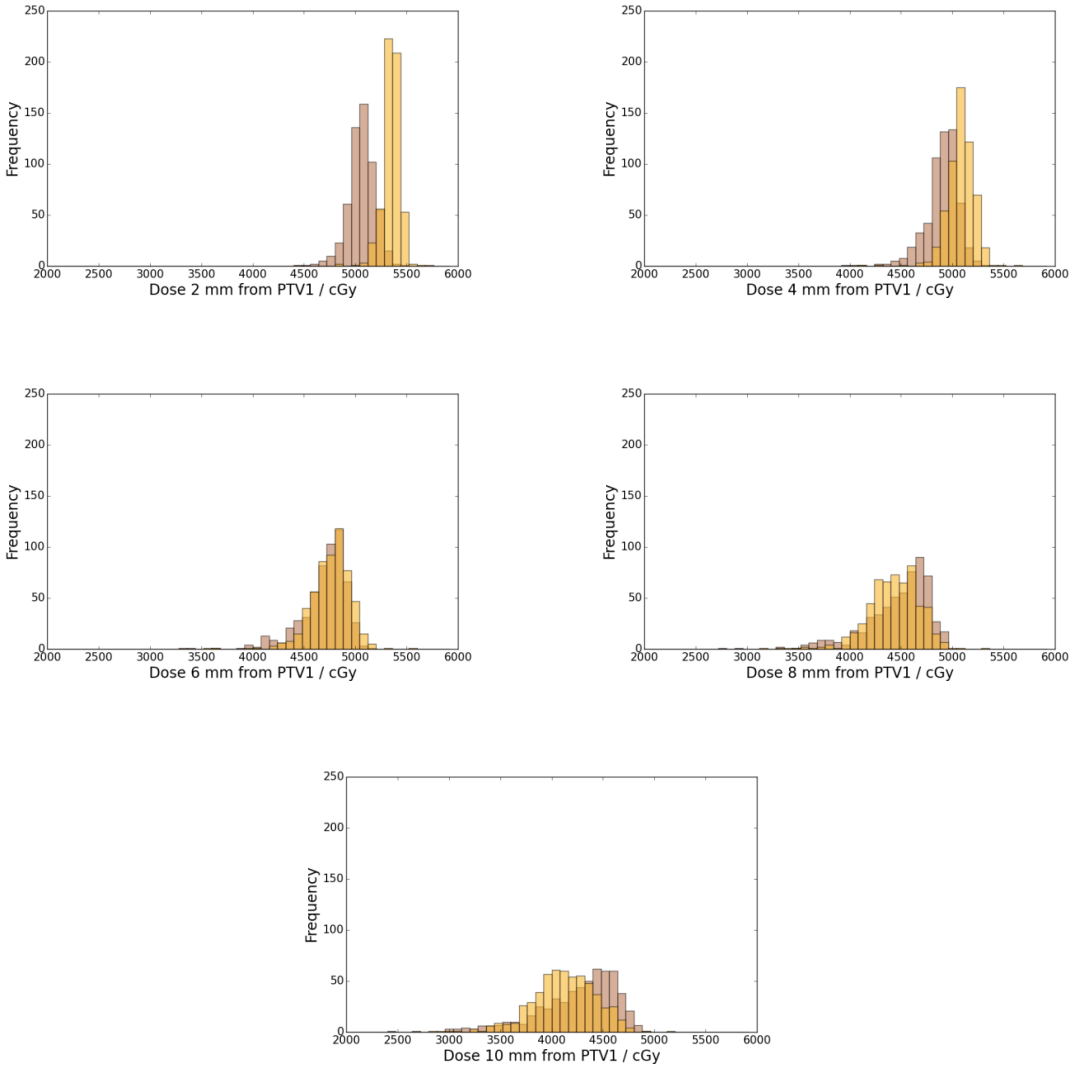

Figure 4.5: Histograms of the 562 bladder and rectum $D_{0.01\,cc}$ at 2 mm, 4 mm, 6 mm, 8 mm and 10 mm from PTV1. The bladder and rectum histograms are shaded yellow and brown respectively.

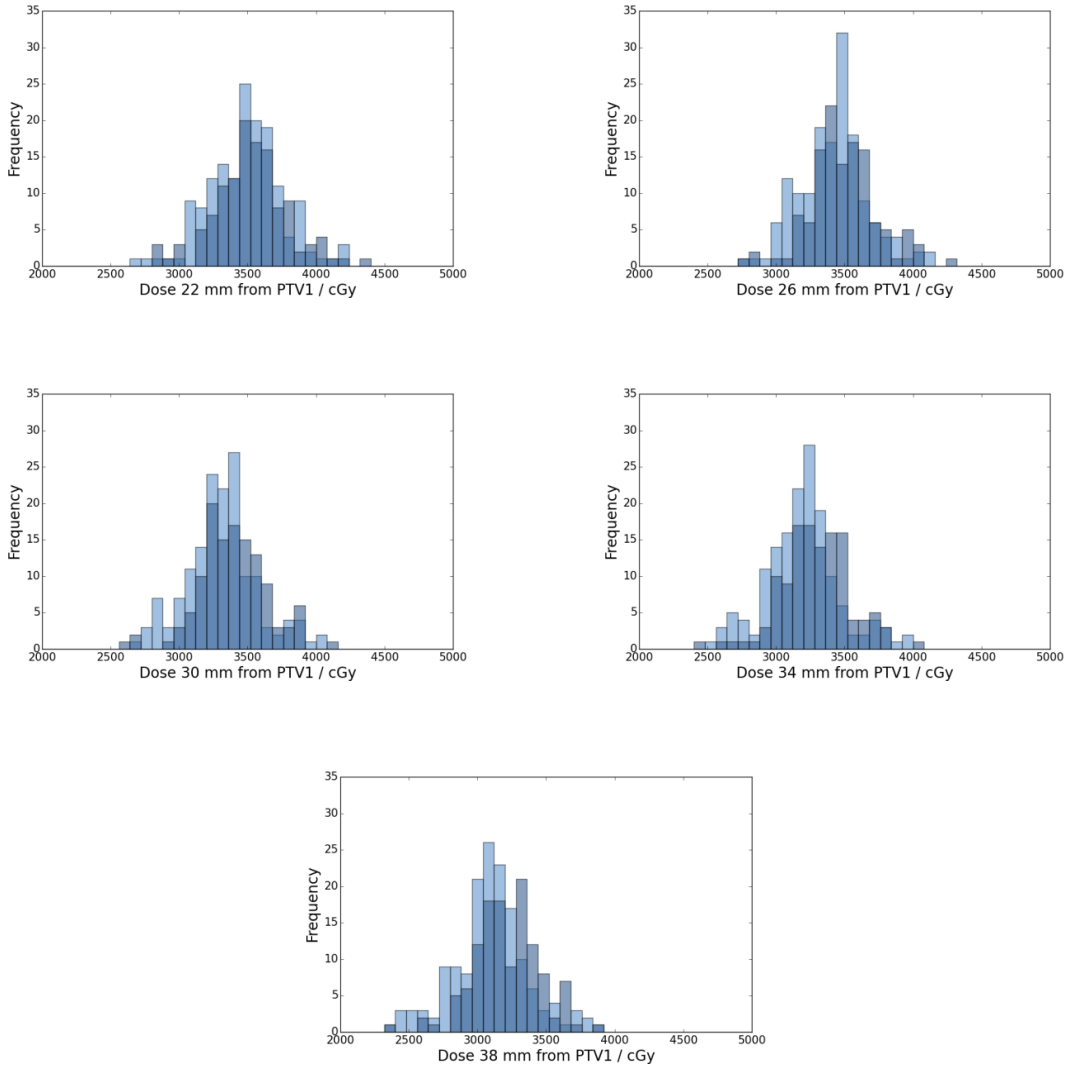

Figure 4.6: Histograms of the 562 Femoral Head $D_{0.01\,cc}$ at 22 mm, 26 mm, 30 mm, 34 mm and 38 mm from PTV1. The right and left femoral head are shaded blue and light blue respectively.

Table 4.1: Skewness calculations performed to benchmark $D_{0.01\,cc}$ distributions for the 562 treatment plans used to train the KB model of the ideal prostate treatment plan.

	Bladder			Rectum		
Distance from PTV1	**Mean / cGy**	**Std / cGy**	**Skewness**	**Mean / cGy**	**Std / cGy**	**Skewness**
2 mm	5350	83	–1.13	5063	140	0.53
4 mm	5081	131	–1.25	4897	164	–1.35
6 mm	4760	194	–1.09	4694	231	–1.67
8 mm	4431	254	–0.77	4475	316	–1.44
10 mm	4116	312	–0.49	4250	393	–1.15
12 mm	3815	360	–0.33	3985	475	–0.85
14 mm	3525	390	–0.19	3708	549	–0.66
16 mm	3275	412	–0.15	3455	608	–0.50
18 mm	3057	419	–0.12	3230	642	–0.31

	FHL			FHR		
Distance from PTV1	**Mean / cGy**	**Std / cGy**	**Skewness**	**Mean / cGy**	**Std / cGy**	**Skewness**
22 mm	3559	264	0.32	3533	257	0.26
24 mm	3491	269	0.10	3463	253	0.44
26 mm	3447	251	0.12	3421	252	0.07
28 mm	3394	248	0.08	3361	258	–0.22
30 mm	3334	270	–0.40	3302	276	–0.47
32 mm	3277	280	–0.42	3260	280	–0.51
34 mm	3213	287	–0.31	3200	297	–0.67
36 mm	3157	301	–0.51	3151	297	–0.65
38 mm	3104	301	–0.43	3110	293	–0.61
40 mm	3559	264	0.32	3533	257	0.26

Figure 4.5 and the skewness calculations in Table 4.1 show negatively skewed distributions of $D_{0.01\,cc}$ for all distances except 2 mm for the bladder and rectum. As discussed in section 4.3, this negative skew is to some degree indicative of how the 562 treatment plans used to train the KB model were not optimised consistently.

It is important to note here that a treatment plan that sits at the 5th percentile at some radial distance from PTV1 does not necessarily sit at the 5th percentile at all radial distances. Care therefore has to be taken with drawing conclusions from the variations demonstrated in Figure 4.5 and Table 4.1. These results only signify that optimisation between patients is inconsistent and does not necessarily mean that any individual treatment plan was uniformly dosimetrically inferior to another.

The skewness analyses presented in Table 4.1 show that generally the PTV1-Bladder distributions exhibit less negative skew than those for PTV1-Rectum. This implies that although both the bladder and rectum doses are not consistently optimised amongst the 562 patients, the rectum shows a greater degree of optimisation variance. At The Christie, prostate radiotherapy patients follow a strict bladder emptying protocol where they void their bladders immediately prior to the acquisition of the treatment planning CT scan and each treatment fraction. This has proved a reasonably reliable means of providing a reproducible bladder size and shape for individual patients[**]. The contents of the rectum, however, are less strictly controlled. Patients are given simple dietary advice prior to the treatment planning CT scan and are asked to follow this throughout treatment. Yahya *et al.* (2013) report that offering such dietary advice is not the optimal approach to managing rectal contents and advocate daily micro enemas instead. Although it is not the purpose of this work to investigate this, it is possible that the more consistent optimisation of the bladder compared to the rectum suggested by the skewed distributions of $D_{0.01\,cc}$ is a result of the way in which the contents of each OAR is managed pre- and during treatment.

The histograms plotted in Figure 4.6 show that over the 562 patients there is no

[**] Although the bladder emptying protocol used at The Christie reproduces bladder size and shape, it is acknowledged here that it is common at many other institutions for a bladder filling protocol to be followed to reduce bladder and (generally) bowel dose.

major difference between the doses to either femoral head. From Table 4.1, the Femoral head $D_{0.01\,cc}$ distributions are less skewed than the bladder and rectum distributions and show a mix of both positive and negative values. Indeed, chi-squared analyses of the femoral head $D_{0.01\,cc}$ distributions indicated that they were reasonably well approximated as normal distributions.

Since the position of the femoral heads (several centimetres from the high dose PTVs) naturally limits the dose they receive, optimisation of clinical treatment plans at The Christie does not require any strict femoral head dose criteria to be met[††]. In section 4.3, it was noted that negatively skewed distributions of $D_{0.01\,cc}$ would be expected if treatment plans were inconsistently optimised because the upper limit is fixed by the minimum dose criteria detailed in Table 3.2 and by the finite target dose. Whilst this logic holds for the bladder and rectum, which sit close to and within the high dose target region, it does not hold for structures in the low dose bath, where typically doses are significantly lower that the prescription and where there are not necessarily any maximum dose criteria to be met.

After running the KB optimisation class-solutions on the 11 test patients, dosimetry data were extracted from Pinnacle[3] and Table 4.2 presents the differences between the KB treatment plans and the original treatment plans that were generated manually and delivered clinically. Paired Student's *t*-tests[‡‡] were used to evaluate the statistical significance of the differences in the reported dose statistics. In Table 4.2, results presented in bold were found to be statistically significant using a threshold for significance of $p < 0.05$.

For the average-KB class-solution, the differences in reported target dose statistics are all within ±1.6 % and mean differences are all within ±0.8 %. PTV1 minimum $D_{1\,cc}$ and $D_{99\,\%}$ were the only target dose statistic differences that were

[††] At the time of writing, the on-going PIVOTALBoost clinical trial gives optimal and mandatory femoral head doses of $V_{4000\,cGy} < 5$ % and < 50 % respectively (The Institute of Cancer Research, 2020). All treatment plans used for this work comfortably met this optimal dose tolerance.

[‡‡] Student's t-test is a parametric test that assumes the samples are taken from normally distributed populations, which is not necessarily the case here. However, for sample sizes greater than ~20, the central limit theorem and Slutsky's theorem imply that the paired Student's *t*-test generates reliable results even for non-normally distributed samples.

statistically significant. However, at 0.8 %, these differences are unlikely to be clinically significant.

For the bladder and rectum, mean differences are all within ±0.9 % although the range of differences is much wider that for the targets: ±5.3 % and ±12.4 % for bladder and rectum respectively. The rectum $V_{5280\,cGy}$ and $V_{5700\,cGy}$ were the only statistically significant OAR dose differences but as noted above the 0.9 % and 1.1 % increases are unlikely to be clinically significant. The wide range of dose differences centred close to zero for the bladder and rectum was expected especially for the low dose DVH statistics. In Figure 4.3, the variation between patients increases with increasing distance (and thus decreasing dose) from PTV1, therefore optimising treatment plans to achieve the average dose gradient means OAR doses in some treatment plans increase and some decrease. The dose difference data in Table 4.2 for the average-KB class-solution are all centred within ±0.9 % of zero, which indicates that the average treatment plan quality has been maintained at a clinically significant level and the plans are optimised consistently.

The 25th percentile-KB class-solution produces treatment plans with on average statistically significant lower bladder and rectum dosimetry and comparable target dosimetry compared to the average-KB class-solution and to the clinical treatment plans. The femoral head $D_{1\,cc}$ for the 25th percentile-KB treatment plans increase compared to the average-KB treatment plans. This is most likely the result of the bladder and rectum doses restricting the contribution of target dose that can be delivered from the anterior-posterior components of the VMAT arc. In order to achieve target coverage, an increased contribution of the dose must be delivered laterally through the femoral heads and hence the femoral head doses increase.

For both KB class-solutions, nine of the 11 treatment plans met all of the dose criteria noted in Table 3.2. For the average-KB class-solution, one of the treatment plans failed to meet the minimum $D_{1\,cc}$ for PTV1 and PTV2 (although the average-KB treatment plan dosimetry was superior to the clinical treatment plan, which also did not meet the minimum $D_{1\,cc}$ criteria) and another failed to meet the rectum $V_{5700\,cGy}$ planning aim. For

the 25th percentile-KB class-solution, all OAR dose criteria were met but two treatment plans failed to achieve the required target dosimetry.

Table 4.2: Average, standard deviation, minimum and maximum percentage differences between KB and manual treatment plans.

	Average KB Class-Solution				25th Percentile KB Class-Solution			
Dose Statistic	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**
PTV1 $D_{99\ \%}$	**0.8**	0.4	0.2	1.3	0.3	0.9	–1.5	1.3
PTV1 min $D_{1\ cc}$	**0.8**	0.5	0.1	1.6	0.2	1.1	–2.0	1.4
PTV2 $D_{99\ \%}$	–0.2	0.4	–1.0	0.4	–0.4	0.7	–2.0	0.4
PTV2 min $D_{1\ cc}$	–0.1	0.4	–1.0	0.4	–0.5	0.7	–2.3	0.4
PTV3 $D_{99\ \%}$	0.1	0.5	–0.7	1.2	–0.1	0.7	–1.7	1.1
PTV3 min $D_{1\ cc}$	0.1	0.5	–0.5	1.2	–0.1	0.8	–2.0	1.1
PTV3 max $D_{1\ cc}$	–0.3	0.5	–1.5	0.2	0.0	0.7	–1.1	1.8
PTV3 median	0.1	0.2	–0.1	0.5	0.1	0.3	–0.5	0.5
FHL max $D_{1\ cc}$	**3.2**	4.3	–3.2	12.7	**5.6**	3.7	–2.7	10.0
FHR max $D_{1\ cc}$	**4.8**	4.2	–0.8	11.2	**5.5**	5.0	–3.6	12.2
Rectum $V_{2460\ cGy}$	0.5	4.2	–6.5	9.5	**–4.7**	4.1	–10.3	2.7
Rectum $V_{3240\ cGy}$	–1.7	6.5	–10.8	12.4	**–6.4**	6.9	–15.1	7.4
Rectum $V_{4080\ cGy}$	–1.0	6.0	–10.1	11.8	**–4.2**	6.0	–14.8	8.1
Rectum $V_{4860\ cGy}$	0.6	2.4	–3.3	5.1	–1.1	2.5	–4.7	3.8
Rectum $V_{5280\ cGy}$	**0.9**	1.1	–0.7	2.5	0.2	1.0	–1.2	1.5
Rectum $V_{5700\ cGy}$	**1.1**	1.1	–0.4	3.0	**0.7**	1.0	–0.6	2.7
Rectum $V_{6000\ cGy}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bladder $V_{4080\ cGy}$	–0.9	2.7	–4.5	5.3	**–2.2**	3.0	–5.4	5.3
Bladder $V_{4860\ cGy}$	–0.4	2.2	–2.8	5.2	–1.5	2.6	–3.8	5.0
Bladder $V_{6000\ cGy}$	–0.2	1.6	–4.6	1.6	0.5	0.9	–1.3	1.6

Figure 4.7 shows a similar plot to Figure 4.4 where the red transparent data points represent the 562 PTV1-OAR gradients from the training data, the solid red data points represent the clinical treatment plans for the 11 test cases and the yellow and blue data points represent the treatment plans generated with the average-KB and 25th percentile-KB class-solutions respectively. The 11 clinical tests cases appear to be consistent with the training data population and both KB class-solutions generate data points that are closer together and offset in the negative-negative direction, which indicates generally steeper and more consistent dose gradients than the clinical treatment plans. As expected the 25th percentile-KB treatment plans generally have steeper dose gradients than the average-KB treatment plans.

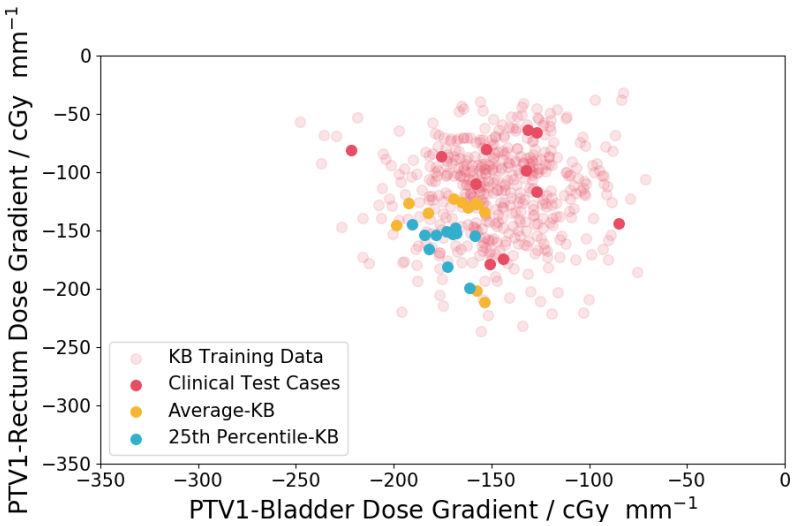

Figure 4.7: Individual PTV1-Rectum and PTV1-Bladder dose gradients plotted against each other for the 562 KB training patients, 11 clinical test cases, average-KB and 25th percentile-KB treatment plans.

When the average-KB treatment plans were reviewed visually, of the nine that met all of the minimum dose criteria, eight were considered clinically acceptable in terms of the dose distribution. In the treatment plan that did not pass the visual inspection, the 5460 cGy isodose line did not conform well enough to PTV2. Figure 4.8 shows a sagittal slice of the PTVs for this patient where an orange 5460 cGy 'hotspot' is visible at the superior extent of PTV1. It was also noted that the 5700 cGy coverage of PTV3 was

generous anteriorly across all average-KB treatment plans. Although the 5700 cGy isodose line did not extend outside PTV2, it generally sat close to the inside surface of PTV2, which is apparent in Figure 4.8 and Figure 4.9, such that it was considered the extreme of what would be clinically acceptable.

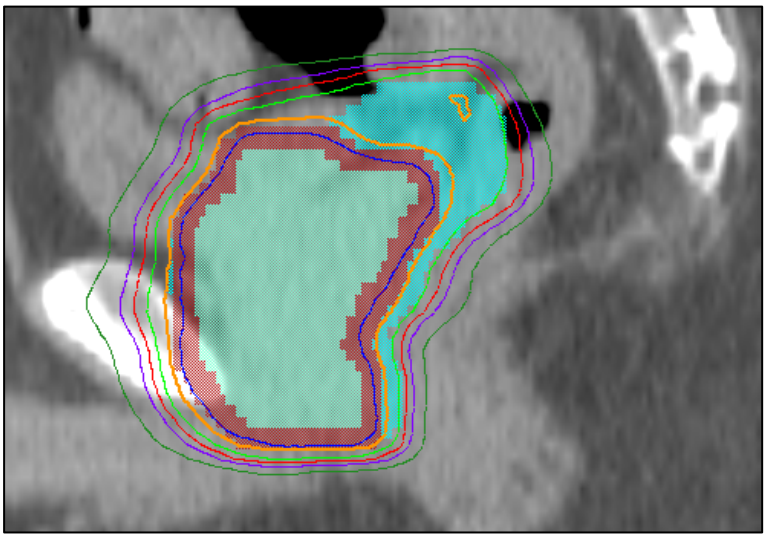

Figure 4.8: A sagittal slice of the average-KB treatment plan that produced 5460 cGy (orange isodose) hotspots in PTV1.

All of the treatment plans generated with the 25th percentile-KB class-solution that met the minimum dose criteria (nine out of 11) were considered clinically acceptable. PTV coverage for treatment plans generated with the 25th percentile-KB class-solution was comparable to that generated with the average-KB class-solution with one exception. This treatment plan was for the same patient for which the average-KB treatment plan also failed to meet the PTV1 and PTV2 minimum $D_{1\,cc}$ dose criteria. As noted above, the average-KB treatment plan for this patient was dosimetrically superior to the clinical treatment plan and so could therefore be considered clinically acceptable. However, when the rectum and bladder were optimised to the 25th percentile of the dose gradients shown in Figure 4.3, PTV coverage became unacceptably compromised.

As expected the PTV1-Rectum and PTV1-Bladder dose gradients were visually steeper in the treatment plans generated with the 25th percentile-KB class-solution but it

was also apparent that this was at the expense of Femoral Head dose. Figure 4.9 shows the same CT slice for one patient where on the left is the average-KB treatment plan and on the right is the 25th percentile-KB treatment plan. The outermost green isodose line, which represents 3155 cGy, covers a smaller proportion of the Rectum and Bladder in the 25th percentile-KB treatment plan but extends further laterally to cover more of the Femoral Heads.

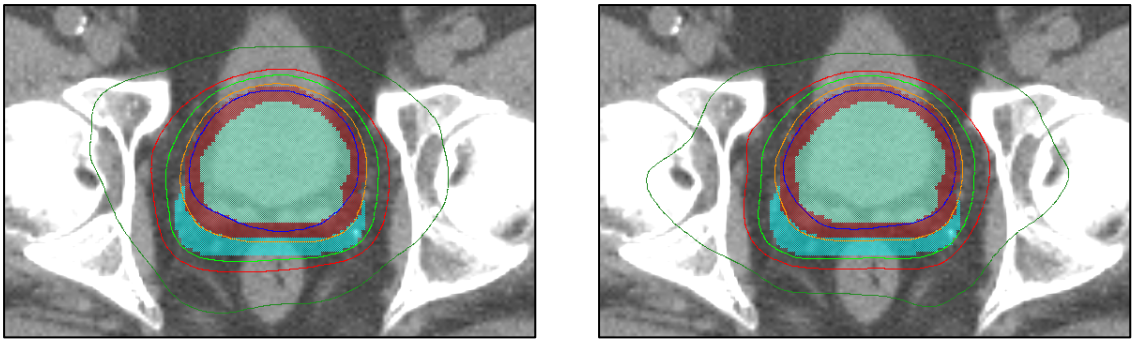

Figure 4.9: A slice of a treatment plans generated with the average-KB class-solution (left) and the 25th percentile-KB class-solution (right).

4.5 Discussion

In this chapter a novel approach to KB treatment planning has been proposed, developed and tested. Unlike traditional KBs that consist of relatively small databases of previously optimised treatment plans the KB developed here consisted of a generalised model of the ideal prostate radiotherapy treatment plan. The model was trained using PTV1-OAR dose gradient data extracted from a large cohort of clinical treatment plans and this was used to create two KB optimisation class-solutions with the aims of maintaining average treatment plan quality (average-KB) and improving it (25th percentile-KB).

It was shown that, amongst the training cohort, optimisation of the high dose fall-off from the PTV1 surface into OARs is extremely inconsistent. Specifically optimising the dose gradient within the OAR has been shown to generate treatment plans of at least comparable quality to manually generated clinical treatment plans but with more

consistently optimised OAR doses.

A major advantage of this novel KB treatment planning approach over the more traditional approaches described in the literature is that it does not involve the difficult process of predicting suitable OAR DVH parameters for optimisation. Instead, the same optimisation parameters can be used for every patient (i.e. the same KB class-solutions) irrespective of OAR size and shape and these optimisation parameters can be carefully selected and refined based on vast numbers of previously optimised clinical treatment plans. Where many KB treatment planning solutions find a role in the clinic as a quality assurance tool for manually generated treatment plans, it is likely that the method developed and tested here has a substantial role to play in the prospective development of new treatment plans.

The test sample for the KB treatment planning approach developed here was limited to 11 cases. Although this is a test sample approximately consistent with similar studies published in the literature, the reliability of conclusions drawn about the success of the KB treatment planning technique are limited. Further work is required to test the KB treatment planning technique more thoroughly on a much larger test patient cohort. In additional, since the aim of the KB class-solutions is to generate clinically acceptable treatment plans without treatment planner intervention, testing optimisation using auto-contours would also be an interesting area for further work because this could provide a fully-automated KB treatment planning workflow.

4.6 Conclusion

The work presented in this chapter has demonstrated the utility of the KB treatment planning model illustrated in Figure 4.1. A large cohort of prostate VMAT radiotherapy treatment plans, generated based on over a decade of clinical experience, was used to train the model and generate two optimisation class-solutions – the first aimed to achieve the average PTV1-OAR dose gradients and the second aimed to achieve the 25th percentile of the PTV1-OAR dose gradients for all patients. These were then tested on 11 new patients where optimisation was performed based on manually drawn target and

OAR contours. When both sets of KB treatment plans were reviewed in terms of the DVH requirements set out in Table 3.2 and checked visually by an experience MPE, clinically acceptable treatment plans were generated for nine of the 11 test patients.

4.7 References

Bzdusek, K., Bystrov, D., Pekar, V., Peters, J., Schadewaldt, N., Schulz, H. & Vik, T. (2012). *Smart Probabilistic Contouring Engine (SPICE)*. The Netherlands: Koninklijke Philips Electronics.

Chanyavanich, V., Das S., Lee, W. & Lo, J. (2011). 'Knowledge-based IMRT treatment planning for prostate cancer', *Medical Physics*, 38(5), pp. 2515-2522.

Dearnaley, D., Syndikus, I., Mossop, H., Vincent, K., Birtle, A., Bloomfield, D., Graham, J., Kirkbride, P., Logue, J., Malik, Z., Money-Kyrle, J., M O'Sullivan, J., Panades, M., Parker, C., Patterson, H., Scrase, C., Staffurth, J., Stockdale, A., Tremlett, T., Bidmead, M., Mayles, M., Naismith, N., South, C., Gao, A., Cruickshank, C., Hassan, S., Pugh, J., Griffin, C. &Hall, E. (2016). 'Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 17(8), pp. 1047-1060.

Djajaputra, D., Wu, Q., Wu, Y. & Mohan, R. (2003). 'Algorithm and performance of a clinical IMRT beam-angle optimisation system', *Physics Medicine Biology*, 48(19), pp. 3191-3212.

Good, D., Lo, J., Lee, R., Wu, J., Yin, F. & Das, S. (2013). 'A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning', *International Journal of Radiation Oncology Biology Physics*, 87(1), pp. 176-181.

Janssen, T., Kusters, M., Wang, Y., Wortel, G., Monshouwer, R., Damen, E. & Petit, S. (In press). 'Independent knowledge-based treatment planning QA to audit Pinnacle autoplanning', *Radiotherapy & Oncology*.

Nwankwo, O., Mekdash, H., Sihono, D., Wenz, F. & Glatting, G. (2015). 'Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning', *Radiation Oncology*, 10(111), pp. 1-5.

Petit, S., Wu, B., Kazhdan, M., Dekker, A., Simari, P., Kumar, R., Taylor. R., Herman, J. & McNutt, T. (2012). 'Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma', *Radiotherapy Oncology*, 102(1), pp. 38-44.

Powis, R., Bird, A., Brennan, M., Hinks, S., Newman, H., Reed, K., Sage, J. & Webster, G. (2017). 'Clinical implementation of a knowledge based planning tool for prostate VMAT', *Radiation Oncology*, 12(81), pp. 1-8.

Wang, Y., Heijmen, B. & Petit, S. (2017). 'Prospective clinical validation of independent DVH prediction for plan QA in automatic treatment planning for prostate cancer patients', *Radiotherapy & Oncology*, 125, pp. 500-506.

Wilkins, A., Mossop, H., Syndikus, I., Khoo, K., Bloomfield, D., Parker, C., Logue, J., Scrase, C., Patterson, H., Birtle, A., Staffurth, J., Malik, Z., Panades, M., Eswar, C., Graham, J., Russell, M., Kirkbride, P., O'Sullivan, J., Gao, A., Cruickshank, C., Griffin, C., Dearnaley, D. & Hall, E. (2015). 'Hypofractionated radiotherapy versus conventionally fractionated radiotherapy for patients with intermediate-risk localised prostate cancer: 2-year patient-reported outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 16(16), pp. 1605-1616.

Wu, B., Ricchetti, F., Sanguineti, G., Kazhdan, M., Simari, P., Chuang, M., Taylor, R., Jacques, R. & McNutt, T. (2009). 'Patient geometry-driven information retrieval for IMRT treatment plan quality control', *Medical Physics*, 36(12), pp. 5497-5505.

Yahya, S., Zarkar, A., Southgate, E., Nightingale, P. & Webster, G. (2013). 'Which bowel

preparation is best? Comparison of a high-fibre diet leaflet, daily microenema and no preparation in prostate cancer patients treated with radical radiotherapy to assess the effect on planned target volume shifts due to rectal distension', *British Journal of Radiology*, 86(1031), pp. 1-9.

Yang, Y., Ford, E., Wu, B., Pinkawa, M., van Triest, B., Campbell, P., Song, D. & McNutt, T. (2013). 'A overlap-volume-histogram based method for rectal dose prediction and automated treatment planning n the external beam prostate radiotherapy following hydrogel injection', *Medical Physics*, 40(1), pp. 1-10.

Yuan, L., Ge, Y., Lee, W., Yin, F., Kirkpatrick, J. & Wu, J. (2012). 'Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans', *Medical Physics*, 39(11), pp. 6868-6878.

Zhu, X., Ge, Y., Li, T., Thongphiew, D., Yen, F. & Wu, Q. (2011). 'A planning quality evaluation tool for prostate adaptive IMRT based on machine learning', *Medical Physics*, 38(2), pp. 719-726.

4.8 Supplementary Material

The tables below present the optimisation objectives used in Pinnacle[3] for the average-KB and 25th percentile-KB class-solutions. The PTV objectives were the same for both in both cases (top table) and OARExp1, 2, 3… refer to successive 2 mm OAR segments.

ROI	Type	Dose / cGy	Volume / %	Weight
PTV3	Min Dose	5950	–	90
PTV3	Max Dose	6100	–	70
PTV2-PTV3	Min Dose	5600	–	100
PTV2-PTV3	Max Dose	5950	–	55
PTV2-PTV3	Max DVH	5900	5	70
PTV1-PTV2	Min Dose	5100	–	95
PTV1-PTV2	Max Dose	5500	–	70
PTV1-PTV2	Max DVH	5450	10	70
Rind_1	Max Dose	5600	–	30
Rind_2	Max Dose	4900	–	30
Rind_3	Max Dose	5600	–	30
Background	Max Dose	4000	–	30

		Dose / cGy		
ROI	**Type**	**Average-KB**	**25th percentile-KB**	**Weight**
RectumExp1	Max Dose	5065	4995	10
RectumExp2	Max Dose	4900	4828	10
RectumExp3	Max Dose	4700	4601	10
RectumExp4	Max Dose	4480	4336	10
RectumExp5	Max Dose	4250	4030	10
RectumExp6	Max Dose	4000	3717	10
RectumExp7	Max Dose	3715	3384	10
RectumExp8	Max Dose	3470	3093	10
RectumExp9	Max Dose	3210	2794	10
BladderExp1	Max Dose	5350	5313	10
BladderExp2	Max Dose	5080	5018	10
BladderExp3	Max Dose	4770	4665	10
BladderExp4	Max Dose	4450	4300	10
BladderExp5	Max Dose	4130	3948	10
BladderExp6	Max Dose	3820	3603	10
BladderExp7	Max Dose	3520	3285	10
BladderExp8	Max Dose	3260	3008	10
BladderExp9	Max Dose	3000	2784	10
FHL	Max EUD	2250	2250	10
FHR	Max EUD	2250	2250	10

Chapter 5

5 Fully-Automated Knowledge-Based Treatment Planning

Joseph Wood[1, 2], Marianne Aznar[2], Philip Whitehurst[1, 2]

[1]The Christie NHS Foundation Trust
[2]The University of Manchester

5.1 Abstract

The purpose of the work presented in this chapter was to unite the ideas of hybrid-planning developed in Chapter 3 with the KB class-solutions developed in Chapter 4 through large scale testing. 131 clinical treatment plans formed a test patient cohort. When optimisation was driven purely by manual-contours, the average-KB class-solution was shown to maintain average treatment plan quality in terms of target coverage and OAR sparing whilst also providing a small reduction in treatment plan complexity. The 25th percentile-KB class-solution improved treatment plan quality with comparable target coverage and lower average OAR doses but this came at the expense of a small

increased treatment plan complexity. For 120 of 131 test cases at least one of the KB class-solutions generated a clinically acceptable treatment plan. When optimisation was driven by manual targets and auto-OARs (the so-called hybrid-plan developed in Chapter 3) clinically acceptable treatment plans were generated for 102 of 131 test cases. Hybrid-planning using the KB class-solutions developed in Chapter 4 can offer significant efficiency savings for busy clinical departments. These efficiency savings could be exploited to enable ART techniques and allow patients earlier access to their cancer treatments.

5.2 Introduction

In many radiotherapy centres, treatment plans are generated manually by expert treatment planners. Optimisation of a treatment dose distribution is an iterative process where the treatment planner specifies appropriate objectives for an optimisation algorithm. Multiple cycles of the optimisation process are typically run until the expert treatment planner is satisfied that the treatment plan is optimal. It is reported in the literature, however, that this manual treatment planning approach is subjective and time consuming (Chanyavanich *et al.*, 2011; Djajaputra *et al.*, 2003) and the results in Chapter 4, which showed large variations in PTV1-OAR dose gradients for clinical treatment plans generated at The Christie, support this claim.

Knowledge-based (KB) treatment planning automatically incorporates prior experience into the treatment plan optimisation process (Nwankwo *et al.*, 2015). This can make treatment planning more consistent within a radiotherapy centre and can generate workflow efficiencies by removing some of the time consuming human interventions.

Nwankwo *et al.* (2014) write that prostate radiotherapy treatment planning is the most common site for KB treatment planning research. However, there is a tendency in the literature for KB databases to consist of a relatively small number of previously optimised treatment plans. When a new treatment plan is to be optimised, it is characterised in some way and the plan with the most similar characteristics is selected

from the database and the same optimisation parameters are used to optimise the new treatment plan (Chanyavanich *et al.*, 2011; Wu *et al.*, 2009; Janssen *et al.*, In press; Wang, Heijmen & Petit, 2017; Petit *et al.*, 2017; Good *et al.*, 2013; Yang *et al.*, 2013; Zhu *et al.*, 2011; Yuan *et al.*, 2012).

There are two major limitations to these approaches. First, is that the need for computational speed restricts the number of previous treatment plans that can be included in the KB database, which limits the likelihood of finding a suitable match for the new treatment plan. Patient cohorts on which approaches to KB treatment planning are tested and reported also tend to be relatively small – typically around 10 patients – and this limits the strength of conclusions that can be drawn from the results (Zhu *et al.*, 2011; Yuan *et al.*, 2012).

Second, is that optimisation of the new treatment plan based only on a single previous treatment plan means that any ways in which the previous treatment plan was suboptimal are propagated to new treatment plans, which will also then be suboptimal. It is perhaps for this reason that where it is reported that KB treatment planning is used in the clinical, it is often only as a tool to quality assurance manual treatment planning and not a primary method of treatment plan generation (Powis *et al.*, 2017).

A novel approach to automated KB treatment planning was developed in Chapter 4 of this thesis. The KB consisted of model of the ideal prostate radiotherapy treatment plan, which was trained using a cohort of 562 clinical treatment plans. The model then informed two optimisation class-solutions that controlled target coverage and PTV1-OAR dose fall-off. The aim of the work presented in this chapter is to test this novel KB approach thoroughly on a large cohort of prostate cancer patients.

Previous work in this thesis has demonstrated that OAR auto-contours with known geometric inaccuracies can be used to drive treatment plan optimisation with minimal impact on the resulting dosimetry reported to manual-contours – the so-called 'hybrid-plan' introduced in Chapter 3. The work presented in this chapter also aims to test driving the novel KB treatment planning class-solutions developed in Chapter 4 with OAR auto-contours – thus (after clinical oncologist outlining of prostate and SV)

providing a fully-automated KB treatment planning workflow.

The final aim of the work presented here is to provide a rigorous analysis of the quality and complexity of treatment plans generated with the KB class-solutions and evaluate how they compare with their manually generated clinical counterparts.

5.3 Method

131 prostate VMAT treatment plans generated at The Christie during the first quarter of 2019 were used to test a fully-automated KB treatment planning workflow. All of these clinical treatment plans were generated using one of three beam models in Pinnacle[3] corresponding to the machine on which the patients were treated: Elekta Agility, Elekta Synergy or Varian Novalis.

As in Chapter 4, the manually drawn OARs were segmented based on 2 mm isotropic expansions of PTV1 and average PTV1-Bladder and PTV1-Rectum dose gradients were calculated based on the $D_{0.01\,cc}$ of each segment. Histograms of $D_{0.01\,cc}$ at each radial distance from PTV1 were generated and skewness was calculated for each histogram using equation 4.1. A plot of the PTV1-Rectum against PTV1-Bladder dose gradients was generated from all 131 patients to assess correlation between attempts to spare the bladder or rectum. The results of these analyses were compared to the training data analyses (see Chapter 4) to ensure the test data belonged to a population consistent with the training data.

All of the 131 patient CT scans were then auto-segmented within Pinnacle[3] using SPICE (Qazi *et al.*, 2011). It was demonstrated in Chapters 2 and 3 that the observed differences in geometric accuracy between SPICE, Mirada RTx and ADMIRE were generally not statistically significant. In addition, differences in the utility of SPICE, Mirada RTx and ADMIRE auto-contours for hybrid-planning was not statistically significantly different. Therefore, for ease of integration into the treatment plan generation workflow, only SPICE was used for the work presented in this chapter. The SPICE OAR auto-contours were also segmented based on the 2 mm expansions of PTV1 noted above and these were used to generate KB hybrid-plans using the average- and

25th percentile-KB class-solutions described in Chapter 4.

KB treatment plans were also generated using the segmented OAR manual-contours for comparison with the KB hybrid-plans. Each new treatment plan was generated with the same TPS beam model as the original clinical treatment plan to ensure fairness in the comparisons. Clinical treatment plan quality is not considered to be different between the three machine types and so the KB class-solutions are universally applicable. For clarity, the flow diagram in Figure 5.1 shows the combinations of contours and class-solutions used to generate the treatment plans and the nomenclature used throughout this chapter.

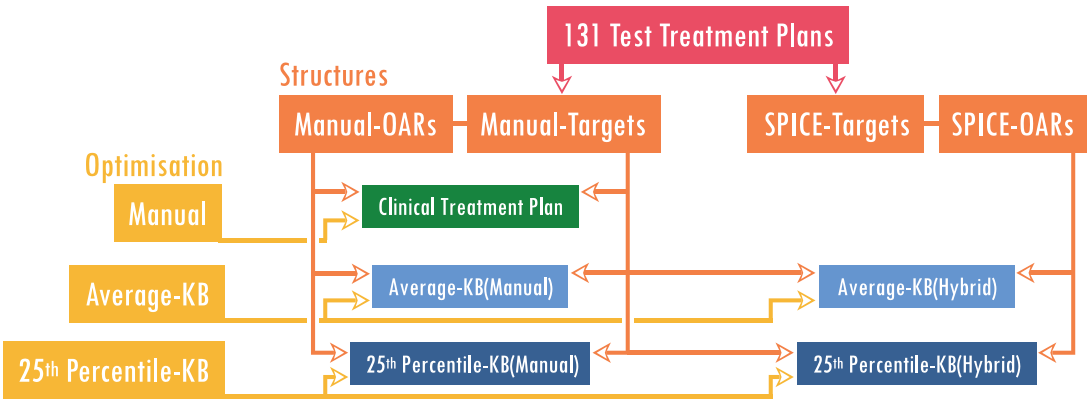

Figure 5.1: Flow diagram showing the combinations of structures and KB class-solutions and treatment plan nomenclature.

Once all of the KB treatment plans were generated, the dosimetry data reported to the manual-contours were extracted from Pinnacle[3]. These were used to calculate the average PTV1-OAR dose gradients, to plot $D_{0.01\,cc}$ histograms for which skewness was calculated using equation 4.1, and to generate plots of PTV1-Rectum against PTV1-Bladder dose gradients. The standard prostate radiotherapy dose statistics presented in Table 3.2 were also calculated for each treatment plan (again, reported to the manual-contours) and these were used to determine how many hybrid-plans met the required minimum dose criteria and to compare the performance of the KB class-solutions for the manual- and hybrid-plans.

A comparison of treatment plan complexity was also performed using the aperture-based metrics discussed in section 1.4. Modulation complexity score (MCS), plan aperture (PA), plan irregularity (PI) and plan modulation (PM) were calculated for each of the 562 treatment plans in the training dataset described in Chapter 4, the 131 clinical test cases and all of the treatment plans generated with the KB class-solutions. Analyses were grouped based on machine type because MLC thickness and machine parameters in the Pinnacle^3 beam models are different between the machines and this can bias aperture-based metrics.

5.4 Results

Figure 5.2 shows plots of mean $D_{0.01\,cc}$ as functions of radial distance from the PTV1 surface within the bladder and rectum for the 131 clinical treatment plans used as a test patient cohort. 25th and 75th percentiles are also plotted as thinner lines and the 5th and 95th percentiles are plotted as dashed lines. Similar plots for the average-KB(Manual), 25th percentile-KB(Manual), average-KB(Hybrid) and 25th percentile-KB(Hybrid) treatment plans are shown in Figure 5.3, Figure 5.4, Figure 5.5 and Figure 5.6 respectively.

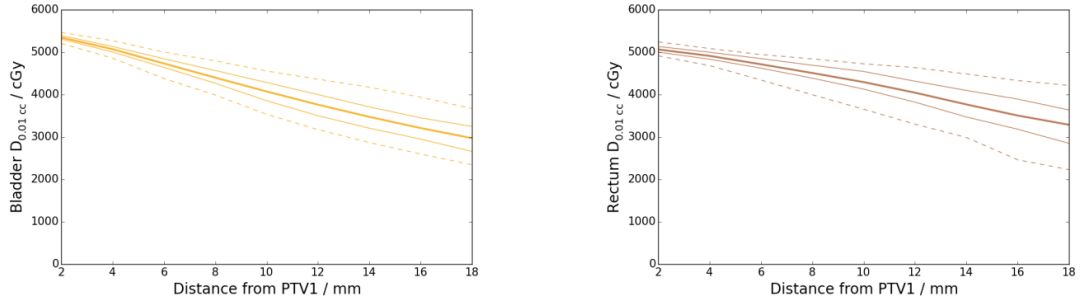

Figure 5.2: Mean dose gradients for the 131 clinical treatment plans used as the test patient cohort within the bladder and rectum. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

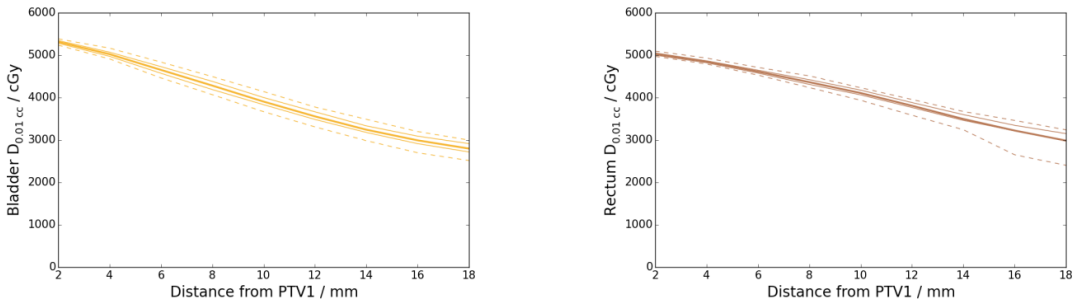

Figure 5.3: Mean dose gradients for the 131 average-KB(Manual) treatment plans within the bladder and rectum. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

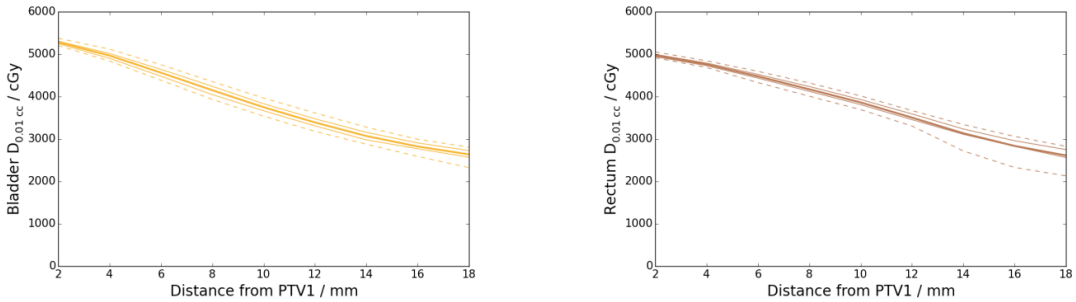

Figure 5.4: Mean dose gradients for the 131 25th percentile-KB(Manual) treatment plans within the bladder and rectum. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

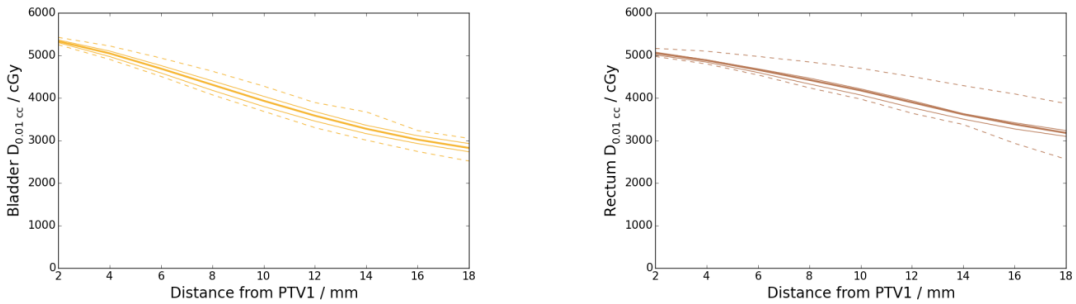

Figure 5.5: Mean dose gradients for the 131 average-KB(Hybrid) treatment plans within the bladder and rectum. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

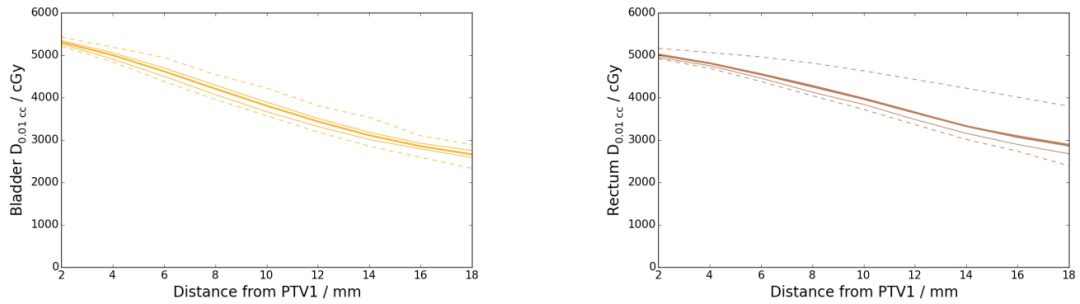

Figure 5.6: Mean dose gradients for the 131 25th percentile-KB(Hybrid) treatment plans within the bladder and rectum. 25th and 75th percentiles are plotted as the thinner lines and 5th and 95th percentiles are plotted as the dashed lines.

It is immediately apparent from Figure 5.2-Figure 5.4 that the variations in dose gradients are significantly reduced when the treatment plans are optimised using either of the KB class-solutions. For the average- and 25th percentile-KB(Hybrid) treatment plans the rectum 95th percentile line sits further from the mean than the corresponding lines for the average- and 25th percentile-KB(Manual) treatment plans. This was caused by gross inaccuracies in rectum auto-contouring with SPICE such that the OAR segments were not appropriate to optimise the actual rectum dose and demonstrates a limitation of this hybrid KB treatment planning workflow using SPICE (and potentially atlas-based auto-segmentation in general).

Table 5.1: Average PTV1-Bladder and PTV1-Rectum dose gradients for the test patient cohort.

	Average Dose Gradient / cGy mm^{-1}	
Treatment Plan	**Bladder**	**Rectum**
Clinical	–152	–115
Average-KB(Manual)	–164	–132
25th percentile-KB(Manual)	–173	–155
Average-KB(Hybrid)	–164	–122
25th percentile-KB(Hybrid)	–174	–140

Straight lines were fit to the average data in the plots above to calculate average PTV1-Bladder and PTV1-Rectum dose gradients, which are presented in Table 5.1. The data for the clinical treatment plans in Table 5.1 were compared to equivalent values for

the training patient cohort (described in Chapter 4) using non-paired Student's *t*-tests and differences in the dose gradients between the cohorts were found not to be statistically significant ($p = 0.13$ and 0.52 for the PTV1-Bladder and PTV1-Rectum respectively). This suggests that the 131 clinical treatment plans in the test cohort are selected from a population consistent with the 562 treatment plans used as training data in Chapter 4.

Both average- and 25th percentile-KB class-solutions for manual- and hybrid-plans generate steeper average PTV1-Bladder and PTV1-Rectum dose gradients than the average of the clinical treatment plans. As expected the 25th percentile-KB class-solution dose gradients are steeper than those for the average-KB class-solution. Since the samples consisted of the same patients but had unequal variances, Welch's *t*-tests were used to assess these differences. For conciseness not all of the results of the statistical analyses are included here as all differences were found to be strongly statistically significant ($p < 0.01$).

A further set of Welch's *t*-tests was performed to evaluate differences between the KB(Manual) and KB(Hybrid) treatment plans generated with the same KB class-solution. This showed that for the Bladder differences in the average dose gradients were not statistically significant (p =0.81 and 0.80 for the average and 25th percentile-KB class-solutions respectively) but for the Rectum differences were statistically significant ($p < 0.01$ for both KB class-solutions). It has been acknowledged above that gross rectum auto-contouring inaccuracies caused some hybrid-plans to be insufficiently optimised. Inclusion of these sub-optimal plans in the statistical analyses, as was done here, will bias the *t*-test towards suggesting a statistically significant difference. Therefore care has to be taken with interpreting this result.

For each treatment plan the PTV1-Bladder and PTV1-Rectum were plotted against each other to assess correlation between the attempts to spare the OARs and these plots are presented in Figure 5.7 and Figure 5.8 for the KB(Manual) and KB(Hybrid) treatment plans respectively. It is clear from these plots that there is no correlation between attempts to spare the bladder or rectum but that consistency of optimisation is

improved dramatically when treatment plans are generated with either KB class-solution.

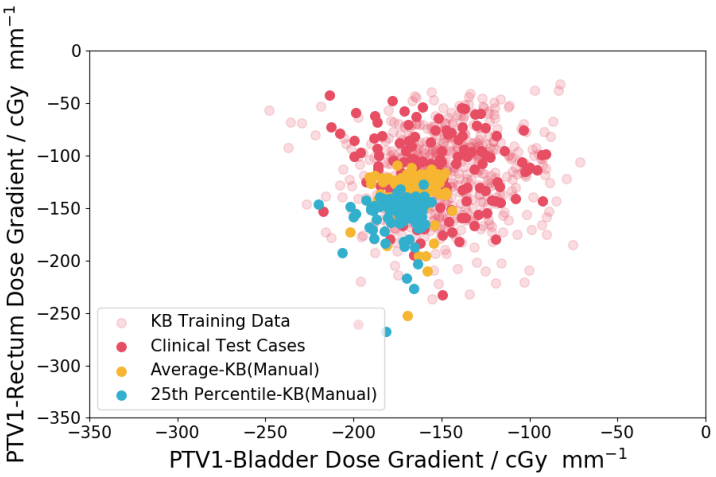

Figure 5.7: Individual PTV1-Rectum and PTV1-Bladder dose gradients plotted against each other for the KB training data (562 patients from Chapter 4) and the 131 clinical test cases, average-KB(Manual) and 25th percentile-KB(Manual).

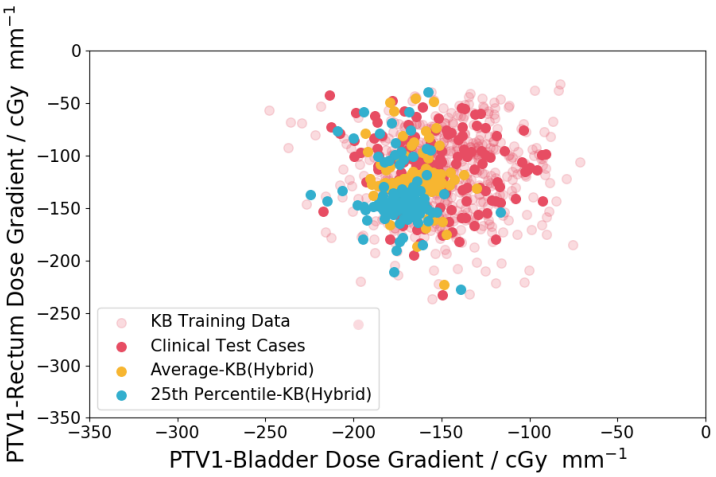

Figure 5.8: Individual PTV1-Rectum and PTV1-Bladder dose gradients plotted against each other for the KB training data (562 patients from Chapter 4) and the 131 clinical test cases, average-KB(Hybrid) and 25th percentile-KB(Hybrid).

Table 5.2 and Table 5.3 present the results of skewness calculations that were performed on the bladder and rectum segmentation data. These results can be compared with those presented in Table 4.1 for the KB training data. Interestingly, it

was noted in Chapter 4 that the PTV1-Rectum distributions in the training dataset were all more negatively skewed than those for PTV1-Bladder. For the test dataset, this difference was not observed.

All of the clinical treatment plans from the training and test cohorts were generated and quality assured following the same protocol – the only difference is that the training data were generated in 2017 and 2018 and the test data in 2019. Identifying the cause of the disappearance of the differences in skewness was beyond the scope of the work presented here and given the consistency of the treatment plans in the training and test datasets, it is reasonable to conclude that it does not present a significant dosimetric or clinical implication for the quality of the treatment plans. It does, however, highlight the susceptibility of manual treatment plan optimisation to subtle changes in treatment planning technique over a prolonged period, which can be unnoticeable even to the most experienced treatment planners.

For the average-KB(Manual) treatment plans, the distributions of $D_{0.01\,cc}$ close to PTV1 are less negatively skewed than corresponding distributions for the clinical treatment plans. However, at larger distances (greater than 10 mm) they become more negatively skewed. This is also apparent in Figure 5.3, where the dashed 5th percentile line shows a steeper gradient at distances above 10 mm – this effect is more pronounced for the rectum than the bladder. The same effect is observed in Figure 5.4 but the distance at which the 25th percentile-KB(Manual) rectum $D_{0.01\,cc}$ distributions become more negatively skewed than the clinical treatment plans is only 4 mm from PTV1.

Although the data suggest relatively large increases in skewness of the $D_{0.01\,cc}$ distributions when treatment plans are optimised using the KB class-solutions, it is also important to consider the widths of the distributions. Table 5.2 and Table 5.3 show that the standard deviations of the $D_{0.01\,cc}$ distributions for the KB treatment plans are all significantly lower than corresponding distributions for the clinical treatment plans. This is also illustrated in Figure 5.9, which shows histograms of the $D_{0.01\,cc}$ distributions at 4 mm and 12 mm from PTV1 for the clinical, average-KB(Manual) and 25th percentile-KB(Manual) treatment plans.

Although the distributions for both 12 mm average- and 25th percentile-KB(Manual) treatment plans retain a negative skew, the overall shape of the distributions is far more preferable to the one for the clinical treatment plans. A similar conclusion can also be drawn for the average- and 25th percentile-KB(Hybrid) treatment plans based on the histograms shown in Figure 5.10. Here the positive tails of the distributions are generated by the same treatment plans that lead to the 95th percentile lines in Figure 5.5 and Figure 5.6 sitting so far from the mean. As noted above, this was due to gross inaccuracies in OAR auto-contouring. If these noisy tails are ignored the majority of the distribution compares well with corresponding KB(Manual) treatment plans.

Table 5.2: Skewness calculations for the segmented OAR $D_{0.01\ cc}$ distributions for the 131 clinical, average-KB(Manual) and 25th percentile-KB(Manual) treatment plans.

		Bladder			Rectum		
Treatment Plan	**Distance from PTV1**	**Mean / cGy**	**Std / cGy**	**Skewness**	**Mean / cGy**	**Std / cGy**	**Skewness**
Clinical	2 mm	5347	77	–0.79	5065	104	–0.06
	4 mm	5071	125	–0.84	4913	125	–0.55
	6 mm	4733	187	–0.68	4715	185	–0.89
	8 mm	4401	261	–0.96	4510	260	–0.93
	10 mm	4073	334	–0.97	4296	328	–0.83
	12 mm	3763	384	–0.42	4046	394	–0.64
	14 mm	3475	417	–0.14	3769	490	–0.87
	16 mm	3211	409	–0.07	3506	578	–0.86
	18 mm	2974	404	0.19	3289	604	–0.72
Average-KB (Manual)	2 mm	5315	45	0.36	5029	39	0.68
	4 mm	5024	80	0.20	4849	44	0.72
	6 mm	4653	116	–0.12	4613	61	–0.30
	8 mm	4284	138	–0.16	4364	90	–0.66
	10 mm	3905	153	–0.51	4108	100	–1.59
	12 mm	3557	166	–0.97	3809	122	–1.74
	14 mm	3243	162	–1.22	3496	208	–4.74
	16 mm	2992	157	–1.45	3223	271	–3.15
	18 mm	2798	164	–1.01	2986	333	–2.91
25th Percentile-KB (Manual)	2 mm	5282	50	0.35	4973	45	0.01
	4 mm	4967	87	0.27	4760	58	–1.15
	6 mm	4563	120	0.07	4478	87	–1.46
	8 mm	4151	139	–0.06	4173	120	–2.11
	10 mm	3750	139	–0.17	3862	132	–2.62
	12 mm	3390	140	–0.34	3505	143	–2.37
	14 mm	3070	140	–0.78	3135	224	–4.07
	16 mm	2824	136	–1.40	2836	260	–3.08
	18 mm	2636	153	–1.41	2612	268	–2.87

Table 5.3: Skewness calculations for the segmented OAR $D_{0.01\ cc}$ distributions for the 131 clinical, average-KB(Hybrid) and 25th percentile-KB(Hybrid) treatment plans.

		Bladder			Rectum		
Treatment Plan	**Distance from PTV1**	**Mean / cGy**	**Std / cGy**	**Skewness**	**Mean / cGy**	**Std / cGy**	**Skewness**
Average-KB (Hybrid)	2 mm	5332	55	0.74	5054	65	1.55
	4 mm	5049	99	0.63	4885	84	1.97
	6 mm	4690	147	0.71	4663	119	1.63
	8 mm	4311	176	0.83	4424	164	1.43
	10 mm	3934	187	0.95	4177	200	1.49
	12 mm	3583	188	0.95	3897	239	1.72
	14 mm	3275	186	0.84	3614	300	0.69
	16 mm	3021	159	0.21	3381	342	0.67
	18 mm	2824	171	0.03	3178	403	–0.25
25th Percentile-KB (Hybrid)	2 mm	5307	68	1.08	5005	90	1.98
	4 mm	5002	115	0.66	4811	121	2.02
	6 mm	4620	172	0.62	4547	165	1.74
	8 mm	4210	192	1.05	4260	215	1.75
	10 mm	3808	218	1.28	3969	258	1.93
	12 mm	3437	211	1.55	3644	308	2.03
	14 mm	3110	210	1.18	3327	373	1.53
	16 mm	2853	190	0.41	3071	425	1.33
	18 mm	2661	201	0.21	2868	473	0.87

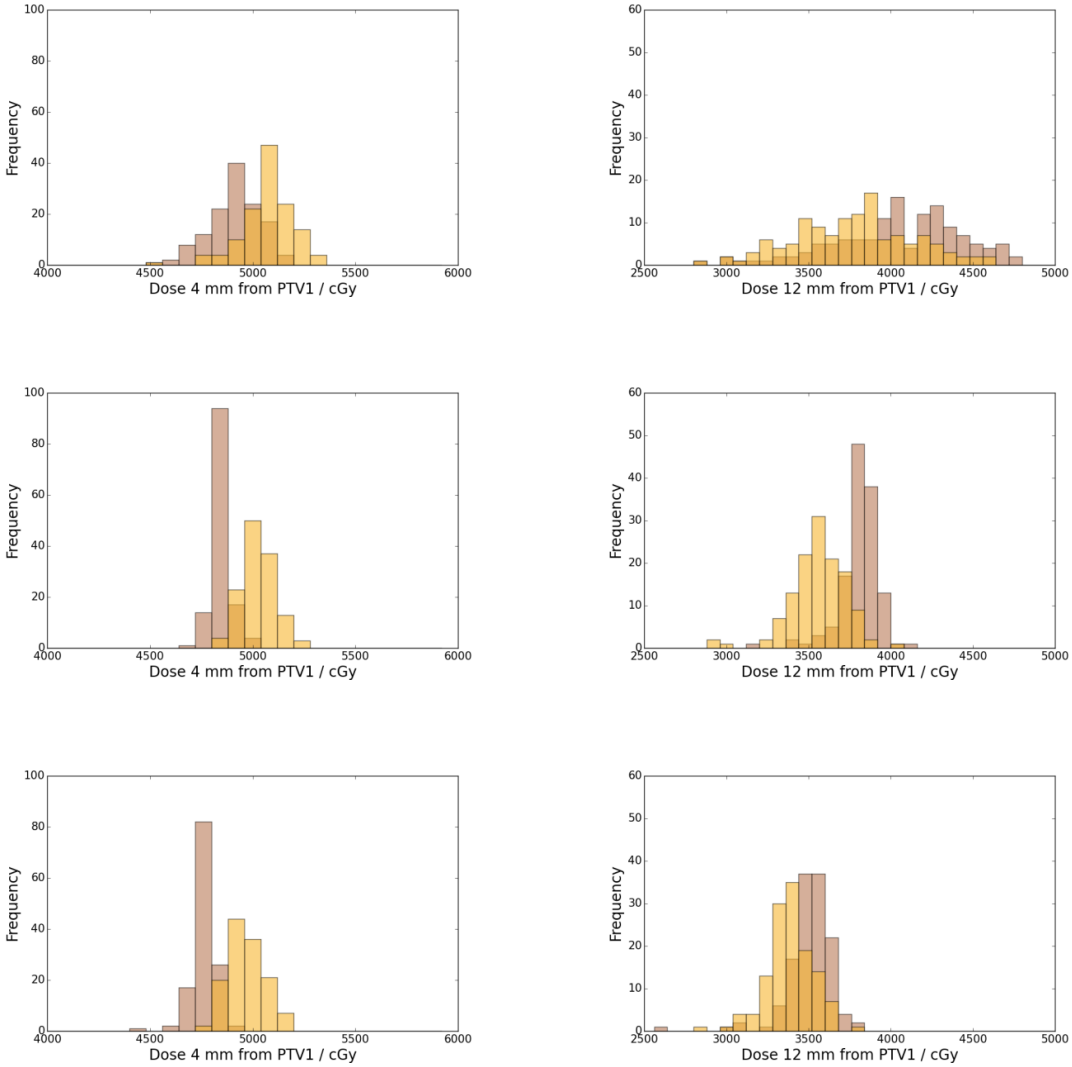

Figure 5.9: Histograms of $D_{0.01\ cc}$ distributions for the Rectum (brown) and Bladder (yellow) segmentations 4 mm and 12 mm from PTV1 for the 131 clinical (top), average-KB(Manual) (middle) and 25th percentile-KB(Manual) (bottom) treatment plans.

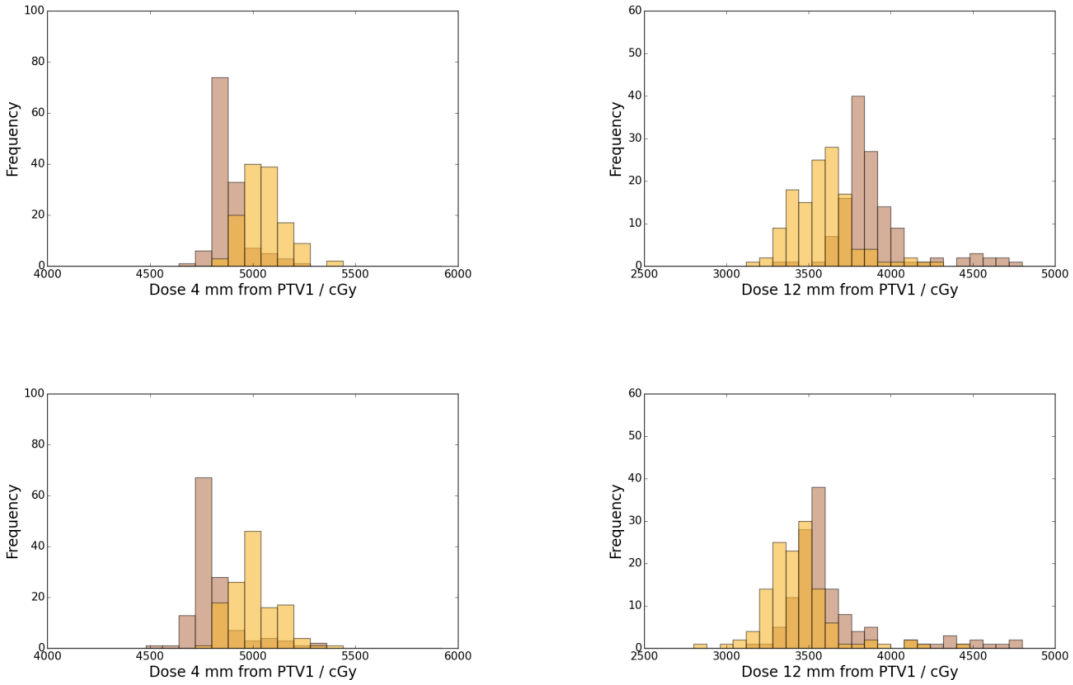

Figure 5.10: Histograms of $D_{0.01\ cc}$ distributions for the Rectum (brown) and Bladder (yellow) segmentations 4 mm and 12 mm from PTV1 for the average-KB(Hybrid) (top) and 25th percentile-KB(Hybrid) (bottom) treatment plans.

Table 5.4 and Table 5.5 present the differences in the standard prostate radiotherapy dose statistics between the 131 clinical treatment plans and corresponding average-KB and 25th percentile-KB treatment plans. Paired Student's *t*-tests were used to assess the statistical significance of each difference in Table 5.4 and Table 5.5 and values presented in bold were found to be statistically significant using a threshold for significance of $p < 0.05$.

A surprising number of the dose statistics show statistically significant average differences between clinical treatment plans and treatment plans generated with the KB class-solutions. For both average-KB(Manual) and 25th percentile-KB(Manual), differences in PTV dose statistics are all within ±0.4 %. It is worth commenting here that a statistically significant difference in a reported dose statistic does not necessarily imply that the difference is clinically significant. As a general rule, a change in dose of less than approximately 2 % is not considered to be clinically significant. Therefore, although the results in Table 5.4 show that coverage of the PTVs is statistically significantly different between the clinical treatment plans and KB(Manual) treatment plans, the differences are not clinically significant.

A similar case can be made for the marginal differences in bladder and rectum dose statistics for the average-KB(Manual) treatment plans of which approximately half are statistically significant. All average differences are within –0.1 % to 0.5 % of the clinical treatment plans with the exception of the rectum $V_{2460\ cGy}$. Femoral head $D_{1\ cc}$ also increased by more than 2 % but as noted previously, at The Christie, femoral head doses are not constrained or reported in clinical treatment plans. The femoral head doses of all KB treatment plans generated for this work were comfortably within the aforementioned 5 % optimal limit of $V_{4000\ cGy}$ specified by the PIVOTALBoost trial.

For the 25th percentile-KB(Manual) treatment plans, the observed reductions in bladder and rectum doses are mostly statistically and clinically significant. This, combined with the comparable PTV coverage noted above, suggests that use of the 25th percentile-KB class-solution generates an average improvement in treatment plan quality over the average-KB class-solution and manual treatment planning.

Similar results are observed for the average-KB(Hybrid) and 25th percentile-KB(Hybrid) treatment plans. Some of the differences in Table 5.5 are greater than corresponding values in Table 5.4. However, it has already been acknowledged that in some cases gross OAR auto-contouring inaccuracies failed to provide suitable contours with which to optimise hybrid treatment plans. Such outliers were not observed in the results shown in Table 5.4 and therefore direct comparison of the differences in Table 5.4 with Table 5.5 is of limit value.

Table 5.4: Average, standard deviation, minimum and maximum percentage differences between KB(Manual) and clinical treatment plans.

	Average-KB(Manual)				25th Percentile-KB(Manual)			
Dose Statistic	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**
PTV1 $D_{99\ \%}$	**0.4**	0.6	–1.3	2.0	**0.3**	0.7	–3.3	1.4
PTV1 min $D_{1\ cc}$	**0.4**	0.7	–1.5	2.1	**0.2**	0.8	–3.6	1.6
PTV2 $D_{99\ \%}$	**–0.3**	0.3	–1.2	0.7	**–0.4**	0.4	–2.3	0.4
PTV2 min $D_{1\ cc}$	**–0.2**	0.3	–1.2	0.9	**–0.3**	0.5	–2.5	0.6
PTV3 $D_{99\ \%}$	**–0.2**	0.4	–1.0	1.0	**–0.2**	0.4	–1.6	0.9
PTV3 min $D_{1\ cc}$	**–0.2**	0.4	–1.1	1.0	**–0.2**	0.4	–1.7	0.8
PTV3 max $D_{1\ cc}$	0.0	0.3	–1.2	0.9	0.1	0.3	–1.2	0.7
PTV3 median	**0.0**	0.1	–0.3	0.3	**0.0**	0.1	–0.3	0.2
FHL max $D_{1\ cc}$	**3.1**	5.5	–13.8	20.8	**3.1**	5.8	–10.8	20.6
FHR max $D_{1\ cc}$	**2.9**	5.5	–9.9	16.9	**3.7**	6.0	–8.8	19.1
Rectum $V_{2460\ cGy}$	**2.2**	5.8	–27.8	16.3	**–3.2**	6.6	–40.8	11.3
Rectum $V_{3240\ cGy}$	0.3	5.8	–18.4	11.5	**–4.5**	6.0	–26.2	7.4
Rectum $V_{4080\ cGy}$	0.2	5.0	–11.7	16.9	**–3.2**	4.8	–16.2	10.7
Rectum $V_{4860\ cGy}$	0.4	2.5	–5.1	10.3	**–1.3**	2.3	–7.0	5.8
Rectum $V_{5280\ cGy}$	**0.5**	1.0	–2.6	3.6	**0.1**	1.1	–2.7	4.1
Rectum $V_{5700\ cGy}$	**0.5**	0.8	–1.2	3.0	**0.3**	0.9	–1.5	4.8
Rectum $V_{6000\ cGy}$	0.0	0.0	0.0	0.1	**0.0**	0.0	0.0	0.1
Bladder $V_{4080\ cGy}$	–0.1	3.7	–7.4	14.5	–1.7	3.6	–11.9	12.3
Bladder $V_{4860\ cGy}$	–0.1	2.5	–5.8	13.0	–1.3	2.5	–6.7	10.0
Bladder $V_{6000\ cGy}$	**0.5**	1.1	–1.9	4.4	**0.8**	1.3	–4.1	5.3

Table 5.5: Average, standard deviation, minimum and maximum percentage differences between KB(Hybrid) and clinical treatment plans.

	Average-KB(Hybrid)				25th Percentile-KB(Hybrid)			
Dose Statistic	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**	**Mean / %**	**SD / %**	**Min. / %**	**Max. / %**
PTV1 $D_{99\ \%}$	**–0.2**	0.6	–2.3	0.9	**0.3**	0.7	–2.4	1.7
PTV1 min $D_{1\ cc}$	**0.9**	0.8	–1.7	3.3	**0.2**	0.9	–3.9	1.7
PTV2 $D_{99\ \%}$	**–0.4**	0.4	–2.4	0.7	**–0.4**	0.5	–4.0	0.7
PTV2 min $D_{1\ cc}$	**–0.4**	0.5	–2.4	0.8	**–0.4**	0.6	–4.1	0.7
PTV3 $D_{99\ \%}$	**–0.3**	0.4	–2.0	0.8	**–0.3**	0.4	–1.6	0.9
PTV3 min $D_{1\ cc}$	**–0.3**	0.4	–1.9	0.7	**–0.3**	0.4	–1.9	0.8
PTV3 max $D_{1\ cc}$	0.0	0.3	–1.2	0.8	**0.1**	0.3	–1.2	1.5
PTV3 median	**0.0**	0.1	–0.3	0.2	0.0	0.1	–0.4	0.3
FHL max $D_{1\ cc}$	**2.1**	4.9	–10.2	20.2	**2.7**	5.5	–12.2	19.3
FHR max $D_{1\ cc}$	**1.8**	5.1	–10.7	18.6	**2.4**	5.1	–12.0	15.8
Rectum $V_{2460\ cGy}$	**6.1**	8.6	–22.5	26.3	–0.2	8.6	–41.1	17.5
Rectum $V_{3240\ cGy}$	**3.0**	8.5	–23.6	28.7	**–2.5**	8.2	–27.3	27.9
Rectum $V_{4080\ cGy}$	**1.4**	6.8	–17.0	27.0	**–2.3**	6.3	–18.9	22.9
Rectum $V_{4860\ cGy}$	**0.8**	3.0	–6.3	10.3	**–0.9**	3.0	–8.2	9.6
Rectum $V_{5280\ cGy}$	**0.8**	1.3	–2.4	6.3	**0.4**	1.3	–3.1	6.0
Rectum $V_{5700\ cGy}$	**1.1**	1.5	–1.4	9.4	**0.9**	1.5	–1.5	8.9
Rectum $V_{6000\ cGy}$	**0.0**	0.0	0.0	0.2	0.0	0.1	0.0	0.7
Bladder $V_{4080\ cGy}$	–0.1	3.8	–8.6	14.6	**–1.5**	4.0	–10.6	15.1
Bladder $V_{4860\ cGy}$	0.2	2.8	–4.8	13.1	**–1.0**	2.8	–5.5	12.5
Bladder $V_{6000\ cGy}$	**0.3**	1.1	–2.6	4.2	**0.9**	1.5	–4.0	7.2

The test dataset consisted of 131 clinical treatment plans of which 123 met the minimum dose criteria set out in Table 3.2. For four of the eight clinical treatment plans that failed to meet the minimum dose criteria, both the average- and 25th percentile-KB(Manual) treatment plans also failed to meet them. For the remaining four, at least one of the KB class-solutions generated a treatment plan that did meet all of the minimum dose criteria and thus would be considered dosimetrically superior.

Of the 131 average- and 25th percentile-KB(Manual) treatment plans, 110 and 114 respectively met all of the minimum dose criteria. When these results are pooled the KB(Manual) treatment planning approach generated dosimetrically acceptable treatment plans in 120 of the 131 test cases, which is a success rate comparable to the manual treatment plan optimisation approach employed clinically.

It is worth recalling here that 131 clinical treatment plans had passed a rigorous quality assurance process, which on average rejects approximately 10 % of treatment plans. The comparison of KB(Manual) and clinical treatment plan success rates is not strictly a fair one because it considers them respectively before and after a quality assurance process and so overestimates the success rate of manual treatment planning.

The number of average- and 25th percentile-KB(Hybrid) treatment plans that met the minimum dose criteria was 86 and 88 respectively. When these results are pooled the KB(Hybrid) treatment planning approach generated dosimetrically acceptable treatment plans in 102 of the 131 test cases. This falls short of the success rates seen with the clinical and KB(Manual) treatment plans and is perhaps unsurprising given it has previously been acknowledged that for some cases gross auto-contouring inaccuracies led to improperly optimised hybrid-plans. It is a success rate of almost 80 %, however, and given that the hybrid-plans can be generated with no human intervention (after the clinician manually outlines the prostate and SV) the KB(Hybrid) treatment planning approach represents a significant potential efficiency saving for treatment planning departments without compromising treatment plan quality.

The box and whisker plots presented in Figure 5.11 show the distributions of total plan MU, PA, PI, PM and MCS for each group of treatment plans. The limits of the boxes

represent the lower and upper quartiles of the data, the central horizontal line represents the median, the extent of the whiskers show the 5th and 95th percentiles and the data points beyond the whiskers are outliers.

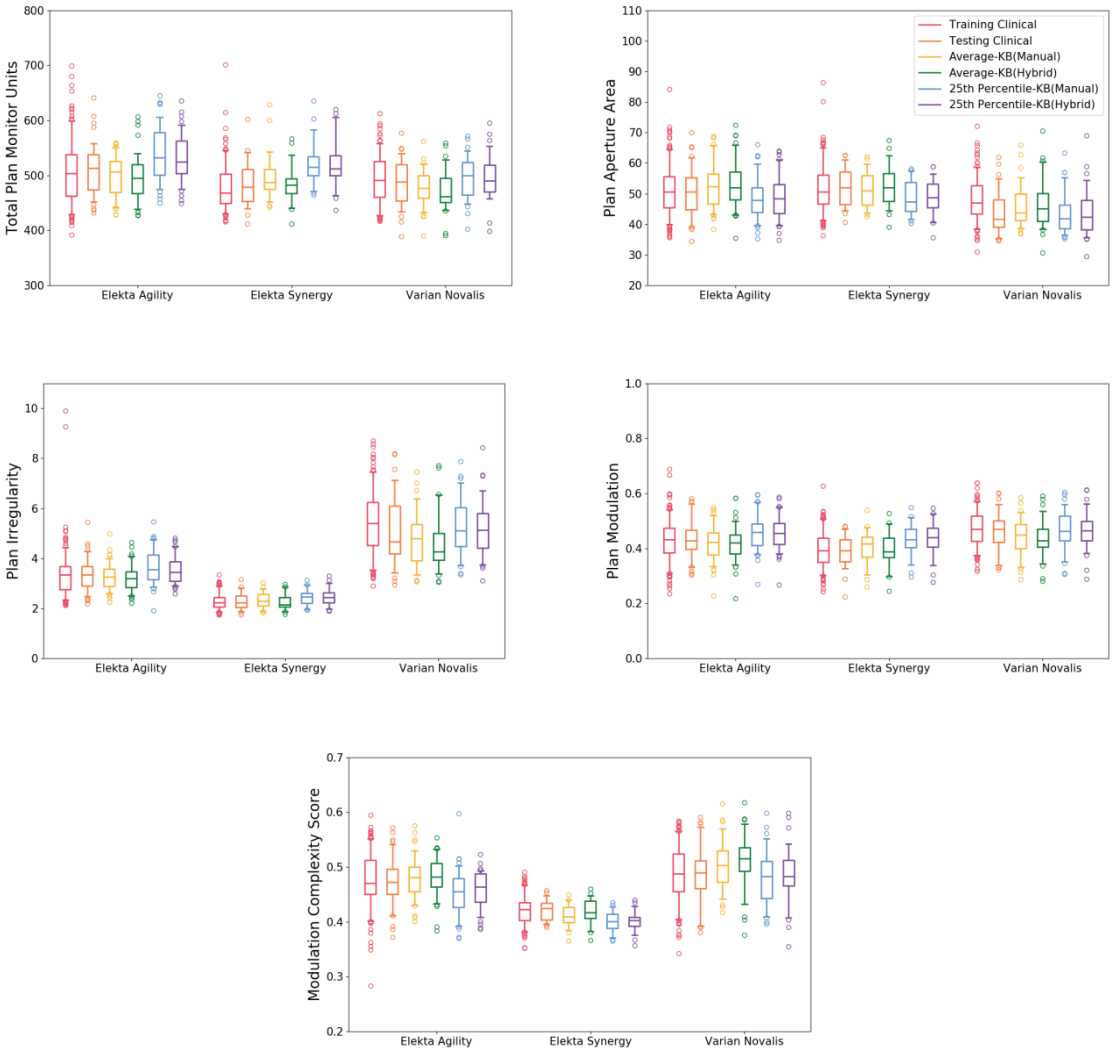

Figure 5.11: Boxplots of MU, Plan Aperture, Plan Irregularity, Plan Modulation and Modulation Complexity Score for the treatment plans evaluated in this chapter.

Unpaired Student's *t*-tests were performed to compare the differences between the clinical treatment plans in the training and test patient cohorts. For all complexity metrics across all three machines types, none of the differences was statistically significant. This supports the earlier result that the 131 treatment plans from the test patient cohort were selected from a population consistent with the training data.

Welch's *t*-tests were performed to compare the differences in the complexity

scores between clinical treatment plans in the test cohort with treatment plans generated with the KB class-solutions. The tests were performed for each of the five complexity scores and each of the three machine types noted above.

For the average-KB(Manual) and (Hybrid) treatment plans, where the differences in complexity scores were statistically significant ($p < 0.05$), the scores indicated that the treatment plans generated with the KB were slightly less complex than the clinical treatment plans. For the 25th percentile-KB(Manual) and (Hybrid) treatment plans, statistically significant increases in complexity were observed for all complexity scores and all machine types.

This analysis implies that the average-KB class-solution maintains average treatment plan quality (at least at a clinically significant level) whilst improving consistency between patients and this is not at the expense of treatment plan complexity. It also implies that the 25th percentile-KB class-solution improves average treatment plan quality and consistency but that this does increase complexity. It should be noted though that there is a large overlap of the ranges of complexity scores for all of the evaluated treatment plans. Therefore small differences in average complexity scores may not present a major challenge for machine delivery.

5.5 Discussion

This work has demonstrated that the KB approach to treatment planning that was proposed and prototyped in Chapter 4 can be used to generate clinically acceptable prostate VMAT treatment plans with a success rate that rivals that of manual treatment planning. The average KB class-solution has been shown to maintain average treatment plan quality over a large test patient population in terms of PTV coverage and OAR sparing whilst improving consistency of optimisation between patients. Using the average-KB class-solution was also shown to reduce treatment plan complexity slightly when assessed using a range of published complexity metrics. When the treatment plan was optimised using the 25th percentile-KB class-solution target coverage was not compromised and OAR sparing was superior to that seen with the average-KB class-

solution and manual treatment planning. This improvement in average treatment plan quality, however, was accompanied by an increase in average complexity.

The same group of patients were also used to generate hybrid-plans with the KB class-solutions based on SPICE OAR auto-contours. Gross auto-contouring inaccuracies led to approximately 20 % of hybrid-plans being insufficiently optimised when dose was reported to accurate manual-contours. However, for the remaining 80 % of cases, at least one of the KB class-solutions generated a clinically acceptable treatment plan. Given that the hybrid treatment plans can be generated with no human intervention (after the clinician manually outlines the prostate and SV) the KB(Hybrid) treatment planning approach represents a significant potential efficiency saving for treatment planning departments. This efficiency saving could be exploited to enable ART strategies and provide patients with earlier access to their cancer treatments.

Approaches to KB treatment planning described in the literature tend to focus on the prediction of appropriate optimisation parameters by using the same parameters as the most similar treatment plan from a database of prior treatment plans. The number of treatment plans that can be included in a database, however, is severely limited for reasons of computational speed. Given the variability in pelvic anatomy between patients, this limitation in the size of the database limits the likelihood of finding a suitable match for the new treatment plan and hence introduces uncertainty into the predicted optimisation parameters. This uncertainty manifests itself in the degree to which the new treatment plan is optimised and this fundamentally limits the utility of the KB treatment planning strategies published in the literature.

The KB treatment planning class-solutions presented here therefore represent a useful and novel approach. 562 previous treatment plans were used to train a KB model of the ideal prostate treatment plan, which is vastly more than is typically included in a KB database of patients. When optimising a new treatment plan, dose gradient information from all 562 treatment plans is used rather than selecting the single previous treatment plan that best matches the new patient. In this work, testing of the KB approach using both manual- and auto-contours, was performed using a large cohort

of 131 patients. This is a much larger test cohort than is typically described in the literature, which is generally of the order of 10 patients. It therefore provides increased confidence in the clinical utility of the approach to KB treatment planning.

It was shown that marginal increases in reported OAR doses are observed when treatment plans are optimised using the average-KB class-solution. Although these increases were not considered clinically significant they were statistically significant so probably represent real marginal increases in OAR doses. It is worth discussing here that since the PTV1-OAR dose gradients were characterised by the $D_{0.01\,cc}$ of the OAR segments, if the gradient is not consistent along the whole superior-inferior extent of PTV1, the method will tend to underestimate the global PTV1-OAR dose gradients. This is illustrated in Figure 5.12, where inferiorly the PTV1-Rectum dose gradient is steeper than at the superior extent in the plane of the SV. For this treatment plan, the PTV1-Rectum dose gradient is characterised in the superior region where it is shallowest since this is where the $D_{0.01\,cc}$ of the OAR segments resides.

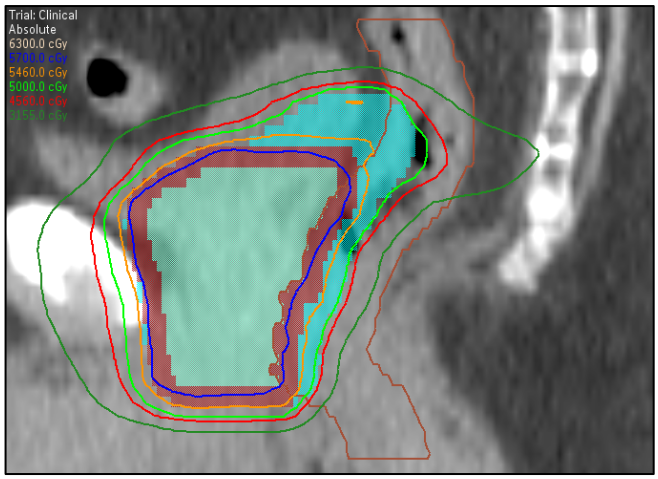

Figure 5.12: Sagittal slice of a clinical treatment plan where the PTV1-Rectum dose gradient is significant steeper at the inferior extent of PTV1 compared with the superior extent at the level of the SV.

The KB class-solutions aimed specifically to optimise the PTV1-OAR dose gradients, which is likely to have made them relatively consistent over the superior-inferior extent of PTV1. The characterisation of the PTV1-OAR dose gradients in the

training dataset, however, is a complex process. Here $D_{0.01\,cc}$ of the 2 mm OAR segments were used but since this tends to underestimate the global dose gradient this could be the cause of the increases in OAR doses when treatment plans are optimised with the average-KB class-solution. Given that the increases in reported OAR doses are not clinically significant and clinically acceptable treatment plans were generated at a rate equivalent to manual treatment planning, it is unlikely that this limitation has a major impact. However, it remains a theoretical limitation and further work could address this by considering alternative characterisations of the PTV1-OAR dose gradients.

5.6 Conclusion

This chapter has presented the results of large scale testing of the KB treatment planning approach developed in Chapter 4. This has also been extended to produce a fully-automated KB treatment planning workflow by incorporating OAR auto-contours.

It has been shown that the average-KB class-solution achieves the intention of making treatment plans more consistent whilst maintaining average treatment plan quality and offering small reductions in treatment plan complexity. It has also been demonstrated that treatment plan quality can be improved using the 25^{th} percentile-KB class-solution but that this tends to result in a small increase in average complexity.

Hybrid-plans generated with manually drawn targets and auto-contoured OARs have shown similar results. Although the number of treatment plans that satisfied all of the minimum dose criteria for prostate VMAT treatment planning was lower for the KB(Hybrid) than KB(Manual) treatment plans. This was largely due to gross errors in the SPICE auto-contours. The instances of auto-contour failures in this work are likely to be limitations specific to SPICE. Although it is unlikely that other auto-contouring packages would not suffer from similar issues, as auto-contouring techniques improve, the success rate of KB(Hybrid) treatment planning will also increase and potentially rival that of KB(Manual) and clinical treatment planning.

5.7 References

Chanyavanich, V., Das S., Lee, W. & Lo, J. (2011). 'Knowledge-based IMRT treatment planning for prostate cancer', *Medical Physics*, 38(5), pp. 2515-2522.

Djajaputra, D., Wu, Q., Wu, Y. & Mohan, R. (2003). 'Algorithm and performance of a clinical IMRT beam-angle optimisation system', *Physics Medicine Biology*, 48(19), pp. 3191-3212.

Good, D., Lo, J., Lee, R., Wu, J., Yin, F. & Das, S. (2013). 'A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning', *International Journal of Radiation Oncology Biology Physics*, 87(1), pp. 176-181.

Nwankwo, O., Sihono, D., Schneider, F. & Wenz, F. (2014). 'A global quality assurance system for personalized radiation therapy treatment planning for the prostate (or other sites)', *Physics in Medicine & Biology*, 59, pp. 5575-5597.

Nwankwo, O., Mekdash, H., Sihono, D., Wenz, F. & Glatting, G. (2015). 'Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning', *Radiation Oncology*, 10(111), pp. 1-5.

Petit, S., Wu, B., Kazhdan, M., Dekker, A., Simari, P., Kumar, R., Taylor. R., Herman, J. & McNutt, T. (2012). 'Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma', *Radiotherapy Oncology*, 102(1), pp. 38-44.

Powis, R., Bird, A., Brennan, M., Hinks, S., Newman, H., Reed, K., Sage, J. & Webster, G. (2017). 'Clinical implementation of a knowledge based planning tool for prostate VMAT', *Radiation Oncology*, 12(81), pp. 1-8.

Yang, Y., Ford, E., Wu, B., Pinkawa, M., van Triest, B., Campbell, P., Song, D. & McNutt, T. (2013). 'A overlap-volume-histogram based method for rectal dose prediction and automated treatment planning n the external beam prostate radiotherapy following hydrogel injection', *Medical Physics*, 40(1), pp. 1-10.

Yuan, L., Ge, Y., Lee, W., Yin, F., Kirkpatrick, J. & Wu, J. (2012). 'Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans', *Medical Physics*, 39(11), pp. 6868-6878.

Zhu, X., Ge, Y., Li, T., Thongphiew, D., Yen, F. & Wu, Q. (2011). 'A planning quality evaluation tool for prostate adaptive IMRT based on machine learning', *Medical Physics*, 38(2), pp. 719-726.

Wang, Y., Heijmen, B. & Petit, S. (2017). 'Prospective clinical validation of independent DVH prediction for plan QA in automatic treatment planning for prostate cancer patients', *Radiotherapy & Oncology*, 125, pp. 500-506.

5.8 Supplementary Material

5.8.1 ESTRO 2020 Poster

Based on the work presented in Chapters 4 and 5 an abstract was submitted to the 2020 European Society for Radiotherapy & Oncology (ESTRO) conference. The abstract was accepted for a poster exhibition and a representation of the poster is shown on the following page.

An automated knowledge-based treatment planning solution for prostate VMAT

J. Wood[1, 2], M. Aznar[2] & P. Whitehurst[1]
[1]The Christie NHS Foundation Trust, United Kingdom. [2]The University of Manchester, United Kingdom.
joe.wood@christie.nhs.uk : @joewood86

Purpose

This work aimed to develop a novel approach to knowledge-based (KB) treatment planning for prostate VMAT.

Method

Converting well defined PTV dose criteria (such as $D_{99\,\%}$ or $D_{1\,cc}$) into objectives for an optimisation algorithm is relatively straightforward. OARs, however, present a greater challenge because they sit in steep dose gradients and their size, shape and position relative to PTVs vary significantly between patients.

Precisely predicting optimal OAR DVH parameters for an individual patient therefore is difficult and uncertainties in the DVH predictions can ultimately become manifest in the degree to which treatment plans are dosimetrically optimal.

For standard-of-care prostate radiotherapy, optimal OAR sparing can be assessed by the PTV-OAR dose gradient (i.e. dose fall-off per unit distance). The acheivable PTV-OAR dose gradient is characterised primarily by delivery machine parameters and not patient anatomy.

On this basis, a model of the ideal prostate treatment plan was developed – see Figure 1, where the isodoses show optimal target coverage and conformity and the colour gradients represent achievable PTV-OAR dose gradients, which should be largely the same for all patients.

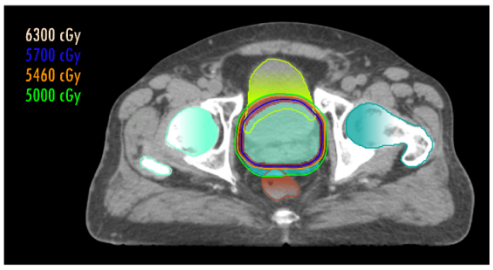

Figure 1: A model of the ideal prostate treatment plan.

A KB of 562 clinical prostate VMAT treatment plans generated and quality assured in 2017 and 2018 according to a well-established clinical protocol were used to train this model. The mean and 25th percentile of the PTV-OAR dose gradients from the training dataset were used to generate two sets of optimisation objectives in the Pinnacle TPS and these were tested on a cohort of 142 patients from 2019 using an automated treatment planning workflow.

The new treatment plans for 10 of the test patients (for which all standardly reported prostate radiotherapy dose criteria were met) were reviewed visually by a medical physics expert experienced in prostate treatment planning.

Results

For 114 of the 142 test cases, at least one of the auto-generated treatment plans (average or 25th percentile) met all of the dose criteria for prostate radiotherapy at The Christie. Furthermore, for 105 of the test cases, at least one of the auto-generated plans was superior to the original clinical plan.

In Figure 2, for each of the 142 test cases, the PTV-Rectum dose gradient is plotted against PTV-Bladder dose gradient. The KB training data are plotted as transparent red circles, the clinical test data are plotted as solid red circles and the auto-generated treatment plan data based on the average and 25 percentile of the KB are plotted as yellow and blue circles respectively.

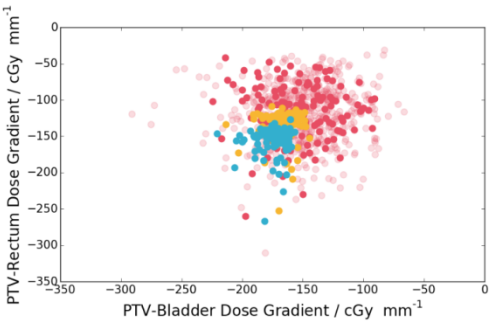

Figure 2: Treatment plans generated using KB dose gradient optimisation (blue and yellow) show more consistent and steeper PTV-OAR dose gradients than corresponding clinical treatment plans (red).

Visual review of 10 auto-generated treatment plans based on the average and 25thpercenticle of the KB dose gradients showed that all were considered clinically acceptable.

Conclusion

A novel approach to KB treatment planning for prostate VMAT has been proposed and tested. Initial results show that auto-generated treatment plans based on the predictions of the model are clinically acceptable in more than 80 % of cases and show superior and more consistent OAR sparing than their corresponding clinical treatment plans.

Although there is scope for refinement of the predictions, this method shows promise for realising significant efficiencies within treatment planning departments and with the transfer of knowledge between centres.

Chapter 6

6 Final Discussion

This thesis began with the statistics that in the UK almost 50, 000 men are diagnosed with prostate cancer every year and that radiotherapy forms an important part of treatment for approximately 30 % of these patients. Figure 1.1 illustrated the typical pathway following which prostate radiotherapy treatments are planned, quality assured and delivered. Each stage of that pathway is typically performed manually and there can often be significant delays between successive stages of the pathway.

When prostate cancer prevalence statistics and the current treatment planning pathway are considered together, it quickly becomes apparent that there exists a problem for modern radiotherapy techniques such as ART. Even marginal increases in prostate radiotherapy treatment planning workload for individual patients can generate wholly unfeasible increases in total workload over the patient population. In order to pursue developments in modern radiotherapy, therefore, the pathway illustrated in

Figure 1.1 needs review and the role of automation needs serious consideration.

Such consideration of automation was the rationale for this thesis. The following sections present a final discussion on how overcoming the problems associated with automation have been addressed and identify areas for future work. Clinical implementation of automation techniques is also discussed.

6.1 Auto-Segmentation as a Clinical Tool

Manual-segmentation of anatomy in CT images is a time consuming and subjective process (Han *et al.*, 2008; Collier *et al.*, 2003). Auto-segmentation using a variety of techniques and algorithms has been extensively described in the literature as having the potential to replace this time consuming task whilst also making segmentation a more consistent exercise. Despite these potential benefits to the clinical treatment planning pathway, the practical utilisation of auto-segmentation in the clinic is limited.

In the literature, assessment of auto-segmentation techniques is often performed in terms of geometric accuracy compared to some 'ground-truth' manual-contours drawn by a trained expert. The work presented in Chapter 2 demonstrated that commonly reported geometric accuracy scores (namely average mean-DTA and DSC) can provide reasonably reliable metrics for male pelvic auto-contours. However, these metrics have their limitations and a major downfall is that they offer no indication of the clinical significance of the geometric inaccuracies they report. It is therefore difficult to determine whether a given auto-segmentation technique is suitable to replace manual-segmentation completely based on geometric accuracy statistics alone.

A better measure of auto-segmentation performance is the utility of auto-contours in the intended clinical setting. Such assessments in the literature typically take the form of time and motion studies where time savings generated by editing auto-contours (as opposed to drawing manual-contours *de novo*) are measured (Teguh *et al.*, 2011; Speight *et al.*, 2014; Lim & Leech, 2016).

Reported time savings associated with manually modifying auto-contours are typically of the order of minutes per patient. When this is viewed in the context of the

typical time between 'Planning CT Scan' and 'Outlining Target Volumes', which may be up to a week, minutes saved are of little practical consequence for the overall pathway. In addition, the added time and financial costs associated with purchasing, installing, maintaining and quality assuring an auto-segmentation medical device (costs often borne by individuals who do not necessarily reap the benefits of the reported time savings) are likely to be a major reason why auto-segmentation is not used widely in clinical practice.

On this basis, the intention of the work presented in Chapter 3 was to assess the clinical utility of auto-contours for use in optimisation where auto- and hybrid-plans were compared against manual-plans[§§]. It was demonstrated that for a number of commercially available auto-segmentation packages, actual target coverage is significantly compromised for auto-plans, which limits the utility of prostate and SV auto-contours. Although the work in Chapter 3 was only based on 11 test cases, auto-plan target coverage was significantly compromised in all cases such that expansion of the test dataset was not necessary.

Hybrid-plans, however, were shown to provide OAR sparing and PTV coverage comparable with manual-plans. The initial testing of the hybrid-plan concept described in Chapter 3 was also only based on 11 patients. Although no significant correlations were found between OAR auto-contour geometric accuracy and hybrid-plan quality, from a dataset of 11 patients it would be unreasonable and counter-intuitive to conclude definitively that hybrid-plan quality is not at all dependent on OAR auto-contour geometric accuracy. Indeed, in Chapter 5, for approximately 20 % of the 131 test cases the hybrid-plan approach failed to generate a clinically acceptable treatment plan due to gross auto-segmentation errors. Therefore, geometric accuracy is clearly important to a degree but OAR auto-contour utility for treatment plan optimisation does not appear to be sensitive to small differences in geometric accuracy performance – especially when the limitations of geometric accuracy metrics are considered.

[§§] Auto-plans were generated using target and OAR auto-contours; manual-plans were generated using target and OAR manual-contours and hybrid-plans were generated with target manual-contours and OAR auto-contours.

This is an important finding and has implications for future evaluations of auto-segmentation techniques. As noted above, reporting of geometric accuracy alone is insufficient to determine the clinical utility of auto-contours for treatment planning. Furthermore, published time and motion studies have suggested potential efficiency savings with auto-segmentation but the potential benefits of such savings can be insufficient to justify adopting it in the clinic.

When considering auto-segmentation as a clinical tool, therefore, it is important to remember that anatomical contours are not ends in themselves. They are simply means by which to optimise and report radiotherapy treatment plans. From a practical standpoint and from the work presented in this thesis, the real utility and benefit of auto-segmentation is in facilitating treatment plan optimisation by removing the manual outlining bottlenecks from the treatment planning pathway. Hybrid-planning has been demonstrated to be a reasonable way to do this for prostate radiotherapy whilst providing a clinically relatable means by which to assess auto-segmentation performance. Future work should be considerate to this idea and pay close attention to how the performance of auto-segmentation is related to its manifestation in the clinical pathway.

It would, of course, be nonsensical to adopt a hybrid-planning approach where treatment plan optimisation is performed manually. To unlock the full potential of current auto-segmentation techniques, auto-contours need to drive an automated treatment planning process and this is the focus of the following section.

6.2 Knowledge-Based Prostate Treatment Planning

It is widely acknowledged in the literature that manual treatment plan optimisation is time consuming and subjective (Djajaputra *et al.*, 2003). Knowledge-based (KB) treatment planning is a means by which prior experience can be used automatically to optimise treatment plans for new patients (Nwankwo *et al.*, 2015). Typically, a KB consists of a database of treatment plans that are considered optimal. When a new treatment plan is to be optimised, the most similar treatment plan is selected from the

KB and the optimisation parameters from this are used.

Selection of the most similar treatment plan is difficult, however, and in practice the typical KB contains around 10-20 treatment plans (Peressutti *et al.*, 2016). For treatment sites in the pelvis, where anatomy size and shape can be extremely variable between patients, this can be an insufficiently wide selection from which to choose a representative case. The work presented in Chapters 4 and 5 developed and tested a new approach to KB treatment planning where the KB consisted of a model of the ideal prostate radiotherapy treatment plan: characterised in terms of ideal target coverage and dose fall-off within OARs. This approach offers an advantage over more conventional approaches to KB treatment planning in that vastly more patients can be used to train the model than can be included in a typical database of treatment plans.

Although it has been stated frequently in this thesis that pelvic anatomy is widely variable between patients, the configuration of anatomical structures within the male pelvis is relatively fixed: the bladder and rectum always sit immediately anterior and posterior respectively to the prostate and SV and the femoral heads sit laterally typically a small number of centimetres from PTV1. The relatively fixed positions of the anatomy allow the KB model of the ideal prostate treatment plan to be reasonably simple such that PTV1-OAR dose gradients can be characterised in terms of distance from PTV1 only. Despite the simplicity of the KB model, it was extensively demonstrated in Chapter 5 that the model provides a reliable way to generate optimal treatment plans.

For many other disease sites, head and neck cancers for example, OAR geometry relative to PTVs is more variable and a greater interplay effect between anatomical structures can exist – whereby dose to one OAR affects dose to another. If the KB treatment planning approach described in Chapter 4 was applied to such disease sites, KB models as simple as the one described in Chapter 4 for prostate treatment planning would not be sufficient. This is a limitation of the ideas developed in this thesis. However, it presents an intriguing and important area for further work. With sufficiently large patient datasets and more sophisticated geometric models, the approach to KB treatment planning developed and tested in the work presented above could be a useful

way of bringing automation into the treatment planning pathway. Furthermore, it could also be an effective way of sharing treatment planning expertise from large, experienced centres with smaller, less experienced centres.

A potential danger of KB treatment planning techniques based on large patient datasets is that current or historic clinical practice becomes engrained into the models and long-term this can make it difficult to adapt and change. This is a topical discussion at The Christie following the work presented in this thesis. It was noted in Chapter 3 that prostate treatment plans are optimised so that the minimum dose to PTV1 is 5000 cGy. This is based on a local protocol at The Christie and is not a widespread practice. At the time of writing, there is on-going discussion amongst the clinical team around whether PTV1[***] should receive a minimum dose of 95 % of 4700 cGy. This is a more typical dose for this volume and this change to the local protocol would bring The Christie in line with clinical practice in many other centres.

This change to clinical practice at The Christie would mean that the trained KB model would no longer be directly appropriate. Given the simplicity of the KB model for prostate treatment planning it would probably be possible and relatively straightforward to scale the dose gradients for the new proposed doses. In addition, the thorough complexity analysis that was performed in Chapter 5 would provide a useful benchmark for gauging the impact of scaled dose gradients on the resultant treatment plans. For future work, this should be an important consideration. Where possible, KB treatment planning should not constrain future developments and changes to clinical practice. In reality this may be difficult to achieve but highlights the importance of keeping KB models as simple and easy to manipulate as possible.

6.3 Fully-Automated Knowledge-Based Treatment Planning

As discussed in section 6.1, the ultimate aim of automation in the treatment planning pathway is to produce treatment plans without human intervention. Not only does this

[***] A slightly modified form of the PTV1 described in Chapter 3, the details of which are not relevant here.

create staffing efficiencies for clinical departments it also enables modern radiotherapy techniques such as ART to be realistically feasible on a large scale. The practicalities of the treatment planning pathway, however, mean that automation of just part of it is of limited value. The fully-automated KB treatment planning workflow described in Chapter 5 therefore represents a significant development in this field. Hybrid-planning using the KB model of the ideal prostate radiotherapy treatment plan has been shown to produce clinically acceptable treatment plans without treatment planner intervention. Contouring of prostate and SV by a clinical oncologist was found still to be required so unfortunately, this bottleneck of the pathway cannot be fully removed as yet. The potential applications of this technique are therefore interesting to consider.

Increased treatment planning workload is an often cited and real barrier to introducing new techniques into the clinic. Basic forms of ART such as plan of the day (PoD) require multiple treatment plans to be generated in advance of treatment starting – this is typically three treatment plans per patient. In the introduction to this thesis, it was stated that even marginal increases in per-patient-workload for a treatment site as large as prostate cancer cannot be accommodated by clinical departments. In practice, this barrier can be compounded by there not being an NHS funding mechanism for ART.

The result is that PoD, and ART more generally, is not feasible for most NHS radiotherapy centres. Therefore, the fully-automated KB treatment planning workflow presented here provides a reasonable means by which an ART strategy like PoD could be routinely implemented (from a treatment planning perspective at least). Chapter 5 demonstrated that automatically generated hybrid-plans were clinically acceptable for 80 % of patients. Generating three treatment plans for each of the 20 % of patients where the hybrid-plans failed would still present an overall 40 % reduction in prostate treatment planning workload. Hybrid-plans were only tested on a large scale with OAR auto-contours generated in SPICE and with alternative auto-segmentation packages it may well be possible to improve the success rate of hybrid-plan generation further.

6.4 Clinical Implementation

It has been noted above that automation of only part of the radiotherapy treatment planning pathway illustrated in Figure 1.1 can be of limited benefit in practice. When an attempt to introduce automation to a clinical workflow is made, it is therefore important to consider carefully how it fits within existing pathways and whether they need to be restructured or redesigned to maximise any benefit. Such a consideration is the focus of this section and Figure 6.1 presents an alternative to Figure 1.1 that incorporates auto-segmentation and KB hybrid-plan generation.

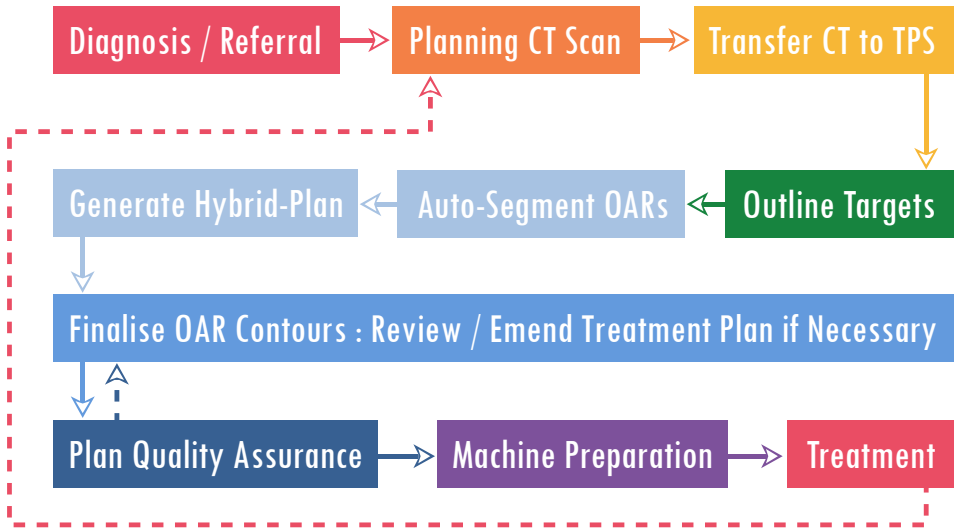

Figure 6.1: An alternative radiotherapy treatment planning pathway that incorporates auto-segmentation of OARs and generation of a hybrid-plan using the KB model of the idea prostate treatment plan.

In Figure 6.1, the lightest blue 'Auto-Segment OARs' and 'Generate Hybrid-Plan' boxes represent a fully automated part of the pathway. In the traditional workflow illustrated in Figure 1.1, a delay of several days can exist between 'Outline Target Volumes' and 'Outline OARs' / 'Treatment Plan Optimisation'. In contrast, the workflow proposed in Figure 6.1 gives an optimised hybrid-plan that is ready for review within approximately 30 minutes of the targets' being outlined. The work in this thesis has shown that for at least 80 % of cases, this hybrid-plan is clinically acceptable and so minimal human intervention is required at the 'Finalise OAR Contours : Review / Emend

Treatment Plan if Necessary' stage. Before any hybrid-plan can be accepted, the OAR auto-contours require correction so that the treatment plan being reviewed is reported to ground-truth anatomy. From the work presented in Chapter 5, for the 20 % of cases where the hybrid-plan is not clinically acceptable, re-optimisation using the KB model driven by the corrected OAR auto-contours will generate a clinically acceptable treatment plan with a success rate equivalent to manual treatment planning. This re-optimisation requires minimal human intervention and could be implemented as a single-click process in most TPSs. It is therefore reasonable to estimate that within 30 minutes of a treatment planner starting the 'Finalise OAR Contours : Review / Emend Treatment Plan if Necessary' stage, a final clinically acceptable treatment plan would be ready for the 'Plan Quality Assurance' stage.

This workflow provides a dramatic reduction in treatment planner involvement up until the quality assurance stage. Maintaining current manual quality assurance processes militates against any errors introduced by the new auto-segmentation and hybrid-planning approaches. The risk of such errors can often present a barrier to adopting new automated processes in the clinic and maintaining a well-established manual quality assurance procedure is a suitable way to ensure the benefits of the new workflow can be exploited whilst keeping risks minimal.

Finally, reducing the involvement of the treatment planner in the pathway would allow patients to flow through the pathway more quickly. As noted above this could be used to generate significant efficiencies for clinical departments, shorten the overall pathway to provide patients with earlier access to their treatments and enable routine ART strategies.

6.5 COVID-19 Impact Statement

The research projects presented in this thesis have been performed over a two year period from September 2018 to September 2020. A substantial proportion has therefore coincided with the global COVID-19 pandemic. Fortunately, the work itself has not been significantly impacted by any of the restrictions imposed nationally by the UK

Government or locally by The University of Manchester and The Christie NHS Foundation Trust. However, the intention during the first half of 2020, prior to the completion of this thesis, was to submit for journal publication articles describing:

1. The development and testing of a model of the ideal prostate treatment plan.
2. A fully automated KB treatment planning workflow.

Unfortunately, at the time of writing, it has not been possible to submit these articles. In the future, however, it certainly is the intention of the author to pursue these publications.

The major impact the COVID-19 situation has been the lack of contact with supervisors during the final six months of the project. Home working has been actively encouraged and although it has been possible to communication via email and video conferencing, these have not always been ideal replacements for face-to-face meetings and discussions – especially during the critical write-up period.

Conclusion

The purpose of this thesis was to investigate the role of automation in radiotherapy treatment planning for prostate cancer. In Chapter 1, the current treatment planning pathway, which incorporates minimal automation, was outlined and described in detail. Throughout the chapters of this thesis, attempts to automate stages of the pathway have been described and tested and this has culminated in a fully automated knowledge-based treatment planning pathway.

Errors in auto-segmentation of target volumes from which PTVs are expanded have been shown to have limited utility for treatment planning and need manual correction prior to treatment plan optimisation. This is unfortunate because it presents a bottleneck to the treatment planning pathway that current automation techniques cannot overcome. This is not true for OAR auto-contours, however. Even where OAR auto-contours contain known geometric inaccuracies, they can still be used reasonably reliably to optimise doses to actual anatomy. Although the OAR auto-contours may need

manual modification for final dose reporting, hybrid-planning allows the optimisation of the treatment plan to proceed prior to this. Therefore, in order to maximise the clinical benefit of OAR auto-segmentation, it needs to be accompanied by automated treatment plan optimisation.

A new method of knowledge-based treatment planning has been developed in this thesis to satisfy this automated treatment plan optimisation goal. The method was based on a large dataset of previously optimised treatment plans from which dose fall-off data were extracted and used to optimise new treatment plans. It has been extensively demonstrated that this knowledge-based method produces clinically acceptable treatment plans with success rate at least equivalent to manual treatment planning.

Combining the ideas of hybrid-planning with the knowledge-based treatment planning method generated clinically acceptable treatment plans in approximately 80 % of cases without human intervention (after the manual outlining of prostate and SV). This success rate could potentially be increased in the future using alternative auto-segmentation techniques.

The benefits of implementing automated knowledge-based hybrid-planning into routine clinical practice for a treatment planning department could be extensive. It would provide significant efficiency savings from a staffing perspective and would mean prostate treatment plans could be generated rapidly. This could overcome some of the barriers currently restricting the routine application of adaptive radiotherapy and potentially allow patients earlier access to their prostate cancer treatments.

D.Clin.Sci Appendix

Below is a summary of the work (in addition to this thesis) that has been undertaken by the author as part of the Doctor of Clinical Science programme. Individual assignment word counts and credits awarded are also provided.

Alliance Manchester Business School A Units

A1: Professionalism & Professional Development in the Healthcare Environment
30 Credits
Assignment 1 (3000 words)
Assignment 2 (3000 words)

A2: Theoretical Foundations of Leadership
20 Credits
Assignment 1 (3000 words)
Assignment 2 (3000 words)

A3: Personal & Professional Development to Enhance Performance
30 Credits
Assignment 1 (1500 words)
Assignment 2 (4000 words)

A4: Leadership & Quality Improvement in the Clinical & Scientific Environment
20 Credits
Assignment 1 (3000 words)
Assignment 2 (3000 words)

A5: Research & Innovation in Health & Social Care
20 Credits
Assignment 1 (3000 words)
Assignment 2 (3000 words)

Medical Physics B Units

B1: Medical Equipment Management
10 Credits
Assignment (3000 words)

B2: Clinical & Scientific Computing
10 Credits
Assignment (4000 words)

B3: Dosimetry
10 Credits
Assignment (4000 words)

B4: Optimisation in Radiotherapy & Imaging
10 Credits
Group Presentation
Assignment (2000 words)

B6: Medical Statistics in Medical Physics
10 Credits
Assignment (3000 words)

B8: Health Technology Assessment
10 Credits
Assignment (3000 words)

B9: Clinical Applications of Medical Imaging Technologies in Radiotherapy Physics
20 Credits
Group Presentation
Assignment (2000 words)

B10a: Advanced Radiobiology
10 Credits
Virtual Experiment
Report

B10c: Novel & Specialised External Beam Radiotherapy
10 Credits
Assignment (2000 words)

B10d: Advanced Brachytherapy
10 Credits
Group Presentation
Assignment (1500 words)

Generic B Units

B5: Contemporary Issues in Healthcare Science
20 Credits
Creative Project
Assignment (1500 words)

B7: Teaching Learning Assessment
20 Credits
20 minute Group Teaching Session

Section C

C1: Innovation Project
70 Credits
Literature Review & Innovation Proposal (5000 words)
Lay Presentation

References

Acosta, O., Mylona, E., Dain, M., Voisin, C., Lizee, T., Rigaud, B., Lafond, C., Gnep, K. & de Crevoisier, R. (2017). 'Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy', *Radiotherapy & Oncology*, 125(3), pp. 492-499.

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J. & Rueckert, D. (2009). 'Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy', *NeuroImage*, 46, pp. 726-738.

Aljabar, P. & Gooding, M. (2017). *The cutting edge: Delineating contours with Deep Learning*. Available at: http://mirada-medical.com/wp-content/uploads/2017/09/mm7121-0DLC-Technical-White-Paper.pdf, (Accessed: 6 November 2017).

Arsigny, V., Commowick, O., Pennec, X. & Ayache, N. (2006). 'A Log-Euclidean Polyaffine Framework for Locally Rigid or Affine Registration', *International Workshop of Biomedical Image Registration*, 4057, pp. 120-127.

Beasley, W., McWilliam, A., Aitkenhead, A., Mackay, R., & Rowbottom, C. (2016). 'The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy', *Journal of Applied Clinical Medical Physics*, 17(2), pp. 41-49.

Beekley Medical (2021). *CT-SPOT 120.* Available at: https://beekley.com/product-details/ct-spot-120, (Accessed: 14 April 2021).

Boylan, C., Golby, C. & Rowbottom, C. (2010). 'A VMAT planning solution for prostate patients using a commercial treatment planning system', *Physics Medicine Biology*, 55(14), pp. 395-404.

Bzdusek, K., Bystrov, D., Pekar, V., Peters, J., Schadewaldt, N., Schulz, H. & Vik, T. (2012). *Smart Probabilistic Contouring Engine (SPICE)*. The Netherlands: Koninklijke Philips Electronics.

Cancer Research UK (2016). *Prostate Cancer TNM Staging*. Available at: http://www.cancerresearchuk.org/about-cancer/prostate-cancer/stages/tnm-staging, (Accessed: 27 February 2018).

Cancer Research UK (2018). *Prostate cancer statistics*. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-Five, (Accessed: 7 March 2018).

Castadot, P., Lee, J., Parraga, A., Geets, X., Macq, B. & Grégoire, V. (2008). 'Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors', *Radiotherapy & Oncology*, 89, pp. 1-12.

Chanyavanich, V., Das S., Lee, W. & Lo, J. (2011). 'Knowledge-based IMRT treatment planning for prostate cancer', *Medical Physics*, 38(5), pp. 2515-2522.

Chung, H., Lee, B., Park, E., Lu, J. & Xia, P. (2008). 'Can all centres plan intensity modulated radiotherapy (IMRT) effectively? An external audit of dosimetric comparisons between three-dimensional conformal radiotherapy and IMRT for adjuvant chemoradiation for gastric cancer', *International Journal of Radiation Oncology Biology Physics*, 71(4), pp. 1167-1174.

Chuter, R., Whitfield, G., McWilliam, A. & Whitehurst, P. (2017). 'Auto-contouring comparison between Mirada, SPICE and ADMIRE for brain SRS patients', *UK Radiation Oncology conference 2017*. Available at: https://www5.shocklogic.com/scripts/jmevent/posterViewer.php?Client_Id=%27PP%27&Project_Id=%27UKRCO17%27&System_Id=1&Session_Code=, (Accessed: 15 February 2018).

Collier, D., Burnett, S., Amin, M., Bilton, S., Brooks, C., Ryan, A., Roniger, D., Tran, D. & Starkschall, G. (2003). 'Assessment of consistency in contouring of normal-tissue anatomic structures', *Journal of Applied Clinical Medical Physics*, 4(1), pp. 17-24.

Dearnaley, D., Syndikus, I., Mossop, H., Vincent, K., Birtle, A., Bloomfield, D., Graham, J., Kirkbride, P., Logue, J., Malik, Z., Money-Kyrle, J., M O'Sullivan, J., Panades, M., Parker, C., Patterson, H., Scrase, C., Staffurth, J., Stockdale, A., Tremlett, T., Bidmead, M., Mayles, M., Naismith, N., South, C., Gao, A., Cruickshank, C., Hassan, S., Pugh, J., Griffin, C. &Hall, E.

(2016). 'Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 17(8), pp. 1047-1060.

Dice, L. (1945). 'Measures of the amount of ecologic association between species', *Ecology*, 26(3), pp. 297-302.

Djajaputra, D., Wu, Q., Wu, Y. & Mohan, R. (2003). 'Algorithm and performance of a clinical IMRT beam-angle optimisation system', *Physics Medicine Biology*, 48(19), pp. 3191-3212.

Du, W., Cho, S., Zhang, X., Hoffman, K. & Kudchadker, R. (2014). 'Quantification of beam complexity in intensity-modulated radiation therapy treatment plans', *Medical Physics*, 41(2), pp. 021716:1-9.

Elekta (2013). *ABAS Atlas-Based Autosegmentation*. Available at: https://www.elekta.com/dam/jcr:7b262e2d-74c7-412c-bd1d-eec08da8183d/ABAS.pdf, (Accessed: 6 November 2017).

Good, D., Lo, J., Lee, R., Wu, J., Yin, F. & Das, S. (2013). 'A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning', *International Journal of Radiation Oncology Biology Physics*, 87(1), pp. 176-181.

Greenham, S., Dean, J., Fu, C., Goman, J., Mulligan, J., Tune, D., Sampson, D., Westhuyzen, J. & McKay, M. (2014). 'Evaluation of atlas-based auto-segmentation software in prostate cancer patients', *Journal of Medical Radiation Sciences*, 61, pp. 151-158.

Gugyeras, D., Farkas, A., Gulyban, A., Csima, M., Hadjiev, J. & Lakosi, F. (2017). 'Multi-Atlas based segmentation (Mirada RTx) vs. manual contouring for OARs in the head and neck region: efficiency time analysis', *South East Europe Technology in Radiation Oncology 2017*. Available at: https://www.researchgate.net/profile/Ferenc_Lakosi/project/Time-sparing-effect-of-multi-ABS-Single-vs-Multi-ABS-Atlas-Based-Segmentation/attachment/5a03823eb53d2fed8ad4336f/AS:558547001327616@1510179390631/download/Seetro2017.pdf?context=projectUpdatesLog, (Accessed: 19 February 2018).

Han, X., Hoogeman, M., Levendag, P., Hibbard, L., Teguh, D., Voet, P., Cowen, A., & Wolf, T. (2008). 'Atlas-based auto-segmentation of head and neck CT images', *Medical Image Computing and Computer-Assisted Intervention*, 11(2), pp. 434-441.

Hardcastle, N., van Elmpt, W., De Ruysscher, D., Bzdusek, K. & Tomé, W. (2013). 'Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy', *Radiation Oncology*, 8(1), pp. 243-250.

Heimann, T. & Meinzer, H. (2009). 'Statistical shape models for 3D medical image segmentation: a review', *Medical Image Analysis*, 13(4), pp. 543-563.

Hosni, A., Rosewall, T., Craig, T., Kong, V., Bayley, A., Berlin, A., Bristow, R., Catton, C., Warde, P. & Chung, P. (2017). 'The effect of bowel preparation regime on interfraction rectal filling variation during image guided radiotherapy for prostate cancer', *Radiation Oncology*, 12(50), pp. 1-6.

Hristov, D., Moftah, B., Charrois, C., Parker, W., Souhami, L. & Podgorsak, E. (2002). 'On the selection of optimization parameters for an inverse treatment planning replacement of a forward planning technique for prostate cancer', *Journal of Applied Clinical Medical Physics*, 3(3), pp. 200-211.

ICRU (1999). *ICRU Report 62*. Available at: https://academic.oup.com/jicru/article-abstract/os32/1/NP/2924047?redirectedFrom=fulltext, (Accessed: 5 December 2017).

ICRU (2010). *ICRU Report 83*. Available at: https://www.fnkv.cz/soubory/216/icru-83.pdf, (Accessed: 30 November 2018).

Janssen, T., Kusters, M., Wang, Y., Wortel, G., Monshouwer, R., Damen, E. & Petit, S. (In press). 'Independent knowledge-based treatment planning QA to audit Pinnacle autoplanning', *Radiotherapy & Oncology*.

Kumar. L., Yadav, G., Raman, K., Bhushan, M. & Pal, M. (2015). 'The dosimetric impact of different photon beam energy on RapidArc radiotherapy planning for cervix carcinoma', *Journal of Medical Physics*, 40(4), pp. 207-213.

Langmack, K., Perry, C., Sinstead, C., Mills, J. & Saunders, D. (2014). 'The utility of atlas-assisted segmentation in the male pelvis is dependent on the interobserver agreement of the structures segmented', *British Journal of Radiology*, 87: 20140299, pp. 1-6.

Lardue, A., Kadir, T. & Gooding, M. (2013). 'Assessment of Contour Fusion Methods for Auto-Contouring of the Male Pelvis', *International Journal of Radiation Oncology Biology Physics*, 87(20), S374.

Larrue, A., Kadir, T. & Gooding, M. (2013). 'SU-E-J-93: Assessment of Contour Fusion Method for Auto-Contouring of Head and Neck CT Images', *Medical Physics*, 40(6), pp. 171.

Li, D., Zang, P., Chai, X., Cui, Y., Li, R. & Xing, L. (2016). 'Automatic multiorgan segmentation in CT images of the male pelvis using region-specific hierarchical appearance cluster models', *Medical Physics*, 43(10), pp. 5426-5436.

Lim, J. & Leech, M. (2016). 'Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck', *ACTA Oncologica*, 55, 7, pp. 799-806.

Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B. & Sánchez, C. (2017). 'A survey on deep learning in medical image analysis', *Medical Image Analysis*, 42,
pp. 60-88.

Low, D., Harms, W., Mutic, S. & Purdy, J. (1998). 'A technique for the quantitative evaluation of dose distributions', *Medical Physics*, 25(5), pp. 656-661.

Mattes, M., Tai, C., Lee, A., Ashamalla, H. & Ikoro, N. (2014). 'The dosimetric effects of photon energy on the quality of prostate volumetric modulated arc therapy', *Practical Radiation Oncology*, 4(1), pp. e39-e34.

Mirada Medical (2017). *RT Imaging. Agile Workflows.* Available at: http://mirada-medical.com/wp-content/uploads/2017/11/mm6908-1-ROW-RT-Imaging-Brochuretoemail.pdf, (Accessed: 11 December 2017).

Mirada Medical (2021). *DLCExport: Automating Radiotherapy Imaging Workflows.* Available at: https://mirada-medical.com/product/dlcexpert/, (Accessed: 15 April 2021).

Moiseenko, V., Liu, M., Kristensen, S., Gelowitz, G. & Berthelet, E. (2007). 'Effect of bladder filling on doses to prostate and organs at risk: a treatment planning study', *Journal of Applied Clinical Medical Physics*, 8(1), pp. 55-68.

National Cancer Registration & Analysis Service (2018). *Chemotherapy, Radiotherapy and Surgical Tumour Resections in England*, Available at: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/main_cancer_treatments, (Accessed: 7 March 2018).

Nijkamp, J., Pos, F., Nuver, T., de Jong, R., Remeijer, P., Sonke, JJ. & Lebesque, J. (2008). 'Adaptive radiotherapy for prostate cancer using kilovoltage cone-beam computed tomography: first clinical results', *International Journal of Radiation Oncology Biology Physics*, 70(1), pp. 75-82.

Nwankwo, O., Sihono, D., Schneider, F. & Wenz, F. (2014). 'A global quality assurance system for personalized radiation therapy treatment planning for the prostate (or other sites)', *Physics in Medicine & Biology*, 59, pp. 5575-5597.

Nwankwo, O., Mekdash, H., Sihono, D., Wenz, F. & Glatting, G. (2015). 'Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning', *Radiation Oncology*, 10(111), pp. 1-5.

Orlandini, L., Coppola, M., Fulcheri, C., Cernusco, L., Wang, P. & Cionini, L. (2017). 'Dose tracking assessment for image-guided radiotherapy of the prostate bed and the impact on clinical workflow', *Radiation Oncology*, 12(78), pp. 1-8.

Özsavas, E., Telatar, Z., Dirican, B., Sağer, Ö & Beyzadeoğlu, M. (2014). 'Automatic Segmentation of Anatomical Structures from CT Scans of Thorax for RTP', *Computational and Mathematical Methods in Medicine*, 2014, pp. 1-14.

Pekar, V., Allaire, S., Kim, J. & Jaffray, A. (2010). *Head and Neck Auto-Segmentation challenge.* Available at: http://www.midasjournal.org/browse/publication/703, (Accessed: 21 November

2017).

Peressutti, D., Schipaanboord, B. & van Soest, J. Lustberg, T., van Elmpt, W., Kadir, T., Dekker, A. & Gooding, M. (2016). 'How Effective Are Current Atlas Selection Methods for Atlas-Based Auto-Contouring in Radiotherapy Planning?', *Medical Physics*, 43(6), pp. 3738-3739.

Petit, S., Wu, B., Kazhdan, M., Dekker, A., Simari, P., Kumar, R., Taylor. R., Herman, J. & McNutt, T. (2012). 'Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma', *Radiotherapy Oncology*, 102(1), pp. 38-44.

Powis, R., Bird, A., Brennan, M., Hinks, S., Newman, H., Reed, K., Sage, J. & Webster, G. (2017). 'Clinical implementation of a knowledge based planning tool for prostate VMAT', *Radiation Oncology*, 12(81), pp. 1-8.

Qazi, A., Pekar, V., Kim, J., Xie, J., Breen, S. & Jaffray, D. (2011). 'Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach', *Medical Physics*, 38(11), pp. 6160-6170.

Seim, H., Kainmueller, D., Heller, M., Lamecker, H., Zachow, S. & Hege, H. (2008). 'Automatic Segmentation of the Pelvic Bones from CT Data Based on a Statistical Shape Model', *Eurographics Workshop on Visual Computing for Biomedicine*. Available at: https://www.zib.de/zachow/publications/vcbm2008.pdf.

Shang, J., Kong, W., Wang, Y., Ding, Z., Yan, G. & Zhe, H. (2014). 'VMAT planning study in rectal cancer patients', *Radiation Oncology*, 9, pp. 1-7.

Sharp, G., Fritscher, K., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H. & Yang, J. (2014). 'Vision 20/20: Perspectives on automated image segmentation for radiotherapy', *Medical Physics*, 41(5), pp. (050902)1-13.

Speight, R., Karakaya, E., Prestwich, R., Sen, M., Lindsay, R., Harding, R. & Sykes, J. (2014). 'Evaluation of atlas based auto-segmentation for head and neck target volume delineation in adaptive/replan IMRT', *Journal of Physics*, 489, pp. 1-4.

Stanley, D., Popp, T., Ha, C., Swanson, G., Eng, T., Papanikolaou, N. & Gutiérrez, A. (2015). 'Dosimetric effect of photon beam energy on volumetric modulated arc therapy treatment plan quality due to body habitus in advanced prostate cancer', *Practical Radiation Oncology*, 5(6), pp. e625-e633.

Stapleford, L., Lawson, J., Perkins, C., Edelman, S., Davis, L., McDonald, M., Waller, A., Schreibmann, E. & Fox, T. (2010). 'Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer', *International Journal of Radiation Oncology Biology Physics*, 77(3), pp. 959-966.

Sykes, J. (2014). 'Reflections on the current status of commercial automated segmentation systems in clinical practice', *Journal of Medical Radiation Sciences*, 61(3), pp. 131-134.

Thomson, D., Boylan, C., Liptrot, T., Aitkenhead, A., Lee, L., Yap, B., Sykes, A., Rowbottom, C. & Slevin, N. (2014). 'Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk', *Radiation Oncology*, 3(9), pp. 1-12.

Teguh, D., Levendag, P., Voet, P., Al-Mamgani, A., Han, X., Wolf, T., Hibbard, L.,Nowak, P., Akhiat, H., Dirkx, M., Hiejmen, B. & Hoogeman, M. (2011). 'Clinical Validation of Atlas-Based Automatic Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck', *International Journal of Radiation Oncology Biology Physics*, 81(4), pp. 950-957.

The Institute of Cancer Research (2020). *PIVOTALboost: A phase III randomised controlled trial of prostate and pelvis versus prostate alone radiotherapy with or without prostate boost.* Available at: https://www.icr.ac.uk/our-research/centres-and-collaborations/centres-at-the-icr/clinical-trials-and-statistics-unit/clinical-trials/pivotalboost, (Accessed: 20 April 2020).

The Royal College of Radiologists, Society & College of Radiographers, Institute of Physics & Engineering in Medicine, National Patient Safety Agency & British Institute of Radiology (2008). *Towards Safer Radiotherapy*. Available at: https://www.rcr.ac.uk/system/files/publication/field_publication_files/Towards_saferRT_final.pdf, (Accessed: 14 February 2019).

The Royal College of Radiologists (2014). *Recommendations for cross-sectional imaging in cancer management, second edition: prostate tumours.* Available at: https://www.rcr.ac.uk/system/files/publication/field_publication_files/BFCR(14)2_15_Prostate.pdf, (Accessed: 10 February 2018).

The Royal College of Radiologists (2016). *Radiotherapy Dose Fractionation Second Edition.* Available at: https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfco163_dose_fractionation_2nd_ed_march2017.pdf, (Accessed: 30 November 2018).

The Royal College of Radiologists (2017). *Radiotherapy target volume definition and peer review.* Available at: https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfco172_peer_review_outlining.pdf, (Accessed: 10 February 2018).

Thor, M., Petersen, J., Bentzen, L., Høyer, M. & Muren, L. (2011). 'Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer', *Acta Oncologica*, 50(6) pp. 918-925.

Voet, P., Dirkx, M., Teguh, D., Hoogeman, M., Levendag, P. & Heijmen, B. (2011). 'Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis', *Radiotherapy & Oncology*, 98, pp. 373-377.

Wang, Y., Heijmen, B. & Petit, S. (2017). 'Prospective clinical validation of independent DVH prediction for plan QA in automatic treatment planning for prostate cancer patients', *Radiotherapy & Oncology*, 125, pp. 500-506.

Warfield, S., Zou, K. & Wells, W. (2004). 'Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation', *IEEE Transactions on Medical Imaging*, 23(7), pp. 903-921.

Webb, S. (2003). 'Use of a quantitative index of beam modulation to characterize dose conformality: Illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy', *Physics Medicine Biology*, 48, pp. 2051-2062

Wilkins, A., Mossop, H., Syndikus, I., Khoo, K., Bloomfield, D., Parker, C., Logue, J., Scrase, C., Patterson, H., Birtle, A., Staffurth, J., Malik, Z., Panades, M., Eswar, C., Graham, J., Russell, M., Kirkbride, P., O'Sullivan, J., Gao, A., Cruickshank, C., Griffin, C., Dearnaley, D. & Hall, E. (2015). 'Hypofractionated radiotherapy versus conventionally fractionated radiotherapy for patients with intermediate-risk localised prostate cancer: 2-year patient-reported outcomes of the randomised, non-inferiority, phase 3 CHHiP trial', *The Lancet*, 16(16), pp. 1605-1616.

Wood, M., Fonseca, A., Sampson, D., Kovendy, A., Westhuyzen, J., Shakespeare, T. & Turnbull, K. (2016). 'Prostate intensity-modulated radiotherapy planning in seven mouse clicks: Development of a class solution for cancer', *Reports of Practical Oncology and Radiotherapy*, 21, pp. 567-570.

Wu, B., Ricchetti, F., Sanguineti, G., Kazhdan, M., Simari, P., Chuang, M., Taylor, R., Jacques, R. & McNutt, T. (2009). 'Patient geometry-driven information retrieval for IMRT treatment plan quality control', *Medical Physics*, 36(12), pp. 5497-5505.

Yahya, S., Zarkar, A., Southgate, E., Nightingale, P. & Webster, G. (2013). 'Which bowel preparation is best? Comparison of a high-fibre diet leaflet, daily microenema and no preparation in prostate cancer patients treated with radical radiotherapy to assess the effect on planned target volume shifts due to rectal distension', *British Journal of Radiology*, 86(1031), pp. 1-9.

Yang, Y., Ford, E., Wu, B., Pinkawa, M., van Triest, B., Campbell, P., Song, D. & McNutt, T. (2013). 'A overlap-volume-histogram based method for rectal dose prediction and automated treatment planning n the external beam prostate radiotherapy following hydrogel injection', *Medical Physics*, 40(1), pp. 1-10.

Yuan, L., Ge, Y., Lee, W., Yin, F., Kirkpatrick, J. & Wu, J. (2012). 'Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans', *Medical Physics*, 39(11), pp. 6868-6878.

Zhu, X., Ge, Y., Li, T., Thongphiew, D., Yen, F. & Wu, Q. (2011). 'A planning quality evaluation tool

for prostate adaptive IMRT based on machine learning', *Medical Physics*, 38(2), pp. 719-726.

Zhu, M., Bzdusek, K. Brink, C., Eriksen, J., Hansen, O., Jensen, H., Gay, H., Thorstad, W., Widder, J., Brouwer, C., Steenbakkers, R., Vanhauten, H., Cao, J., McBrayne, G., Patel, S., Cannon, D., Hardcastle, N., Tomé, W., Guckenberg, M. & Parikh, P. (2013). 'Multi-institutional Quantitative Evaluation and Clinical Validation of Smart Probabilistic Image Contouring Engine (SPICE) Autosegmentation of Target Structures and Normal Tissues on Computer Tomography Images in the Head and Neck, Thorax, Liver, and Male Pelvis Areas', *International Journal of Radiation Oncology Biology Physics*, 87(4), pp. 809-816.

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J. & Rueckert, D. (2009). 'Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy', *NeuroImage*, 46, pp. 726-738.

Collier, D., Burnett, S., Amin, M., Bilton, S., Brooks, C., Ryan, A., Roniger, D., Tran, D. & Starkschall, G. (2003). 'Assessment of consistency in contouring of normal-tissue anatomic structures', *Journal of Applied Clinical Medical Physics*, 4(1), pp. 17-24.

Dice, L. (1945). 'Measures of the amount of ecologic association between species', *Ecology*, 26(3), pp. 297-302.

Gugyeras, D., Farkas, A., Gulyban, A., Csima, M., Hadjiev, J. & Lakosi, F. (2017). 'Multi-Atlas based segmentation (Mirada RTx) vs. manual contouring for OARs in the head and neck region: efficiency time analysis', *South East Europe Technology in Radiation Oncology 2017*. Available at: https://www.researchgate.net/profile/Ferenc_Lakosi/project/Time-sparing-effect-of-multi-ABS-Single-vs-Multi-ABS-Atlas-Based-Segmentation/attachment/5a03823eb53d2fed8ad4336f/AS:558547001327616@1510179390631/download/Seetro2017.pdf?context=projectUpdatesLog, (Accessed: 19 February 2018).

Han, X., Hoogeman, M., Levendag, P., Hibbard, L., Teguh, D., Voet, P., Cowen, A., & Wolf, T. (2008). 'Atlas-based auto-segmentation of head and neck CT images', *Medical Image Computing and Computer-Assisted Intervention*, 11(2), pp. 434-441.

Langmack, K., Perry, C., Sinstead, C., Mills, J. & Saunders, D. (2014). 'The utility of atlas-assisted segmentation in the male pelvis is dependent on the interobserver agreement of the structures segmented', *British Journal of Radiology*, 87: 20140299, pp. 1-6.

Larrue, A., Kadir, T. & Gooding, M. (2013). 'SU-E-J-93: Assessment of Contour Fusion Method for Auto-Contouring of Head and Neck CT Images', *Medical Physics*, 40(6), pp. 171.

Lim, J. & Leech, M. (2016). 'Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck', *ACTA Oncologica*, 55, 7, pp. 799-806.

Qazi, A., Pekar, V., Kim, J., Xie, J., Breen, S. & Jaffray, D. (2011). 'Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach', *Medical Physics*, 38(11), pp. 6160-6170.

Sharp, G., Fritscher, K., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H. & Yang, J. (2014). 'Vision 20/20: Perspectives on automated image segmentation for radiotherapy', *Medical Physics*, 41(5), pp. (050902)1-13.

Speight, R., Karakaya, E., Prestwich, R., Sen, M., Lindsay, R., Harding, R. & Sykes, J. (2014). 'Evaluation of atlas based auto-segmentation for head and neck target volume delineation in adaptive/replan IMRT', *Journal of Physics*, 489, pp. 1-4.

Teguh, D., Levendag, P., Voet, P., Al-Mamgani, A., Han, X., Wolf, T., Hibbard, L.,Nowak, P., Akhiat, H., Dirkx, M., Hiejmen, B. & Hoogeman, M. (2011). 'Clinical Validation of Atlas-Based Automatic Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck', *International Journal of Radiation Oncology Biology Physics*, 81(4), pp. 950-957.

Voet, P., Dirkx, M., Teguh, D., Hoogeman, M., Levendag, P. & Heijmen, B. (2011). 'Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis', *Radiotherapy & Oncology*, 98, pp. 373-377.

Zhu, M., Bzdusek, K. Brink, C., Eriksen, J., Hansen, O., Jensen, H., Gay, H., Thorstad, W., Widder, J., Brouwer, C., Steenbakkers, R., Vanhauten, H., Cao, J., McBrayne, G., Patel, S., Cannon, D., Hardcastle, N., Tomé, W., Guckenberg, M. & Parikh, P. (2013). 'Multi-institutional

Quantitative Evaluation and Clinical Validation of Smart Probabilistic Image Contouring Engine (SPICE) Autosegmentation of Target Structures and Normal Tissues on Computer Tomography Images in the Head and Neck, Thorax, Liver, and Male Pelvis Areas', *International Journal of Radiation Oncology Biology Physics*, 87(4), pp. 809-816.