Development of Pareto Guided Automated Radiotherapy Treatment Planning and its Application to Prostate Cancer

A thesis submitted to the University of Manchester for the degree of Doctor of Clinical Science in the Faculty of Biology, Medicine and Health

2021

Philip A. Wheeler

School of Medical Sciences

Table of Contents

Word count: 22484

List of Figures

List of Tables

1 Abstract

Background and purpose: Current automated planning methods do not allow for the intuitive exploration of clinical trade-offs during calibration. This work introduces a novel automated planning solution, which aimed to address this problem through incorporating Pareto navigation techniques into the calibration process. The efficacy of this new solution was evaluated for prostate cancer patients with and without elective nodal irradiation.

Materials and methods: The developed automated planning methodology was as follows. For each tumour site a set of planning goals is defined. Utilising Pareto navigation techniques an operator calibrates the solution through intuitively exploring different treatment options: selecting the optimum balancing of competing planning goals for the given site. Once calibrated, fully automated plan generation is possible, with specific algorithms implemented to ensure trade-off balancing of new patients is consistent with that during calibration. Using the proposed methodology the system was calibrated for prostate and seminal vesicle (PSV), and prostate and pelvic node (PPN) treatments. For 40 randomly selected patients (20 PSV and 20 PPN) automatically generated plans ($VMAT_{Auto}$) were compared against plans created by expert dosimetrists under clinical conditions ($VMAT_{Clinical}$) and no time pressures ($VMAT_{Ideal}$). Plans were compared through quantitative comparison of dosimetric parameters and blind review by an oncologist.

Results: Upon blind review 39/40 and 33/40 $VMAT_{Auto}$ plans were considered preferable or equal to $VMAT_{Clinical}$ and $VMAT_{Ideal}$ respectively, with all deemed clinically acceptable. Dosimetrically, $VMAT_{Auto}$, $VMAT_{Clinical}$ and $VMAT_{Ideal}$ were similar, with observed differences generally of low clinical significance. Compared to $VMAT_{Clinical}$, $VMAT_{Auto}$ reduced hands-on planning time by 94% and 79% for PSV and PPN respectively. Total planning time was significantly reduced from 22.2 mins to 14.0 mins for PSV, with no significant reduction observed for PPN.

Conclusions: An automated planning methodology with a Pareto navigation based calibration has been developed, enabling the complex balancing of competing trade-offs to be intuitively incorporated into automated protocols. It was successfully applied to two sites of differing complexity and robustly generated high quality plans in an efficient manner.

2 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.

3 Copyright Statement

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 s/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.

4 Acknowledgements

I would like to thank the following people whose input and support throughout this research were invaluable:

The project team at Velindre Cancer Centre whose collaborative working led to both the successful completion of this research and the positive impact on patient care through clinical introduction of the automated solution developed in this work.

My supervisors at both Velindre Cancer Centre and University of Manchester for their continued support throughout the project. Special thanks to Prof Geraint Lewis whose input in the final stages was key in getting this thesis over the line.

Sue Campbell and Jim Fitzgibbon who represented the views of the patient and public throughout the project.

Finally to my family, Lucie, Luca and Sofia…we did it…whoop…whoop!

5 The Author

The author of this thesis is a state registered Clinical Scientist (CS18020) and Medical Physics Expert (932) working at Velindre Cancer Centre. Previous education includes:

- **MSc Radiation Physics with Medical Applications (Distinction)**
  University College London (UCL)

- **BSc Physics (Class I)**
  University of Warwick

6 Submission Format

This thesis has been constructed in the Manchester University journal format in accordance with the requirements of School of Medical Sciences.

Following journal format guidance the thesis is separated into three main sections. Section 1 (chapter 7) presents an introduction to the field of advanced radiotherapy planning, a review of the literature and discussion of key gaps in the existing evidence base. Section 2 (chapters 8-9) presents two published journal articles which fully describe the body of research work undertaken. Finally section 3 (chapter 10) presents a critical appraisal of the research as a whole.

The content of the journal articles presented in this thesis are slightly modified from the original published versions. As per journal format guidelines, their presentation has been edited to align with the overall thesis format such that a single coherent body of work is created. In addition, for each journal article the contribution of each author has been explicitly defined. The full changes made to the articles were:

- Inclusion of a table defining author contributions.
- Formatting changed to match thesis, including section\figure\table numbering.
- Figure 6 caption amended to include definitions of DMPO and imax.
- Addition of paragraph in section 8.4.2.3 of article 1 to justify the values of variables used in the PB-AIO optimsation framework.
- Details on how the navigation patient was chosen added to the methods section of article 1.
- Supplementary information included as thesis appendices.
- Referencing sections removed; a single reference list is used across the whole thesis (excluding appendices).

The reader is pointed to the original published versions to access high resolution images of the article's figures.

As part of the requirements of the Doctorate in Clinical Science this thesis also includes an innovation proposal (Appendix D). The purpose of the proposal is to conceive and present an innovation that has the potential to make a positive contribution to healthcare. In this case,

the proposed innovation is utilisation of automated planning to improve plan quality assurance within radiotherapy clinical trials.

Finally Appendix E presents a summary of the all the additional modules undertaken as part of the Doctorate in Clinical Science.

7 Introduction

7.1 Overview

Radiotherapy is a technologically advanced, clinically effective treatment, which is indicated for half of all cancer patients at some point during their care [1]. All patients undergoing radiotherapy require a bespoke, personalised, treatment plan to be generated prior to treatment. The effectiveness of radiotherapy is highly dependent on the quality of this treatment plan. Poor quality radiotherapy negatively affects patient outcomes: it increases the risks of treatment failure; increases overall mortality; and detrimentally impacts the patient's quality of life [2–5].

Plan generation is complex and resource intensive, requiring specialist radiotherapy simulation software operated by expert staff to ensure plans are safe and clinically effective. Current planning methods are heavily reliant on time consuming manual interactions by the operator. This not only hinders the efficiency of the process, but also forms a dependence on the quality of the resultant plan with the expertise of the operator. This leads to substantial variations in plan quality both at intra- and inter-institutional level [6].

In recent years there has been an increased focus on making the radiotherapy planning process more automated and intuitive to the operator, with the aims of improving quality and efficiency, and reducing variation. This section reviews the current state-of-the-art in radiotherapy plan generation. The review is limited to intensity modulated radiotherapy (IMRT), which is widely acknowledged as the standard of care for radical radiotherapy due to its superior tissue sparing over conventional techniques [7,8]. Within this thesis the term IMRT is defined as being inclusive of specific modulated delivery techniques such as volumetric arc therapy (VMAT). Through this review, the differing plan generation methods are critiqued and key gaps in the literature identified, which informed the aims and objectives of the presented doctoral research project.

7.2 Background

7.2.1 The Radiotherapy Planning Process

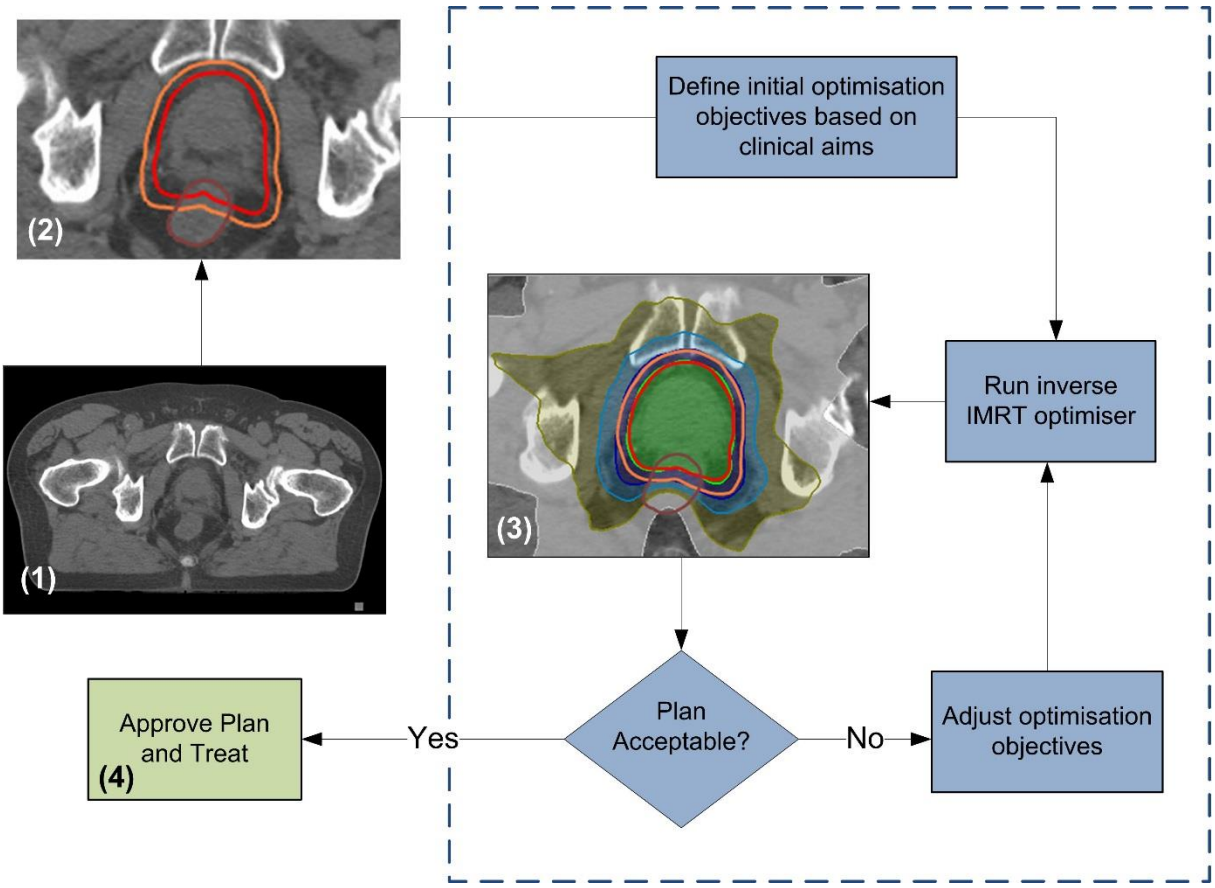


Figure 1: *Overview of the treatment planning process. (1) Patient is CT scanned. (2) Radiotherapy planning targets and organs at risk are delineated on the CT. (3) Treatment plan generated using Inverse IMRT optimisation: operator defines optimisation objectives, inverse optimiser generates treatment plan, operator reviews plan, if plan is unacceptable objectives are adjusted and plan re-optimised until a suitable plan is obtained. (4) Final plan is approved and used to treat patient.*

Radiotherapy plans are created on patient computed tomography (CT) scans using specialist computer simulation treatment planning software (TPS) and aim to maximise the radiation dose to the cancer, whilst avoiding sensitive organs. Figure 1 presents an overview of the planning process for IMRT treatments. Due to the high number of delivery variables, complex IMRT plans are generated using inverse optimisation algorithms, where plans are automatically generated based on a set of user defined constraints and objectives. Whilst the plan generation process is automated, it requires constraints and objectives to be appropriately defined such that a clinically optimal plan is produced. This is a non-trivial problem. In practice, the inverse optimisation process is highly iterative, requiring expert operators using manual trial and error techniques to hone in on objectives and constraints

which yield clinically acceptable plans. Traditional inverse manual planning (MP) is resource intensive, dependent on operator experience [9] and may yield plans which are clinically acceptable but not optimal.

The resultant radiotherapy plans are defined by a series of treatment control points, or segments, which are delivered sequentially. For each segment the following machine parameters are defined:

- the quantity of radiation to be delivered, specified in terms of monitor units (MU)
- the machine orientation (i.e. gantry and collimator angle)
- The x-ray aperture, as defined by the individual positioning of the machine's multi-leaf collimators (MLCs))

Importantly, when generating radiotherapy plans, the resultant dose distribution is not the only factor to consider, the deliverability of the plans in terms of dosimetric accuracy is also vital to ensure treatment efficacy. Plans where a high proportion of the MU is delivered through multiple small or complex apertures are considered to have a high plan complexity (or modulation) and are typically more challenging to model and deliver accurately. Plan complexity can be quantified through a range of different metrics, the most simplistic being plan MU due to more complex plans (with correspondingly smaller apertures) requiring more MU to deliver the same dose prescription [10].

7.2.2 Pareto Optimality

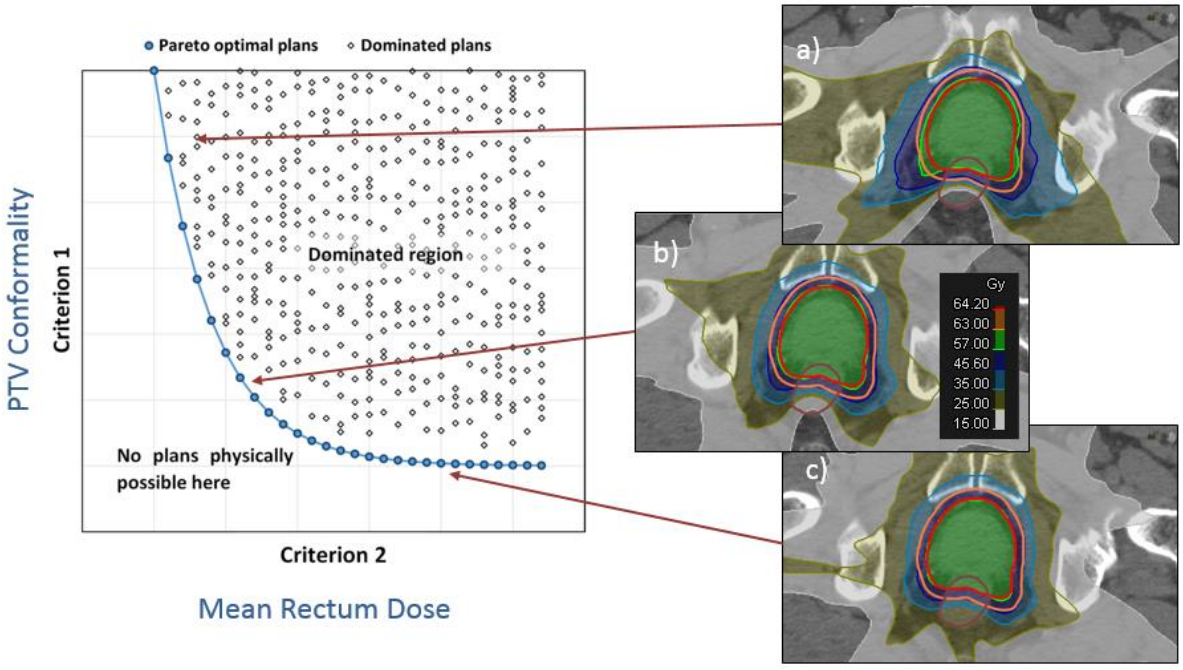


Figure 2 : *Schematic diagram demonstrating the concept of Pareto optimality for prostate cancer treatments. When considering the dose metrics of PTV conformality and mean rectum dose, all three plans (a, b and c) lie on the Pareto front and are mathematically optimal, it is the responsibility of the decision maker to determine which plan is clinically optimal. (Graph representing Pareto front taken from Hussein et al.* [11]*).*

When considering the quality of IMRT plans, an important concept is that of Pareto optimality. Plans are considered Pareto optimal when improvement of one objective can only be made at the detriment to another. If this condition is not true the solution is considered 'dominated' and thus sub-optimal. For a given optimisation problem there is an infinite set of Pareto optimal plans, which define the 'Pareto front' (Figure 2). Whilst any solution which lies on the Pareto front is mathematically optimal, for radiotherapy plan generation the desired solution is the position on the front which is considered clinically optimal by the treating oncologist. Thus there are two fundamental aims within the plan generation process: firstly, the clinically optimal position on the Pareto front must be identified, ideally in an intuitive manner, and secondly, plan generation methods must adequately generate plans which are both Pareto optimal and lie within the clinically desirable region of the front.

7.3 Review of Advanced Planning Techniques

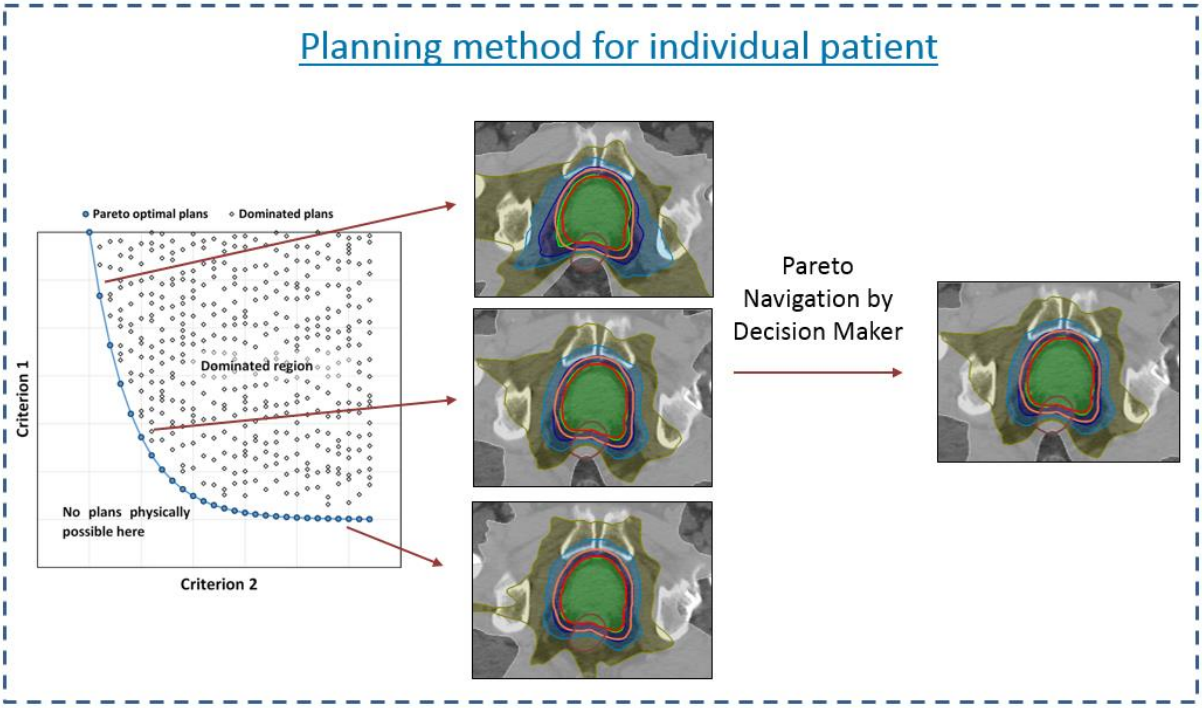


Figure 3: *Schematic representation of the 'a posteriori' advanced planning technique. For each patient a database of Pareto optimal plans is automatically generated. Using an appropriate GUI the decision maker navigates the Pareto surface and selects the clinically optimal plan for that individual patient. (Graph representing Pareto front taken from Hussein et al.* [11]*).*

Within the literature advanced planning techniques fall under two broad methodologies, defined by whether the clinical decision making is made before (*a priori*) or after (*a posteriori*) the plan generation process. *A posteriori* methodologies (Figure 3) assume that the clinically optimal position on the Pareto front for a given patient is either unknown, or cannot be reliably obtained automatically. *A posteriori* techniques therefore generate a range of Pareto optimal plans on a per patient basis, which are reviewed by the decision maker who selects the most clinically optimum solution. In contrast, *a priori* methodologies assume that the clinically optimum position on the Pareto front is known for all patients within a given treatment site and can be obtained automatically. For *a priori* methods (Figure 4) a single plan is automatically produced for an individual patient, with site specific templates or plan generation methodologies 'calibrated' *a priori* such that the single plan lies within the clinically desirable region of the Pareto front. As *a priori* methodologies generate a single plan fully automatically, for the remainder of this document they will be defined by the more general terminology 'automated planning' (AP). Furthermore, in the context of AP

the term ‘calibration’ is defined as the process of protocol-specific modification of any adjustable AP parameter or input, including the training of machine learning models, such that a clinically acceptable AP solution for a given treatment protocol is obtained.

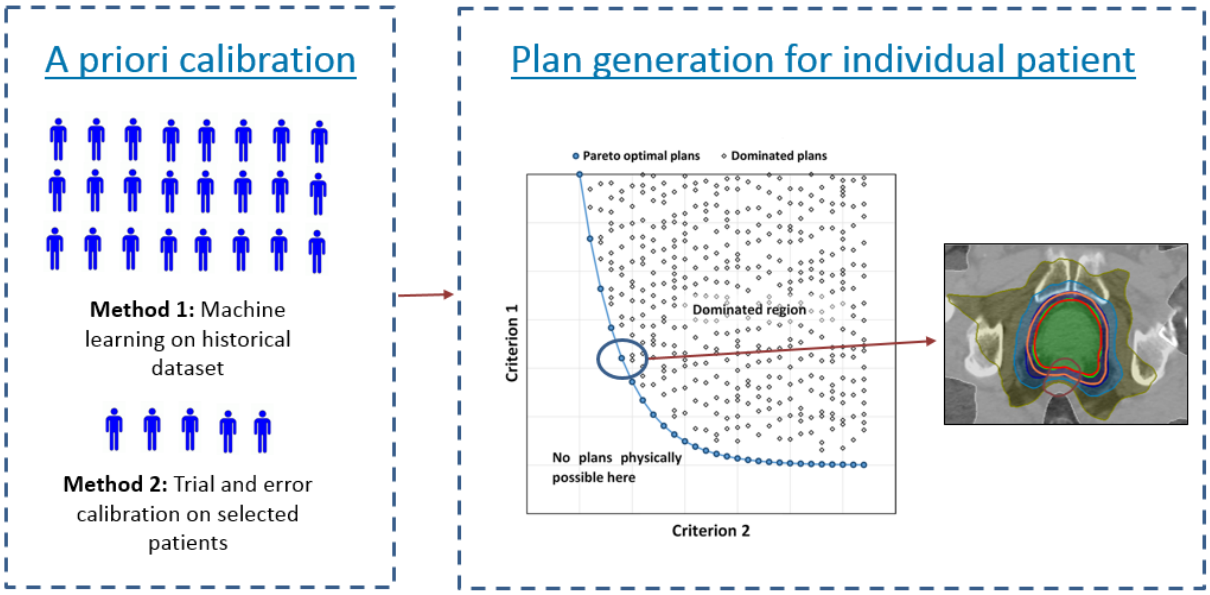


Figure 4: *Schematic representation of the a priori advanced planning technique. For a given treatment site an a priori calibration is performed using machine learning or trial and error. The aim of the calibration is to target automated solutions towards the clinical desirable part of the Pareto front. Following calibration a single automated plan can be generated for new patients. (Graph representing Pareto front taken from Hussein et al.* [11]*).*

What follows is a review of the currently implemented *a posteriori and AP* techniques presented in the literature. Studies within this review were identified though the literature search methodology presented in Appendix A for publications between 2018-2021, with studies pre 2018 identified from the review performed by Hussein *et al* [11].

7.3.1 Posteriori Optimisation

As discussed, the aim of *a posteriori* methodologies is to automatically generate a database of Pareto optimal plans which sample the Pareto front. The decision maker (e.g. oncologist or dosimetrist) then interactively navigates the Pareto front using a navigation star [12] or sliders [13] to select the clinically optimum solution. During navigation, convex combinations (a linear combination where all coefficients are non-negative and sum to 1) of neighbouring plans provide real time estimates of the dose distribution and DVHs at the navigated point on the Pareto front, which informs the decision making. Once the clinically optimum point on the front is determined by the operator, the navigated solution is converted into a

deliverable treatment plan. Within the literature this form of optimisation is termed *posteriori* multi-criteria optimisation ($MCO_{posteriori}$).

Craft *et al* [13] and Thieke *et al.* [12] provided the first clinical evaluations of IMRT $MCO_{posteriori}$ solutions through the application to individual cancer cases of brain, prostate, lung and paraspinal meningioma. In these 'proof of principle studies' no meaningful comparison with standard manual planning was made. $MCO_{posteriori}$ has since been incorporated into a number of commercial planning systems leading to a range of clinical studies evaluating its efficacy against MP for prostate, head and neck, brain, spine, lung and pancreatic cancer [14–25].

With the exception of a single study where $MCO_{posteriori}$ led to marginal increases in rectal doses for prostate and pelvic node radiotherapy [17], all studies reported that $MCO_{posteriori}$ yields plans of comparable [14,16,18–20] or superior quality [15,19,21–25] to MP, but with significant reductions in planning time of approximately 70% - 90% [20,22,23]. Several studies highlight that the low resource, intuitive nature of $MCO_{posteriori}$ could enable planning to be driven by oncologists, which in turn should yield plans more congruent with the oncologist's clinical aims [23,26]. This hypothesis was tested in Craft *et al.*'s [23] seminal paper evaluating the application of $MCO_{posteriori}$ to pancreatic and glioblastoma patients. The study demonstrated that intuitively navigating differing treatment options through $MCO_{posteriori}$ led the oncologist to select trade-offs which compromised PTV homogeneity to a greater extent than MP such that organ at risk (OAR) doses were reduced. Across all 10 patients $MCO_{posteriori}$ plans were considered superior to MP. For head and neck cancer, Kierkels *et al.* [20] also demonstrated that $MCO_{posteriori}$ enabled less experienced planners to efficiently produce high quality treatment plans. This is considered a key advantage over MP where plan quality has been shown to correlate with planner experience [9].

Despite the conceptual advantages of $MCO_{posteriori}$ there are some key limitations in the current implementations of this technique. Firstly, the navigated point on the Pareto front is represented by a 3D dose distribution and as such contains no information on the treatment plan parameters required to deliver that dose. A separate step, which typically utilises the inverse optimisation engine, is therefore required to convert the navigated dose distribution into a deliverable plan. This however is not a lossless process and depending on the implementation method can lead to substantial, clinically significant detriments in plan

quality [14,27]. Secondly, during Pareto surface navigation the navigated dose distribution is approximated through the convex combination of neighbouring Pareto plans. Depending on the number of plans which sample the Pareto front this can result in an 'approximation error' in the displayed navigated dose, which is propagated across to the final segmented plan. The approximation error is defined as the difference between the navigated/displayed dose and the dose of its corresponding Pareto optimal plan that, by definition, lies on the Pareto front. Sampling strategies such as the 'sandwich algorithm', which target areas of high approximation error when generating the Pareto plan set, aid in reducing this error [28]. However, especially for complex tumour sites where the dimensionality of the problem is high, the error may be significant. Typical approximation errors are <5% and <15% for adequality sampled (n > 75) and sparsely sampled (n = number of objectives +1) datasets respectively [29].

When reviewing the literature on $MCO_{posteriori}$ what is most apparent is the low number of reported clinical studies for a technique which has clear theoretical advantages and was released as a commercial solution in 2012. In Hussein *et al.*'s systematic review on innovations in IMRT planning [11] only five clinical studies were identified for $MCO_{posteriori}$ [14–16,20,25] compared to 73 for AP. This current review identified an additional seven studies [17–19,21–24]. Importantly, of these 12 studies, four have dataset sizes of <10 patients per tumour site, five studies use between 10-19 patients, three studies have datasets of 20-30 patients and one utilises multiple datasets containing between 7-20 patients across 4 different clinical sites. Aside from the low patient numbers, five of the studies have a strong probability of bias, with $MCO_{posteriori}$ plans generated in a different planning system to the MP comparator [14,15,20,23]. Furthermore, one study provides insufficient detail on patient selection and results to provide any meaningful confidence in the study's conclusion [18]. In summary, since the first $MCO_{posteriori}$ commercial product was released in 2012, only 2 high quality studies (datasets of 20 or more patients, comparators generated within the same planning system), have been identified [21,22]. Both studies are for the clinical site of head and neck and demonstrate improved plan quality through reduction in OAR doses with $MCO_{posteriori}$.

It is assumed that the limitations discussed in this review and in the literature [14,27] may be hindering the realisation of the theoretical benefits of the technique, resulting in the low

number of supporting publications. Interestingly, despite these limitations, when comparing $MCO_{posteriori}$ to MP only one negative result has been published for the site of prostate and pelvic nodes [17]. It is hypothesised this could be due to an underlying publication bias, where negative results are either not submitted for publication or not accepted.

7.3.2 Automated Planning

Automated planning (AP) has been successfully implemented using a range of differing techniques, which fall into 3 broad categories: knowledge based planning (KBP), sequential ε-constraint planning (εc), which is also defined as *a priori* multi-criteria optimisation, and protocol based automatic iterative optimisation (PB-AIO) [11].

7.3.2.1 Knowledge Based Planning

KBP utilises databases of historical treatment plans to inform the optimisation of new patients. Within automated planning the most common and effective methods train machine learning algorithms on previously treated patients, which then predict the optimum dose volume histograms (DVH) [30–32] or dose distributions [33–35] for new patients. The predicted DVH or dose distribution is then utilised during the automated planning process to tailor the plan optimisation to the individual patient. Common optimisation methods include use of DVH line objective functions or voxel based dose mimicking functions [34,36], which seek to drive the dose towards the predicted DVH or voxel dose respectively. Alternatively the predicted DVH or dose distribution can be used as an input to an 'inverse optimisation pipeline' [37] that outputs a set of optimisation objectives, which are used in standard treatment planning optimisers to generate a deliverable plan [32,35].

KBP is the most widely validated automated solution in the literature, forming 83% of all clinical studies identified by Hussein *et al.* [11] in their systematic review on automation. The majority of these studies utilise the commercial KBP product RapidPlan (Varian, Palo Alto), which is based on a DVH prediction and line objective function methodology. KBP has been applied to head and neck, prostate, cervical, lung, spinal metastasis, breast cancer, upper gastrointestinal (GI) and lower GI cancers, with all studies reporting KBP maintaining or slightly improving plan quality [11,38]. While the quality of plans generated with KBP has been comprehensively evaluated, there remains minimal evidence on the technique's potential efficiency saving compared to MP. A systematic review on KBP identified only 6 studies (out of 73), which contained some form of timing analysis, with the authors concluding that current evidence provides only 'preliminary' data on efficiency [38].

When compared to MP, KBP has a substantive supportive evidence base, however the technique has some key limitations. KBP generally requires large historical datasets, which

may not be present for novel techniques or prescriptions. A systematic review of KBP studies identified that across all clinical sites the mean size of training datasets was ~100 patients [38]. More specifically for prostate cancer, Boutilier *et al* demonstrate that for DVH prediction algorithms, which are the most common KBP approach, datasets of >75 patients are required to ensure all OAR DVHs are modelled appropriately [39]. Models are also strongly dependent on the optimality and consistency of plans in the training dataset [40], which is not guaranteed as they are predominantly generated using traditional manual planning methods with known limitations [6,9]. Hussein *et al.* [41] highlight that KBP solutions require considerable tuning to deliver suitable solutions. Resultant plans are also highly dependent on the makeup of the training patient datasets, with Tol *et al.* [42] demonstrating that two independent 30 patient models, populated with different patients from a 60 patient head and neck cohort, yielded markedly different results for individual patients. Furthermore, even if models are trained against Pareto optimal patient datasets, modelling uncertainty results in clinically relevant discrepancies in predicted metrics [43]. Finally, by implementing DVH or voxel based mimicking objectives the prioritisation of individual ROIs is removed, thus the optimiser may make clinically relevant compromises to critical OARs (e.g. maximum dose to rectum) in order to improve the global mimicking of dose. It is considered that these limitations hinder the optimality of KBP solutions, leading to only small improvements in plan quality when compared to MP as evidenced by a KBP systematic review and data synthesis by Ge and Wu [38].

7.3.2.2 Sequential ε-constraint Planning

Sequential εc planning utilises an optimisation approach where clinical objectives are minimised in strict sequential order, generating treatment plans fully autonomously. This technique has been implemented within the automated solution 'iCycle' [44] and evaluated for prostate, head and neck, spinal metastasis, gastric, cervical, rectal and lung cancers [45–53]. Optimisations involving iCycle are based on a tumour site specific 'wish list' containing user defined hard constraints (which must not be violated) and objectives, which are ordered in terms of priority. iCycle begins with a pre-optimisation in fluence space, where objectives are minimised in sequential order to generate a Pareto optimal plan. Based on this pre-optimisation, Lagrange multipliers are utilised to convert the result into an equivalent weighted sum optimisation problem, which is optimised in the commercial

planning system Monaco (Elekta, Stockholm) to generate the final deliverable plan. The two-stage process is fully automated.

εc has been evaluated in both single institutional [45–51,53] and multi institutional settings [52]. Across all the reported studies, on a population basis εc was considered equivalent or superior to MP, however for lung cancer 15% of AP required manual intervention to be clinically acceptable [45]. Unlike KBP, in some cases marked improvements in dose metrics were observed, for example, when applied to prostate and pelvic node cancer, mean bladder and rectum doses were reduced by 10.7 Gy and 4.5 Gy respectively [49]. It is likely these observed improvements are due to εc algorithms driving solutions towards Pareto optimality, in contrast to KBP which nominally aims to reproduce the quality of the historical datasets. Interestingly though, in the five studies where the plan monitor units (MU) were reported, εc resulted in a substantial increase in MU (29%) for four studies [47,49,50,52]. A moderate reduction in MU (13%) was reported for one study, which was unique in that it was a simple clinical situation (rectal cancer) with only one organ at risk [53]. Craft *et al.* have previously demonstrated a link between OAR sparing and MU [54], suggesting that the observed improvement in plan quality, for the complex cases at least, may partially be due to increased modulation and not solely due to superiority of the technique. In terms of efficiency, the εc process is fully automated and therefore offers a substantial reduction in staff 'hands on' time compared to manual approaches.

εc does have a number of limitations. Firstly, generating the 'wish list' is an iterative process, reliant on trial and error [11]. This process is analogous to traditional MP, but at the patient cohort level, therefore as with MP the resultant solution may not optimally align to oncologists' clinical preference. This iterative process is also time-consuming. Secondly, plan generation time scales linearly with the number of criteria in the 'wish list'; for complex treatments (e.g. head and neck) planning time can be in the order of hours [55]. Thirdly, due to the two-stage plan generation process, there may be clinically relevant discrepancies between the Pareto optimal fluence based plan, and the final segmented solution. Finally, it is assumed that a single 'wish list' is optimum for every patient in the treatment cohort, which may not be the case as even if plans are Pareto optimal, they may not be clinically optimal for that individual patient.

Recent work by van Haveren *et al.* [56,57] has however sought to address limitations in the optimisation and calibration efficiency of the εc methodology. In terms of optimisation speed, a lexicographic reference point method (LRPM) has been developed where optimisation criteria defined in the 'wish list' are optimised within a single optimisation problem, rather than the sequential approach used with εc. The LRPM method has been applied to prostate and head and neck with planning time reduced by a factor of 10 and 21 respectively at no detriment to plan quality [56,57]. For calibration, a methodology has been developed whereby LRPM solutions are automatically configured using a dataset of historical treatment plans, with the purpose of removing the manual iterative process of protocol calibration. The method has been validated for prostate and head and neck, and demonstrated acceptable solutions can be calibrated automatically, with limited dataset sizes ($n > 9$) [58,59]. Whilst this approach is promising in terms of improving the time burden of calibrations, the methodology still required a user preferences template as an input, with the authors acknowledging that this input will need iterative adaptation, resulting in a semi-automated, not automated approach [58]. Furthermore, as with KBP, the balancing of competing trade-offs will be based on a historical dataset, which may not be fully aligned with the clinician's clinical preference.

7.3.2.3 Protocol Based Automatic Iterative Optimisation

For PB-AIO solutions the general principle is to load an initial set of objectives/constraints into the planning system's native optimiser, then iteratively adapt these parameters during the optimisation process to drive the solution towards a clinically acceptable plan. Whilst a broad range of automation methodologies can be categorised as PB-AIO, across the literature there are two core implementation strategies. The first ($PB\text{-}AIO_A$) utilises a trial and error approach, akin to the process undertaken by expert staff during manual planning [60–63]. The broad concept is as follows. For a given treatment site a set of hard and soft clinical goals is defined either within a site specific template, or for single site solutions, hard coded into the automated algorithms. Based on these goals an initial optimisation is performed that aims to meet all hard clinical goals, with soft goals largely ignored. Multiple optimisation rounds follow where the objectives for each soft clinical goal are iteratively adapted to minimise the dose to normal tissue, whilst not violating the hard constraints. If hard constraints are violated, the objectives are reset to values just before the violation

occurred. During this process auxiliary ROIs may be generated to improve the geometric specificity of the optimisation, for example to minimise regions of over/under dose. This approach has been implemented within the commercial planning system Pinnacle (Philips Radiation Oncology Systems, Fitchburg) (PB-AIO$_{Pinnacle}$). The second methodology (PB-AIO$_B$) loads a set of predefined optimisation objectives. During the optimisation process the value or weight of objectives related to soft clinical goals are then iteratively updated such that either the distance between the objective and the DVH line is held constant [64,65], or the resultant objective function value nominally equals a predefined target (OV$_T$) [66–68]. The effect of both approaches is to ensure the soft objectives form a significant but not dominant contribution to the composite objective function value. Using this approach soft objectives are minimised, without compromising the higher weighted hard objectives.

PB-AIO has been successfully applied to breast, oesophagus, head and neck, liver, Hodgkin lymphoma, prostate, brain, oesophagus, gynaecological, rectal and lung treatments [61,63–86]. Of the clinical evaluation studies identified in this review 17/25 evaluated PB-AIO$_{Pinnacle}$ [69–75,77–86]. Across all 25 studies, 13 demonstrated superiority of PB-AIO compared to MP in terms of plan quality [61,63,65,66,68,69,73,74,79–82,85], 11 demonstrated non-inferiority [64,67,70,71,75–78,83,84,86] and one yielded mixed results where only 75% of plans were considered clinically acceptable upon review [72]. With approximately 50% of the identified studies yielding superiority over MP these results, as with εc, are suggestive that the active minimisation of OAR doses is preferential to KBP, which only seeks to mimic previous treatments.

The limitations of PB-AIO are as follows. Firstly, many of the PB-AIO solutions presented in the literature are specific to a single site [61,66,68,87], with only PB-AIO$_{Pinnacle}$ being easily applicable to a wide range of differing clinical sites. Secondly, as with εc, calibrating an acceptable PB-AIO solution is an iterative process, with no intuitive method to balance competing trade-offs. For PB-AIO$_B$ solutions, balancing is defined through the manual specification of either OV$_T$ or the weights within the initial set of optimisation objectives, both of which have minimal relevance to clinical prioritisation. For PB-AIO$_{Pinnacle}$, instead of objective weights, clinical goals are assigned priorities (low, medium, high), however, correctly setting these values is still iterative and requires expert input from treatment planners [73]. Finally, there is evidence that PB-AIO$_{Pinnacle}$, which forms 68% of the clinical

evaluations in the literature, may yield suboptimal results for a significant cohort of clinical patients. In a lung study by Vanderstraeten *et al.* [72] only 75% of patients were clinically approvable. Additionally, Zhang *et al.* [70] identify a correlation between anatomical separation (brainstem vs CTV) and the optimality of the resultant plan. Finally an audit of PB-AIO$_{Pinnalce}$ plan quality across two institutions using a KBP audit tool identified suboptimal clinical plans in 8% (centre A) and 25% (centre B) of clinical prostate plans [88].

7.3.2.4 Summary of Automated Methodologies

	KBP	εc	PB-AIO
Methodology	Does prediction via machine learning	Minimisation of clinical objectives in strict sequential order	Iterative adaptation of parameters during plan optimisation
Clinical Application	Head and neck Prostate Lung Cervical Spinal metastasis Breast cancer Upper GI Lower GI	Head and neck Prostate Lung Spinal metastasis Gastric Cervical Rectal	Head and neck Prostate Lung Breast Oesophagus Liver Hodgkin lymphoma Brain Oesophagus Gynaecological Rectal
Plan Quality vs MP	Equivalent or marginally superior	Equivalent or superior	Equivalent or superior
Training Dataset Requirements	Large datasets (typically n > 75) Delineated CT & treatment plan	Small datasets (n ≤ 10) Delineated CT	Small datasets (n ≤ 10) Delineated CT
Training/Calibration Methodology	Machine learning on historical datasets	Trial and error or machine learning on historical datasets	Trial and error
Advantages	Commercially available Conceptually simple	Generates Pareto optimal plans	Commercially available Actively minimises OAR doses
Disadvantages	Quality is dependent on training dataset OAR doses not actively minimised. Calibration via machine learning is an iterative process	Not commercially available Calibration via trial and error is challenging and time consuming Calibration via machine learning is an iterative process	Calibration via trial and error is challenging and time consuming Methodology can yield suboptimal solutions for individual patients

Table 1: *Summary of the different automated planning methodologies*

7.3.3 Hybrid Approaches to Advanced Planning

The advanced planning approaches discussed in section 7.3.1 and 7.3.2 are not mutually exclusive and combinations of advanced techniques may offer potential advantages. What follows is a review of the different hybrid advanced planning solutions in the literature.

For KBP, which is strongly reliant on the quality of the training dataset, utilising advanced techniques to generate a 'gold standard' training dataset has the potential to yield improved models. Miguel-Chumacero *et al.* [21] and Wall *et al.* [89] presented data for head and neck, and prostate cancer respectively, comparing KBP models trained on MP and $MCO_{posterior}$ datasets. Their results demonstrated that training with $MCO_{posteriori}$ led to substantial reductions in OAR doses, with rectum and parotid mean doses reduced by 9.4 Gy and 5.7 Gy respectively. Lin *et al.* [90] utilised a constrained hierarchal optimisation (CHO), a methodology analogous to εc, to automatically generate a KBP training dataset for prostate bed radiotherapy. Whilst KBP models based on the CHO dataset generated equivalent plans compared to a traditional MP dataset, CHO based datasets could be autonomously generated, enabling the potential of rapid adaptability to changes in radiotherapy protocols or planning techniques. Finally Wang *et al.* [91] and Chatterjee *et al.* [92] presented an iterative approach to developing KBP models, where models trained on MP datasets are subsequently refined through selected re-training against their own KBP output if the KBP is superior to the original MP. In this way, it is hypothesised that KBP models will trend towards Pareto optimality. This method was shown to yield models with increased OAR sparing and reduced prediction uncertainty [92] when compared to models trained on a MP dataset. However, this method was also prone to overfitting, leading to inconsistent results on novel patients [91].

Use of $MCO_{posteriori}$ techniques downstream of KBP generation to improve the quality of the final plan has also been investigated for head and neck cancer [21]. Results demonstrated this hybrid approach improved plan quality compared to KBP trained on MP (KBP_{MP}) or $MCO_{posteriori}$ (KBP_{MCO}) datasets, with the largest improvements observed when comparing to KBP_{MP}.

For PB-$AIO_{Pinnacle}$, KBP techniques have been implemented upstream of the core automated planning engine with the aim of predicting dose metrics for individual patients and tailoring

the automated planning protocol based on these results. This technique has been successfully applied to hepatocellular and oesophageal cancer, with result demonstrating non-inferiority to MP [93,94]

For εc type methodologies (including LRPM), as discussed in section 7.3.2.2 a KBP approach has been developed to automatically calibrate a 'wish list' based on historical treatment plans. Furthermore, through automatically adjusting parameters within an εc 'wish list', Pareto fronts for individual patients can be generated. Whilst this has not yet been implemented as an $MCO_{posteriori}$ solution, the method has been utilised to evaluate the impact of trade-off balancing on KBP models [40] and rectal toxicity for prostate cancer [95] .

Finally PB-$AIO_{Pinnacle}$ has been used upstream of the $MCO_{posteriori}$ planning process to generate individualised $MCO_{posteriori}$ templates (used for Pareto dataset generation) based on the DVH obtained from AP [93]. In this way, the sampling of the Pareto surface is tailored to the individual, focusing on the clinically relevant areas. When compared to $MCO_{posteriori}$ plans generated with population based templates, this method yielded plans of similar quality but with a 10.5% and 8.4% reduction in dose to the glottis larynx and pharyngeal constrictor muscles respectively.

7.3.4 Application of Advanced Techniques to Prostate Cancer

Prostate cancer accounts for 13% of all new cancer cases in the UK [96] and radiotherapy is a key treatment modality in the management of this disease. At Velindre Cancer Centre prostate cancer accounts for ~19% of all radical radiotherapy treatments. It is therefore a good candidate site to pilot and evaluate the novel automated solution that was developed within this project. What follows is a summary of the evidence on the application of advanced techniques to intact prostate and seminal vesicle radiotherapy, both with (PPN) and without (PSV) elective nodal irradiation.

For PSV radiotherapy a thorough summary comparing AP with MP has been presented by Heijmen *et al.* [52]; with 12 studies identified as demonstrating small differences between automated and manual plans [41,51,61,82,97–104], and only Heijmen *et al.'s* more recent multi-centre study showing the overall dosimetric superiority of automation through reduced rectum doses [52]. Since the publication of this summary a further 11 studies have been identified, five demonstrating the dosimetric superiority [63,74,105–107] and six the

non-inferiority [108–113] of AP compared to MP. For $MCO_{posteriori}$, excluding studies with less than 10 patients, only 2 studies comparing $MCO_{posteriori}$ with MP for PSV have been identified. Both studies were performed on only 10 patients and concluded that $MCO_{posteriori}$ was non-inferior to MP [14,17].

For the more complex situation of PPN radiotherapy, only five studies which present results comparing MP with AP have been identified [49,74,75,92,108]. Of these studies three demonstrate dosimetric superiority [49,74,92] of AP though reduced OAR doses and two non-inferiority [75,108]. In some cases the reduction in OAR doses was substantial; Buschmann *et al.* showed AP reduced mean bladder and rectum doses by 10.7 Gy and 4.5 Gy respectively [49]. Across these five studies, two had low sample sizes ($n \leq 12$) [75,108] and one allowed manual correction of the AP post generation [74], therefore only two of the five identified studies are considered be considered high quality and robust [49,92]. For $MCO_{posteriori}$, only one clinical study with a sample size ≥ 10 was identified. The study concluded that $MCO_{posteriori}$ generated high quality treatment plans with minimal workload, however OAR sparing was slightly improved with MP [17].

In summary, for PSV there is a wide body of evidence demonstrating non-inferiority, and on some occasions superiority, of AP compared to MP. For PPN, the results are similarly supportive of AP, but the evidence base is substantially smaller with only five studies identified. For $MCO_{posteriori}$, the evidence base is very weak with only three studies identified across both PSV and PPN, all with a sample size of 10 patients. The studies indicate $MCO_{posteriori}$ yields plans of a clinically acceptable quality, yet due to the low number of studies and small sample size no robust assessment of the technique's non-inferiority to MP can be made.

7.4 Discussion: Key Gaps and Unanswered Questions

In terms of advanced planning techniques, there is a clear gap in the utilisation of $MCO_{posteriori}$ techniques within the AP calibration process. In all the AP approaches presented in the literature $MCO_{positeriori}$ has only been utilised as part of the calibration process for KBP (via an $MCO_{posteriori}$ database). In the two studies which investigated this approach, both indicated substantial improvements in the KBP model with $MCO_{posteriori}$. For both εc and PB-AIO however, no solution has yet been developed which enables the intuitive navigation of

competing trade-offs during the set-up phase for a given clinical site. The approaches currently rely heavily on trial and error to align solutions to oncologists' clinical preference, which is not an optimal approach. Implementing $MCO_{posteriori}$, during the setup phase of automated solutions could reduce this reliance on trial and error, and help ensure solutions are congruent with oncologists' treatment aims. This innovation is likely to be especially powerful for ɛc and PB-AIO methodologies, where only small calibration datasets are required, as the $MCO_{posteriori}$ approach could focus on accuracy (e.g. minimising the approximation error via generation of large Pareto datasets), rather than efficiency (required for $MCO_{posteriori}$ in the clinical environment or in the generation of large KBP training databases).

In reviewing the evaluative studies for both $MCO_{posteriori}$ and AP a number of deficiencies were observed. As discussed in section 7.3.1, for $MCO_{posteriori}$ there are limited studies evaluating its efficacy and questions over publication bias. There is therefore a clear need to conduct further robust studies on $MCO_{posteriori}$ that aim to publish results regardless of the outcome. Across both MCO and AP techniques the majority of studies are single institutional, with only a small number of multicentre studies identified [52,97,111,114,115]. The potential benefit of minimising inter-institutional variation though AP has also not been adequately explored. A further weakness in the evidence base is the majority of evaluative studies use the clinically treated MP as the comparator, meaning that results could be biased by local institutional planning expertise, which is known to vary, and suboptimal planning due to clinical time pressures. Finally, there are few studies directly comparing differing advanced planning techniques, with only six studies identified [19,21,94,114,116,117]. Whilst direct comparisons are complex due to many solutions being implemented within differing planning systems, to better understand their relative advantage there is a real need for robust multi-solution studies to be undertaken.

More specifically for the site of prostate cancer, there is a dearth of high quality studies evaluating the efficacy of $MCO_{posteriori}$ for both PSV and PPN. Furthermore, whilst there is a substantive evidence in support of AP for PSV, for PPN there is scope for a strengthening of the evidence base, with only 5 studies identified across all AP techniques.

In summary, within the literature the following key gaps on advanced planning techniques have been identified:

- No εc or PB-AIO technique utilises $MCO_{posteriori}$ during the calibration process.
- There are a small number of studies evaluating the application of advanced techniques to PPN radiotherapy.
- There is a dearth of high quality $MCO_{posteriori}$ evaluative studies.
- There are a small number of multi-institutional studies.
- There are minimal studies comparing differing advanced techniques.

The work presented in this thesis aimed to address the first two of these gaps in evidence. Firstly through developing an automated planning solution based on a $MCO_{posteriori}$ calibration process and secondly through evaluating the solution for both PSV and PPN radiotherapy treatments. What follows are two published journal articles detailing the work undertaken. The first article details, in full, the developed algorithms and presents an initial assessment of their application to PSV plan generation. The second article presents a comprehensive clinical study for both PSV and PPN treatments, comparing the new automated technique with plans generated manually by expert dosimetrists.

8 Journal Article 1: Utilisation of Pareto Navigation Techniques to Calibrate a Fully Automated Radiotherapy Treatment Planning Solution

Published in: *Physics and Imaging in Radiation Oncology, Volume 10, April 2019, Pages 41-48*

8.1 Author List and Contribution

Author	Contribution
Philip A. Wheeler[1]	Designed and built EdgeVcc. Designed study. Calibrated the PSV EdgeVcc solution with Rosemary Holmes. Performed all evaluation work including data analysis. Author of paper (including all figures, tables and supplementary information).
Michael Chu[1]	Advised on translational aspects of EdgeVcc; ensuring alignment with requirements of the treatment planning clinical service.
Rosemary Holmes[1]	Developed data analysis package in python, which was utilised as part of the statistical and graphical analysis. Calibrated the PSV EdgeVcc solution with Philip Wheeler.
Maeve Smyth[1]	Developed prototype Pareto navigation module, which was adapted and incorporated into EdgeVcc by Philip Wheeler.
Rhydian Maggs[1]	Advised on translational aspects of EdgeVcc; ensuring alignment with requirements of the treatment planning clinical service.
Emiliano Spezi[2]	Advised on study design, provided academic oversight of the project
John Staffurth[3]	Provided clinical governance across the study, including advising and reviewing the quality of generated plans.
David G. Lewis[1]	Advised on study design, provided academic oversight of the project with a focus on statistical methods
Anthony E. Millin[1]	Advised on study design, provided academic oversight of the project, advised on the design of EdgeVcc from a technical perspective to ensure safe translation to the clinical environment.

[1] Department of Medical Physics, Velindre Cancer Centre, Cardiff, United Kingdom.
[2] School of Engineering, Cardiff University, Cardiff, United Kingdom.
[3] School of Medicine, Cardiff University, Cardiff, United Kingdom.

All authors contributed to the review and revision of the manuscript.

8.2 Abstract

Background and purpose: Current automated radiotherapy planning solutions do not allow for the intuitive exploration of different treatment options during protocol calibration. This work introduces an automated planning solution, which aims to address this problem through incorporating Pareto navigation techniques into the calibration process.

Materials and methods: For each tumour site a set of planning goals is defined. Utilising Pareto navigation techniques an operator calibrates the solution through intuitively exploring different treatment options: selecting the optimum balancing of competing planning goals for the given site. Once calibrated, fully automated plan generation is possible, with specific algorithms implemented to ensure trade-off balancing of new patients is consistent with that during calibration. Using the proposed methodology the system was calibrated for prostate and seminal vesicle treatments. The resultant solution was validated through quantitatively comparing the dose distribution of automatically generated plans ($VMAT_{Auto}$) against the previous clinical plan, for ten randomly selected patients.

Results: $VMAT_{Auto}$ yielded statistically significant improvements in: PTV conformity indices, high dose bladder metrics, mean bowel dose, and the majority of rectum dose metrics. Of particular note was the reduction in mean rectum dose (median 25.1Gy vs. 27.5 Gy), rectum $V_{24.3Gy}$ (median 41.1% vs. 46.4%), and improvement in the conformity index for the primary PTV (median 0.86 vs. 0.79). Dosimetric improvements were not at the cost of other dose metrics.

Conclusions: An automated planning methodology with a Pareto navigation based calibration has been developed, which enables the complex balancing of competing trade-offs to be intuitively incorporated into automated protocols.

8.3 Introduction

Inverse radiotherapy planning is a time consuming, iterative process where optimal plan quality is not guaranteed [6]. A solution to this problem is automated planning (AP) where high quality treatment plans are generated fully autonomously [47,49–51,66,67,72,79,118,119]. AP has been implemented using a range of methodologies, which can be categorised within the following three broad domains: knowledge based planning

(KBP), sequential ε-constraint optimisations (εc) and protocol based automatic iterative optimisations (PB-AIO).

KBP utilises information from previously treated patients to inform the optimisation of future patients. The most common methods use machine learning algorithms, trained on databases of historical treatment plans, to predict the achievable dose distribution [33,34] or dose volume histograms [30,31] for new patients. This information is utilised during the inverse optimisation process to generate a plan whose dose distribution best matches that predicted.

εc generates plans according to a list of prioritised clinical goals, which are minimised in strict sequential order under the condition that lower priority goals must not compromise higher priority goals. Through the appropriate selection and ordering of goals, a single prioritised list can generate desirably balanced plans for individual patients within a given treatment site [49,51].

Finally, PB-AIO techniques load an initial set of objectives (either hard coded or derived from a site specific template) into the planning system's native optimiser. During the optimisation process, automated algorithms iteratively adjust objectives based on information from the optimised dose distribution to tailor the plan to the desired clinical aims [61,66,68]. Specific examples include: regularly updating objective positions such that a constant distance below the corresponding DVH line is maintained [65] and modulating objective weights such that the function's objective value (OV) tends towards a target OV during the course of the optimisation [66,67]. Through implementing these methods of dynamic objective adjustment it has been shown that a single set of initial objectives can yield plans with minimised organ at risk (OAR) doses and consistently balanced trade-offs across all patients within a given treatment site [65–67]. In this manuscript, we define objectives whose weight and position are modified in this specific manner as 'dynamic objectives'.

A key challenge in all three approaches is adequately and intuitively capturing the oncologist's experience and decision making during the calibration process, such that automated plans are congruent with clinical preference. KBP is dependent on the optimality of large datasets of previous clinical plans, which is not guaranteed, and both εc and PB-AIO rely on trial and error to develop and refine automated protocols [11]. In this paper a novel

solution to this problem is proposed through integrating Pareto navigation techniques directly into the calibration process.

Pareto navigation enables operators to intuitively explore differing treatment plan options such that an informed choice can be made on the optimal balancing of competing trade-offs [12,13]. Navigation is performed on a pre-calculated set of Pareto optimal plans, which aim to sample clinically relevant parts of the Pareto surface. In this regard, a treatment plan is considered Pareto optimal when improvement of a given trade-off is only possible at the detriment of another, with the Pareto surface being represented by an infinite set of such plans. On an individual patient basis, intuitively exploring the Pareto surface through Pareto navigation has shown to reduce the need for trial and error, and yield plans more congruent with oncologists' clinical preferences [23]. It is expected that incorporating Pareto navigation into the automated planning calibration process will yield similar benefits at the patient cohort level.

The purpose of this work is to present the methodology of EdgeVcc (Experience Driven plan Generation Engine by Velindre Cancer Centre): a PB-AIO based automated planning solution, designed to be applicable across a range of radiotherapy treatment sites and uniquely calibrated using Pareto navigation principles. The first section of this paper provides a detailed description of the proposed methodology alongside its associated algorithms. The second section presents an example of its application to the tumour site of prostate and seminal vesicles (PSV).

8.4 Methods and Materials

8.4.1 Patient Dataset

20 patients previously treated at Velindre Cancer Centre between July and December 2015 were randomly selected into a calibration (n=10) and validation (n=10) cohort. Patients were planned on computed tomography scans of 3 mm slice thickness with prostate, seminal vesicles, rectum, bladder and bowel delineated. Prostate + seminal vesicles were expanded 10 mm isotropically to form the planning target volume PTV48 and prostate expanded by 5mm (6 mm craniocaudally) to form PTV60, with the PTV's suffix denoting its prescribed dose in Gy. All study patients were previously treated with VMAT on Elekta Agility linear accelerators with treatment plans generated manually in Oncentra Masterplan (v4.3, Elekta

Ltd, Crawley) using a single 6MV 360˚ arc, simultaneous integrated boost technique. Treatments were prescribed for 20 fractions, and manually planned to local clinical goals (supplementary table 1) using a class solution based methodology.

8.4.2 Automated System

8.4.2.1 System Overview

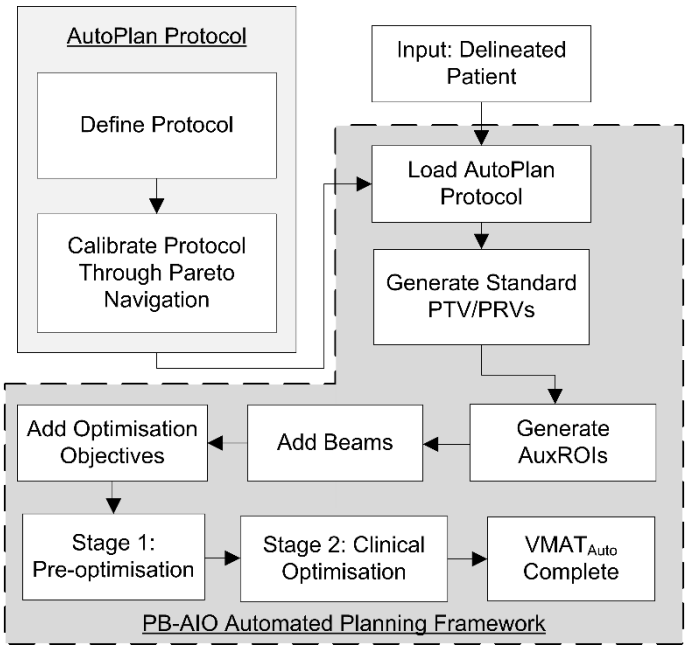


Figure 5: *Flowchart depicting the workflow of the proposed solution, with all items within the PB-AIO framework (as represented by the dashed area) fully automated. For each tumour site a calibrated AutoPlan protocol is required.*

The proposed solution (Figure 5) was developed in the treatment planning system RayStation (Raysearch Laboratories, Stockholm) using custom python scripts. For each tumour site a set of planning goals is defined within an 'AutoPlan protocol' (section 8.4.2.2). On a selected calibration patient(s) Pareto navigation techniques are utilised to derive a set of planning goal weighting factors, which correspond to a clinically desirable point on the Pareto surface for the given patient (section 8.4.2.4). The weighting factors are stored within the AutoPlan protocol, which then forms the input for automated planning of new patients. Automated plan generation is based on a PB-AIO framework (section 8.4.2.3) that utilises 'dynamic objectives' to ensure OAR doses are minimised and trade-off balancing for new patients is consistent with that selected during protocol calibration.

8.4.2.2 AutoPlan Protocol

The treatment modality, beam arrangement, standard PTV and planning volume at risk (PRV) margins, and planning goals are defined within the AutoPlan protocol. Planning goals guide the optimisation process and are stratified into three priority levels: primary normal tissue goals (P_1), target goals (P_2) and trade-off goals (P_3). The optimisation methodology aims to meet goals in order of their priority, with compromise to target goals permissible by P_1 but not P_3. Trade-off goals are assigned a group number, which determines the order in which they are explored during the calibration process. Goals of the same parameter type and clinical relevance (e.g. low dose rectum objectives) are grouped to reduce degrees of freedom during calibration. The planning goals for PSV are presented in the supplementary table 2.

Planning goals are designed to be clinically intuitive, with no specification of weighting factors required. Weighting factors are instead derived through two distinct processes. For P_1 and P_2, the clinical preference across all tumour sites when balancing conflicting goals is explicitly defined: target coverage is compromised to maintain normal tissue goals. Conflicting goals are therefore explicitly handled through region of interest (ROI) retraction algorithms, enabling weights to be defined by simple hard coded algorithms. In contrast, conflicting P_3 trade-offs require careful balancing for each tumour site; a complex process requiring specialist clinical judgment. Weights are therefore derived through utilisation of Pareto navigation techniques (Section 8.4.2.4).

8.4.2.3 PB-AIO Framework

The following PB-AIO framework is used to generate both the final automated plans and those utilised during the calibration process.

Auxiliary Optimisation Volumes

Following PTV and PRV creation, a standard set of auxiliary optimisation ROIs (AuxROIs) are generated according to the algorithms detailed in the supplementary file S1. AuxROIs have two purposes. For conformity related planning goals AuxROIs enable a higher level of geometric specificity. For target goals, in line with ICRU 83 [120], AuxROIs subdivide each PTV into three sub-volumes to avoid conflicting planning aims: $PTV_{sv\text{-}1}$ is retracted from the skin surface and proximal primary OARs, $PTV_{SV\text{-}2}$ consists of areas of PTV within the skin

surface or extending into air, and PTV_{SV-3} is the PTV volume not covered by PTV_{SV-1} or PTV_{SV-2}, which represents parts of PTV proximal to primary OARs. It is through this subdivision that P_2 goals are compromised for P_1 goals and IMRT flash is secured for superficial PTVs.

Initial Optimisation Objectives

Scaling Factor	Description
F_V	Scales objective weight according the volume of the corresponding ROI
F_T	Scales objective weight according to the objective's target dose level (D_T). This removes an unwanted dependency of RayStation's objective functions on D_T
F_C	A hardcoded constant utilised to reduce the weight of PTV sub-volume objectives to avoid skin boosting and reduce conflicts within the PTV/OAR overlap region.
F_N	F_N enables $w_{initial}$ to be modified for an individual planning goal. The purpose is to bring $w_{initial}$ closer to the anticipated final weight for dynamic objectives

Table 2: *Summary of objective weight scaling factors*

Following treatment beam/arc definition an initial set of optimisation objectives are loaded into RayStation's native optimiser. Optimisation objectives are derived from the defined planning goals according to the algorithms specified in supplementary file S2, with the initial weight, $w_{initial}$, for the i^{th} objective defined by:

$$w^i_{initial} = w^i_{nom} F^i_V F^i_T F^i_C F^i_N \qquad \text{Eq. (1)}$$

where, $w^i_{nom}$is the nominal weight of the planning goal from which the objective is derived, and F^i_V, F^i_T, F^i_C and F^i_N are optimisation objective specific scaling factors. Each scaling factor is summarised in Table 2, with full definitions provided in the supplementary file S3. For P_1 and P_2 goals, w^i_{nom} is an empirically derived hardcoded value (supplementary table 3), intended to be common across all treatment sites. For P_3 goals, w^i_{nom} is generated through the Pareto navigation calibration process.

Plan Optimisation

The employed optimisation algorithms (Figure 6) consist of two stages: a pre-optimisation, which sets initial P_3 objective target values (T^i_{p3}) and a main optimisation, which generates the final clinical plan. During the optimisation process, P_3 goals are implemented as 'dynamic objectives' with the aim of minimising OAR doses and keeping trade-off balancing across patients consistent. In the following description of the methodology, we define Δ^i by the equation:

$$\Delta^i = \frac{D^i_{p3} - T^i_{p3}}{x^i} \qquad \text{Eq. (2)}$$

Where D^i_{p3} is the current value of the planning goal's corresponding dose parameter (c.f. supplementary table 2) and x^i equals D_{Presc} for dose objectives and V^i_{ROI} for volume objectives, where V^i_{ROI} is the volume of the i[th] objective's corresponding ROI.

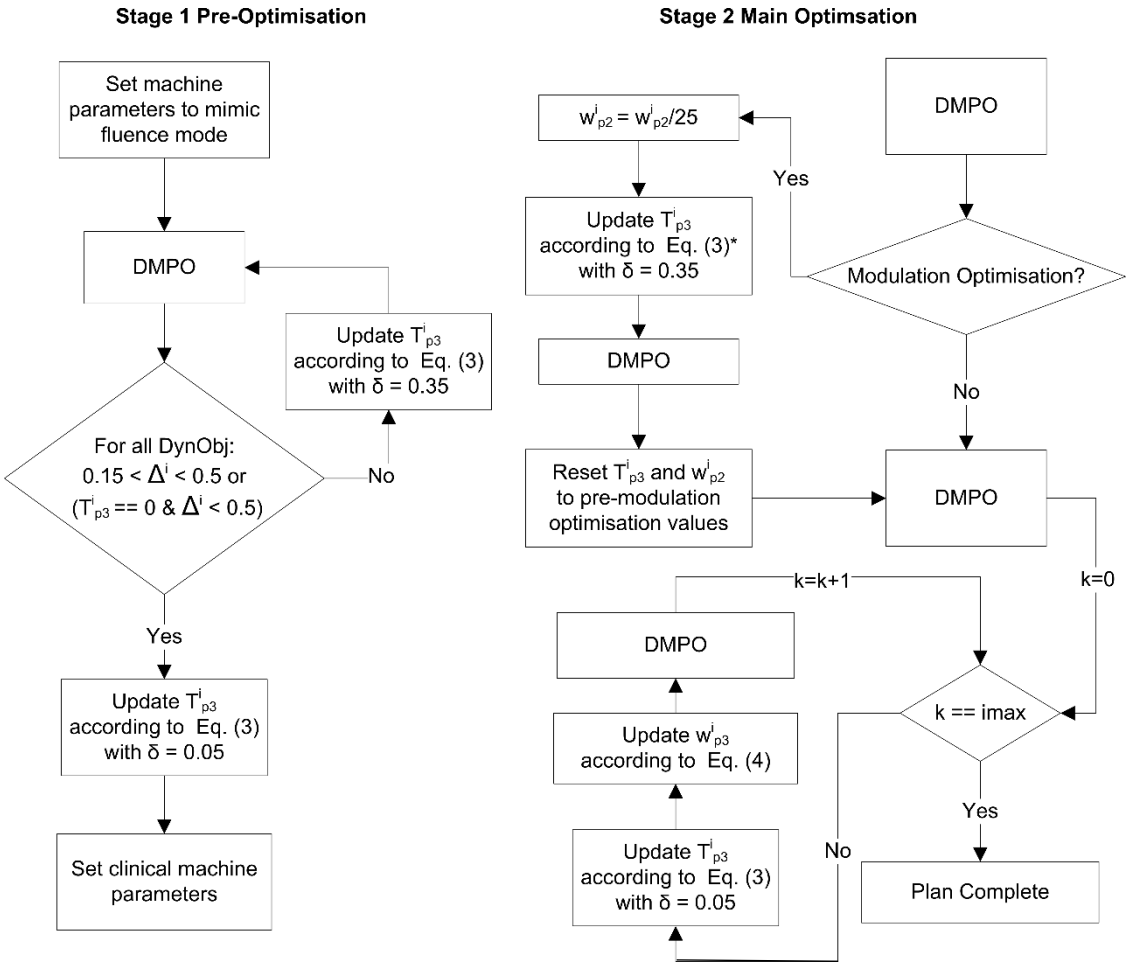


Figure 6: *Flowchart of the stage 1 and stage 2 optimisation algorithms, where: w^i_{p2}, T^i_{p3}, w^i_{p3}, Δ^i and δ are defined in the main manuscript, imax specifies the number of optimisation loops to perform and was set to 6, DynObj is an abbreviation for dynamic objectives and DMPO indicates a direct machine parameter optimisation where optimisations are bound by the treatment machine's physical limits. *During the modulation optimisation, for Eq. 3, D^i_{p3} is calculated from the final stage one pre-optimisation distribution.*

For stage one, a fluence-based optimisation, which allows beam intensity to be modulated with minimal physical limits, is performed. Following the optimisation, if Δ^i does not lie within the range [0.15-0.5] (or [0.0-0.5] if T^i_{p3}=0) for each dynamic objective, T^i_{p3} is updated according to equation 3 (with the variable δ set to 0.35) and the optimisation rerun.

$$T_{p3}^{i} = D_{p3}^{i} - x^{i}\delta \qquad \text{Eq. (3)}$$

The process is repeated until Δ^i lies within the specified bounds across all dynamic objectives. Bounding Δ^i in this manner ensures P_3 optimisation objectives are a significant but not dominant component of the composite objective function, resulting in P_3 goals being minimised without significantly compromising P_1 or P_2 goals. The resulting dose distribution, which is generated within 1-2 minutes, provides an approximate prediction of the final, fully optimised, clinical solution. Based on this distribution, T_{p3}^{i} for the main optimisation is set according to Eq. 3 with δ=0.05. Fluence-based VMAT optimisations are not possible in RayStation, therefore stage one treatment arcs are approximated through 15 equi-spaced static IMRT fields.

For stage two, a preliminary direct machine parameter optimisation (DMPO), where optimisations are bound by the machine's physical limits, is executed to generate an initial set of segments. An optional modulation optimisation is performed where P_2 objective weights (w_{p2}^{i}) are reduced by a factor of 25 and, using the stage one pre-optimisation distribution as the reference dose, T_{p3}^{i} set according to Eq.3 with δ=0.35. This prioritises the minimisation of P_3 objectives during the initial phases of the plan generation process and results in a reduction of OAR doses at the lower dose levels in the final clinical plan. This however is at the expense of increased modulation and MU. After the modulation optimisation, objective positions and weights are reverted to their original values. Finally, in the main stage of the plan generation process, multiple DMPOs are performed with T_{p3}^{i} and w_{p3}^{i} adjusted after each round. Using the stage two dose distribution as the reference, T_{p3}^{i} is updated according to Eq. 3 (with δ=0.05) and w_{p3}^{i} according to:

$$w_{p3}^{i} = \frac{OV_t^i w_c^i}{OV_c^i} \qquad \text{Eq. (4)}$$

where, w_c^i, OV_c^i and OV_t^i are the current weight, current OV and target OV of the ith objective respectively, with OV_t^i derived from w_{norm}^{i} according to the supplementary file S4. A function's OV is defined as the product of its weight and function, therefore this iterative weight adjustment ensures that across the multiple DMPOs, OV_c^i tends towards OV_t^i.

The values for δ and Δ^i bounds within the optimisation algorithm were based on previous experience developing VMAT class solutions, where it was observed that performing

multiple fluence based optimisations until Δ^i lay within the range [0.15-0.5] yielded a distribution where OAR doses were minimised with target coverage not excessively compromised. This fluence distribution represented an approximate prediction of the optimal OAR DVH for a given patient, with DVH objective positions then set to 5% below the predicted value (i.e. δ=0.05) and a full optimisation performed to create a deliverable plan. This process is very much akin to KBP approaches, but with the fluence optimisation used to predict DVHs and inform DVO positioning rather than machine learning. The parameters during the class solution development were derived through trial and error and were chosen as a starting point for the algorithms in this work and not modified. For the modulation optimisation the chosen values of 25 and 0.35 for the weight reduction factor and δ respectively were derived through trial and error.

8.4.2.4 AutoPlan Protocol Calibration

A flowchart of the calibration process, used to generate w_{nom}^{i} for each P_3 planning goal, is provided in the supplementary file S5. A calibration patient data set, consisting of typically 10-20 delineated patients, is defined. From this dataset a single 'navigation patient' is selected and the Pareto navigation process started.

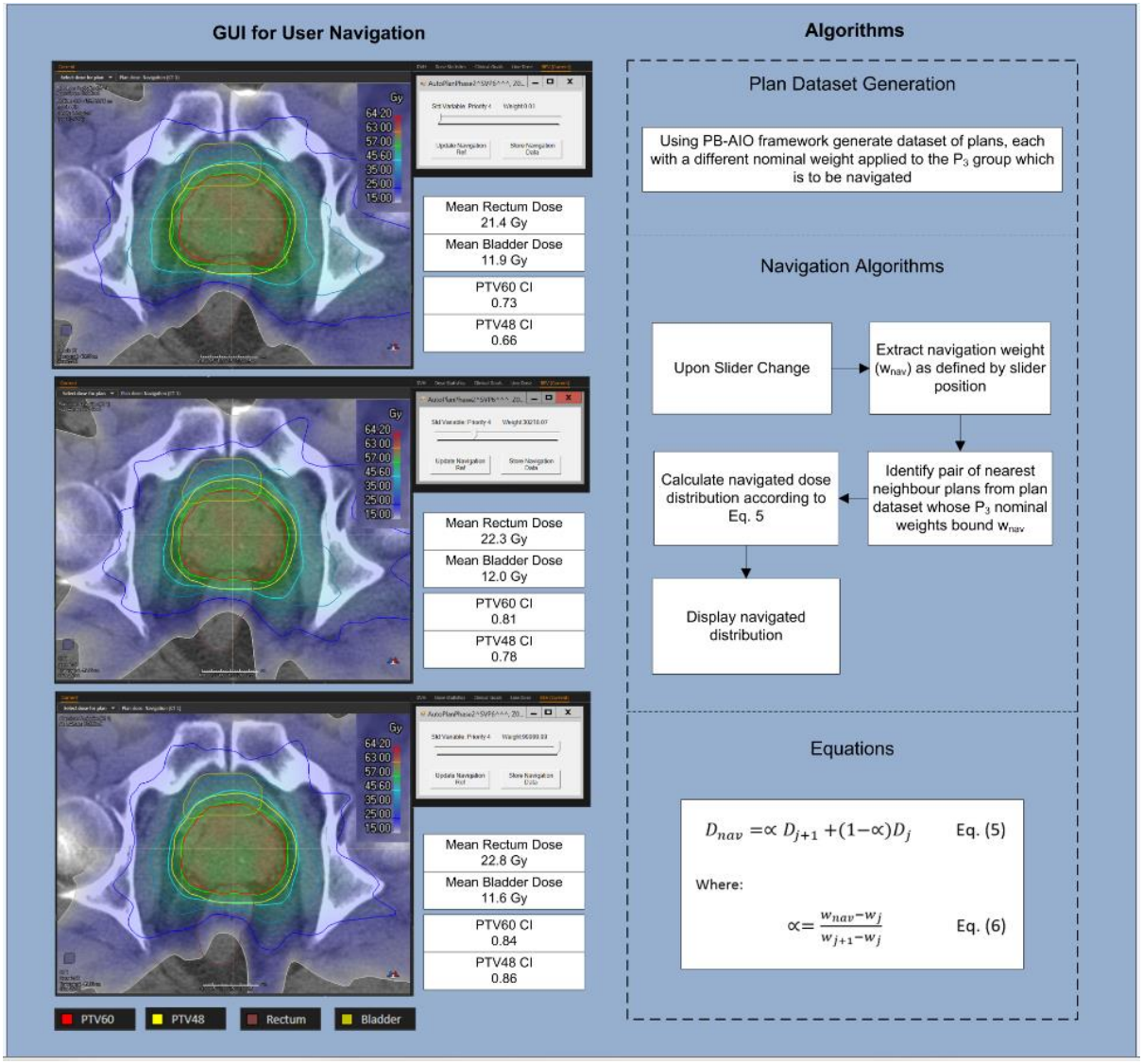


Figure 7: *(LHS) Screenshots demonstrating using the slider GUI to navigate through different weighted options for the PSV conformality goal (P_3 group 4). The displayed DVH metrics, which are not part of the calibration GUI, demonstrate the trade-off between the Paddick's conformity index (CI) for both PTVs and organ at risk mean doses. Isodose legend is enlarged for clarity. (RHS) Algorithms associated with the navigation module where: w_j and w_{j+1} are the nominal weights of the nearest neighbour plans j and j+1 respectively, whose weights bound the navigation weight, w_{nav}; D_j and D_{j+1} correspond to the dose distribution of plan j and j+1 respectively; and D_{nav} represents the estimated navigated dose distribution.*

Initially all P_3 nominal weights are set to zero. For the first P_3 group, multiple plans (nominally five) are generated, each with a different value of w_{nom}^{i} applied to the group. The operator uses a slider to navigate through convex combinations of the differently weighted plans, with the navigated dose distribution and associated DVH updated in real time to inform the decision-making (Figure 7). The operator selects what they consider to be the optimum group weighting and the navigated weight is stored in the AutoPlan protocol. The process is then repeated for the next group using the updated protocol. Once all groups are navigated a final 'rebalancing' navigation is performed on a set of plans with differing factors (range 0.25 to 1.25) applied globally to all P_3 nominal weights. This process allows the ratio of P_3 weights, and P_1 and P_2 weights to be explored, ensuring a solution can be selected where higher priority goals are not compromised. Once this first calibration round is complete the solution is tested across all calibration patients, with amendments to planning goals or additional navigations (on selected P_3 groups) performed as required to refine the solution.

Generating one set of navigation plans for a P_3 group takes 1-3 hours, depending on plan complexity, and each group must be optimised and navigated sequentially. Navigations are initially performed on a single patient, however where there are large inter-patient anatomical variations, repeat navigations over population outliers may be required to ensure the solution is robust across the whole patient cohort. When navigating over multiple patients the operator decides whether the P_3 group weighting is based a particular patient, or averaged over multiple patients. The calibration process can be considered equivalent to navigating the Pareto surface one dimension (or P_3 group) at a time, with operators using clinical experience and expertise to balance competing trade-offs.

8.4.3 Application to Prostate Cancer

All 10 calibration patients alongside their previous clinically approved treatment plan ($VMAT_{Clinical}$) were available during the AutoPlan protocol calibration for PSV. The navigation patient was selected following a visual review of all calibration patients and chosen based on its anatomy being reasonably representative of the set. Once successfully calibrated a final automated plan ($VMAT_{Auto}$) was generated for all study patients using identical arc configurations to $VMAT_{Clinical}$. To assess the efficacy of the calibrated automated solution, plan quality was quantitatively compared to $VMAT_{Clinical}$ using the local clinical goals,

alongside D98%, D2% and Paddick's Conformity Index [121] for each target volume. The statistical significance of any differences was assessed using two-sided Wilcoxon matched-paired signed-rank tests.

8.5 Results

8.5.1 AutoPlan Protocol Calibration

Protocol calibration was performed by one physicist with a radiation oncologist providing clinical input on trade-off prioritisation prior to calibration. 15 individual trade-off navigations were required to calibrate the AutoPlan protocol. All navigations were performed on a single patient, with planning goals manually modified twice after reviewing results across all patients in the calibration dataset. Key planning goal updates included the addition of bowel and low dose bladder planning goals. The final nominal weights are presented in supplementary table 3.

8.5.2 Comparison with $VMAT_{Clinical}$

		$VMAT_{Auto}$			$VMAT_{Clinical}$			
	Metric	Median	Range		Median	Range		p value
PTV60	D98% (Gy)	57.8	57.7 to 58.0		57.9	57.6 to 58.3		0.17
	D2% (Gy)	61.7	61.6 to 61.7		61.7	61.3 to 62.2		0.33
	CI	**0.86**	**0.85**	**to 0.87**	**0.79**	**0.76**	**to 0.82**	**0.01**
PTV48	D98% (Gy)	46.8	46.5 to 47.4		47.0	46.6 to 47.7		0.06
	D2% (Gy)	**58.8**	**58.5**	**to 59.2**	**59.4**	**58.8**	**to 59.8**	**0.01**
	CI	**0.84**	**0.82**	**to 0.87**	**0.77**	**0.75**	**to 0.79**	**0.01**
Rectum	V24.3Gy (%)	**41.1**	**27.3**	**to 63.6**	**46.4**	**30.6**	**to 66.1**	**0.01**
	V40.5Gy (%)	24.2	15.3 to 39.6		24.6	16.4 to 41.2		0.06
	V52.7Gy (%)	**11.0**	**4.7**	**to 16.5**	**12.8**	**5.6**	**to 18.9**	**0.01**
	V60.8Gy (%)	0.1	0.0 to 0.5		0.0	0.0 to 0.3		0.09
	DMean (Gy)	**25.1**	**17.5**	**to 32.3**	**27.5**	**20.5**	**to 34.0**	**0.01**
Bladder	V40.5Gy (%)	15.3	8.8 to 31.5		15.4	8.9 to 31.0		0.39
	V52.7Gy (%)	**7.8**	**3.0**	**to 16.8**	**8.0**	**3.3**	**to 17.9**	**0.04**
	V56.8Gy (%)	**5.3**	**2.1**	**to 11.7**	**5.7**	**2.3**	**to 12.8**	**0.03**
	DMean (Gy)	18.7	13.1 to 30.7		19.1	13.9 to 31.2		0.51
Bowel	V36.5Gy (cm^3)	0.0	0.0 to 0.7		0.0	0.0 to 0.9		0.27
	V44.6Gy (cm^3)	0.0	0.0 to 0.0		0.0	0.0 to 0.1		0.32
	DMean (Gy)	**6.4**	**3.4**	**to 11.2**	**7.5**	**3.9**	**to 14.1**	**0.01**
External	D1.8cm^3 (Gy)	61.6	61.5 to 61.8		61.7	61.2 to 62.4		0.33
Plan MU	MU	**600**	**582**	**to 653**	**546**	**496**	**to 627**	**0.01**

Statistical significance: results where p<=0.05 are presented in bold
CI: Paddick's Conformity Index for the specified PTV.

Table 3: *Dosimetric comparison of $VMAT_{Auto}$ and $VMAT_{Clinical}$ for the validation patient cohort*

A summary of the quantitative comparison of $VMAT_{Auto}$ and $VMAT_{Clinical}$ for the validation cohort is presented in Table 3. In comparison with the previous clinical plans, $VMAT_{Auto}$

yielded statistically significant ($p<0.05$) improvements in: PTV conformity indices, high dose bladder metrics, mean bowel dose, and the majority of rectum dose metrics. Of particular note was the reduction in mean rectum dose (median 25.1 Gy vs. 27.5 Gy), rectum $V_{24.3Gy}$ (median 41.1% vs 46.4%), and improvements in CI_{PTV60} (median 0.86 vs. 0.79) and CI_{PTV48} (median 0.84 vs. 0.77). Dosimetric improvements were not at the expense of other dose metrics, with observed detriments either statistically or clinically insignificant. In terms of modulation, $VMAT_{Auto}$ yielded plans with a median MU 10% higher than $VMAT_{Clinical}$.

Extending the comparison across all 20 study patients yielded similar results (supplementary tables 4 and 5), with all treatment plans meeting the locally defined mandatory goals for clinical acceptability.

8.6 Discussion

To the authors' knowledge this paper presents the first automated planning solution that directly incorporates Pareto navigation techniques into the calibration process. Compared to clinical practice (Table 3), $VMAT_{Auto}$ consistently yielded plans with improved conformity and reduced organ at risk doses, with no clinically relevant compromise to other dose metrics. As $VMAT_{Clinical}$ and $VMAT_{Auto}$ are generated in differing planning systems (Oncentra vs RayStation), these results are not intended to form a robust assessment as to their relative efficacy, this is the subject of future work. Instead they provide sound evidence that directly calibrating automated solutions through Pareto navigation is feasible and yields plans of improved dosimetric quality compared to current clinical practice.

Utilisation of Pareto navigation within the calibration process was observed to have two main benefits. Firstly, exploring differently weighted options via a sliding interface and live dose distribution allowed trade-off options to be explored in a visually intuitive manner. Secondly, an automated solution was derived in a time efficient manner with minimal trial and error. These advantages have been demonstrated on a per-patient basis by a number of studies [20,23,26] and this work indicates that similar benefits can be realised by applying this technique at a patient cohort level.

A potential weakness of Pareto navigation is that there can be clinically relevant discrepancies between the navigated dose distribution and that of the final deliverable plan [27]. To minimise these discrepancies the following approaches were adopted. Firstly, the

Pareto surface was sampled one dimension at a time to reduce interpolation errors during navigation (whilst maintaining a reasonable computational cost). Secondly, the Pareto dataset was populated with deliverable plans, ensuring navigations were performed on clinically achievable solutions. By utilising these approaches, discrepancies throughout the calibration process were of negligible clinical significance (supplementary table 6). This navigation methodology does however have a potential weakness, in that by limiting the navigation to one P_3 group at a time a full exploration of the Pareto surface is not performed. Whilst this was not considered a problem for the relatively simple site of PSV cancer, for more complex sites, navigation of multiple P_3 groups in parallel may be required to derive the most clinically desirable solution.

An interesting finding from this study was that Pareto navigation across a single patient appears sufficient for a successful calibration. Whilst modifications to planning goals were required after reviewing results across all calibration patients, these adjustments were due to deficiencies in the original set of planning goals, which were highlighted by differing patient geometries (e.g. proximity of bowel to PTV demonstrating requirement for bowel planning goals), rather than an inappropriate calibration. The ability to calibrate automated solutions against small patient cohorts (5-10 subjects) is in-line with examples in the literature for εc [49] and PB-AIO [72] solutions, and should enable a more efficient automation of novel techniques or protocols than KBP solutions, which require the manual generation of large patient datasets for each change in clinical practice.

The implemented calibration methodology requires algorithms that balance trade-offs consistently across differing patients. Dynamically adapting trade-off objective positions and weights during the optimisation was hypothesised to fulfil this function. By implementing ‘dynamic objectives’ a single calibrated AutoPlan protocol was found to yield appropriately balanced, clinically acceptable plans across all 20 study patients. In terms of robustness to patient geometry, even when PTV/OAR overlap differed considerable from the navigation patient, trade-off balancing was observed to be appropriate (Figure 8).

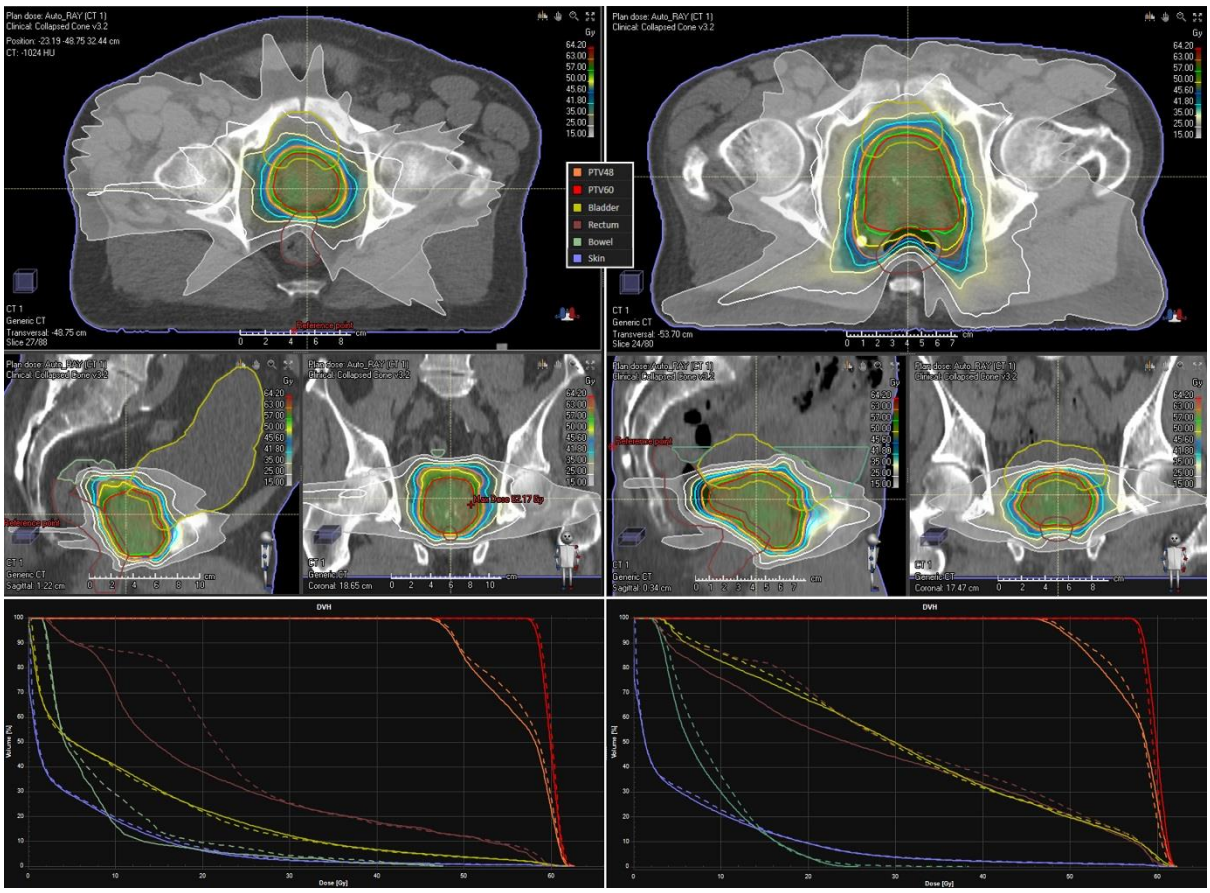


Figure 8: *DVH and dose distributions for the navigation patient (LHS) and patient 7 in the validation cohort (RHS), demonstrating the robustness of the automated solution to different anatomy. For both patients the DVH results for* $VMAT_{Clinical}$ *are provided for reference (dashed line). For patient 7 the overlap of rectum with PTV60 and PTV48 was 9% and 24% respectively (c.f. 4% and 13% respectively for the navigation patient), and for bladder 8% and 19% respectively (c.f. 1% and 4% for the navigation patient). The PTV/OAR overlaps for patient 7 were all greater than the* 89^{th} *percentile when considering the overlaps of all 20 study patients.*

In summary, a novel automated planning solution has been developed, which for the first time directly incorporates Pareto navigation into the calibration process. The solution has been successfully calibrated for the site of PSV, yielding clinically acceptable, appropriately balanced treatment plans, fully autonomously.

8.7 Acknowledgements

This research is funded by Health and Care Research Wales and sponsored by Velindre NHS Trust. The authors would like to thank Mrs Sue Campbell and Dr Jim Fitzgibbon who represented the views of the patients and public throughout the research.

8.8 Conflicts of Interest

The authors declare that they have no conflicts of interest.

9 Journal Article 2: Evaluating the Application of Pareto Navigation Guided Automated Radiotherapy Treatment Planning to Prostate Cancer

Published in: *Radiotherapy and Oncology, Volume 141, December 2019, Pages 220-226*

9.1 Author List and Contribution

Author	Contribution
Philip A. Wheeler[1]	Designed and built EdgeVcc. Designed study. Calibrated the EdgeVcc PSV solution with Rosemary Holmes. Calibrated the EdgeVcc PPN solution. Performed all evaluation work including data analysis. Author of paper (including all figures, tables and supplementary information).
Michael Chu[1]	Advised on translational aspects of EdgeVcc; ensuring alignment with requirements of the treatment planning clinical service.
Rosemary Holmes[1]	Developed data analysis package in python, which was utilised as part of the statistical and graphical analysis. Calibrated the EdgeVcc PSV solution with Philip Wheeler.
Owain W Woodley[1]	Generated all manual plans for PPN
Ceri S Jones[1]	Generated all manual plans for PSV
Rhydian Maggs[1]	Advised on translational aspects of EdgeVcc; ensuring alignment with requirements of the treatment planning clinical service.
John Staffurth[2]	Provided clinical governance across the study. Responsible for the clinical review of all PSV study plans
Nachi Palaniappan[3]	Responsible for the clinical review of all PPN study plans
Emiliano Spezi[4]	Advised on Study design, provided academic oversight of the project
David G. Lewis[1]	Advised on study design, provided academic oversight of the project with a focus on statistical methods
Sue Campbell[5]	Patient and public research partner. Ensured the research was aligned with the interests of the patients and public
Jim Fitzgibbon[6]	Patient and public research partner. Ensured the research was aligned with the interests of the patients and public
Anthony E. Millin[1]	Advised on study design, provided academic oversight of the project, advised on the design of EdgeVcc from a technical perspective to ensure safe translation to the clinical environment.

[1] Department of Medical Physics, Velindre Cancer Centre, Cardiff, United Kingdom
[2] School of Medicine, Cardiff University, Cardiff, United Kingdom
[3] Department of Oncology, Velindre Cancer Centre, Cardiff, United Kingdom
[4] School of Engineering, Cardiff University, Cardiff, United Kingdom
[5] Velindre Cancer Centre, Cardiff, United Kingdom.
[6] Wales Cancer Research Centre, Cardiff University, Cardiff, United Kingdom

All authors contributed to the review and revision of the manuscript.

9.2 Highlights

- An automated treatment planning solution, which is uniquely calibrated using Pareto navigation principles, has been evaluated for prostate cancer patients with and without elective nodal irradiation.
- Pareto navigation enabled the intuitive exploration of clinical trade-offs during protocol calibration, ensuring automated solutions were closely aligned to oncologists' clinical preference
- Upon evaluation, automated plans were considered non-inferior to manual planning by expert dosimetrists under no time pressures.

9.3 Abstract

Background and purpose: Current automated planning methods do not allow for the intuitive exploration of clinical trade-offs during calibration. Recently a novel automated planning solution, which is calibrated using Pareto navigation principles, has been developed to address this issue. The purpose of this work was to clinically validate the solution for prostate cancer patients with and without elective nodal irradiation.

Materials and methods: For 40 randomly selected patients (20 prostate and seminal vesicles (PSV) and 20 prostate and pelvic nodes (PPN)) automatically generated plans ($VMAT_{Auto}$) were compared against plans created by expert dosimetrists under clinical conditions ($VMAT_{Clinical}$) and no time pressures ($VMAT_{Ideal}$). Plans were compared through quantitative comparison of dosimetric parameters and blind review by an oncologist.

Results: Upon blind review 39/40 and 33/40 $VMAT_{Auto}$ plans were considered preferable or equal to $VMAT_{Clinical}$ and $VMAT_{Ideal}$ respectively, with all deemed clinically acceptable. Dosimetrically, $VMAT_{Auto}$, $VMAT_{Clinical}$ and $VMAT_{Ideal}$ were similar, with observed differences generally of low clinical significance. Compared to $VMAT_{Clinical}$, $VMAT_{Auto}$ reduced hands-on planning time by 94% and 79% for PSV and PPN respectively. Total planning time was significantly reduced from 22.2 mins to 14.0 mins for PSV, with no significant reduction observed for PPN.

Conclusions: A novel automated planning solution has been evaluated, whose Pareto navigation based calibration enabled clinical decision-making on trade-off balancing to be

intuitively incorporated into automated protocols. It was successfully applied to two sites of differing complexity and robustly generated high quality plans in an efficient manner.

9.4 Introduction

Intensity modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) treatment plan generation is a complex process, traditionally performed manually by medical physicists or specialist dosimetrists. Manual methods can be time consuming and dependent on the treatment planner's experience [6]. A solution to this problem is automated planning, where high quality plans are generated autonomously with minimal operator interaction [49–51,66,67,79,118,119].

A key challenge in automated planning is incorporating treatment planners' or oncologists' clinical experience and decision-making within the autonomous process. A number of different methods have been employed: knowledge based planning (KBP) utilises databases of previous clinical plans to correlate the relationship between patient geometry and the resultant dose distribution, which then informs the optimisation of new patients [34,36,98,118,122]; sequential ε-constraint planning (εc) optimises plans based on a list of clinically prioritised goals [44,46,49–51,123]; and protocol based automatic iterative optimisation (PB-AIO) adapts optimisation parameters during the planning process, tailoring the optimisation to the individual patient [61,66,68,72]. Whilst these techniques have been successfully applied to automated planning, a method for intuitive exploration of different 'trade-off' options during their calibration has not yet been demonstrated.

Recently we developed a fully automated treatment planning solution, which is uniquely calibrated using Pareto navigation principles. This novel calibration process allows differing trade-off options to be intuitively explored, ensuring clinical experience and decision-making can be effectively incorporated into the autonomous plan generation process. Utilisation of Pareto navigation techniques on a per patient basis has been shown to improve congruence between the oncologist's clinical preference and the final clinical plan [23], improve efficiency [20,23,124], and enable novice operators to generate high quality plans [20]. It is anticipated that utilising such an approach to inform and calibrate an automated solution would have similar benefits and provide significant advantages over current methods, which are reliant on trial and error, or calibration against historical datasets.

In a previous publication we presented in detail the algorithms behind our automated approach, demonstrated the calibration process for the tumour site of prostate and seminal vesicles (PSV), and presented results from a limited proof of principle pilot study on 10 patients [125]. The objective of this study was to additionally calibrate the solution for the complex site of prostate and pelvic nodes (PPN), and for both PSV and PPN perform a comprehensive clinical evaluation comparing this new automated technique with plans generated manually by expert dosimetrists. It is hypothesised that this novel approach to calibration will result in high quality plans that are closely aligned with oncologist clinical preferences.

9.5 Methods and Materials

9.5.1 Patient Selection and Planning Protocol

Calibration for the tumour site of PPN was performed on a dataset of 20 previously treated patients at Velindre Cancer Centre; 10 randomly selected from patients treated between July and December 2015 and 10 selected from a previous research database of patients treated between June and September 2014. The subsequent evaluative study was performed on an independent validation dataset of 40 subjects (20 PSV and 20 PPN) which were randomly selected from patients treated at Velindre Cancer Centre between January and June 2016.

Patients were planned on computed tomography scans with 3mm slice thickness. Prostate, seminal vesicles (SV), rectum, bladder, bowel, pelvic nodes (PPN only) and an optional pelvic node boost volume covering gross nodal disease (PPN only) were delineated prior to planning. The following planning target volumes (PTV) were subsequently generated: prostate, pelvic nodes and pelvic node boost expanded by 5 mm (6 mm craniocaudally) to form PTV60, PTV44 (PPN only) and PTV50 (PPN only) respectively; and prostate + SV expanded isotropically by 10 mm to form PTV48. For automated plan generation an additional volume, BowelBagRegion, was manually delineated for PPN, with details provided in the supplementary file S1.

Treatments were prescribed for 20 fractions using a simultaneous integrated boost (SIB) technique, with the PTV's suffix denoting its prescribed dose in Gy. The local clinical planning goals, adapted from the UK clinical trial PIVOTAL [126], are detailed in the supplementary file

S2. All plans in this study were generated within RayStation (v4.99, Raysearch Laboratories, Stockholm) using identical computer clients, treatment units (Elekta Agility, Elekta Ltd, Crawley) and VMAT arc configurations (6MV single 360° arc for PSV; 6MV dual 360° arc for PPN).

9.5.2 Automated System Overview

Automated planning was performed using EdgeVcc: an 'in-house' automated treatment planning solution, implemented within RayStation using its scripting functionality. This section provides an brief overview of the system, with full technical details provided by Wheeler *et al* [125].

Prior to automated plan generation a site-specific 'AutoPlan protocol' must be created and calibrated. The AutoPlan protocol specifies the treatment modality, beam arrangement and planning goals for a given tumour site. Planning goals are split into three priority levels: primary normal tissue goals (P_1), target goals (P_2) and trade-off goals (P_3). The planning goals used in this study for PPN are presented in the supplementary file S3.

Planning goals do not require any user defined optimisation weighting factors, instead weights are automatically assigned during plan generation through one of two processes. For P_1 and P_2 goals, where the handling of competing clinical trade-offs is explicitly defined (i.e. target coverage is compromised to maintain normal tissue goals), weights are derived from a set of hard-coded nominal weights, which are common to all tumour sites. When derived, weighting factors are scaled according to the volume of their corresponding region of interest to account for the observation that to obtain the same effect, small volumes require lower weighting factors than large volumes. For P_3 goals, interaction between conflicting trade-offs is complex, site specific and requires careful balancing of competing clinical demands. P_3 nominal weights are therefore derived through an intuitive Pareto navigation based calibration process, where the operator sequentially explores differently weighted options of each P_3 goal using an interactive slider GUI, with DVHs and dose distributions updated in real-time to inform the decision-making. The calibration is initially performed on a single patient, with the resultant solution tested against the remaining patients in the calibration cohort to ensure robustness against the whole population. Where there are large inter-patient anatomical variations, repeat navigations over population

outliers may be required to improve the robustness of the solution. In this situation the operator decides if the final weighting is based a particular patient, or averaged over multiple patients. Once calibrated, a single high quality treatment plan can be automatically generated for delineated patients within that tumour site.

Treatment plans are generated using RayStation's native optimiser with optimisation objectives derived from the defined planning goals. Plan optimisation is based on a PB-AIO framework where the target values and weights of P_3 related objectives are dynamically adjusted during the optimisation, such that the plan is tailored to the individual patient. Implementation of 'dynamic objectives' ensures P_3 goals are always acted on by the optimiser and thus minimised, and additionally is hypothesised to enable a common set of calibration weights to be applicable across all patients for a given site.

9.5.3 Automated Plan Generation

Using the calibration patient dataset an AutoPlan protocol for PPN was created and calibrated. The final PPN protocol and previously calibrated PSV protocol [125] were used to generate a single automated plan ($VMAT_{Auto}$) for each patient in the corresponding independent validation datasets. Plans were reviewed for clinical acceptability, with manual dose scaling performed if required. All work was performed by a single clinical scientist (PW).

9.5.4 Study Design and Statistical Analysis

To benchmark $VMAT_{Auto}$, experienced IMRT/VMAT dosimetrists (CJ for PSV; OW for PPN) generated two manual treatment plans ($VMAT_{Clinical}$ & $VMAT_{Ideal}$) for each patient in the validation dataset. $VMAT_{Clinical}$ was generated under simulated clinical conditions following standard protocols, which utilise an efficient template-based class-solution methodology. As per clinical practice the dosimetrist ceased optimising once a clinically acceptable plan was generated. Then, in the absence of time pressure, the dosimetrist used their knowledge and expertise to improve $VMAT_{Clinical}$ as far as they deemed possible to produce an 'ideal' treatment plan, $VMAT_{Ideal}$.

Prior to manual plan generation and the calibration of both AutoPlan protocols, all operators were briefed on trade-off prioritisation via discussions with a consultant oncologist assigned to each clinical site (JS for PSV; NP for PPN). For all three techniques operator hands-on and total planning times were recorded.

$VMAT_{Auto}$ was compared to both $VMAT_{Clinical}$ and $VMAT_{Ideal}$ in terms of plan quality and planning efficiency. Plan quality was quantitatively assessed using local clinical planning goals; and D98%, D2% and Paddick's Conformity Index (CI) [121] for each target volume. Two-sided Wilcoxon matched-paired signed-rank tests assessed the statistical significance of any differences in plan quality and timing metrics. In addition, a blinded qualitative assessment was performed by the assigned oncologist to: score overall plan quality using a five point scale (1-unacceptable, 2-poor, 3-satisfactory, 4-good, 5-excellent); establish the clinical acceptability of each plan; and rank the trio of plans in order of preference, with clinically equivalent plans given equal rank.

9.6 Results

Calibration for the complex site of PPN was challenging and iterative due to the high number of competing trade-offs and large inter-patient variability in OAR volumes. Over 40 individual navigations across six patients were performed. During PPN calibration the hard coded P_1 nominal weight for primary conformality goals was considered suboptimal and manually increased to match the weight for P_1 primary OAR goals. The post calibration nominal weights are presented in the supplementary file S4.

		SVP			PPN		
	Metric	$VMAT_{Auto}$	$VMAT_{Clinical}$	$VMAT_{Ideal}$	$VMAT_{Auto}$	$VMAT_{Clinical}$	$VMAT_{Ideal}$
PTV60	D98% (Gy)	57.9 ± 0.1	57.8 ± 0.2	**57.7 ± 0.1**	57.8 ± 0.1	**58.0 ± 0.1**	**57.9 ± 0.1**
	D2% (Gy)	61.6 ± 0.1	**61.7 ± 0.2**	**61.7 ± 0.2**	61.7 ± 0.1	**61.9 ± 0.2**	**61.9 ± 0.2**
	CI	0.86 ± 0.01	**0.84 ± 0.03**	**0.88 ± 0.02**	0.82 ± 0.02	**0.81 ± 0.03**	0.81 ± 0.03
PTV50	D98% (Gy)				48.2 ± 0.2	48.0 ± 0.3	47.9 ± 0.2
	D2% (Gy)				52.3 ± 1.7	52.0 ± 1.0	52.1 ± 0.9
	CI				0.41 ± 0.05	0.41 ± 0.07	0.42 ± 0.07
PTV48	D98% (Gy)	46.8 ± 0.5	46.8 ± 0.4	**46.5 ± 0.3**	46.6 ± 0.6	46.7 ± 0.4	46.6 ± 0.4
	D2% (Gy)	58.9 ± 0.2	59.0 ± 0.3	**58.6 ± 0.3**	59.5 ± 0.3	**59.6 ± 0.3**	**59.6 ± 0.3**
	CI	0.85 ± 0.01	**0.82 ± 0.01**	**0.87 ± 0.01**	0.65 ± 0.05	**0.59 ± 0.07**	**0.60 ± 0.07**
PTV44	D98% (Gy)				42.3 ± 0.1	**42.4 ± 0.1**	**42.4 ± 0.1**
	D2% (Gy)				47.4 ± 1.6	**47.8 ± 1.7**	47.7 ± 1.8
	CI				0.82 ± 0.02	**0.81 ± 0.02**	**0.81 ± 0.02**
Rectum	V24.3Gy (%)	36.7 ± 10.1	**40.8 ± 11.1**	**38.0 ± 9.3**	53.3 ± 9.3	**59.3 ± 7.3**	56.2 ± 8.1
	V40.5Gy (%)	20.4 ± 7.2	20.4 ± 7.4	20.0 ± 7.2	24.0 ± 6.1	23.8 ± 6.4	**23.1 ± 6.5**
	V52.7Gy (%)	8.5 ± 3.7	8.1 ± 3.6	**8.0 ± 3.5**	10.5 ± 3.0	**10.0 ± 2.9**	**9.6 ± 2.9**
	V60.8Gy (%)	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 0.1	0.1 ± 0.1	**0.0 ± 0.1**	**0.0 ± 0.1**
	DMean (Gy)	22.7 ± 3.9	**25.1 ± 3.5**	**23.4 ± 3.5**	29.5 ± 2.7	**30.4 ± 2.6**	29.7 ± 2.6
Bladder	V40.5Gy (%)	19.2 ± 10.7	18.3 ± 9.6	**17.4 ± 9.5**	24.7 ± 10.4	23.7 ± 8.5	23.7 ± 8.5
	V52.7Gy (%)	8.8 ± 5.9	**8.3 ± 5.2**	**7.9 ± 5.2**	7.4 ± 4.8	**7.6 ± 4.8**	7.5 ± 4.7
	V56.8Gy (%)	6.1 ± 4.2	**5.7 ± 3.8**	**5.6 ± 3.9**	4.9 ± 3.1	**5.3 ± 3.5**	**5.3 ± 3.5**
	DMean (Gy)	23.0 ± 9.1	**22.2 ± 8.6**	**21.6 ± 8.6**	33.0 ± 3.9	**31.3 ± 3.5**	**31.1 ± 3.5**
Bowel	V36.5Gy (cc)	0.9 ± 2.0	0.9 ± 1.9	0.7 ± 1.6	48.6 ± 35.9	**53.9 ± 38.7**	**51.2 ± 38.0**
	V44.6Gy (cc)	0.3 ± 0.7	0.3 ± 0.8	0.3 ± 0.8	3.6 ± 6.5	3.5 ± 6.0	3.3 ± 5.6
	V52.7Gy (cc)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DMean (Gy)	8.6 ± 4.7	8.4 ± 4.7	**7.7 ± 4.2**	18.7 ± 2.6	**19.6 ± 2.6**	**19.3 ± 2.4**
Patient Outline	D1.8cm^3 (Gy)	61.6 ± 0.1	**61.7 ± 0.2**	**61.7 ± 0.2**	61.7 ± 0.1	**61.9 ± 0.3**	**61.9 ± 0.3**
Plan MU	MU	616 ± 43	**563 ± 58**	**575 ± 57**	714 ± 60	**695 ± 69**	711 ± 68
Planning Time	Hands on time (mins)	1.3 ± 0.3	**22.2 ± 5.3**	**85.4 ± 39.9**	4.4 ± 0.5	**20.6 ± 6.3**	**65.4 ± 21.1**
	Total time (mins)	14.0 ± 1.4	**22.2 ± 5.3**	**85.4 ± 39.9**	36.4 ± 3.1	41.8 ± 11.4	**200.0 ± 53.1**
Plan Quality	Score	5.0 ± 0.2	4.6 ± 0.5	4.9 ± 0.3	5.0 ± 0.2	4.8 ± 0.4	4.8 ± 0.4
Plan Ranking vs	Plans Superior (%)		5%	15%		0%	20%
$VMAT_{Auto}$	Plans Equivalent (%)		35%	55%		35%	15%
	Plans Inferior (%)		60%	30%		65%	65%

Statistical significance: $VMAT_{Clinical}$ and $VMAT_{Ideal}$ dosimetric and timing data are presented in bold where statistically significant differences ($p<0.05$) with $VMAT_{Auto}$ are observed.

CI: Paddick's Conformity Index for the specified PTV.

Table 4: *Dosimetric comparison of $VMAT_{Auto}$, $VMAT_{Clinical}$ and $VMAT_{Ideal}$ for the treatment sites PSV and PPN (mean ± standard deviation)*

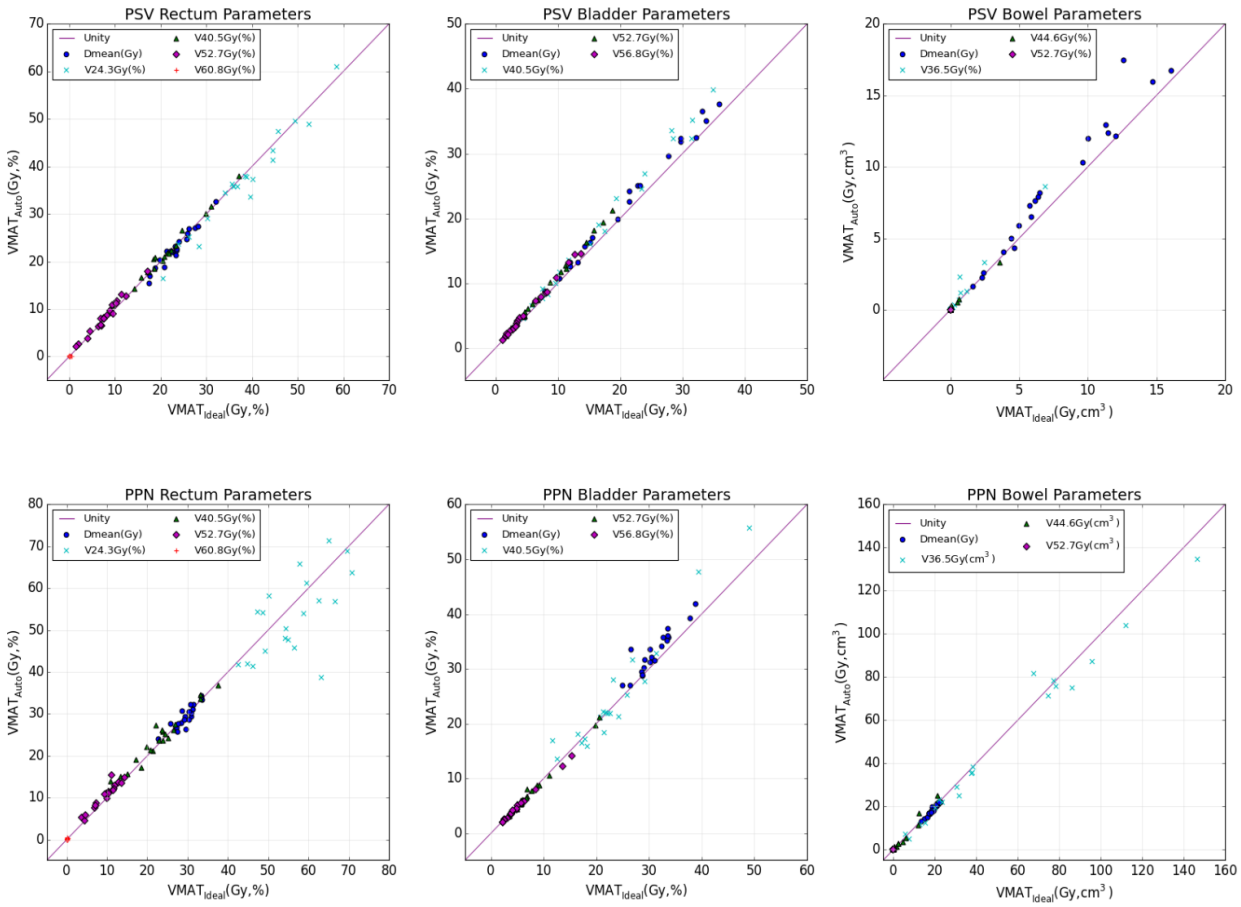


Figure 9: *Comparison of rectum, bladder and bowel dosimetric plan parameters between automatically generated plans ($VMAT_{Auto}$) and plans generated by expert dosimetrists under no time pressure ($VMAT_{Ideal}$).*

39/40 $VMAT_{Auto}$ plans were generated with no user intervention; for one PPN patient the plan MU was scaled by 0.3% to ensure PTV44 D99% was within the local clinical planning goal. A summary of the quantitative plan comparison is presented in Table 4 and Figure 9, with example dose distributions presented in Figure 10. For both PPN and PSV, $VMAT_{Ideal}$ led to small reductions in all OAR metrics when compared to $VMAT_{Clinical}$ and across all three techniques observed differences were generally considered of low clinical significance. For PSV $VMAT_{Auto}$, the noteworthy statistically significant ($p<0.05$) differences with $VMAT_{Ideal}$ and $VMAT_{Clinical}$ were: reductions in rectum mean dose and V24.3Gy, increases in the majority of bladder metrics, improved (increased) CI compared to $VMAT_{Clinical}$, and decreased CI compared to $VMAT_{Ideal}$. For PPN $VMAT_{Auto}$ the noteworthy differences ($p<0.05$) were: reduction in bowel V36.5Gy; increased mean bladder dose; increased PTV48 CI; and when compared to $VMAT_{Clinical}$ only, decreased rectum V24.3Gy. For PSV, automation led to a

moderate increase in plan MU of 7% and 9% compared to $VMAT_{Ideal}$ and $VMAT_{Clinical}$ respectively, which may be indicative of increased modulation.

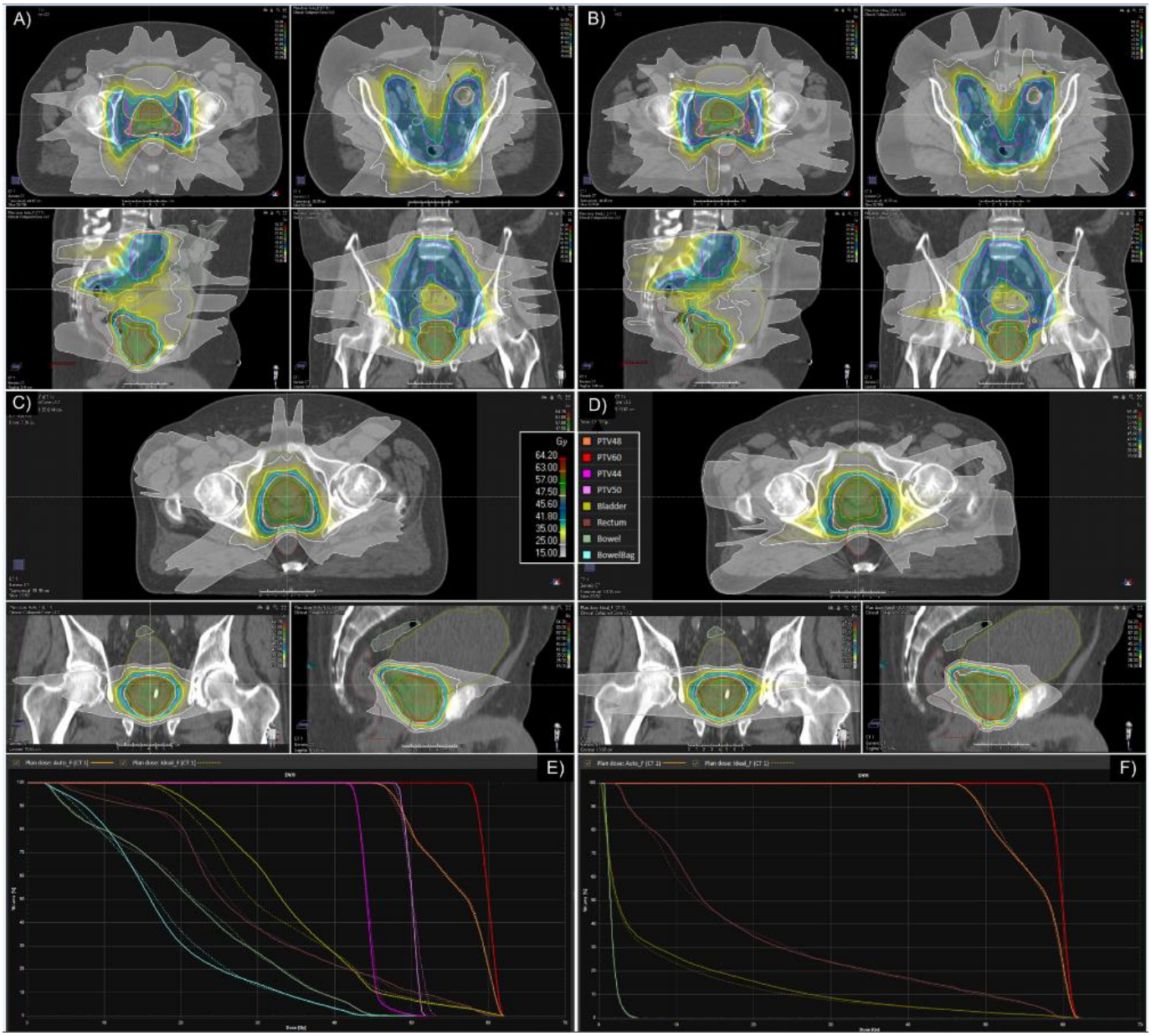


Figure 10: *DVH and dose distributions for patient 1 in the PPN and PSV validation cohort. (A) PPN $VMAT_{Auto}$ dose distribution. (B) PPN $VMAT_{Ideal}$ dose distribution. (C) PSV $VMAT_{Auto}$ dose distribution. (D) PSV $VMAT_{Ideal}$ dose distribution. (E) PPN DVH for $VMAT_{Auto}$ (solid line) and $VMAT_{Ideal}$ (dotted line). (F) PSV DVH for $VMAT_{Auto}$ (solid line).*

All 120 plans were considered acceptable upon blind review by the oncologist, with plan quality scores either good (4) or excellent (5). Analysis of the plan ranking determined that 39/40 and 33/40 of $VMAT_{Auto}$ plans were considered preferable or equal to $VMAT_{Clinical}$ and $VMAT_{Ideal}$ respectively. When compared to $VMAT_{Clinical}$, hands-on planning time was significantly reduced by 94% and 79% for PSV and PPN respectively. Total planning time was

significantly reduced from 22.2mins to 14.0mins for PSV, with no significant reduction observed for PPN.

9.7 Discussion

In this study a novel automated treatment planning solution, which is directly calibrated using Pareto navigation principles, has been robustly validated for prostate cancer. The resultant automated protocols were rigorously evaluated against plans generated by expert dosimetrists, with favourable results towards automation. Furthermore the solution's robustness to treatment site complexity was validated through application to PPN; a treatment site with up to four PTV prescription levels and wide inter-patient OAR volume variation.

In our previous work we demonstrated that for the simple site of PSV, Pareto navigation enabled both the intuitive exploration of competing trade-offs and the creation of a high quality solution in a time efficient manner; benefits which are congruent with Pareto navigation applied on a per patient basis [20,23,124]. In this study the generalisability and versatility of the calibration methodology was demonstrated through successful application to PPN, a site of significant complexity. As with PSV, the intuitive exploration of trade-offs was considered a key benefit in ensuring alignment between the final automated solution and the oncologist's clinical aims. However, due to wide variations in inter-patient anatomy the calibration was more iterative and challenging, with additional navigations required over population outliers. This is in contrast to PSV where navigation over a single patient was sufficient for successful protocol calibration [125].

During the calibration process several potential improvements in the implemented methodology were identified. Firstly, the hard-coded objective weights for P_1 and P_2, which were based on previous clinical experience, may need further refining, as evidenced by the requirement to increase the nominal weight for P_1 primary conformality goals for PPN. Secondly, challenges during the PPN calibration indicated that the optimum calibration weights for a given patient were still correlated with anatomical geometry, even when objective positions and weights were dynamically adjusted. Further work will include assessing and correcting for this correlation using machine learning.

The evaluative study demonstrated that when compared to manual planning under clinical conditions, $VMAT_{Auto}$ was the superior technique both in terms of quality and efficiency. In addition, results indicate $VMAT_{Auto}$ is non-inferior to manual planning by expert dosimetrists under no time pressure. In general, dosimetric differences between $VMAT_{Ideal}$ and $VMAT_{Auto}$ were small, which was considered supportive evidence that implementation of 'dynamic objectives' within the automated planning process were yielding plans which were, or were near to, Pareto optimal.

Interestingly clinical preference towards automation was stronger for the more complex site of PPN. It is hypothesised that for PPN the high degrees of freedom within the optimisation problem made the manual trial and error exploration of trade-offs difficult. In contrast, implementation of Pareto navigation techniques allowed intuitive exploration of these trade-offs and whilst calibration was challenging, this approach resulted in plans more closely aligned to the clinician's preference. Improved congruence with the clinician's clinical preference is a key benefit of Pareto navigation, which has been demonstrated on a per-patient level [23] and this work supports the hypothesis that similar benefits can be realised by applying this technique at a patient cohort level.

A potential limitation of this study is its tightly controlled study design, in that for each treatment site all manual planning was performed by a single treatment planner, and guidance on trade-off balancing and the subsequent blind review was performed by a single oncologist. The study was designed such that inter-observer bias was minimised, however as a consequence results may not be directly translatable to clinical practice where inter-observer variability in manual plan quality and oncologist trade-off preferences may be significant.

Compared to existing methods of calibrating automated solutions, Pareto navigation presents a clear alternative. For both εc and PB-AIO, automated solution calibrations are reliant on trial and error. It is envisaged that the methods presented in this study would enhance many of the existing εc and PB-AIO solutions and bring the advantages of intuitive trade-off exploration into the wider field of automated planning. When compared to KBP, the employed calibration methodology benefits from having no requirement for a database of reference treatment plans. Automated solutions are therefore not influenced by the quality or quantity of historical plans and new techniques can be developed without the

time consuming manual creation of a training dataset. In addition, it is envisaged that due to flexibility in the calibration process this approach is ideal for successful implementation in radiotherapy centres with differing clinical protocols.

When comparing to previously published studies, for the tumour site of PSV a thorough summary has recently been presented by Heijmen et al [52]; with 12 studies identified as demonstrating small differences between automated and manual plans [41,51,61,82,97–104], and only their more recent multi-centre study showing the overall dosimetric superiority of automation through reduced rectum doses [52]. For PPN, to the authors' knowledge two studies have been published. The first being a methodological paper presenting results from a single patient [60], which will not be discussed further, and the second a 30 patient evaluative study comparing automated planning using εc with manual planning under no time pressures [49]. The study demonstrated a clear preference towards automated planning, with notable improvements in a wide range of dosimetric parameters. Direct comparison between these examples in the literature and results from the study presented in this manuscript is not possible or appropriate due to the wide range of confounding factors including: patient selection criteria, planner and institutional expertise, and clinical protocol complexity. However, what can be ascertained is that results from this study, which demonstrate that automated planning is non-inferior to expert manual planning, are consistent with existing literature and supportive of Pareto navigation guided automated planning. Furthermore, in a recent review on automated planning by Hussein et al [11] only two out of the 81 identified evaluative studies were for complex pelvis treatments (SIB technique with nodal irradiation) [49,127]. Our work builds on this limited evidence base, providing further data in support of automation for even the most complex tumour sites.

9.8 Conclusions

EdgeVcc is a versatile new automated planning solution whose unique Pareto navigation based calibration methodology enabled clinical decision-making on trade-off balancing to be intuitively incorporated within automated protocols. It has been successfully applied to two sites of differing complexity and robustly generates high quality plans in an efficient manner.

9.9 Acknowledgements

This research is funded by Health and Care Research Wales and sponsored by Velindre NHS Trust.

9.10 Conflicts of Interest

The authors declare that they have no conflicts of interest.

10 Critical Appraisal of the Research

10.1 Overview

Advanced radiotherapy planning techniques have the potential to transform both the quality and efficiency of radiotherapy. This project focused on integrating two advanced techniques, through developing an automated solution that was calibrated using $MCO_{posteriori}$ methods. The rationale behind this integrative approach being that MCO would provide a more intuitive calibration methodology than existing methods and yield automated solutions that were aligned with oncologists' treatment wishes.

The first proffered paper presented the full methodology of the developed automated solution, alongside a proof of principle validation for prostate cancer patients without nodal irradiation. The second paper presented the results of a robust study that evaluated the technique for prostate patients with and without elective nodal irradiation.

What follows is a critical appraisal of the research conducted in this project, which includes: evaluating its contribution to theory and clinical practice, appraising the developed automated solution and research process taken, and finally summarising clinical implementation of the solution and opportunities for further research.

10.2 Contribution to Theory and Practice

10.2.1 Contribution to Theory

The key contribution of this work is demonstrating that incorporation of Pareto navigation techniques natively into an automated planning solution's calibration process is feasible and yields results which are non-inferior to manual planning performed under no time pressures. Whilst $MCO_{posteriori}$ techniques have been utilised to populate datasets on which KBP solutions are trained [21,89], this is the first automated solution which has been purpose built around the concept of a Pareto navigation calibration process. Developing a purpose built solution, rather than merging two existing implementations of $MCO_{posteriori}$ and AP, has a key advantage as implementation of the Pareto navigation concept could be tailored to match the requirements of an AP calibration process (where navigation accuracy is of highest importance), rather than individual patient planning (where efficiency is a key requirement). This is important as for existing $MCO_{posteriori}$ commercial solutions the approximation error can reach clinically significant levels [27], which could result in large

systematic errors in the resulting AP solution. For AP calibration, it is therefore preferable to sacrifice computational efficiency to minimise the approximation error as far as reasonably practicable.

The developed Pareto navigation methodology utilised three core methods for reducing the approximation error to a negligible level. Firstly, the Pareto plan dataset that samples the Pareto surface was generated using deliverable plans as opposed to undeliverable fluence based plans, which at the time of development was the method implemented within RayStation's commercial $MCO_{posteriori}$ solution. Whilst this substantially increased the Pareto dataset generation time (100 iterations for PSV typically takes ~20 seconds for fluence vs ~120 seconds for deliverable optimisations), it ensured navigation was performed on plans that were accurate representations of the potential clinical solutions, rather than idealised non-deliverable distributions. Secondly, when not using advanced sampling methodologies (e.g. sandwich algorithm [28]) the sampling requirements of multi-dimensional navigation increases to the power of the number of dimensions. Therefore in this initial implementation, Pareto navigation was performed one dimension (or trade-off) at a time. This substantially reduced the number of plans required to accurately sample the Pareto surface, ensuring Pareto navigation errors could be minimised without impractical computational times. Finally, generation of the Pareto dataset was fully customisable, in that the operator could choose the range of weights and number of plans to be used for a given trade-off when generating the dataset. Unlike commercial solutions of $MCO_{posteriori}$, where Pareto databases typically sample the whole Pareto front for each trade-off, the customisable approach enables operators to focus on the clinically relevant parts of the Pareto front, reducing the search space and therefore the sampling requirements of the problem. Through implementing these three methods, when using a sampling frequency of 5 plans per trade-off the approximation error of Pareto navigation during the calibration process was kept at low levels (<0.9 Gy or <1.2% volume across all dose metrics (supplementary table 6)). This ensured the risk of potential systematic errors in calibration due to errors in the Pareto navigation process was minimised. These results are in contrast to the clinically relevant approximation errors highlighted by McGarry et al. and Kryodio et al. [14,27] and underlie the importance of tailoring the Pareto navigation approach to the specific requirements of AP.

A second key contribution of this work is strengthening the evidence base on utilising dynamic objectives to both improve the optimality of the resultant treatment plan and reduce the dependence of the optimisation process on individual patient anatomy. As discussed in section 7.3.2.3, utilisation of dynamic objectives is the key component of a whole subcategory of PB-AIO solutions (PB-AIO$_B$). This work hypothesised that dynamic objectives could be utilised for two functions. Firstly, by adapting the position of objectives based on the individual patient's dose distribution, OAR doses could be automatically minimised and driven towards Pareto optimality. Secondly, by adapting objective weights to drive the objective value (OV) towards a target OV, the relative importance of the objectives, and thus the balancing of trade-offs, could be kept consistent between patients. This should enable a common set of objectives to be utilised across all patients within a single treatment site. Results of the evaluative study presented in section 9 provided supportive evidence on this hypothesis, with the nominally equivalent OAR doses between VMAT$_{Auto}$ and VMAT$_{Ideal}$ evidence of Pareto optimality and alignment of AP plans with oncologist preference (Table 4) indicative of consistent trade-off balancing across all study patients. This is in direct contrast to standard 'static' objectives implemented within commercial systems, where optimum objective weights for a given patient exhibit a strong correlation with patient anatomy [128] and therefore must be adapted on a patient-by-patient basis.

The third key contribution this work makes to theory is building on the evidence base that calibrating on small patient datasets is feasible ($n < 10$ for PSV, $n < 20$ for PPN). This is in direct contrast to KBP solutions, which form 83% of the evidence base [11], where datasets of > 75 patients are generally utilised for DVH prediction using regression methods [39]. Calibration on small datasets ensures AP solutions can be dynamic and adaptable to novel techniques or clinical protocols, without the resource intensive requirements to generate and curate large training datasets. Calibration on small datasets has been demonstrated for both εc [49] and PB-AIO [72] solutions, and this work strengthens this evidence base. However, this work also provides initial evidence that for simple sites (i.e. PSV), calibrating on a single patient may yield clinically optimal solutions. To the author's knowledge this has not yet been demonstrated in the literature. Whilst calibrating on a single patient may initially seem counter-intuitive and unfeasible due to natural variations in patient anatomy, it is hypothesised that as dynamic objectives ensure trade-off balancing is consistent

between patients, only one patient is required to capture the intended balancing for a given site. Further work is however required to fully test this hypothesis as it is possible these results represent a false positive and that calibrations based on different patients are not robust across the whole patient cohort.

The final contribution this work makes is providing robust evidence that automation of a highly complex site (PPN) is feasible. As highlighted in section 7.3.4, five studies have been identified which present evidence comparing AP with MP for PPN [49,74,75,92,108], with only two considered to provide high quality and robust data (n ≥ 20, no further manual optimisation of AP plans post generation) [49,92]. Furthermore as these two studies evaluate εc and KBP solutions, there is minimal robust evidence in the use of PB-AIO methodologies for planning complex PPN treatments. The evaluative study presented in the second paper (section 9) has a reasonable sample size (n = 20), uses a gold standard comparator (plans generated by expert planners under no time pressure) and includes a blinded qualitative review of plan quality by an oncologist. This work therefore provides high quality evidence on the efficacy of AP for PPN; filling a gap in the literature for PB-AIO solutions and building on the limited evidence across all AP methodologies. In addition, with patients treated with up to four SIB prescriptions, the study's treatment protocol represents a higher complexity than all five identified studies, which were limited to 1-2 SIB volumes.

10.2.2 Contribution to Practice

The automated planning solution developed in this work has two core implications for local practice, relating to both the quality and efficiency of the service.

Firstly, when compared to current clinical practice, implementation of AP would yield substantial efficiency savings with no sacrifice to plan quality. The evaluative study presented in the second paper provided clear, robust evidence on this matter, demonstrating that hands-on planning time could be reduced by 79% - 94%, with no compromise in quality for even the most complex situations (PPN with four SIB volumes). These efficiency savings would not only release resources within a busy treatment planning department, but also have the potential to aid in streamlining the planning pathway with the aim of reducing the time from decision to treat to start of treatment. Reduction of this key metric could have a substantial impact on patients' care due to its strong correlation with

treatment outcome for many cancer sites [129]. In addition, a second area of potential efficiency saving is the ability to readily adapt the solution to new prescriptions or treatment protocols. For example, updating to a new prescription requires no calibration, just manual amendment to the existing AutoPlan protocol and introduction of a new OAR (e.g. penile bulb or femoral heads) would generally only require Pareto navigation across that individual OAR to recalibrate the solution. Furthermore, new solutions can be automatically run over a large set of test patients to ensure their suitability for clinical use. This is in contrast to the current template-based class solution approach, where adjustments are iterative and time consuming, and testing the resultant solution on multiple patients prior to clinical release must be done manually.

Secondly, a key hypothesis of the developed approach was that AP plans are Pareto optimal and due to the Pareto navigation process, highly aligned with the clinical aims of the treatment. The evaluative study provides initial evidence in support of this hypothesis (Table 4); AP dose metrics are closely aligned with $VMAT_{ideal}$ (indicating Pareto optimality) and oncologist preference is skewed toward AP plans. Locally, based on the data comparing $VMAT_{Auto}$ and $VMAT_{Clinical}$ (Table 4), any potential improvement in clinical practice through introduction of AP is considered small and likely of minimal clinical significance. In comparison with some studies in the literature, which demonstrate substantial improvements in plan quality with AP [49,52], this may be considered a negative result with respect to the optimality of the algorithms developed in this study. On the contrary, since the introduction of VMAT at Velindre Cancer Centre, substantial work has been performed on developing robust, patient tailored planning methodologies [130]. It is therefore instead considered that these results reflect the high quality of local clinical planning methods, which yield plans of nominal equivalence to both $VMAT_{Auto}$ and $VMAT_{Ideal}$. This work therefore contributes to practice, not by improving treatment quality, but rather by providing an effective audit of current clinical practice and confidence that introduction of AP will maintain the same high quality of treatments provided.

10.3 Advantages and Limitations of the Developed Automated Approach

10.3.1 Advantages

The key novel and advantageous aspect of this work is the integration of Pareto navigation into the AP calibration process. Whilst the non-inferiority of AP has been demonstrated

across a wide range of publications [11], there are clear challenges when it comes to the calibration methods implemented within the various AP approaches. Prior to this work two core methods were utilised for calibrating AP solutions: simple trial and error, and use of machine learning to train AP algorithms on historical patients. Fundamentally both have the same weakness as they, explicitly or implicitly, heavily rely on trial and error, either when adjusting calibration parameters, or when generating plans for use in a calibration dataset. The key issue with trial and error approaches is that the search space is large and feedback loops are slow (at least 5 minutes per optimisation loop). The process is therefore prone to be time-consuming and/or deliver sub-optimal solutions due to the solution space not being fully explored. Utilising $MCO_{posteriori}$ within the calibration process offers a clear alternative. By generating a set of Pareto plans *a priori*, the search space is pre-populated, enabling the operator to navigate seamlessly across the different treatment options with feedback given in real-time. In this study, whilst navigation was limited to one trade-off at a time, the methodology enabled the operator to intuitively navigate the impact of varying the emphasis of a given trade-off through real-time updates to the whole 3D dose distribution and DVHs. Figure 7, provides a good example of the benefits of the interface. In this example the operator is navigating through different weightings of the conformality trade-off, with the mean rectum dose increasing as a consequence of a tighter (improved) conformality. The key decision the operator must make is how to optimally balance this trade-off. This decision is complex, it involves not only an assessment of the clinical acceptability/optimality of the current navigated distribution, but also an understanding of the rate of change of the different dose metrics as the balancing of any trade-off is adjusted. The rate of change is key for any decision making; for example a detriment in the Paddick's conformality index of 0.05 may be acceptable if the mean rectum dose reduces by 0.5 Gy but likely unacceptable if the rectum dose reduces by 0.05 Gy. In the example in Figure 7, the rate of change in the mean rectum dose was small as the conformality was increased, therefore a calibration with a high conformality was selected. Current trial and error methods provide no readily available information on the rate of change of a metric and therefore decision making is hindered. It is the author's view that at present Pareto navigation is the only method which provides the operator with ready access to this key information when calibrating an automated solution (via both the DVH and whole 3D dose distribution); ensuring that informed decisions can be made on the appropriate balancing of trade-offs.

A second key advantage of the developed approach is it can be calibrated on small datasets. As discussed in section 10.2.1, the ability to calibrate on small datasets is in-line with existing PB-AIO and εc solutions [49,72], but offers a key advantage over KBP solutions where large datasets are required [39]. Small calibration datasets ensure that automated solutions can be dynamic and adaptable to an ever-changing technological environment, without the need to continually update and curate more substantial patient datasets. Whilst a range of studies have demonstrated broad scope KBP models can be created that are suitable across multiple institutions or differing protocols [108,111,115], in these publications the models have been developed specifically for transferability. Clinically implemented KBP solutions are unlikely to be developed in this manner and therefore their suitability and optimality to differing protocols cannot be guaranteed. Furthermore, for novel protocols it is highly likely that existing KBP models would be completely unsuitable, with a full model generation required. For example, in 2019 the phase III PACE-B trial reported results that established the non-inferiority of extreme hypo-fractionated prostate radiotherapy (EHRT) (40 Gy in 5#) when compared to a conventional hypo-fractionated technique (60 Gy in 20#) [131]. The two treatment protocols differed substantially in terms of their PTV and OAR prioritisation and therefore KBP implementation of AP for EHRT would require creation of a new model trained on a dataset of >40 EHRT patient plans. This is in stark contrast to the approach developed in this body of work (and existing PB-AIO and εc solutions) where new AP solutions can be created using small patient datasets ($5 \leq n \leq 20$) with no requirement for a reference treatment plan.

Having no requirement for reference treatment plans during the AP calibration process is another key advantage of the developed approach, which again is in-line with PB-AIO and εc solutions, and is in contrast to KBP solutions where reference plans are a fundamental requirement. Unlike KBP, automated solutions will therefore not be influenced by the quality of planning databases. This is vitally important to ensure the optimality of an AP solution, as calibration datasets are generally populated with manually produced clinical plans that are: (i) prone to inter-observer variation [6], (ii) may generate plans that are not fully aligned with oncologist treatment aims [23], and (iii) can yield clinically suboptimal results [132]. Within the KBP based literature there is a growing evidence base both on the dependence of the AP solutions on the underlying database [40,90,133] and on the development of a range

of strategies to mitigate this issue. Strategies include calibrating using multiple training loops, which aim to incrementally improve the quality in the underlying dataset [91,92] and populating databases using $MCO_{posteriori}$ [21,89] or alternative AP solutions [43,90,134] in a bid to train on Pareto optimal plans. Therefore, whilst KBP offers a conceptually attractive AP approach (i.e. learn from historical plans to treat novel patients), the non-trivial requirement to train on large, high quality datasets, leads to substantial and significant disadvantages over the approach developed in this work.

A further potential benefit of the developed approach is utilisation of dynamic objectives to drive plans towards Pareto optimality. As discussed in section 10.2.2, results of the evaluative study provide indicative evidence that dynamic objectives fulfil this function, which is key to delivering optimal plans for individual patients. This approach is in-line with PB-AIO and εc methodologies, and offers a key advantage over KBP, where OAR doses are not minimised to achieve Pareto optimality, but rather to match predicted doses based on the training dataset.

Finally, whilst the automated solution developed in this body of work is implemented within RayStation, a key advantage of the approach taken is that the core concepts (calibration via Pareto navigation and implementation of dynamic objectives) are generalisable to differing planning systems or automated solutions. In terms of calibration via Pareto navigation, this could theoretically be applied to any automated system where plan generation is based on an input template (i.e. PB-AIO and εc). The key requirements for implementation are developing methods to: generate a range of plans (Pareto dataset) using different input template parameters, perform real-time navigation of the dataset and reduce the approximation error to clinically negligible levels. In terms of dynamic objectives, there are already a range of examples within the literature where similar methodologies have been implemented across differing systems, generally via scripting [64–68]. It is therefore expected that the approach developed in this work would be readily implementable in a range of planning systems either via scripting, or for maximum benefit through native integration in a commercial optimiser.

10.3.2 Disadvantages

A key limitation of the implemented approach is that Pareto navigation could only be performed on one trade-off at a time, meaning that a full exploration of the Pareto surface was not performed. For the simple site of PSV, this was not considered a key issue as there were relatively few competing trade-offs, and balancing those trade-offs with a single slider was straightforward. However for the complex site of PPN, trade-off balancing was more complex and there were occasions where navigating multiple competing trade-offs simultaneously would have been advantageous to derive the most clinically desirable solution. This was most notable in the abdominal area where bladder, bowel and PTV conformity were challenging to balance effectively. Extending the Pareto navigation functionality to enable multi-dimensional navigation is a key area where the developed solution could be improved.

A further disadvantage of the approach is the time required to generate a treatment plan. On an individual patient basis, the measured plan generation time of 14.0 mins and 36.4 mins for PSV and PPN respectively is in-line with manual planning and perfectly suitable for standard clinical practice. However, radiotherapy is rapidly evolving and online adaptive radiotherapy (where patients are scanned, planned and treated within a single treatment slot) is likely to become a significant part of the radiotherapy pathway for a range of cancer indications in the future. For online adaptive radiotherapy, plan generation time needs to be in the order of minutes, an order of magnitude quicker than is possible with the currently implemented methodology. In addition, planning time also limits the number of plans which can be realistically generated within a Pareto plan dataset. For single slider navigations this is not a major issue as 5 plans can be generated within 70 and 182 minutes for PSV and PPN respectively. However, for multi-dimensional navigation with no advanced sampling algorithms the computational cost increases to the power of the number of dimensions. For PSV a 3 and 4 dimensional navigation would require 29.2 and 145.8 hours to generate a suitable Pareto plan set; current plan generation times therefore significantly hinder the number of dimensions that can be navigated in parallel. Native optimisations within RayStation are actually relatively fast (for PSV, ~120 seconds per 100 iterations), the primary reason for the slow generation time with the developed method is instead related to updating the dynamic objectives at multiple intervals as the optimisation progresses. This is

a surprisingly slow process using the scripting interface, taking ~0.25 seconds to read or write a single objective parameter. With each objective having multiple parameters (e.g. weight, dose level and volume) and ~26 objectives required for a typical SVP optimisation, updating all dynamic objectives within the optimisation problem is a time consuming process, taking ~20 seconds. There is therefore no simple method to speed up the plan generation process without significant changes to RayStation's scripting implementation, which is outside of the user's control. Importantly though, the core automated planning methodology (i.e. implementation of dynamic objectives) is not inherently slow. In fact, it is assumed that if dynamic objectives were implemented natively within a treatment planning optimiser, automated plan generation time would be no slower than a standard optimisation round (e.g. for PSV, ~120 second per 100 iterations).

A final potential limitation in the developed approach is the assumption that through implementing dynamic objectives, a standard set of calibration weights is suitable and optimal for all patients in a given treatment protocol. For PSV, experience during the calibration and results of the evaluative study provide indicative evidence that this assumption is valid. The solution was successfully calibrated on a single patient and robust across a range of patient anatomies (Figure 8). However for PPN, where there is a large inter-patient variation in the anatomy in the abdominal cavity, it was considered that the challenges observed during the calibration were indicative that the optimum calibration weights for a given patient were still correlated with anatomical geometry. Nevertheless, even if the calibration weights were not fully optimum for all patients, results from the evaluative study demonstrated AP was still considered non-inferior to expert planners operating under no time pressure. It is therefore assumed that by refining the calibration across a larger dataset, the solution could be tailored toward the average anatomy of the patient population and therefore remain robust across the whole spread of anatomical variation. It is also important to note that when compared to existing AP approaches, these results do not point to any inferiority of the developed solution. For example, Cagni *et al.* [43] demonstrated that even when calibrated against a Pareto optimal dataset, for 62% of prostate cancer patients KBP yielded modelling errors corresponding to a deviation of ±10% in the predicted NTCP for rectal bleeding. Similarly, for PB-AIO Janssen *et al.* [88] identified

that 8%-25% of prostate plans generated using Pinicale's AutoPlan across two institutions were considered suboptimal.

10.4 Strengths and Weaknesses of the Research Process

10.4.1 Strengths

The key strength of the work undertaken in this project is that the automated system was developed by clinical experts within a commercial system. It has been built from the ground up with a focus on practical, broad scope clinical implementation and therefore has the potential for rapid translation into the patient care pathway. Whilst this report has focused on presenting the methodology of the developed solution and evaluating its application to prostate cancer, at Velindre Cancer Centre additional work has been undertaken in a wide programme of work which has built a substantive evidence base on the solution's generalisability. To date the solution has been robustly evaluated for PSV, PPN, prostate EHRT, head and neck cancer, and PSV at two external centres [135]. Across the 134 plans evaluated in these studies all $VMAT_{Auto}$ plans were deemed clinically acceptable, with 93% considered equivalent or better than $VMAT_{Manual}$ upon review by a clinical oncologist. Furthermore, as will be discussed in section 10.5.1, the automated solution has been clinically implemented for prostate EHRT, rectum, and head and neck patients. Additional solutions have also been developed and validated for lung, anus, oesophagus and prostate bed cancers, to support a widespread implementation during 2021 and early 2022.

A further strength of the research is the sound design used in the main evaluative study (section 9.5.4). There are two aspects of the study design which are considered particularly strong. Firstly, utilisation of two comparators ($VMAT_{Clinical}$ and $VMAT_{Ideal}$) ensured that robust evidence was generated not only on the likely impact AP would have on clinical practice, but also on the optimality of automated plans when compared to a gold standard ($VMAT_{Ideal}$). This is in contrast to a high proportion of published studies where previous clinical plans are selected as the comparator. As already discussed, there is significant evidence that clinical plans generated manually are prone to unwarranted variation and sub optimal quality [6,132]. Therefore, unlike the study presented in this work, the published studies only using clinical plans as a comparator cannot provide sound conclusions as to the optimality of the automated approach. Secondly, this study included a blind qualitative review by an experienced oncologist, which is considered vital to ensure the clinical suitability of

automated plans. As with the discussion on comparators, a significant proportion of published literature has no qualitative assessment of the generated plans, with only quantitative analysis of dose metrics given. This is considered a real weakness in studies comparing AP with MP, as it is easy for dose metrics to hide sub-optimality as they remove important spatial information on the dose distribution. For example, consider two plans, both with the PTV D99% equalling 95% of the prescribed dose (a common dose metric constraint). In Plan A the area of PTV receiving a dose <95% is limited to the periphery of the volume but in Plan B the under-dose is central and over an area where there is known disease. In terms of dose metrics the plans are equivalent, however a qualitative review of the 3D dose-distribution would likely highlight Plan B as being sub-optimal and possibly, clinically unapprovable.

The final core strength of this work is that there has been substantive patient involvement throughout the project, from conception to clinical implementation, with two patient research partners (RP) recruited as part of the project delivery team. Patient involvement had a number of benefits. Firstly it ensured that the team was continually focused on the end goal of delivering patient benefit, which can sometimes be lost in a highly technical project. Secondly, co-producing the research with patient representatives added a weight and validity to the body of work, which would not be possible otherwise; acknowledgement of the impact of the work via both professional colleagues and representatives of those who use the service adds a level of significance to the research. Finally, with RPs being external to the department and organisation, their participation in the project was also considered to add an additional layer of external governance to the process, with RPs given opportunity through the project to communicate any concerns they had in the direction of the work or the manner in which it was undertaken.

10.4.2 Weaknesses

The key weakness of the research process as a whole was that whilst the evaluative study provided robust evidence on the efficacy of the developed approach compared to MP, minimal evidence was generated that justified and validated the individual methodological approaches adopted within the solution. Key examples include: the use of a variety of hard coded weight scaling factors (Table 2) including a volume scaling factor (f_v) for P_2 target objectives, implementation of an optional ‘modulation optimisation’ (Figure 6), the values of

parameters and ranges within the PB-AIO framework (Figure 6) and most importantly calibration using Pareto navigation. The research process as a whole instead focused on the AP output as providing the overall justification of the approach undertaken, rather than critically analysing each component separately. This approach had the advantage that an automated solution could be constructed in an efficient manner, relying on previous planning experience or publications to justify implementing a method rather than a time consuming critical analysis of each component. However, it also led to a number of weaknesses. Firstly, which is especially the case for the weight scaling factors, it is highly likely the method or factors are not fully optimised. Secondly, it is feasible that the implemented methods do not fulfil their intended function or have unintended consequences. For example, the volume scaling factor f_v may have the unintended consequence of diluting the OAR objective weights for patient with large PTVs, resulting in shift in the balancing of the plan away from OAR minimisation and towards PTV coverage. Thirdly, when disseminating the work, readers are reliant on the author's observations, rather than scientific results when considering the influence, impact and benefits of the implemented methodologies. This is particularly relevant for the novel aspect of calibration through Pareto navigation. As an operator, the successful application of this methodology is clear; plans generated on novel patients exhibit similar balancing of competing trade-offs as the navigation patients. However to the reader, there is no substantial evidence demonstrating the impact different navigations have on the resultant automated solution, only that the overall automated planning methodology yielded plans which were non-inferior to MP.

One approach that would have made this research stronger, without excessive additional work, would have been to perform a sensitivity analysis on the different hardcoded parameters or methodologies. In this approach, the values of a range of variables are modified and the impact on the distribution explored. For example, by assessing the effect of the output if the Δ^i bounds during the stage 1 optimisation (Figure 6) are changed from [0.15 – 0.5] to [0.25 – 0.6] and [0.05 – 0.4] the importance of the tuning of these parameters on the distribution is explored. Specifically for the Pareto navigation calibration process, which is the core novelty of this work, a simple approach would have been to calibrate the solution on a single patient according to two different preferences of trade-off balancing,

one focusing on conformality and homogeneity, and the other on driving down OAR doses. Through comparing the distributions of these two calibrations for both the calibration patient and a selection of novel patients, the impact of Pareto navigation on the resultant solution could have been clearly demonstrated.

An additional weakness of the research is that the patient dataset is relatively small, and may not reflect the full diversity of the patient cohort due to the patient selection criteria (for example, patients with prosthetic hips were excluded from the study). Furthermore, whilst plans were generated by expert planners under no time pressure, there is the potential this is not the best comparator. Plans generated in this manner are still reliant on trial and error, with the quality limited by the skills and experience of the individual operator. This research could therefore be strengthened by running additional studies across multiple institutions and through including $MCO_{posteriori}$ and alternative AP techniques as comparators. In this way results would not be limited to the tools, techniques or experience within the local institution. A final weakness of this research is that Pareto Navigation is performed one trade-off at a time. It is acknowledged that this is a simplified approach, chosen for a first stage implementation, however it could be challenged whether this is true Pareto navigation as the full search space cannot be explored by the operator. Extension of the approach to multiple dimensions should alleviate this potential criticism.

10.5 Clinical Implementation and Further Research

10.5.1 Clinical Implementation

A key challenge with any research in the field of healthcare is its timely translation into clinical practice for the benefit of patients. Following the strong results of the evaluative study a programme of work was put in place to implement the automated planning methods developed in this project at Velindre Cancer Centre. This included validating the software for clinical use according to best practice (i.e. documentation, risk assessments, independent code review and software testing) and developing automated protocols for new sites. To date automation has been implemented for the sites of rectum, prostate EHRT, and head and neck patients. In addition, solutions have been developed for prostate bed, anus, oesophagus, lung and prostate cancer patients, with clinical implementation of these groups expected by early 2022. Finally, following a successful multi-centre study for prostate cancer

[135], discussions are underway on the feasibility of clinical implementation at external NHS centres.

Clinical implementation is a key success of this work. It is testament to the value of embedding research within the clinical workforce who have a deep understanding of the key challenges in healthcare and can establish novel, yet practical solutions to help resolve them.

10.5.2 Further Research

Throughout this critical appraisal a number of areas for future research have been identified. These included: additional validations across new clinical sites, external institutions and differing comparators (e.g. $MCO_{Posteriori}$ and alternative AP solutions); extension to multi-dimensional Pareto navigation; and further investigation into individual components of the methodology (e.g. validation or refinement of scaling factors). As discussed, as part of a continuing program of work on this subject at Velindre Cancer Centre, the automated solution has already been calibrated across a range of new treatment sites. Furthermore, the extension to multi-dimensional Pareto navigation and its validation across two external institutions has been completed for prostate cancer. Results demonstrated that multi-dimensional navigation was feasible (albeit with a substantially increased Pareto dataset generation time due the extra dimensions) and upon blind review by an oncologist, the resultant AP solutions were considered equivalent or better than manually generated clinical plans in 37/40 cases. Interestingly, unlike the results presented in this work, automation led to a substantial reduction in rectal doses at the lower dose levels across both institutions, with the V24.3 Gy dose metric reduced by more than 8% [135]. This was considered to provide supportive evidence on the benefit of using dynamic objectives to drive down OAR doses towards optimality. Due to the growing evidence base supporting both the optimality and generalisability of the developed approach, future research on refining the methodology's individual components is now not considered necessary. Therefore future research will instead focus on more novel work relating to Pareto navigation, machine learning and comparison with alternative advanced techniques.

In terms of Pareto navigation, the updated tool enables efficient multi-dimensional exploration of trade-offs, where each position on the Pareto front is represented by a specific set of automated planning calibration factors (w_{nom}, section 8.4.2.4). Through

running the calibration process independently across a database of patients, the ideal set of automated planning calibration factors for each individual patient can be determined. If the developed automated solution operated perfectly, these factors would be consistent across the patient population. However, based on experience calibrating the solution for PPN it is hypothesised there is an underlying correlation between the optimum calibration factors and individual patient anatomy. Future research would therefore look at using machine learning to identify and correct for these correlations, thereby improving the quality of the automated solution. Furthermore as the data on the ideal calibration factors also represent the clinically optimal point on the Pareto front for each patient, sampling this data across a dataset also has the effect of sampling the clinically relevant parts of the Pareto front for a given patient cohort. Further work could therefore investigate using this data on novel patients to limit the Pareto front to only include clinically relevant solutions, thereby significantly reducing the search space and sampling requirements of the Pareto dataset. This could significantly improve the efficiency of $MCO_{posteriori}$ solutions or, by maintaining the number of plans in a Pareto dataset, be utilised to significantly reduce the approximation error.

A final area of further research is strengthening the evidence base on the comparison of advanced optimisation techniques, which was identified as a key area of weakness in the literature (section 7.4). Robust comparison of advanced techniques is however challenging as differing solutions are implemented within different planning systems, leading to a range of confounding factors that limit the validity of the results. For example, differences in plan quality maybe completely unrelated to the implemented advanced approach but rather represent differences in the linac treatment planning models, dose calculation engines or the quality of the systems native optimisation engine. In this regard, an obvious piece of work to undertake is comparison of the developed automated approach with RayStation's $MCO_{posteriori}$ solution. As plans would be generated within the same treatment planning system, there would be minimal confounding factors and results would provide sound evidence on the comparative effectiveness of AP and $MCO_{posteriori}$; filling a key gap in the literature on both the comparison of advanced techniques and utility of $MCO_{posteriori}$.

10.6 Summary and Key Implications and Recommendations

In this work a versatile automated planning solution has been developed, which uniquely utilises Pareto navigation to enable the intuitive calibration of automated protocols. It has been robustly evaluated across two sites of differing complexity (PSV and PPN) and efficiently generates high quality plans, which are non-inferior to expert planners.

Results from this work have led to a broad scope clinical implementation at Velindre Cancer Centre, with a view to improve the quality and efficiency of the treatment planning service. To date, automated solutions have been developed for PSV, PPN, prostate bed, prostate EHRT, anus, rectum, oesophagus, lung and head and neck, which demonstrates the generalisability and versatility of the developed approach.

In this section the developed solution and results of the research have been critically appraised. Based on this analysis the work presented in this thesis is considered to have the following key implications for theory and practice:

- Pareto Navigation is an effective calibration tool for automated planning solutions and presents a clear alternative to existing calibration methods.
- Dynamic objectives are effective in both ensuring the Pareto optimality of generated plans and reducing the dependence of optimisation parameters on patient anatomy. There is a strong case that commercial planning systems should support dynamic objectives natively within their inverse optimiser.
- Automated planning is non-inferior to manual planning and there is a strong rational for its rollout across UK radiotherapy services to ensure the quality and efficiency of radiotherapy plan generation.

11 References

[1] Delaney G, Jacob S, Featherstone C, Barton M. The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines. Cancer 2005;104:1129–37. https://doi.org/10.1002/cncr.21324.

[2] Fairchild A, Straube W, Laurie F, Followill D. Does quality of radiation therapy predict outcomes of multicenter cooperative group trials? A literature review. Int J Radiat Oncol Biol Phys 2013;87:246–60. https://doi.org/10.1016/j.ijrobp.2013.03.036.

[3] Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials. J Natl Cancer Inst 2013;105:387–93. https://doi.org/10.1093/jnci/djt001.

[4] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy. Radiother Oncol 2012;105:4–8. https://doi.org/10.1016/j.radonc.2012.08.008.

[5] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02. J Clin Oncol 2010;28:2996–3001. https://doi.org/10.1200/JCO.2009.27.4498.

[6] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. https://doi.org/10.1016/j.prro.2011.11.012.

[7] Staffurth J. A review of the clinical evidence for intensity-modulated radiotherapy. Clin Oncol 2010;22:643–57. https://doi.org/10.1016/j.clon.2010.06.013.

[8] Veldeman L, Madani I, Hulstaert F, De Meerleer GO, Mareel M, De Neve W. Evidence behind use of intensity-modulated radiotherapy: a systematic review of comparative clinical studies. Lancet Oncol 2008;9:367–75. https://doi.org/10.1016/S1470-2045(08)70098-6.

[9] Batumalai V, Jameson MG, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. Pract Radiat Oncol 2013;3:e99–106. https://doi.org/10.1016/j.prro.2012.06.009.

[10] Kamperis E, Kodona C, Hatziioannou K, Giannouzakos V. Complexity in Radiation Therapy: It's Complicated. Int J Radiat Oncol Biol Phys 2020;106:182–4. https://doi.org/10.1016/j.ijrobp.2019.09.003.

[11] Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations. Br J Radiol 2018;91:20180270. https://doi.org/10.1259/bjr.20180270.

[12] Thieke C, Küfer KH, Monz M, Scherrer A, Alonso F, Oelfke U, et al. A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. Radiother Oncol 2007;85:292–8. https://doi.org/10.1016/j.radonc.2007.06.020.

[13] Craft D, Halabi T, Shih HA, Bortfeld T. An Approach for Practical Multiobjective IMRT Treatment Planning. Int J Radiat Oncol Biol Phys 2007;69:1600–7. https://doi.org/10.1016/j.ijrobp.2007.08.019.

[14] McGarry CK, Bokrantz R, O'Sullivan JM, Hounsell AR. Advantages and limitations of navigation-based multicriteria optimization (MCO) for localized prostate cancer IMRT planning. Med Dosim 2014;39:205–11. https://doi.org/10.1016/j.meddos.2014.02.002.

[15] Wala J, Craft D, Paly J, Zietman A, Efstathiou J. Maximizing dosimetric benefits of IMRT in the treatment of localized prostate cancer through multicriteria optimization planning. Med Dosim 2013;38:298–303. https://doi.org/10.1016/j.meddos.2013.02.012.

[16] Ghandour S, Matzinger O, Pachoud M. Volumetric-modulated arc therapy planning using multicriteria optimization for localized prostate cancer. J Appl Clin Med Phys 2015;16:258–69. https://doi.org/10.1120/jacmp.v16i3.5410.

[17] Buschmann M, Seppenwoolde Y, Wiezorek T, Weibert K, Georg D. Advanced optimization methods for

whole pelvic and local prostate external beam therapy. Phys Medica 2016;32:465–73. https://doi.org/10.1016/j.ejmp.2016.03.002.

[18] Guerrero M, Fellows Z, Mohindra P, Badiyan S, Lamichhane N, Snider JW, et al. Multicriteria optimization: Site-specific class solutions for VMAT plans. Med Dosim 2020;45:7–13. https://doi.org/10.1016/j.meddos.2019.04.003.

[19] Lu L, Sheng Y, Donaghue J, Liu Shen Z, Kolar M, Wu QJ, et al. Three IMRT advanced planning tools: A multi-institutional side-by-side comparison. J Appl Clin Med Phys 2019;20:65–77. https://doi.org/10.1002/acm2.12679.

[20] Kierkels RG, Visser R, Bijl HP, Langendijk JA, van 't Veld AA, Steenbakkers RJ, et al. Multicriteria optimization enables less experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. Radiat Oncol 2015;10. https://doi.org/10.1186/s13014-015-0385-9.

[21] Miguel-Chumacero E, Currie G, Johnston A, Currie S. Effectiveness of Multi-Criteria Optimization- based Trade-Off exploration in combination with RapidPlan for head & neck radiotherapy planning. Radiat Oncol 2018;13:1–13. https://doi.org/10.1186/s13014-018-1175-y.

[22] Xiao J, Li Y, Shi H, Chang T, Luo Y, Wang X, et al. Multi-criteria optimization achieves superior normal tissue sparing in intensity-modulated radiation therapy for oropharyngeal cancer patients. Oral Oncol 2018;80:74–81. https://doi.org/10.1016/j.oraloncology.2018.03.020.

[23] Craft DL, Hong TS, Shih HA, Bortfeld TR. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2012;82:83–90. https://doi.org/10.1016/j.ijrobp.2010.12.007.

[24] Zieminski S, Khandekar M, Wang Y. Assessment of multi-criteria optimization (MCO) for volumetric modulated arc therapy (VMAT) in hippocampal avoidance whole brain radiation therapy (HA-WBRT). J Appl Clin Med Phys 2018;19:184–90. https://doi.org/10.1002/acm2.12277.

[25] Kamran SC, Mueller BS, Paetzold P, Dunlap J, Niemierko A, Bortfeld T, et al. Multi-criteria optimization achieves superior normal tissue sparing in a planning study of intensity-modulated radiation therapy for RTOG 1308-eligible non-small cell lung cancer patients. Radiother Oncol 2016;118:515–20. https://doi.org/10.1016/j.radonc.2015.12.028.

[26] Müller BS, Shih HA, Efstathiou JA, Bortfeld T, Craft D. Multicriteria plan optimization in the hands of physicians: a pilot study in prostate cancer and brain tumors. Radiat Oncol 2017;12:168. https://doi.org/10.1186/s13014-017-0903-z.

[27] Kyroudi A, Petersson K, Ghandour S, Pachoud M, Matzinger O, Ozsahin M, et al. Discrepancies between selected Pareto optimal plans and final deliverable plans in radiotherapy multi-criteria optimization. Radiother Oncol 2016;120:346–8. https://doi.org/10.1016/j.radonc.2016.05.018.

[28] Craft DL, Halabi TF, Shih H a, Bortfeld TR. Approximating convex pareto surfaces in multiobjective radiotherapy planning. Med Phys 2006;33:3399–407. https://doi.org/10.1118/1.2335486.

[29] Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? Phys Med Biol 2008;53:2785–96. https://doi.org/10.1088/0031-9155/53/11/002.

[30] Yuan L, Ge Y, Lee WR, Yin FF, Kirkpatrick JP, Wu QJ. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. Med Phys 2012;39:6868. https://doi.org/10.1118/1.4757927.

[31] Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. Med Phys 2012;39:7446–61. https://doi.org/10.1118/1.4761864.

[32] Babier A, Boutilier JJ, McNiven AL, Chan TCY. Knowledge-based automated planning for oropharyngeal cancer. Med Phys 2018;45:2875–83. https://doi.org/10.1002/mp.12930.

[33] Mcintosh C, Purdie TG. Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning. Phys Med Biol 2017;62:415–31. https://doi.org/10.1088/1361-6560/62/2/415.

[34] Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys 2018. https://doi.org/10.1002/mp.13271.

[35] Babier A, Mahmood R, McNiven AL, Diamant A, Chan TCY. Knowledge-based automated planning with three-dimensional generative adversarial networks. Med Phys 2020;47:297–306. https://doi.org/10.1002/mp.13896.

[36] McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Phys Med Biol 2017;62:5926–44. https://doi.org/10.1088/1361-6560/aa71f8.

[37] Babier A, Boutilier JJ, Sharpe MB, McNiven AL, Chan TCY. Inverse optimization of objective function weights for treatment planning using clinical dose-volume histograms. Phys Med Biol 2018;63:105004. https://doi.org/10.1088/1361-6560/aabd14.

[38] Ge Y, Wu QJ. Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches. Med Phys 2019;46:2760–75. https://doi.org/10.1002/mp.13526.

[39] Boutilier JJ, Craig T, Sharpe MB, Chan TCY. Sample size requirements for knowledge-based treatment planning. Med Phys 2016;43:1212–21. https://doi.org/10.1118/1.4941363.

[40] Wang Y, Heijmen BJM, Petit SF. Knowledge-based dose prediction models for head and neck cancer are strongly affected by interorgan dependency and dataset inconsistency. Med Phys 2019;46:934–43. https://doi.org/10.1002/mp.13316.

[41] Hussein M, South CP, Barry M a., Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol 2016;120:473–9. https://doi.org/10.1016/j.radonc.2016.06.022.

[42] Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol Biol Phys 2015;91:612–20. https://doi.org/10.1016/j.ijrobp.2014.11.014.

[43] Cagni E, Botti A, Wang Y, Iori M, Petit SF, Heijmen BJM. Pareto-optimal plans as ground truth for validation of a commercial system for knowledge-based DVH-prediction. Phys Medica 2018;55:98–106. https://doi.org/10.1016/j.ejmp.2018.11.002.

[44] Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys 2012;39:951–63. https://doi.org/10.1118/1.3676689.

[45] Della Gala G, Dirkx MLP, Hoekstra N, Fransen D, Lanconelli N, van de Pol M, et al. Fully automated VMAT treatment planning for advanced-stage NSCLC patients. Strahlentherapie Und Onkol 2017;193:402–9. https://doi.org/10.1007/s00066-017-1121-1.

[46] Voet PWJ, Dirkx MLP, Breedveld S, Fransen D, Levendag PC, Heijmen BJM. Toward fully automated multicriterial plan generation: A prospective clinical study. Int J Radiat Oncol Biol Phys 2013;85:866–72. https://doi.org/10.1016/j.ijrobp.2012.04.015.

[47] Buergy D, Sharfo AWM, Heijmen BJM, Voet PWJ, Breedveld S, Wenz F, et al. Fully automated treatment planning of spinal metastases - A comparison to manual planning of Volumetric Modulated Arc Therapy for conventionally fractionated irradiation. Radiat Oncol 2017;12:1–7. https://doi.org/10.1186/s13014-017-0767-2.

[48] Sharfo AWM, Breedveld S, Voet PWJ, Heijkoop ST, Mens JWM, Hoogeman MS, et al. Validation of fully automated VMAT plan generation for library-based plan-of-the-day cervical cancer radiotherapy. PLoS One 2016;11:1–13. https://doi.org/10.1371/journal.pone.0169202.

[49] Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Goldner G, Georg D, et al. Automated volumetric modulated arc therapy planning for whole pelvic prostate radiotherapy. Strahlentherapie Und Onkol 2018;194:333–42. https://doi.org/10.1007/s00066-017-1246-2.

[50] Sharfo AWM, Stieler F, Kupfer O, Heijmen BJM, Dirkx MLP, Breedveld S, et al. Automated VMAT planning for postoperative adjuvant treatment of advanced gastric cancer. Radiat Oncol 2018;13:74.

https://doi.org/10.1186/s13014-018-1032-z.

[51] Voet PWJ, Dirkx MLP, Breedveld S, Al-Mamgani A, Incrocci L, Heijmen BJM. Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. Int J Radiat Oncol Biol Phys 2014;88:1175–9. https://doi.org/10.1016/j.ijrobp.2013.12.046.

[52] Heijmen B, Voet P, Fransen D, Penninkhof J, Milder M, Akhiat H, et al. Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy – An international multi-center validation for prostate cancer. Radiother Oncol 2018;128:343–8. https://doi.org/10.1016/j.radonc.2018.06.023.

[53] Bijman R, Rossi L, Janssen T, de Ruiter P, Carbaat C, van Triest B, et al. First system for fully-automated multi-criterial treatment planning for a high-magnetic field MR-Linac applied to rectal cancer. Acta Oncol (Madr) 2020;59:926–32. https://doi.org/10.1080/0284186X.2020.1766697.

[54] Craft D, Süss P, Bortfeld T. The Tradeoff Between Treatment Plan Quality and Required Number of Monitor Units in Intensity-modulated Radiotherapy. Int J Radiat Oncol Biol Phys 2007;67:1596–605. https://doi.org/10.1016/j.ijrobp.2006.11.034.

[55] Breedveld S, Craft D, van Haveren R, Heijmen B. Multi-criteria optimization and decision-making in radiotherapy. Eur J Oper Res 2019;277:1–19. https://doi.org/10.1016/j.ejor.2018.08.019.

[56] van Haveren R, Ogryczak W, Verduijn GM, Keijzer M, Heijmen BJM, Breedveld S. Fast and fuzzy multi-objective radiotherapy treatment plan generation for head and neck cancer patients with the lexicographic reference point method (LRPM). Phys Med Biol 2017;62:4318–32. https://doi.org/10.1088/1361-6560/62/11/4318.

[57] van Haveren R, Breedveld S, Keijzer M, Voet P, Heijmen B, Ogryczak W. Lexicographic extension of the reference point method applied in radiation therapy treatment planning. Eur J Oper Res 2017;263:247–57. https://doi.org/10.1016/j.ejor.2017.04.062.

[58] van Haveren R, Heijmen BJM, Breedveld S. Automatically configuring the reference point method for automated multi-objective treatment planning. Phys Med Biol 2019;64:35002. https://doi.org/10.1088/1361-6560/aaf9fe.

[59] van Haveren R, Heijmen BJM, Breedveld S. Automatic configuration of the reference point method for fully automated multi-objective treatment planning applied to oropharyngeal cancer. Med Phys 2020;47:1499–508. https://doi.org/10.1002/mp.14073.

[60] Xhaferllari I, Wong E, Bzdusek K, Lock M, Chen J. Automated IMRT planning with regional optimization using planning scripts. J Appl Clin Med Phys 2013;14:4052. https://doi.org/10.1120/jacmp.v14i1.4052.

[61] Winkel D, Bol GH, van Asselen B, Hes J, Scholten V, Kerkmeijer LGW, et al. Development and clinical introduction of automated radiotherapy treatment planning for prostate cancer. Phys Med Biol 2016;61:8587–95. https://doi.org/10.1088/1361-6560/61/24/8587.

[62] Guo C, Zhang P, Gui Z, Shu H, Zhai L, Xu J. Prescription Value-Based Automatic Optimization of Importance Factors in Inverse Planning. Technol Cancer Res Treat 2019;18:1533033819892259. https://doi.org/10.1177/1533033819892259.

[63] Ayala R, Ruiz G, Valdivielso T. Automatizing a nonscripting TPS for optimizing clinical workflow and reoptimizing IMRT/VMAT plans. Med Dosim 2019;44:409–14. https://doi.org/10.1016/j.meddos.2019.02.006.

[64] Tol JP, Dahele M, Delaney AR, Doornaert P, Slotman BJ, Verbakel WFAR. Detailed evaluation of an automated approach to interactive optimization for volumetric modulated arc therapy plans. Med Phys 2016;43:1818–28. https://doi.org/10.1118/1.4944063.

[65] Tol JP, Dahele M, Peltola J, Nord J, Slotman BJ, Verbakel WFAR. Automatic interactive optimization for volumetric modulated arc therapy planning. Radiat Oncol 2015;10:75. https://doi.org/10.1186/s13014-015-0388-6.

[66] Song Y, Wang Q, Jiang X, Liu S, Zhang Y, Bai S. Fully automatic volumetric modulated arc therapy plan generation for rectal cancer. Radiother Oncol 2016;119:531–6. https://doi.org/10.1016/j.radonc.2016.04.010.

[67] Quan EM, Chang JY, Liao Z, Xia T, Yuan Z, Liu H, et al. Automated Volumetric Modulated Arc Therapy Treatment Planning for Stage III Lung Cancer: How Does It Compare With Intensity-Modulated Radio Therapy? Int J Radiat Oncol Biol Phys 2012;84:e69–76. https://doi.org/10.1016/j.ijrobp.2012.02.017.

[68] Zhang X, Li X, Quan EM, Pan X, Li Y. A methodology for automatic intensity-modulated radiation treatment planning for lung cancer. Phys Med Biol 2011;56:3873–93. https://doi.org/10.1088/0031-9155/56/13/009.

[69] Hazell I, Bzdusek K, Kumar P, Hansen CR, Bertelsen A, Eriksen JG, et al. Automatic planning of head and neck treatment plans. J Appl Clin Med Phys 2016;17. https://doi.org/10.1120/jacmp.v17i1.5901.

[70] Zhang Q, Ou L, Peng Y, Yu H, Wang L, Zhang S. Evaluation of automatic VMAT plans in locally advanced nasopharyngeal carcinoma. Strahlentherapie Und Onkol 2021;197:177–87. https://doi.org/10.1007/s00066-020-01631-x.

[71] Zhang Q, Peng Y, Song X, Yu H, Wang L, Zhang S. Dosimetric evaluation of automatic and manual plans for early nasopharyngeal carcinoma to radiotherapy. Med Dosim 2020;45:e13–20. https://doi.org/10.1016/j.meddos.2019.05.005.

[72] Vanderstraeten B, Goddeeris B, Vandecasteele K, van Eijkeren M, De Wagter C, Lievens Y. Automated Instead of Manual Treatment Planning? A Plan Comparison Based on Dose-Volume Statistics and Clinical Preference. Int J Radiat Oncol Biol Phys 2018;102:443–50. https://doi.org/10.1016/j.ijrobp.2018.05.063.

[73] Speer S, Klein A, Kober L, Weiss A, Yohannes I, Bert C. Evaluation of head and neck cancer patient plans created by the Pinnacle3 scripting and Auto-Planning functions. Strahlentherapie Und Onkol 2017;193:656–65. https://doi.org/10.1007/s00066-017-1150-9.

[74] McConnell K a., Marston T, Zehren BE, Lirani A, Stanley DN, Bishop A, et al. Dosimetric Evaluation of Pinnacle's Automated Treatment Planning Software to Manually Planned Treatments. Technol Cancer Res Treat 2018;17:153303381878006. https://doi.org/10.1177/1533033818780064.

[75] Cilla S, Ianiro A, Romano C, Deodato F, Macchia G, Buwenge M, et al. Template-based automation of treatment planning in advanced radiotherapy: a comprehensive dosimetric and clinical evaluation. Sci Rep 2020;10:423. https://doi.org/10.1038/s41598-019-56966-y.

[76] Yang Y, Shao K, Zhang J, Chen M, Chen Y, Shan G. Automatic Planning for Nasopharyngeal Carcinoma Based on Progressive Optimization in RayStation Treatment Planning System. Technol Cancer Res Treat 2020;19. https://doi.org/10.1177/1533033820915710.

[77] Creemers IHP, Kusters JMAM, van Kollenburg PGM, Bouwmans LCW, Schinagl DAX, Bussink J. Comparison of dose metrics between automated and manual radiotherapy planning for advanced stage non-small cell lung cancer with volumetric modulated arc therapy. Phys Imaging Radiat Oncol 2019;9:92–6. https://doi.org/10.1016/j.phro.2019.03.003.

[78] Wang H, Chen H, Shao Y, Gu H, Duan Y, Feng A, et al. A new normal portion of the esophagus (NPE) sparing technique for upper and middle esophagus cancer by SIB-IMRT automatic planning. Int J Radiat Res 2019;17:549–57.

[79] Hansen CR, Bertelsen A, Hazell I, Zukauskaite R, Gyldenkerne N, Johansen J, et al. Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans. Clin Transl Radiat Oncol 2016;1:1–7. https://doi.org/10.1016/j.ctro.2016.08.001.

[80] Hansen CR, Nielsen M, Bertelsen AS, Hazell I, Holtved E, Zukauskaite R, et al. Automatic treatment planning facilitates fast generation of high-quality treatment plans for esophageal cancer. Acta Oncol (Madr) 2017;56:1495–500. https://doi.org/10.1080/0284186X.2017.1349928.

[81] Li X, Wang L, Wang J, Han X, Xia B, Wu S, et al. Dosimetric benefits of automation in the treatment of lower thoracic esophageal cancer: Is manual planning still an alternative option? Med Dosim 2017;42:289–95. https://doi.org/10.1016/j.meddos.2017.06.004.

[82] Nawa K, Haga A, Nomoto A, Sarmiento R a., Shiraishi K, Yamashita H, et al. Evaluation of a commercial automatic treatment planning system for prostate cancers. Med Dosim 2017;42:203–9. https://doi.org/10.1016/j.meddos.2017.03.004.

[83] Gallio E, Giglioli FR, Girardi A, Guarneri A, Ricardi U, Ropolo R, et al. Evaluation of a commercial automatic treatment planning system for liver stereotactic body radiation therapy treatments. Phys Medica 2018;46:153–9. https://doi.org/10.1016/j.ejmp.2018.01.016.

[84] Chen K, Wei J, Ge C, Xia W, Shi Y, Wang H, et al. Application of auto-planning in radiotherapy for breast cancer after breast-conserving surgery. Sci Rep 2020;10:10927. https://doi.org/10.1038/s41598-020-68035-w.

[85] Marrazzo L, Meattini I, Arilli C, Calusi S, Casati M, Talamonti C, et al. Auto-planning for VMAT accelerated partial breast irradiation. Radiother Oncol 2019;132:85–92. https://doi.org/10.1016/j.radonc.2018.11.006.

[86] Clemente S, Oliviero C, Palma G, D'Avino V, Liuzzi R, Conson M, et al. Auto- versus human-driven plan in mediastinal Hodgkin lymphoma radiation treatment. Radiat Oncol 2018;13:202. https://doi.org/10.1186/s13014-018-1146-3.

[87] Quan EM, Li X, Li Y, Wang X, Kudchadker RJ, Johnson JL, et al. A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment. Int J Radiat Oncol Biol Phys 2012;83:1169–78. https://doi.org/10.1016/j.ijrobp.2011.09.015.

[88] Janssen TM, Kusters M, Wang Y, Wortel G, Monshouwer R, Damen E, et al. Independent knowledge-based treatment planning QA to audit Pinnacle autoplanning. Radiother Oncol 2019;133:198–204. https://doi.org/10.1016/j.radonc.2018.10.035.

[89] Wall PDH, Carver RL, Fontenot JD. An improved distance-to-dose correlation for predicting bladder and rectum dose-volumes in knowledge-based VMAT planning for prostate cancer. Phys Med Biol 2018;63:15035. https://doi.org/10.1088/1361-6560/aa9a30.

[90] Lin YH, Hong LX, Hunt MA, Berry SL. Use of a constrained hierarchical optimization dataset enhances knowledge-based planning as a quality assurance tool for prostate bed irradiation. Med Phys 2018;45:4364–9. https://doi.org/10.1002/mp.13163.

[91] Wang M, Li S, Huang Y, Yue H, Li T, Wu H, et al. An interactive plan and model evolution method for knowledge-based pelvic VMAT planning. J Appl Clin Med Phys 2018;19:491–8. https://doi.org/10.1002/acm2.12403.

[92] Chatterjee A, Serban M, Faria S, Souhami L, Cury F, Seuntjens J. Novel knowledge-based treatment planning model for hypofractionated radiotherapy of prostate cancer patients. Phys Medica 2020;69:36–43. https://doi.org/10.1016/j.ejmp.2019.11.023.

[93] Wang J, Chen Z, Li W, Qian W, Wang X, Hu W. A new strategy for volumetric-modulated arc therapy planning using AutoPlanning based multicriteria optimization for nasopharyngeal carcinoma. Radiat Oncol 2018;13:94. https://doi.org/10.1186/s13014-018-1042-x.

[94] Ling C, Han X, Zhai P, Xu H, Chen J, Wang J, et al. A hybrid automated treatment planning solution for esophageal cancer. Radiat Oncol 2019;14:232. https://doi.org/10.1186/s13014-019-1443-5.

[95] Bijman R, Rossi L, Sharfo AW, Heemsbergen W, Incrocci L, Breedveld S, et al. Automated Radiotherapy Planning for Patient-Specific Exploration of the Trade-Off Between Tumor Dose Coverage and Predicted Radiation-Induced Toxicity—A Proof of Principle Study for Prostate Cancer. Front Oncol 2020;10:943. https://doi.org/10.3389/fonc.2020.00943.

[96] Cancer Research UK. Prostate cancer statistics n.d. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer (accessed August 3, 2021).

[97] Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: An example application to prostate cancer planning. Int J Radiat Oncol Biol Phys 2013;87:176–81. https://doi.org/10.1016/j.ijrobp.2013.03.015.

[98] Yang Y, Ford EC, Wu B, Pinkawa M, van Triest B, Campbell P, et al. An overlap-volume-histogram based method for rectal dose prediction and automated treatment planning in the external beam prostate radiotherapy following hydrogel injection. Med Phys 2013;40:11709. https://doi.org/10.1118/1.4769424.

[99] Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the pre-clinical validation of a commercial model-based optimisation engine : Application to volumetric modulated arc therapy for patients with lung or prostate cancer. Radiother Oncol 2014;113:385–91. https://doi.org/10.1016/j.radonc.2014.11.009.

[100] Nwankwo O, Mekdash H, Sihono DSK, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. Radiat Oncol 2015;10:111. https://doi.org/10.1186/s13014-015-0416-6.

[101] Kubo K, Monzen H, Ishii K, Tamura M, Kawamorita R, Sumida I, et al. Dosimetric comparison of RapidPlan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. Phys Medica 2017;44:199–204. https://doi.org/10.1016/j.ejmp.2017.06.026.

[102] Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. PLoS One 2017;12:1–13. https://doi.org/10.1371/journal.pone.0178034.

[103] Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys 2011;38:2515–22. https://doi.org/10.1118/1.3574874.

[104] Yang J, Zhang P, Zhang L, Shu H, Li B, Gui Z. Particle swarm optimizer for weighting factor selection in intensity-modulated radiation therapy optimization algorithms. Phys Medica 2017;33:136–45. https://doi.org/10.1016/j.ejmp.2016.12.021.

[105] Amaloo C, Hayes L, Manning M, Liu H, Wiant D. Can automated treatment plans gain traction in the clinic? J Appl Clin Med Phys 2019;20:29–35. https://doi.org/10.1002/acm2.12674.

[106] Rossi L, Sharfo AW, Aluwini S, Dirkx M, Heijmen B. First fully automated planning solution for robotic radiosurgery – comparison with automatically planned volumetric arc therapy for prostate cancer. Acta Oncol (Madr) 2018;57:1–9. https://doi.org/10.1080/0284186X.2018.1479068.

[107] Wall PDH, Fontenot JD. Evaluation of complexity and deliverability of prostate cancer treatment plans designed with a knowledge-based VMAT planning technique. J Appl Clin Med Phys 2020;21:69–77. https://doi.org/10.1002/acm2.12790.

[108] Bossart E, Duffy M, Simpson G, Abramowitz M, Pollack A, Dogan N. Assessment of specific versus combined purpose knowledge based models in prostate radiotherapy. J Appl Clin Med Phys 2018;19:209–16. https://doi.org/10.1002/acm2.12483.

[109] Scaggion A, Fusella M, Roggio A, Bacco S, Pivato N, Rossato MA, et al. Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution. Phys Medica 2018;53:86–93. https://doi.org/10.1016/j.ejmp.2018.08.016.

[110] Masi K, Archer P, Jackson W, Sun Y, Schipper M, Hamstra D, et al. Knowledge-based treatment planning and its potential role in the transition between treatment planning systems. Med Dosim 2018;43:251–7. https://doi.org/10.1016/j.meddos.2017.10.001.

[111] Panettieri V, Ball D, Chapman A, Cristofaro N, Gawthrop J, Griffin P, et al. Development of a multicentre automated model to reduce planning variability in radiotherapy of prostate cancer. Phys Imaging Radiat Oncol 2019;11:34–40. https://doi.org/10.1016/j.phro.2019.07.005.

[112] Kaderka R, Mundt RC, Li N, Ziemer B, Bry VN, Cornell M, et al. Automated Closed- and Open-Loop Validation of Knowledge-Based Planning Routines Across Multiple Disease Sites. Pract Radiat Oncol 2019;9:257–65. https://doi.org/10.1016/j.prro.2019.02.010.

[113] Li X, Zhang J, Sheng Y, Chang Y, Yin F-F, Ge Y, et al. Automatic IMRT planning via static field fluence prediction (AIP-SFFP): a deep learning algorithm for real-time prostate treatment planning. Phys Med Biol 2020;65:175014. https://doi.org/10.1088/1361-6560/aba5eb.

[114] Wu B, Kusters M, Kunze-busch M, Dijkema T, McNutt T, Sanguineti G, et al. Cross-institutional knowledge-based planning (KBP) implementation and its performance comparison to Auto-Planning Engine (APE). Radiother Oncol 2017;123:57–62. https://doi.org/10.1016/j.radonc.2017.01.012.

[115] Fogliata A, Nicolini G, Clivio A, Vanetti E, Laksar S, Tozzi A, et al. A broad scope knowledge based model

for optimization of VMAT in esophageal cancer: validation and assessment of plan quality among different treatment centers. Radiat Oncol 2015;10:220. https://doi.org/10.1186/s13014-015-0530-5.

[116] Krayenbuehl J, Zamburlini M, Ghandour S, Pachoud M, Tanadini-Lang S, Tol J, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. Radiat Oncol 2018;13:170. https://doi.org/10.1186/s13014-018-1113-z.

[117] Smith A, Granatowicz A, Stoltenberg C, Wang S, Liang X, Enke CA, et al. Can the Student Outperform the Master? A Plan Comparison Between Pinnacle Auto-Planning and Eclipse knowledge-Based RapidPlan Following a Prostate-Bed Plan Competition. Technol Cancer Res Treat 2019;18:1533033819851763. https://doi.org/10.1177/1533033819851763.

[118] Wu B, McNutt T, Zahurak M, Simari P, Pang D, Taylor R, et al. Fully automated simultaneous integrated boosted-intensity modulated radiation therapy treatment planning is feasible for head-and-neck cancer: A prospective clinical study. Int J Radiat Oncol Biol Phys 2012;84:e647–53. https://doi.org/10.1016/j.ijrobp.2012.06.047.

[119] Purdie TG, Dinniwell RE, Fyles A, Sharpe MB. Automation and Intensity Modulated Radiation Therapy for Individualized High-Quality Tangent Breast Treatment Plans. Int J Radiat Oncol Biol Phys 2014;90:688–95. https://doi.org/10.1016/j.ijrobp.2014.06.056.

[120] International Commission on Radiation Units and Measurements. Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy IMRT (Report 83). J ICRU 2010;10. https://doi.org/10.1093/jicru/10.1.report83.

[121] Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. J Neurosurg 2000;93:219–22.

[122] Wu B, Pang D, Simari P, Taylor R, Sanguineti G, McNutt T. Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: a head-and-neck case study. Med Phys 2013;40:21714. https://doi.org/10.1118/1.4788671.

[123] Sharfo AW, Voet P, Breedveld S, Mens JW, Hoogeman M, Heijmen B. Comparison of VMAT and IMRT strategies for cervical cancer patients using automated planning. Radiother Oncol 2015;114:395–401. https://doi.org/10.1016/j.radonc.2015.02.006.

[124] Liao X, Lang J, Li N, Wang P, Li J, Yang J, et al. Dosimetric comparisons of IMRT planning using MCO and DMPO techniques. Technol Heal Care 2017;25:107–14. https://doi.org/10.3233/THC-171312.

[125] Wheeler PA, Chu M, Holmes R, Smyth M, Maggs R, Spezi E, et al. Utilisation of Pareto navigation techniques to calibrate a fully automated radiotherapy treatment planning solution. Phys Imaging Radiat Oncol 2019;10:41–8. https://doi.org/10.1016/j.phro.2019.04.005.

[126] Dearnaley D, Griffin C, Harris V, Lewis R, Mayles P, Scrase C, et al. OC-0155: First toxicity results of a phase II randomised trial of prostate and pelvis versus prostate alone radiotherapy. Radiother Oncol 2017;111:S59–60. https://doi.org/10.1016/S0167-8140(15)30260-7.

[127] Wu H, Jiang F, Yue H, Li S, Zhang Y. A dosimetric evaluation of knowledge-based VMAT planning with simultaneous integrated boosting for rectal cancer patients. J Appl Clin Med Phys 2016;17:78–85. https://doi.org/10.1120/jacmp.v17i6.6410.

[128] Lee T, Hammad M, Chan TCY, Craig T, Sharpe MB. Predicting objective function weights from patient anatomy in prostate IMRT treatment planning. Med Phys 2013;40:121706. https://doi.org/10.1118/1.4828841.

[129] Chen Z, King W, Pearcey R, Kerba M, Mackillop WJ. The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. Radiother Oncol 2008;87:3–16. https://doi.org/10.1016/j.radonc.2007.11.016.

[130] Wheeler PA, Chu M, Mazurek AR, Maggs RH, Jadon R, Staffurth J, et al. OC-0384: Development of fully optimised single iteration VMAT class solutions and their clinical application. Radiother Oncol 2014;111:S149. https://doi.org/10.1016/S0167-8140(15)30489-8.

[131] Brand DH, Tree AC, Ostler P, van der Voet H, Loblaw A, Chu W, et al. Intensity-modulated fractionated radiotherapy versus stereotactic body radiotherapy for prostate cancer (PACE-B): acute toxicity findings

from an international, randomised, open-label, phase 3, non-inferiority trial. Lancet Oncol 2019;20:1531–43. https://doi.org/10.1016/S1470-2045(19)30569-8.

[132] Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, et al. Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126. Int J Radiat Oncol Biol Phys 2015;92:228–35. https://doi.org/10.1016/j.ijrobp.2015.01.046.

[133] Wall PDH, Carver RL, Fontenot JD. Impact of database quality in knowledge-based treatment planning for prostate cancer. Pract Radiat Oncol 2018;8:437–44. https://doi.org/10.1016/j.prro.2018.03.004.

[134] Tol JP, Dahele M, Gregoire V, Overgaard J, Slotman BJ, Verbakel WFAR. Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance. Radiother Oncol 2019;130:75–81. https://doi.org/10.1016/j.radonc.2018.10.005.

[135] Wheeler P, West N, Powis R, Maggs R, Chu M, Pearson RA, et al. OC-0183 Multi-Institutional Evaluation of a Pareto Navigation Guided Automated Planning Solution. Radiother Oncol 2019;133:S92–3. https://doi.org/10.1016/S0167-8140(19)30603-6.

Appendix A – Literature Review Search Strategy

1 Structured Search:

1. knowledge-based.mp.
2. knowledge based.mp.
3. KBP.mp.
4. (Auto* adj5 planning).mp.
5. Pareto.mp.
6. multicriteria optimi*.mp.
7. multi-criteria optimi*.mp.
8. MCO.mp.
9. Artificial Intelligence/
10. AI.mp.
11. neural network.mp.
12. Deep Learning/
13. RapidPlan.mp.
14. AutoPlan.mp.
15. IMRT planning.mp.
16. IMRT treatment planning.mp.
17. VMAT planning.mp.
18. VMAT treatment planning.mp.
19. intensity modulated radiotherapy planning.mp.
20. intensity modulation radiotherapy treatment planning.mp.
21. volume* modulated arc therapy planning.mp.
22. volume* modulated arc therapy planning.mp.
23. radiotherapy planning.mp.
24. radiotherapy treatment planning.mp.
25. radiation therapy planning.mp.
26. radiation therapy treatment planning.mp.
27. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
28. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 25 or 26
29. **27 and 28**

2 Databases Searched:

1. Ovid MEDLINE(R) <1946 to March Week 2 2021>
2. Ovid EMBASE <1974 to Week 11 2021>

Appendix B – Journal Article 1: Supplementary Information

1 Supplementary Information S1: Algorithms for AuxROI Generation

1.1 Abbreviations

Abbreviations	Definition
D_{Presc}	Prescribed treatment dose
D_{PTV}	Prescribed treatment dose of individual PTV
T_{p1a}	Dose target for primary OAR goals (priority one) as defined in the AutoPlan protocol
T_{p1b}	Dose target for primary conformity goals (priority one) as defined in the AutoPlan protocol
z	Distance specified for primary conformity objectives (priority one) as defined in the AutoPlan protocol

1.2 AuxROI Generation

1.2.1 Priority 2 AuxROIs

The following AuxROI generation algorithm is run for each PTV dose level:

Generate AuxPTV$_{SV-1}$

1. $TmpROI_A$ = PTV expanded 1 mm isotropically with option (specified by user) not to expand into specifically defined ROIs
2. $TmpROI_B$ = $TmpROI_A$ retracted from PTVs with higher dose prescriptions and skin surface by 4 mm
3. *AuxPTV$_{SV-1}$* = $TmpROI_B$ retracted away x cm (x/3 cm craniocaudally) from each primary OAR specified in the AutoPlan protocol where x is defined by the following equation:

$$x = \frac{(D_{PTV} - T_{p1a})}{0.5 D_{Presc}}$$

Generate AuxPTV$_{SV-2}$

1. $TmpROI_A$ = PTV retracted from skin surface by 4 mm
2. *AuxPTV$_{SV-2}$* = $TmpROI_A$ subtracted from original PTV

Generate AuxPTV$_{SV-3}$

1. $TmpROI_A$ = PTV retracted from skin surface by 4 mm
2. *AuxPTV$_{SV-3}$* = *AuxPTV$_{SV-1}$* subtracted from $TmpROI_A$

1.2.2 Priority 3 AuxROIs

The following AuxROI generation algorithm is run for each priority 3 dose fall off goal:

If $PO_{FallOffType}$ == "Normal Tissue Fall Off"

Generate $AuxConf_A$

1. $TmpROI_A$ = PTV defined by field 'ROI Name' expanded 15 mm isotropically
2. $AuxConf_A$ = PTV defined by field 'ROI Name' subtracted from $TmpROI_A$

Else If $PO_{FallOffType}$ == "Normal Tissue Fall Off (Sup)"

Generate $AuxConf_B$

1. $TmpROI_A$ = PTV defined by field 'ROI Name' expanded 30 mm isotropically
2. $AuxConf_B$ =Delete all $TmpROI_A$ contours except those superior to defined PTV

Else If $PO_{FallOffType}$ == "Normal Tissue Fall Off (Inf)"

Generate $AuxConf_C$:

1. $TmpROI_A$= PTV defined by field 'ROI Name' expanded 30mm isotropically
2. $AuxConf_C$ =Delete all $TmpROI_A$ contours except those inferior to defined PTV

1.2.3 Priority 1 AuxROIs

The following AuxROI generation algorithm is run for each priority 1 primary conformality objective:

Generate $AuxConf_D$

1. $TmpROI_A$ = Summation of all PTVs expanded by x cm (x/3 cm caudally) where x is defined by the following equation:

$$x = z + \frac{\max(0, (D_{PTV} - T_{p1b}))}{0.33 D_{Presc}}$$

2. $AuxConf_D$ = $TmpROI_A$ subtracted from body outline

2 Supplementary Information S2: Derivation of Initial Optimisation Objectives

2.1 Abbreviations

Abbreviations	Definition
D_{Presc}	Prescribed treatment dose
D_{PTV}	Prescribed treatment dose of individual PTV
T_{p2}	Dose target for P2 target goals as defined in the AutoPlan protocol
T_{p3HDL}	High dose level target for P3 dose fall-off goals as defined in the AutoPlan protocol
T_{p3LDL}	Low dose level target for P3 dose fall-off goals as defined in the AutoPlan protocol
T_{p3DG}	Dose gradient target for P3 dose fall-off goals as defined in the AutoPlan protocol
R_{HDL}	Dose fall off objective parameter: high dose level
R_{LDL}	Dose fall off objective parameter: low dose level
R_r	Dose fall off objective parameter: distance

2.2 Description of Dose Fall off Objectives

The proposed solution utilises RayStation's dose fall off objectives during automated plan generation. As these objectives are atypical, a brief description is provided below:

For a standard max dose objective function a single target dose (D_T) is applied to all voxels within the ROI. For dose fall off objectives the target dose of a given voxel is instead dependent on its radial distance from the PTV, with D_T calculated according to the following rules: at the PTV surface $D_T = R_{HDL}$; at a radial distance equal to or greater than R_r, $D_T = R_{LDL}$; and for radial distances between the PTV surface and R_r, D_T is calculated though linear interpolation of R_{HDL} and R_{LDL}.

2.3 Derivation of Initial Optimisation Objectives

Initial optimisation objectives are derived from the defined planning goals according to the following rules:

1. For planning goals where an associated AuxROI(s) has been generated, optimisation objectives are applied to the AuxROI(s) not the ROI defined in the planning goal.
2. For PTVs where $D_{PTV} < D_{presc}$, P2 D_{max} planning goals are implemented using dose fall-off optimisation objectives in order to reduce conflict between PTVs of differing dose prescriptions. Optimisation objective parameters are assigned according to the following table:

Dose Fall-Off Objective Parameter	Assigned Value
R_{HDL} (Gy)	$1.5D_{Presc}$
R_{LDL} (Gy)	T_{p2}
R_r (cm)	$\frac{(D_{Presc} - D_{PTV})}{0.25D_{Presc}}$

3. For P3 dose fall-off goals, the parameters of the corresponding optimisation objective are assigned according to the following table:

Dose Fall-Off Objective Parameter	Assigned Value
R_{HDL} (Gy)	T_{p3HDL}
R_{LDL} (Gy)	T_{p3LDL}
R_r (cm)	$\frac{(T_{p3_{HDL}} - T_{p3_{LDL}})}{T_{p3_{DG}}D_{Presc}}$

4. For all dose fall off objectives 'adapt to target dose levels' is enabled
5. For all other planning goals a direct translation to RayStation's native optimisation objectives is performed.

3 Supplementary Information S3: Definition and Description of F_V^i, F_T^i, F_C^i and F_N^i Scaling Factors

3.1 Abbreviations

Abbreviations are consistent with the main manuscript

3.2 Definition and Description of F_V^i

Previous planning experience indicated that optimum objective weights were dependant on the volume of their corresponding ROI (V_{ROI}^i), with small volumes requiring lower weighing. F_V^i enables correction of this dependency through scaling the objective weight according to:

$$F_V^i = \left(1 - f_v^i\right) + \left(\frac{f_v^i V_{ROI}^i}{100}\right)$$

where, f_v^i is hard coded for each priority level (supplementary table 3), with values determined through previous clinical planning experience and initial testing of the proposed methodology. When f_v^i equals zero, no scaling is applied. When f_v^i equals 1.0, weights are scaled according to V_{ROI}^i, normalised to a reference volume of 100 cm^3. By selecting a value of f_v^i within the range 0-1 a combination of these two methods can be selected.

3.3 Definition and Description of F_T^i

RayStation objective functions (excluding dose-fall-off objectives, c.f. section 2.2) include a normalisation denominator equal to the target dose, D_T^i, squared. The resultant objective function value is therefore inversely proportional to $(D_T^i)^2$; a dependency which was considered unintuitive and undesirable as it leads to high penalties for functions with low D_T^i. F_T^i scales the objective weight according to:

$$F_T^i = \left(\frac{D_T^i}{D_{Presc}}\right)^2$$

where D_{Presc} is the prescribed treatment dose. This replaces the objective function's $(D_T^i)^2$ denominator with D_{Presc}^2, thereby removing this unwanted dependency. For dose-fall-off objectives the denominator is a more complex function, however a similar dependency on D_T^i exists and therefore F_T^i is universally applied across all objective types for simplicity.

3.4 Definition and Description of F_C^i

F_C^i is a hardcoded constant that enables objective weights to be individually adjusted. F_C^i is applied to PTV sub-volume objectives in order to avoid skin boosting and reduce conflicts within the PTV/OAR overlap region. Based on previous planning experience, F_C^i is set at 0.001 and zero for min dose objectives applied to $PTV_{SV\text{-}2}$ and $PTV_{SV\text{-}3}$ respectively. For all other objectives F_C^i is set to unity.

3.5 Definition and Description of F_N^i

P_3 objective weights are modulated during the optimisation process. For some objectives, especially those applied to ROIs with a large volume, the final modulated objective weight (w_f^i) maybe orders of magnitude different from its initial value. F_N^i is a user defined constant, stored within the AutoPlan protocol, which allows $w_{initial}^i$ to be adjusted such that this difference is reduced and the objective weight's convergence to w_f^i is more rapid. By default F_N^i is set to unity and was not adjusted during the course of this study.

4 Supplementary Information S4: Definition of OV_t^i and Limits on w_{p3}^i Adjustments

4.1 Abbreviations

Abbreviations are consistent with the main manuscript.

4.2 Definition of OV_t^i

OV_t^i is defined by the follow equation:

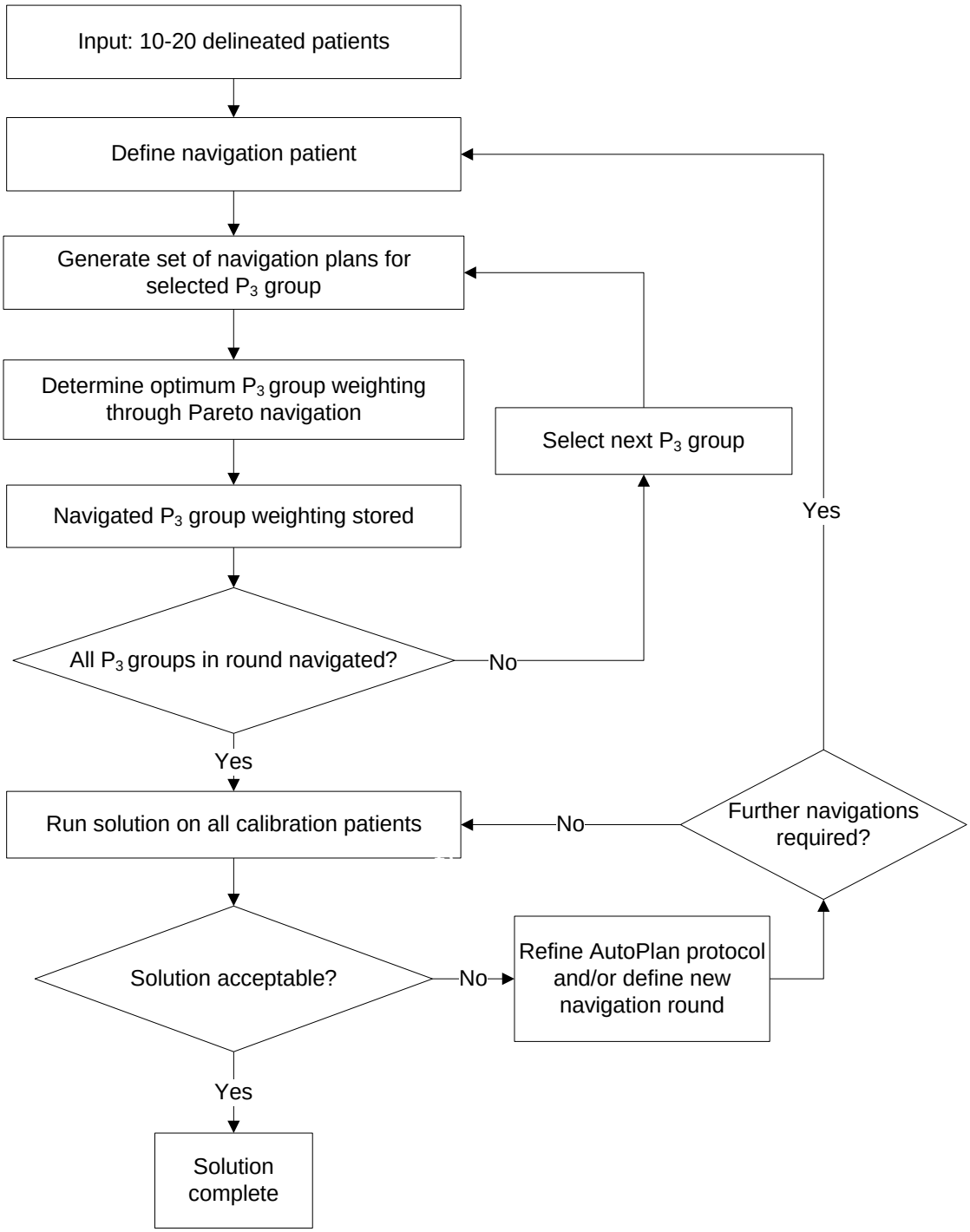
$$OV_t^i = \frac{w_{nom}^i}{10{,}000}$$

The denominator (10,000) corresponds to the nominal weight required to obtain an OV of 1.0 for a standard dose volume objective (DVO) where Δ^i=0.05 and serves to normalise the correspondence between OV_t^i and w_{nom}^i.

4.3 Limits on w_{p3}^i adjustments

During the plan generation process, to ensure w_{p3}^i stays within reasonable bounds for easily achievable objectives, its value is limited to $20w_{initial}^i$. This limit is not applied to P_3 dose-fall-off objectives as the correspondence between $w_{initial}^i$ and the final dynamically adjusted w_{p3}^i value was observed to be less predictable.

5 Supplementary Information S5: Flowchart of the Calibration Process



6 Supplementary Information: Tables

Supplementary Table 1
Local Clinical Planning Goals for PSV

Priority 1: Primary OAR Goals

ROI Name	Dose Parameter	Actionable*
Bowel	D0.1 cm^3	≤52.7 Gy

Priority 2: Target and Max Dose Goals

ROI Name	Dose Parameter	Actionable*
All PTVs	D99%	≥95% of PTV prescription
Patient Outline	D1.8 cm^3	≤107% of Prescribed Dose

Priority 3: Secondary OAR Goals

ROI Name	Dose Parameter	Optimal	Actionable*
Rectum	V24.3 Gy	≤80%	-
Rectum	V32.4 Gy	≤65%	-
Rectum	V40.5 Gy	≤50%	≤60%
Rectum	V48.6 Gy	≤35%	≤50%
Rectum	V52.7 Gy	≤30%	≤30%
Rectum	V56.8 Gy	≤15%	≤15%
Rectum	V60.8 Gy	≤3%	≤5%
Bladder	V40.5 Gy	≤50%	-
Bladder	V48.6 Gy	≤25%	-
Bladder	V52.7 Gy	-	≤50%
Bladder	V56.8 Gy	≤5%	≤35%
Bowel	V36.5 Gy	≤78 cm^3	≤158 cm^3
Bowel	V40.5 Gy	≤17 cm^3	≤110 cm^3
Bowel	V44.6 Gy	≤14 cm^3	≤28 cm^3
Bowel	V48.6 Gy	≤0.5 cm^3	≤6 cm^3

*Deviations from actionable planning goals are permissible if approved by the treating oncologist.

Supplementary Table 2

Planning goals for the PSV AutoPlan protocol

Priority 1: Primary OAR Goals

ROI Name	Dose Parameter	Target (Gy)
Bowel	Dmax	51.0

Priority 1: Primary Conformality Goals

ROI Name	Dose Parameter	Target (Gy)	Distance (cm)
PTV48	Dmax	37.4	1.5

Priority 2: Target Goals

ROI Name	Dose Parameter	Target ($\%_{Presc,PTV}$)
PTV60	Dmin	96.5
PTV60	Dmax	102.5
PTV60	D50% max	99.5
PTV48	Dmin	96.5
PTV48	Dmax	105.0

Priority 3: Trade-off Goals (Standard)

ROI Name	Dose Parameter	Target (Gy or $\%_{Vol}$)	Group
Rectum	V23.4Gy (%)	0.0	1
Rectum	V31.5Gy (%)	0.0	1
Rectum	V39.6Gy (%)	0.0	1
Rectum	V47.7Gy (%)	0.0	1
Rectum	V51.8Gy (%)	0.0	1
Rectum	V55.9Gy (%)	0.0	1
Rectum	Dmean (Gy)	5.0	2
Bladder	V30.0Gy (%)	0.0	3
Bladder	V39.6Gy (%)	0.0	3
Bladder	V47.7Gy (%)	0.0	3
Bladder	V51.8Gy (%)	0.0	3
Bladder	V55.9Gy (%)	0.0	3
Bladder	Dmax (Gy)	54.0	3
Rectum	Dmax (Gy)	60.0	6
Bowel	V36.0Gy (%)	0.0	7
Bowel	V45.6Gy (%)	0.0	7

Priority 3: Trade-off Goals (Dose Fall Off)

ROI Name	Fall Off Type	High Dose Level (Gy)	Low Dose Level (Gy)	Dose Gradient ($\%_{Presc}$ cm^{-1})	Group
PTV48	Normal Tissue Falloff	57.0	40.8	50%	4
PTV48	Intra PTV Falloff	54.0	52.8	50%	5

Abbreviations: $\%_{Presc,\ PTV}$ = % of individual PTV prescription dose; $\%_{Presc}$ = % of overall treatment prescription; $\%_{Vol}$ = % volume of ROI.

Notes: Priority 3 target = 0.0 by default, but can be specified if desired. The target is dynamically adjusted during optimisation and therefore initial values have negligible impact plan quality, but may decrease planning time if correctly defined.

Supplementary Table 3

f^i_v and w^i_{nom} values for the calibrated PSV AutoPlan protocol planning goals

Priority	Type/Group	w^i_{nom}*	f^i_v
Priority 1	Primary OAR Goals	**1000**	**1.00**
	Primary Conformality Goals	**50**	**0.00**
Priority 2	Target Goals	**250**	**0.75**
Priority 3	Group 1	1.04	**0.00**
	Group 2	4.04	**0.00**
	Group 3	0.583	**0.00**
	Group 4	35.0	**0.00**
	Group 5	19.8	**0.00**
	Group 6	1.08	**0.00**
	Group 7	0.762	**0.00**

* Rounded to 3 significant figures

Notes: Hard coded values which are common across all tumour sites are presented in bold

Table 4: Dosimetric data from validation patient cohort

		$VMAT_{Auto}$									
	Metric	Patient Val01	Patient Val02	Patient Val03	Patient Val04	Patient Val05	Patient Val06	Patient Val07	Patient Val08	Patient Val09	Patient Val10
PTV60	D98% (Gy)	57.79	57.87	57.73	57.82	57.79	57.87	57.84	57.92	57.79	57.96
	D2% (Gy)	61.68	61.67	61.71	61.64	61.58	61.63	61.68	61.58	61.63	61.67
	CI	0.87	0.85	0.85	0.86	0.86	0.85	0.86	0.86	0.86	0.87
	HI	0.065	0.063	0.066	0.064	0.063	0.063	0.064	0.061	0.064	0.062
PTV48	D98% (Gy)	46.53	47.15	46.62	46.86	46.74	47.09	46.84	46.52	46.59	47.41
	D2% (Gy)	58.55	59.21	58.71	58.81	58.70	59.05	58.99	58.78	58.67	58.96
	CI	0.84	0.82	0.84	0.84	0.85	0.87	0.83	0.86	0.83	0.87
	HI	0.233	0.223	0.236	0.227	0.230	0.222	0.225	0.242	0.236	0.213
Rectum	V24.3Gy (%)	45.47	53.39	42.82	34.33	37.20	27.27	50.41	63.57	36.98	39.42
	V32.4Gy (%)	35.55	39.95	32.43	23.62	29.51	20.95	41.48	51.62	27.03	30.86
	V40.5Gy (%)	27.08	28.92	24.28	16.79	22.39	15.34	33.01	39.62	18.14	24.12
	V48.6Gy (%)	15.30	17.45	15.49	10.56	13.91	9.98	22.61	21.28	7.96	17.63
	V52.7Gy (%)	10.28	13.46	11.78	8.13	9.47	7.20	16.51	12.99	4.73	13.78
	V56.8Gy (%)	5.85	8.00	7.06	4.43	5.49	3.79	10.73	7.90	2.41	8.36
	V60.8Gy (%)	0.37	0.02	0.29	0.01	0.10	0.01	0.48	0.05	0.00	0.07
	DMean (Gy)	25.82	29.08	25.88	21.17	22.66	17.47	28.26	32.26	21.45	24.41
Bladder	V40.5Gy (%)	23.22	10.88	15.35	16.14	18.55	15.33	31.53	8.80	13.43	9.28
	V48.6Gy (%)	16.60	7.60	9.96	10.88	12.28	10.72	21.26	4.45	8.17	6.47
	V52.7Gy (%)	12.27	6.13	7.30	8.26	8.83	8.20	16.80	2.99	5.58	5.06
	V56.8Gy (%)	8.49	4.34	4.79	5.80	6.14	5.77	11.72	2.07	3.54	3.78
	DMean (Gy)	26.41	14.60	17.36	27.71	22.56	19.00	30.73	14.78	18.38	13.10
Bowel	V36.5Gy (cm^3)	0.10	0.00	0.63	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	V40.5Gy (cm^3)	0.00	0.00	0.22	0.19	0.00	0.00	0.00	0.00	0.00	0.00
	V44.6Gy (cm^3)	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	7.31	8.99	11.20	7.75	4.54	5.50	8.24	3.81	3.40	4.23
External	D1.8cm^3 (Gy)	61.67	61.78	61.62	61.65	61.53	61.58	61.71	61.57	61.63	61.70
Plan MU	MU	600.6	599.3	614.0	653.0	598.1	583.8	611.9	586.9	617.3	581.5

Table 4 (continued): Dosimetric data from validation patient cohort

		$VMAT_{Manual}$									
	Metric	Patient Val01	Patient Val02	Patient Val03	Patient Val04	Patient Val05	Patient Val06	Patient Val07	Patient Val08	Patient Val09	Patient Val10
PTV60	D98% (Gy)	57.97	58.06	57.95	58.22	57.86	58.28	57.64	57.95	57.90	57.66
	D2% (Gy)	61.74	62.20	61.27	61.96	61.74	61.68	61.51	61.29	61.82	62.07
	CI	0.80	0.82	0.76	0.80	0.78	0.78	0.79	0.79	0.80	0.81
	HI	0.063	0.069	0.056	0.062	0.065	0.056	0.065	0.056	0.065	0.074
PTV48	D98% (Gy)	46.65	47.13	47.01	47.04	47.07	47.74	47.69	46.59	46.82	47.03
	D2% (Gy)	59.31	59.79	59.21	59.59	59.35	59.54	59.47	58.80	59.22	59.71
	CI	0.78	0.76	0.77	0.75	0.77	0.79	0.76	0.77	0.78	0.79
	HI	0.238	0.231	0.230	0.232	0.227	0.212	0.212	0.233	0.236	0.229
Rectum	V24.3Gy (%)	50.84	55.81	49.88	38.31	42.97	30.61	59.91	66.10	42.54	41.83
	V32.4Gy (%)	36.90	39.53	34.71	23.93	31.31	22.56	46.13	52.47	28.22	30.17
	V40.5Gy (%)	28.26	28.64	25.43	16.42	23.85	16.67	36.51	41.15	18.42	23.64
	V48.6Gy (%)	18.12	17.21	16.93	10.64	15.82	10.60	25.50	23.42	8.33	18.17
	V52.7Gy (%)	12.62	13.08	13.20	8.29	10.75	7.94	18.87	14.99	5.62	14.54
	V56.8Gy (%)	7.35	8.19	8.46	5.43	6.30	4.72	12.90	10.04	3.36	9.64
	V60.8Gy (%)	0.02	0.31	0.00	0.00	0.00	0.01	0.29	0.00	0.00	0.00
	DMean (Gy)	27.61	29.72	28.99	25.07	26.87	20.50	32.07	33.97	23.39	27.40
Bladder	V40.5Gy (%)	22.79	9.66	15.34	15.40	19.31	16.02	31.01	8.95	12.87	9.21
	V48.6Gy (%)	16.18	6.78	10.28	10.27	12.48	11.04	22.00	4.79	8.20	6.07
	V52.7Gy (%)	13.03	5.59	7.73	8.24	9.73	8.89	17.94	3.27	5.74	5.09
	V56.8Gy (%)	9.51	4.05	5.45	5.97	6.67	6.61	12.83	2.27	3.70	3.63
	DMean (Gy)	26.56	13.86	17.94	25.41	23.59	19.55	31.18	14.28	18.60	13.93
Bowel	V36.5Gy (cm^3)	0.39	0.00	0.86	0.59	0.00	0.00	0.03	0.00	0.00	0.00
	V40.5Gy (cm^3)	0.09	0.00	0.32	0.13	0.00	0.00	0.00	0.00	0.00	0.00
	V44.6Gy (cm^3)	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	8.93	9.26	14.08	8.97	5.53	6.08	9.13	3.86	3.90	5.18
External	D1.8cm^3 (Gy)	61.72	62.38	61.16	61.98	61.70	61.62	61.55	61.25	61.83	62.12
Plan MU	MU	545.5	626.9	515.7	608.1	518.8	503.7	549.0	496.3	562.1	546.6

Table 4 (continued): Dosimetric data from validation patient cohort

		$VMAT_{Auto}$			$VMAT_{Manual}$			
	Metric	Median	Min	Max	Median	Min	Max	p value
PTV60	D98% (Gy)	57.83	57.73	57.96	57.95	57.64	58.28	0.17
	D2% (Gy)	61.66	61.58	61.71	61.74	61.27	62.20	0.33
	CI	0.86	0.85	0.87	0.79	0.76	0.82	0.01
	HI	0.064	0.061	0.066	0.064	0.056	0.074	0.72
PTV48	D98% (Gy)	46.79	46.52	47.41	47.04	46.59	47.74	0.06
	D2% (Gy)	58.79	58.55	59.21	59.41	58.80	59.79	0.01
	CI	0.84	0.82	0.87	0.77	0.75	0.79	0.01
	HI	0.228	0.213	0.242	0.230	0.212	0.238	0.65
Rectum	V24.3Gy (%)	41.12	27.27	63.57	46.42	30.61	66.10	0.01
	V32.4Gy (%)	31.64	20.95	51.62	33.01	22.56	52.47	0.02
	V40.5Gy (%)	24.20	15.34	39.62	24.64	16.42	41.15	0.06
	V48.6Gy (%)	15.39	7.96	22.61	17.07	8.33	25.50	0.01
	V52.7Gy (%)	11.03	4.73	16.51	12.85	5.62	18.87	0.01
	V56.8Gy (%)	6.46	2.41	10.73	7.77	3.36	12.90	0.01
	V60.8Gy (%)	0.06	0.00	0.48	0.00	0.00	0.31	0.09
	DMean (Gy)	25.12	17.47	32.26	27.50	20.50	33.97	0.01
Bladder	V40.5Gy (%)	15.34	8.80	31.53	15.37	8.95	31.01	0.39
	V48.6Gy (%)	10.34	4.45	21.26	10.27	4.79	22.00	0.72
	V52.7Gy (%)	7.75	2.99	16.80	7.99	3.27	17.94	0.04
	V56.8Gy (%)	5.28	2.07	11.72	5.71	2.27	12.83	0.03
	DMean (Gy)	18.69	13.10	30.73	19.08	13.86	31.18	0.51
Bowel	V36.5Gy (cm^3)	0.00	0.00	0.70	0.00	0.00	0.86	0.27
	V40.5Gy (cm^3)	0.00	0.00	0.22	0.00	0.00	0.32	0.29
	V44.6Gy (cm^3)	0.00	0.00	0.04	0.00	0.00	0.06	0.32
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	na, zero error
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	na, zero error
	DMean (Gy)	6.41	3.40	11.20	7.50	3.86	14.08	0.01
External	D1.8cm^3 (Gy)	61.64	61.53	61.78	61.71	61.16	62.38	0.33
Plan MU	MU	600.0	581.5	653.0	546.0	496.3	626.9	0.01

Table 5: Dosimetric data from all study patients. Patient Cal01 represents the navigation patient.

		$VMAT_{Auto}$																			
	Metric	Patient Cal01	Patient Cal02	Patient Cal03	Patient Cal04	Patient Cal05	Patient Cal06	Patient Cal07	Patient Cal08	Patient Cal09	Patient Cal10	Patient Val01	Patient Val02	Patient Val03	Patient Val04	Patient Val05	Patient Val06	Patient Val07	Patient Val08	Patient Val09	Patient Val10
PTV60	D98% (Gy)	57.88	57.87	57.97	57.71	57.82	57.80	57.92	57.91	57.80	57.94	57.79	57.87	57.73	57.82	57.79	57.87	57.84	57.92	57.79	57.96
	D2% (Gy)	61.62	61.63	61.58	61.72	61.58	61.66	61.60	61.70	61.70	61.59	61.68	61.67	61.71	61.64	61.58	61.63	61.68	61.58	61.63	61.67
	CI	0.86	0.87	0.87	0.86	0.85	0.84	0.87	0.85	0.86	0.86	0.87	0.85	0.85	0.86	0.86	0.85	0.86	0.86	0.86	0.87
	HI	0.062	0.063	0.060	0.067	0.063	0.064	0.061	0.063	0.065	0.061	0.065	0.063	0.066	0.064	0.063	0.063	0.064	0.061	0.064	0.062
PTV48	D98% (Gy)	46.76	46.69	46.85	47.35	46.52	46.46	46.58	47.11	46.80	46.74	46.53	47.15	46.62	46.86	46.74	47.09	46.84	46.52	46.59	47.41
	D2% (Gy)	58.76	58.84	59.01	58.95	58.47	59.02	58.89	58.94	58.95	59.11	58.55	59.21	58.71	58.81	58.70	59.05	58.99	58.78	58.67	58.96
	CI	0.84	0.84	0.84	0.85	0.86	0.86	0.84	0.87	0.88	0.86	0.84	0.82	0.84	0.84	0.85	0.87	0.83	0.86	0.83	0.87
	HI	0.230	0.230	0.233	0.215	0.239	0.239	0.233	0.220	0.227	0.231	0.233	0.223	0.236	0.227	0.230	0.222	0.225	0.242	0.236	0.213
Rectum	V24.3Gy (%)	31.78	59.39	26.03	33.99	45.39	29.94	44.79	19.25	31.44	27.19	45.47	53.39	42.82	34.33	37.20	27.27	50.41	63.57	36.98	39.42
	V32.4Gy (%)	22.92	47.04	17.05	25.79	36.80	22.41	35.74	13.38	24.26	19.17	35.55	39.95	32.43	23.62	29.51	20.95	41.48	51.62	27.03	30.86
	V40.5Gy (%)	17.33	35.81	11.09	19.01	27.96	16.38	27.10	9.61	18.76	13.74	27.08	28.92	24.28	16.79	22.39	15.34	33.01	39.62	18.14	24.12
	V48.6Gy (%)	12.71	22.94	7.01	12.87	16.49	9.40	16.85	6.06	13.29	9.33	15.30	17.45	15.49	10.56	13.91	9.98	22.61	21.28	7.96	17.63
	V52.7Gy (%)	9.85	15.75	5.24	9.47	11.43	5.70	12.39	4.39	10.36	7.18	10.28	13.46	11.78	8.13	9.47	7.20	16.51	12.99	4.73	13.78
	V56.8Gy (%)	5.74	8.13	2.75	5.16	6.40	3.02	8.46	2.29	5.83	4.15	5.85	8.00	7.06	4.43	5.49	3.79	10.73	7.90	2.41	8.36
	V60.8Gy (%)	0.00	0.00	0.00	0.17	0.04	0.02	0.54	0.00	0.00	0.07	0.37	0.02	0.29	0.01	0.10	0.01	0.48	0.05	0.00	0.07
	DMean (Gy)	21.48	30.66	18.24	19.74	26.23	19.79	24.86	13.06	19.64	19.63	25.82	29.08	25.88	21.17	22.66	17.47	28.26	32.26	21.45	24.41
Bladder	V40.5Gy (%)	6.44	36.84	21.04	12.43	16.67	6.34	19.33	4.60	11.81	25.63	23.22	10.88	15.35	16.14	18.55	15.33	31.53	8.80	13.43	9.28
	V48.6Gy (%)	4.04	25.29	11.85	8.16	7.88	4.02	12.85	3.10	6.52	18.39	16.60	7.60	9.96	10.88	12.28	10.72	21.26	4.45	8.17	6.47
	V52.7Gy (%)	3.09	19.89	8.74	6.01	4.54	3.05	9.44	2.38	4.25	15.12	12.27	6.13	7.30	8.26	8.83	8.20	16.80	2.99	5.58	5.06
	V56.8Gy (%)	2.09	13.61	6.27	4.24	2.90	2.14	6.07	1.61	2.22	11.15	8.49	4.34	4.79	5.80	6.14	5.77	11.72	2.07	3.54	3.78
	DMean (Gy)	12.02	35.81	27.99	19.42	23.24	10.50	24.18	8.83	20.63	29.65	26.41	14.60	17.36	27.71	22.56	19.00	30.73	14.78	18.38	13.10
Bowel	V36.5Gy (cm^3)	1.04	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.27	0.10	0.00	0.63	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	V40.5Gy (cm^3)	0.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.22	0.19	0.00	0.00	0.00	0.00	0.00	0.00
	V44.6Gy (cm^3)	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	7.12	10.27	7.29	3.09	7.70	3.67	2.53	1.21	1.96	8.89	7.31	8.99	11.20	7.75	4.54	5.50	8.24	3.81	3.40	4.23
External	D1.8cm^3 (Gy)	61.64	61.69	61.62	61.70	61.49	61.55	61.65	61.63	61.67	61.62	61.67	61.78	61.62	61.65	61.53	61.58	61.71	61.57	61.63	61.70
Plan MU	MU	647.0	560.7	668.1	628.5	615.0	675.7	553.6	583.6	639.7	675.6	600.6	599.3	614.0	653.0	598.1	583.8	611.9	586.9	617.3	581.5

Table 5 (continued): Dosimetric data from all study patients. Patient Cal01 represents the navigation patient.

		$VMAT_{Manual}$																			
	Metric	Patient Cal01	Patient Cal02	Patient Cal03	Patient Cal04	Patient Cal05	Patient Cal06	Patient Cal07	Patient Cal08	Patient Cal09	Patient Cal10	Patient Val01	Patient Val02	Patient Val03	Patient Val04	Patient Val05	Patient Val06	Patient Val07	Patient Val08	Patient Val09	Patient Val10
PTV60	D98% (Gy)	58.25	58.29	57.70	57.97	58.67	58.07	57.95	57.79	58.32	58.35	57.97	58.06	57.95	58.22	57.86	58.28	57.64	57.95	57.90	57.66
	D2% (Gy)	61.84	61.85	61.57	61.95	62.13	61.38	61.73	61.78	61.87	61.89	61.74	62.20	61.27	61.96	61.74	61.68	61.51	61.29	61.82	62.07
	CI	0.79	0.80	0.81	0.78	0.70	0.76	0.82	0.78	0.77	0.80	0.80	0.82	0.76	0.80	0.78	0.78	0.79	0.79	0.80	0.81
	HI	0.060	0.059	0.064	0.066	0.058	0.055	0.063	0.067	0.059	0.059	0.063	0.069	0.056	0.062	0.065	0.056	0.065	0.056	0.065	0.074
PTV48	D98% (Gy)	47.16	47.14	46.64	47.61	47.20	47.10	46.80	47.35	47.05	47.11	46.65	47.13	47.01	47.04	47.07	47.74	47.69	46.59	46.82	47.03
	D2% (Gy)	59.35	59.44	59.39	59.65	59.55	59.35	59.32	59.44	59.76	59.62	59.31	59.79	59.21	59.59	59.35	59.54	59.47	58.80	59.22	59.71
	CI	0.76	0.77	0.79	0.75	0.74	0.75	0.79	0.74	0.77	0.78	0.78	0.76	0.77	0.75	0.77	0.79	0.76	0.77	0.78	0.79
	HI	0.227	0.227	0.238	0.216	0.237	0.224	0.231	0.218	0.229	0.227	0.238	0.231	0.230	0.232	0.227	0.212	0.212	0.233	0.236	0.229
Rectum	V24.3Gy (%)	36.87	58.40	59.80	39.43	47.81	36.21	47.21	22.03	38.51	32.81	50.84	55.81	49.88	38.31	42.97	30.61	59.91	66.10	42.54	41.83
	V32.4Gy (%)	22.85	45.91	32.93	27.37	36.21	24.63	35.84	14.37	24.02	18.86	36.90	39.53	34.71	23.93	31.31	22.56	46.13	52.47	28.22	30.17
	V40.5Gy (%)	16.73	36.18	13.30	20.55	28.98	18.07	27.35	10.01	18.39	13.09	28.26	28.64	25.43	16.42	23.85	16.67	36.51	41.15	18.42	23.64
	V48.6Gy (%)	12.55	23.64	7.18	13.87	18.50	10.66	18.27	6.31	13.54	9.09	18.12	17.21	16.93	10.64	15.82	10.60	25.50	23.42	8.33	18.17
	V52.7Gy (%)	10.57	17.39	5.32	10.59	13.73	7.14	13.73	4.67	11.11	7.40	12.62	13.08	13.20	8.29	10.75	7.94	18.87	14.99	5.62	14.54
	V56.8Gy (%)	7.23	10.62	3.02	6.43	9.06	4.23	9.39	2.78	7.14	4.81	7.35	8.19	8.46	5.43	6.30	4.72	12.90	10.04	3.36	9.64
	V60.8Gy (%)	0.01	0.08	0.00	0.00	0.06	0.00	0.00	0.01	0.01	0.00	0.02	0.31	0.00	0.00	0.00	0.01	0.29	0.00	0.00	0.00
	DMean (Gy)	25.28	31.78	25.52	22.26	29.21	23.50	26.91	14.73	23.67	22.96	27.61	29.72	28.99	25.07	26.87	20.50	32.07	33.97	23.39	27.40
Bladder	V40.5Gy (%)	6.15	34.81	22.03	12.99	17.03	6.38	19.41	4.75	12.10	23.71	22.79	9.66	15.34	15.40	19.31	16.02	31.01	8.95	12.87	9.21
	V48.6Gy (%)	3.86	24.64	11.39	8.49	9.29	4.21	13.10	3.28	6.22	18.70	16.18	6.78	10.28	10.27	12.48	11.04	22.00	4.79	8.20	6.07
	V52.7Gy (%)	3.05	19.87	8.81	6.41	5.67	3.30	9.88	2.57	4.27	15.74	13.03	5.59	7.73	8.24	9.73	8.89	17.94	3.27	5.74	5.09
	V56.8Gy (%)	2.11	14.50	6.31	4.81	3.68	2.28	6.52	1.74	2.44	11.35	9.51	4.05	5.45	5.97	6.67	6.61	12.83	2.27	3.70	3.63
	DMean (Gy)	11.74	35.75	29.99	20.64	22.81	10.32	23.80	9.10	22.43	26.79	26.56	13.86	17.94	25.41	23.59	19.55	31.18	14.28	18.60	13.93
Bowel	V36.5Gy (cm^3)	1.81	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.39	0.00	0.86	0.59	0.00	0.00	0.03	0.00	0.00	0.00
	V40.5Gy (cm^3)	1.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.09	0.00	0.32	0.13	0.00	0.00	0.00	0.00	0.00	0.00
	V44.6Gy (cm^3)	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	8.53	9.84	8.50	3.40	10.76	3.58	2.66	1.58	2.27	9.11	8.93	9.26	14.08	8.97	5.53	6.08	9.13	3.86	3.90	5.18
External	D1.8cm^3 (Gy)	61.87	61.93	61.61	61.91	61.96	61.24	61.79	61.72	61.83	61.92	61.72	62.38	61.16	61.98	61.70	61.62	61.55	61.25	61.83	62.12
Plan MU	MU	547.0	578.5	492.7	533.9	543.3	482.6	509.0	564.6	536.2	568.3	545.5	626.9	515.7	608.1	518.8	503.7	549.0	496.3	562.1	546.6

Table 5 (continued): Dosimetric data from all study patients. Patient Cal01 represents the navigation patient.

		$VMAT_{Auto}$			$VMAT_{Manual}$			
	Metric	Median	Min	Max	Median	Min	Max	p value
PTV60	D98% (Gy)	57.86	57.71	57.97	57.97	57.64	58.67	0.01
	D2% (Gy)	61.63	61.58	61.72	61.80	61.27	62.20	0.05
	CI	0.86	0.84	0.87	0.79	0.70	0.82	0.00
	HI	0.063	0.060	0.067	0.062	0.055	0.074	0.22
PTV48	D98% (Gy)	46.75	46.46	47.41	47.08	46.59	47.74	0.00
	D2% (Gy)	58.91	58.47	59.21	59.44	58.80	59.79	0.00
	CI	0.85	0.82	0.88	0.77	0.74	0.79	0.00
	HI	0.230	0.213	0.242	0.229	0.212	0.238	0.31
Rectum	V24.3Gy (%)	37.09	19.25	63.57	42.76	22.03	66.10	0.00
	V32.4Gy (%)	28.27	13.38	51.62	30.74	14.37	52.47	0.02
	V40.5Gy (%)	20.70	9.61	39.62	22.10	10.01	41.15	0.01
	V48.6Gy (%)	13.60	6.06	22.94	14.85	6.31	25.50	0.00
	V52.7Gy (%)	10.07	4.39	16.51	10.93	4.67	18.87	0.00
	V56.8Gy (%)	5.78	2.29	10.73	7.18	2.78	12.90	0.00
	V60.8Gy (%)	0.03	0.00	0.54	0.00	0.00	0.31	0.08
	DMean (Gy)	22.07	13.06	32.26	26.20	14.73	33.97	0.00
Bladder	V40.5Gy (%)	15.34	4.60	36.84	15.37	4.75	34.81	0.63
	V48.6Gy (%)	9.06	3.10	25.29	9.78	3.28	24.64	0.94
	V52.7Gy (%)	6.72	2.38	19.89	7.07	2.57	19.87	0.00
	V56.8Gy (%)	4.57	1.61	13.61	5.13	1.74	14.50	0.00
	DMean (Gy)	20.03	8.83	35.81	21.53	9.10	35.75	0.50
Bowel	V36.5Gy (cm^3)	0.00	0.00	1.27	0.00	0.00	1.81	0.40
	V40.5Gy (cm^3)	0.00	0.00	0.72	0.00	0.00	1.11	0.50
	V44.6Gy (cm^3)	0.00	0.00	0.34	0.00	0.00	0.53	0.11
	V48.6Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	na, zero error
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	na, zero error
	DMean (Gy)	6.31	1.21	11.20	7.29	1.58	14.08	0.00
External	D1.8cm^3 (Gy)	61.63	61.49	61.78	61.81	61.16	62.38	0.06
Plan MU	MU	612.9	553.6	675.7	544.4	482.6	626.9	0.00

Table 6: Comparison of navigation and segmented dosimetric data from the first round of navigations on P_3 groups 1-6.

		Navigation Doses						Segmented Doses					
	Navigation Group	Group1	Group2	Group3	Group4	Group5	Group6	Group1	Group2	Group3	Group4	Group5	Group6
PTV60	D98% (Gy)	58.30	58.23	58.02	57.85	57.85	57.79	58.04	58.04	57.82	57.85	57.75	57.73
	D2% (Gy)	61.14	61.14	61.11	61.49	61.44	61.45	61.27	61.42	61.40	61.50	61.59	61.55
	CI	0.69	0.69	0.72	0.85	0.87	0.87	0.69	0.71	0.73	0.85	0.87	0.87
	HI	0.048	0.049	0.052	0.061	0.060	0.061	0.054	0.057	0.060	0.061	0.064	0.064
PTV48	D98% (Gy)	47.07	47.26	47.16	46.67	46.83	46.73	46.74	47.01	46.80	46.66	46.71	46.69
	D2% (Gy)	59.72	59.82	60.28	59.14	58.74	58.76	60.23	60.03	60.24	58.98	58.62	58.72
	CI	0.60	0.61	0.65	0.85	0.85	0.85	0.60	0.62	0.66	0.85	0.84	0.84
	HI	0.228	0.226	0.242	0.238	0.229	0.232	0.244	0.235	0.248	0.235	0.229	0.231
Rectum	V24.3Gy (%)	29.49	27.38	28.74	31.52	32.14	32.34	30.01	28.17	29.92	31.71	31.95	32.39
	V32.4Gy (%)	22.10	21.12	21.65	22.93	22.96	23.09	22.08	21.53	22.59	23.09	22.96	23.28
	V40.5Gy (%)	17.01	16.75	16.78	17.04	17.13	17.16	16.90	16.84	17.05	17.22	17.25	17.52
	V48.7Gy (%)	12.50	12.55	12.48	12.33	12.37	12.26	12.49	12.59	12.46	12.33	12.40	12.34
	V52.7Gy (%)	9.55	9.44	9.46	9.42	9.27	9.23	9.53	9.66	9.32	9.42	9.28	9.29
	V56.8Gy (%)	5.68	5.72	5.76	5.67	5.71	5.59	5.79	5.73	5.70	5.77	5.67	5.63
	V60.8Gy (%)	0.00	0.00	0.02	0.00	0.00	0.00	0.03	0.01	0.00	0.01	0.00	0.00
	DMean (Gy)	22.89	20.56	21.11	22.61	23.01	23.13	21.96	20.58	21.37	22.71	22.79	23.04
Bladder	V40.5Gy (%)	9.78	9.73	6.77	6.02	5.99	6.03	10.09	9.80	6.72	6.06	6.10	6.12
	V48.7Gy (%)	6.53	6.67	4.12	3.63	3.57	3.56	6.76	6.44	4.10	3.61	3.59	3.59
	V52.7Gy (%)	4.89	5.19	3.05	2.76	2.69	2.67	5.02	4.92	3.05	2.74	2.68	2.69
	V56.8Gy (%)	2.98	3.24	1.90	1.79	1.72	1.71	3.09	3.08	1.88	1.80	1.71	1.72
	DMean (Gy)	13.90	13.14	12.17	11.85	12.13	12.43	13.65	13.29	11.92	11.87	12.35	12.30
Bowel	V36.5Gy (cm^3)	2.42	2.21	2.29	1.24	1.28	1.08	2.44	2.33	2.19	1.15	1.50	1.27
	V40.5Gy (cm^3)	2.20	2.05	1.51	0.80	0.71	0.70	1.92	1.96	1.30	0.73	0.73	0.73
	V44.6Gy (cm^3)	1.03	0.80	0.80	0.36	0.51	0.45	0.82	0.82	0.76	0.38	0.43	0.40
	V48.7Gy (cm^3)	0.21	0.02	0.02	0.00	0.00	0.00	0.11	0.18	0.04	0.00	0.00	0.00
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	7.94	7.82	8.02	7.23	7.12	7.06	8.05	8.18	8.06	7.21	7.35	7.09
External	D1.8cm^3 (Gy)	61.16	61.14	61.20	61.50	61.46	61.47	61.29	61.44	61.43	61.51	61.61	61.57

Table 6 (continued): Comparison of navigation and segmented dosimetric data from the first round of navigations on P_3 groups 1-6.

		Approximation Error (Navigation - Segmented)								
	Navigation Group	Group1	Group2	Group3	Group4	Group5	Group6	Median	Min	Max
PTV60	D98% (Gy)	0.26	0.19	0.20	0.01	0.10	0.05	0.15	0.01	0.26
	D2% (Gy)	-0.13	-0.29	-0.29	-0.01	-0.15	-0.10	-0.14	-0.29	-0.01
	CI	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	-0.02	0.00
	HI	-0.007	-0.008	-0.008	0.000	-0.004	-0.003	-0.01	-0.01	0.00
PTV48	D98% (Gy)	0.33	0.25	0.36	0.01	0.12	0.03	0.18	0.01	0.36
	D2% (Gy)	-0.52	-0.22	0.04	0.16	0.12	0.04	0.04	-0.52	0.16
	CI	0.00	-0.02	0.00	0.00	0.00	0.01	0.00	-0.02	0.01
	HI	-0.016	-0.009	-0.006	0.003	0.001	0.001	0.00	-0.02	0.00
Rectum	V24.3Gy (%)	-0.52	-0.79	-1.18	-0.19	0.19	-0.05	-0.35	-1.18	0.19
	V32.4Gy (%)	0.02	-0.41	-0.93	-0.16	0.00	-0.19	-0.18	-0.93	0.02
	V40.5Gy (%)	0.11	-0.09	-0.27	-0.18	-0.11	-0.36	-0.15	-0.36	0.11
	V48.7Gy (%)	0.02	-0.04	0.02	0.00	-0.03	-0.08	-0.01	-0.08	0.02
	V52.7Gy (%)	0.03	-0.22	0.14	0.00	-0.01	-0.06	-0.01	-0.22	0.14
	V56.8Gy (%)	-0.10	-0.01	0.07	-0.11	0.04	-0.04	-0.02	-0.11	0.07
	V60.8Gy (%)	-0.03	-0.01	0.02	-0.01	0.00	0.00	0.00	-0.03	0.02
	DMean (Gy)	0.92	-0.02	-0.26	-0.10	0.22	0.09	0.03	-0.26	0.92
Bladder	V40.5Gy (%)	-0.31	-0.07	0.05	-0.04	-0.11	-0.08	-0.08	-0.31	0.05
	V48.7Gy (%)	-0.24	0.23	0.02	0.02	-0.02	-0.03	0.00	-0.24	0.23
	V52.7Gy (%)	-0.13	0.26	0.01	0.02	0.01	-0.01	0.01	-0.13	0.26
	V56.8Gy (%)	-0.12	0.16	0.02	-0.01	0.02	-0.01	0.00	-0.12	0.16
	DMean (Gy)	0.25	-0.16	0.25	-0.03	-0.23	0.13	0.05	-0.23	0.25
Bowel	V36.5Gy (cm^3)	-0.03	-0.12	0.10	0.09	-0.22	-0.19	-0.07	-0.22	0.10
	V40.5Gy (cm^3)	0.28	0.09	0.21	0.06	-0.02	-0.03	0.08	-0.03	0.28
	V44.6Gy (cm^3)	0.21	-0.02	0.04	-0.03	0.08	0.05	0.04	-0.03	0.21
	V48.7Gy (cm^3)	0.10	-0.15	-0.02	0.00	0.00	0.00	0.00	-0.15	0.10
	V52.7Gy (cm^3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	DMean (Gy)	-0.11	-0.36	-0.04	0.02	-0.23	-0.03	-0.08	-0.36	0.02
External	D1.8cm^3 (Gy)	-0.13	-0.30	-0.22	-0.01	-0.15	-0.11	-0.14	-0.30	-0.01

Appendix C – Journal Article 2: Supplementary Information

1 Supplementary Information S1: Additional Manually Delineated AuxROI Required for PPN

1.1 Overview

The following details the additional anatomically related BowelBagRegion volume which was manually delineated for all patients prior to automated plan generation.

1.2 AuxROI Purpose

Bowel is delineated as standard for all patients treated at our centre. During the course of treatment bowel may move within the abdominal cavity and fall outside regions spared by the IMRT/VMAT optimiser. Our clinical practice is therefore to delineate an AuxROI which corresponds to the abdominal cavity, and reduce dose to this region during the optimisation. This ensures that dose is minimised across the whole abdominal cavity and therefore plans are more robust to bowel movement during treatment. BowelBagRegion is delineated manually and does not need to be accurately defined to fulfil its purpose.

1.3 AuxROI Delineation

Abdominal cavity and tissue anterior to the cavity delineated. Inferior boundary defined as two slices superior to PTV60 and superior boundary defined as at least two slices superior to PTV44. Delineated volume retracted from skin surface and all PTVs to create final ROI (Figure S1).

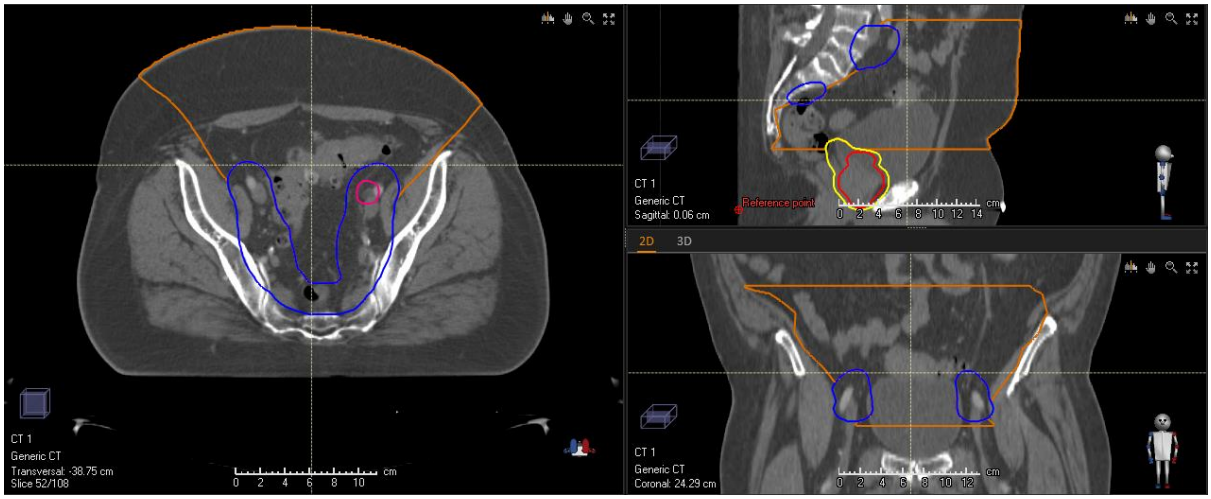


Figure S1: Example BowelBagRegion AuxROI (Brown) required for PPN automated planning. PTV60 (red), PTV50 (pink), PTV48 (yellow) and PTV44 (blue) ROIs are also shown.

2 Supplementary Information S2: Local Clinical Planning Goals for PSV and PPN

Supplementary Table S2
Local Clinical Planning Goals for PSV and PPN

Priority 1: Primary OAR Objectives

ROI Name	Dose Parameter	Actionable*
Bowel	D0.1 cm^3	≤52.7 Gy

Priority 2: Target and Max Dose Objectives

ROI Name	Dose Parameter	Actionable*
All PTVs	D99%	≥95% of PTV prescription
Patient Outline	D1.8 cm^3	≤107% of Prescribed Dose

Priority 3: Secondary OAR Objectives

ROI Name	Dose Parameter	Optimal	Actionable*
Rectum	V24.3 Gy	≤80%	-
Rectum	V32.4 Gy	≤65%	-
Rectum	V40.5 Gy	≤50%	≤60%
Rectum	V48.6 Gy	≤35%	≤50%
Rectum	V52.7 Gy	≤30%	≤30%
Rectum	V56.8 Gy	≤15%	≤15%
Rectum	V60.8 Gy	≤3%	≤5%
Bladder	V40.5 Gy	≤50%	-
Bladder	V48.6 Gy	≤25%	-
Bladder	V52.7 Gy	-	≤50%
Bladder	V56.8 Gy	≤5%	≤35%
Bowel	V36.5 Gy	≤78 cc	≤158 cc
Bowel	V40.5 Gy	≤17 cc	≤110 cc
Bowel	V44.6 Gy	≤14 cc	≤28 cc
Bowel	V48.6 Gy	≤0.5 cc	≤6 cc
Bowel	V52.7 Gy	≤0.0 cc	≤0.0 cc

*Deviations from actionable planning goals are permissible if approved by the treating oncologist.

3 Supplementary Information S3: Planning goals for the PPN AutoPlan Protocol

Supplementary Table S3
Planning goals for the PPN AutoPlan protocol

Priority 1: Primary OAR Goals

ROI Name	Dose Parameter	Target (Gy)
Bowel_1mm	Dmax	51.0

Priority 1: Primary Conformality Goals

ROI Name	Dose Parameter	Target (Gy)	Distance (cm)
PTV48	Dmax	43.2	1.5
PTV44	Dmax	37.4	1.5

Priority 2: Target Goals

ROI Name	Dose Parameter	Target ($\%_{Presc,PTV}$)
PTV60	Dmin	96.5
PTV60	Dmax	102.5
PTV60	D50% max	99.5
PTV50	Dmin	96.5
PTV50	Dmax	102.5
PTV48	Dmin	96.8
PTV48	Dmax	105.0
PTV44	Dmin	96.8
PTV44	Dmax	102.7

Priority 3: Trade-off Goals (Standard)

ROI Name	Dose Parameter	Target (Gy or $\%_{Vol}$)	Group
Rectum	V23.4Gy (%)	0.0	12
Rectum	V31.5Gy (%)	0.0	12
Rectum	V39.6Gy (%)	0.0	1
Rectum	V47.7Gy (%)	0.0	1
Rectum	V51.8Gy (%)	0.0	1
Rectum	V55.9Gy (%)	0.0	1
Rectum	Dmax (Gy)	58.8	1
Rectum	Dmean (Gy)	5.0	2
Bladder	V24.0Gy (%)	0.0	3
Bladder	V31.8Gy (%)	0.0	3
Bladder	V39.6Gy (%)	0.0	3
Bladder	V47.7Gy (%)	0.0	3
Bladder	V51.8Gy (%)	0.0	3
Bladder	Dmax (Gy)	55.9	3
Bowel	V36.0Gy (%)	0.0	6
Bowel	V40.9Gy (%)	0.0	6
Bowel	V43.8Gy (%)	0.0	6
Bowel	Dmax (Gy)	48.6	6
BowelBagRegion	V19.8Gy (%)	0.0	11
BowelBagRegion	V28.8Gy (%)	0.0	11
BowelBagRegion	V36.0Gy (%)	0.0	8
BowelBagRegion	V40.9Gy (%)	0.0	8
BowelBagRegion	V43.8Gy (%)	0.0	8
BowelBagRegion	V48.6Gy (%)	0.0	8

Priority 3: Trade-off Goals (Dose Fall Off)

ROI Name	Fall Off Type	High Dose Level (Gy)	Low Dose Level (Gy)	Dose Gradient ($\%_{Presc}$ cm^{-1})	Group
PTV48	Normal Tissue Falloff	54.0	36.5	50.0	4
PTV48	Intra PTV Falloff	54.0	52.8	50.0	5
PTV48	Normal Tissue Falloff (Inf)	54.0	31.2	50.0	13
PTV44	Normal Tissue Falloff	54.0	33.4	50.0	9
PTV44	Normal Tissue Falloff	54.0	39.6	50.0	14
PTV44	Intra PTV Falloff	54.0	44.9	50.0	15

Abbreviations: $\%_{Presc,\ PTV}$ = % of individual PTV prescription dose; $\%_{Presc}$ = % of overall treatment prescription; $\%_{Vol}$ = % volume of ROI.

Notes: Bowel_1mm ROI generated automatically by EdgeVcc through isotropic expansion of Bowel ROI by 1mm.
Priority 3 target = 0.0 by default, but can be specified if desired. The target is dynamically adjusted during optimisation and therefore initial values have negligible impact plan quality, but may decrease planning time if correctly defined.

4 Supplementary Information S4: Post Calibration Nominal Weights for PPN AutoPlan Protocol

Supplementary Table S4

Nominal weights (w_n) for the calibrated PPN AutoPlan protocol planning goals

Priority	Type/Group	w_n*
Priority 1	Primary OAR Goals	1000
	Primary Conformality Goals**	1000
Priority 2	Target Goals	250
Priority 3	Group 1	1.23
	Group 2	4.65
	Group 3	0.500
	Group 4	70.0
	Group 5	1.73
	Group 6	6.30
	Group 7	7.78
	Group 8	24.00
	Group 9***	365
	Group 10	0.800
	Group 11	24.0
	Group 12	5.00
	Group 13	1.26
	Group 14***	31.9
	Group 15	4.27

* Rounded to 3 significant figures

**Nominal weight manually increased during PPN calibration to match Primary OAR nominal weight

***For PTV44 normal tissue fall off goals, the final dynamically adjusted weight was observed to be orders of magnitude smaller than the initial weights (w_i) loaded into the optimiser. This discrepancy was reduced by setting the constant F_N to 0.01. Details of F_N are provided by wheeler *et al* [1]

5 Supplementary Information References

[1] Wheeler P a., Chu M, Holmes R, Smyth M, Maggs R, Spezi E, et al. Utilisation of Pareto navigation techniques to calibrate a fully automated radiotherapy treatment planning solution. Phys Imaging Radiat Oncol 2019;10:41–8. doi:10.1016/j.phro.2019.04.005.

Appendix D – Innovation Proposal

1 Project Title

ASSURE-RT: AI driven quality a**SSU**rance for **R**adioth**ER**apy clinical **T**rials: ensuring safe, personalised treatments for UK trial patients.

2 Lay Summary

All patients undergoing radiotherapy (RT) require a bespoke, personalised, treatment plan to be generated prior to treatment. RT plans aim to maximise the radiation dose to the cancer, whilst avoiding sensitive organs. Plan generation is however, complex and the quality of a patient's plan depends on where they are treated and who created the plan.

Variation in quality is a key issue for clinical trials. It can affect the safety and effectiveness of the treatment, but also influence the results of the trial. To address this problem, a national RT trials quality assurance group (RTTQA) has been commissioned, which provides an independent quality assurance service across all UK RT trials.

RTTQA's key role is to ensure the quality of RT plans used within trials. Reviewing treatment quality on individual plans is however a time-consuming manual process. As such it is performed on less than 1 in 10 trial patients. This is not sufficient to ensure the quality of all plans within a trial.

ASSURE-RT is a proposed software platform that uses artificial intelligence to objectively assess the quality of a patient's RT plan. ASSURE-RT would be used within trials to assess the quality of each plan, with results stored in a large database and passed on to the treating centre in time for a patient's plan to be improved.

ASSURE-RT should enable all trial patients to receive a personalised quality check of their plan prior to treatment, leading to safer, more effective treatments. Furthermore, analysis of results in the large database would enable the objective assessment of treatment quality across the UK. Poor performance could be highlighted, best practice identified and information shared to drive quality at a national level.

3 The Heath Care Problem

All patients undergoing RT require a bespoke, personalised, treatment plan to be generated prior to treatment. RT plans are created on patient CT scans using specialist computer

simulation software and aim to maximise the radiation dose to the cancer, whilst avoiding sensitive organs. Plan generation, typically performed manually by expert staff, is highly complex and time consuming, and there exists substantial variations in plan quality both at intra- and inter-institutional level [1].

The effectiveness of RT is highly dependent on the quality of the treatment plan. Poor quality RT negatively affects patient outcomes: it increases the risks of treatment failure; increases overall mortality; and detrimentally impacts the patient's quality of life [2–6]. This is a significant issue within RT clinical trials, where standardisation of plan quality is paramount; not only to ensure treatment efficacy, but also that clinical trial outcomes reflect differences in randomisation schedules, rather than variances in planning practice and quality.

There is growing evidence highlighting the potential magnitude of this problem. For prostate cancer, Moore *et al* [7] demonstrated that that the risk of grade 2 rectal complications within a phase III clinical trial could be reduced by 4.7%, solely through improvements in planning practice. Similarly, Tol *et al* showed that for 16% of patients within a phase III head and neck (H&N) trial, the probability of a clinically significant reduction (>75%) in salvia flow rate could be reduced by >10% [8]. More generally, across the whole RT pathway, a systematic review by Weber et al [4] concluded that poor quality RT had a significant impact on the primary study end-point in 62.5% of the studies identified in the review.

In the UK a national RT trials quality assurance group (RTTQA) has been commissioned, at a cost of £1,135,000 per year, to provide a centralised independent quality assurance (QA) service across all UK RT trials. The core part of this service is governing the safety and optimality of RT treatment plans. Reviewing treatment quality on individual patients is however a time-consuming manual process. Therefore reviews are mainly limited to 'pre-trial benchmark' cases, which centres must complete prior to enrolling patients, and prospective reviews on a small number of on-trial patients (<10%). Whilst an improvement on standard practice, current RTTQA practices are insufficient to ensure the consistency and quality of all RT treatment plans within UK trials.

4 The Proposed Innovation

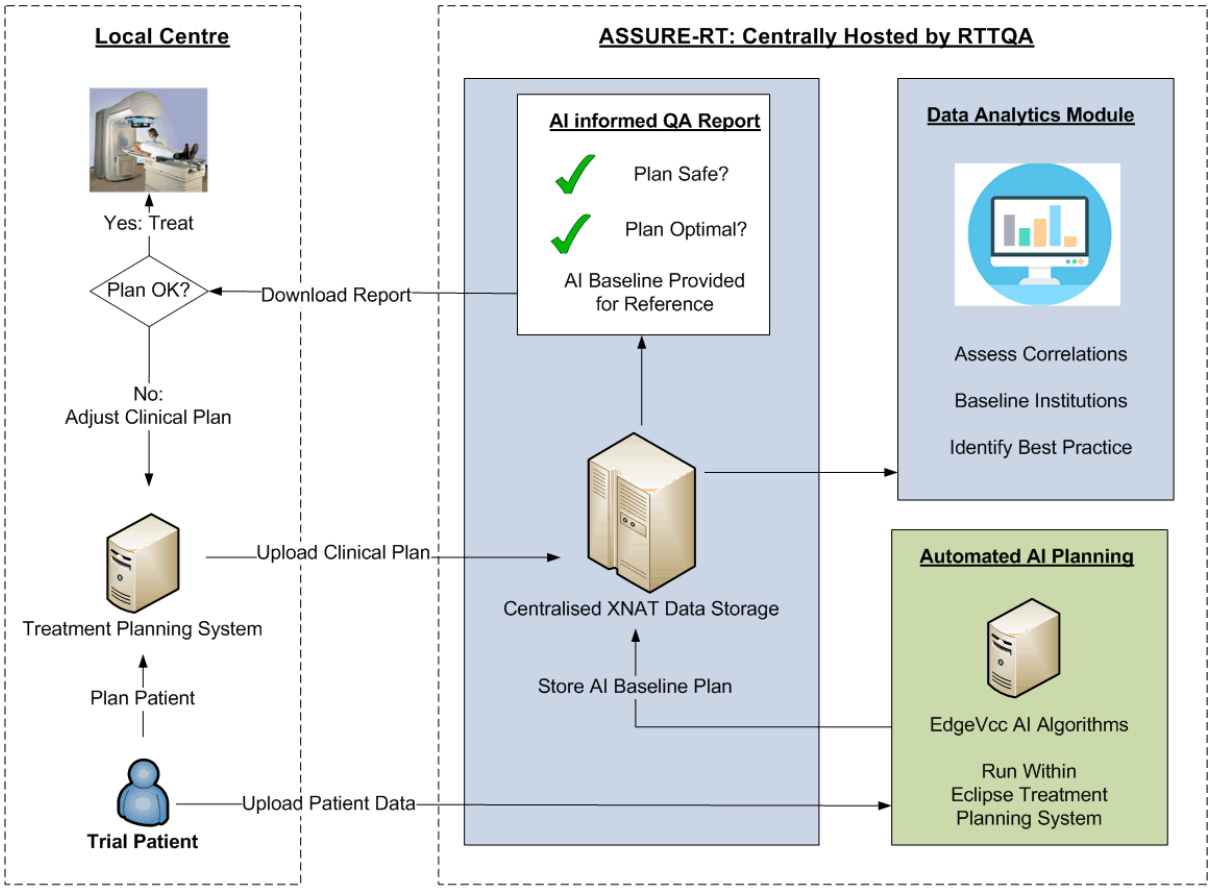


Figure 1: Overview of the ASSURE-RT platform. ASSURE-RT is hosted by RTTQA and utilises Eclipse (green) and XNAT (blue) software

To help resolve this issue, we propose to develop ASSURE-RT: an AI driven plan QA platform for RT clinical trials. ASSURE-RT (Figure 1) will utilise validated AI algorithms (EdgeVcc), developed by Velindre Cancer Centre (VCC) [9,10], to automatically generate a personalised, gold standard RT reference plan for trial patients uploaded to the platform. Clinical plans will be baselined against this AI plan using a range of established plan quality metrics, with unwarranted variations in quality identified and flagged. Results will be reported back to the treating centre, enabling corrective action to be taken prior to the patient's treatment. This process will be automated, providing the potential to deliver timely, patient specific QA for plans of all recruited trial patients in the UK.

In addition to individual patient QA, ASSURE-RT will store all baseline data within the RTTQA centralised database. A data analytics module will enable the rapid analysis of treatment quality at an institutional and national level. With nearly every UK RT centre participating in trials, this database will act as an auditing tool for UK RT practice. Poor performance can be

highlighted, best practice identified and information shared to drive quality at a national level.

5 Benefit of the Innovation

Due to the resource intensive nature of current manual reviewing techniques, RTTQA only performs individual prospective reviews on <10% of trial patients. Furthermore, the turnaround time for prospective reviews is 72hrs, which can put increased time pressure on the patient pathway. The ASSURE-RT platform will provide a paradigm shift in the operational efficiency of RTTQA. It will enable AI-informed reviews on potentially all trial patients, but require no additional revenue costs outside of RTTQA's current budget. Furthermore, turnaround times will be reduced from 72hrs to less than 24hrs, ensuring feedback is timely and can be acted on before the start of treatment.

Within trials, breaches in protocols can be frequent (a 2012 systematic review identified a 28% failure rate [4]) and are correlated with poor outcomes [2–6]. Timely trial QA is a powerful tool to detect and intervene on protocol breaches prior to the patient being treated. ASSURE-RT will enable the provision of timely prospective plan QA for potentially all UK RT trial patients. In addition to protocol breaches, retrospective benchmarking studies highlight a wide variation in the optimality of treatment plans within trials [11,12]. Sub-optimal plans can result in a higher treatment toxicity or lower chance of cure. Through comparison against the AI baseline plan, ASSURE-RT will identify clinically significant deviations in quality such that corrective action can be taken prior to treatment.

Finally, results from RT clinical trials lead to sustained change to UK practice [13]. By improving the consistency and quality of treatment plans, ASSURE-RT will help deliver higher quality trial data on which future UK practice is based. By collating baseline results into RTTQA's centralised database, the ASSURE-RT platform provides a unique opportunity to deliver UK wide auditing capability. The data analytics module will enable regular comparative analysis of plan quality across the UK; enabling poor performing institutions to initiate quality improvement drives and high performing institutions to share best practice. RT clinical trials also have a track record of delivering sustained improvements in treatment planning practice across the UK for non-trial patients [14,15]. It is expected that

improvements in trial plan quality delivered through ASSURE-RT will therefore propagate through to standard UK practice as centres apply best practice across their service.

6 Implementation of the Innovation

6.1 Patient and Public Involvement (PPI)

Over the last 5 years PPI representatives have been actively involved in the development and evaluation of EdgeVcc. For development and implementation of the innovation it is proposed an experienced PPI lead be recruited to maximise engagement opportunities. Activities should include: co-producing PPI plans with established VCC Patient and Carer Liaison Groups; developing role descriptions for PPI participants; inducting, mentoring and developing patient and public participants; establishing PPI evaluation, monitoring and impact reporting methods

6.2 System Development and Evaluation

Appendix A provides an estimate of the resources required to develop the ASSURE-RT platform and evaluate it for two clinical sites (prostate and H&N). An example work plan is also provided. In summary, a 3 year project costing an estimated £850k is proposed. The project would be a collaboration between VCC, RTTQA and an academic partner (who are experts in the X-NAT architecture required for ASSURE-RT) and would generate high quality, real world evidence on the efficacy of the innovation. The evaluation would take the form of a before and after study, where ASSURE-RT is implement within running prostate and H&N clinical trials. Time series analysis across a range of established dose metrics would be performed to monitor the impact ASSURE-RT has on RT plan quality within these trials. If results are supportive of the innovation the following adoption strategy will implemented.

6.3 Adoption Strategy

RTTQA's nationally agreed remit to provide a centralised QA service for all UK RT trials provides a unique opportunity to adopt novel trial QA technologies at scale and at pace. The programme of work detailed in section 6.2 will establish all the necessary processes and infrastructure required to enable implementation of ASSURE-RT across a broad spectrum of clinical trials at a national level. As such, widespread adoption of ASSURE-RT does not require substantial financial investment, changes to infrastructure, or adoption of significantly new clinical practices, all of which are key barriers in the adoption of

technologies within the NHS. Furthermore, due to the expected efficiency savings on current practice, adoption of ASSURE-RT would not require additional year on year revenue funding.

Successful, sustained adoption of ASSURE-RT is therefore relatively straight forward and requires the following strategic steps:

Adoption by Clinical Trial Chief Investigators (CI): For widespread adoption CIs must endorse and adopt ASSURE-RT. RTTQA and CIs work in close collaborative partnerships, with RTTQA providing expert advice on RT plan QA requirements within trials. We will utilise these established collaborations to engage CIs throughout the project. We expect strong CI support and already have received statements of support from the CIs of two large UK led trials.

Active Engagement from UK RT Centres: For UK centres to enrol patients into clinical trials they must adhere to RTTQA's plan QA protocols. Adoption of ASSURE-RT by RTTQA will naturally lead to adoption by UK centres. RTTQA has close collaborative partnerships with UK centres and this will be utilised throughout the course of the implementation to actively engage centres and ensure the benefit of ASSURE-RT to a centre's quality of care is maximised.

6.4 Barriers to Adoption

RTTQA's centralised structure and national remit, alongside support from key trial CIs, result in minimal barriers to adoption. Automated planning is also widely trusted within the field of RT, ensuring incorporation of AI within the RTTQA processes will be understood and accepted.

The key barrier to successful adoption, is engagement from participating centres. The concept of ASSURE-RT relies on UK centres utilising baseline results to actively improve patient care. Our view is that centres inherently want to deliver good patient care, and by engaging and collaborating at an early stage we can remove these barriers and ensure the success of the ASSURE-RT platform.

7 References

[1] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. doi:10.1016/j.prro.2011.11.012.

[2] Fairchild A, Straube W, Laurie F, Followill D. Does quality of radiation therapy predict outcomes of multicenter cooperative group trials? A literature review. Int J Radiat Oncol Biol Phys 2013;87:246–60. doi:10.1016/j.ijrobp.2013.03.036.

[3] Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials. J Natl Cancer Inst 2013;105:387–93. doi:10.1093/jnci/djt001.

[4] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy. Radiother Oncol 2012;105:4–8. doi:10.1016/j.radonc.2012.08.008.

[5] Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704 - A phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. Int J Radiat Oncol Biol Phys 2012;82:809–16. doi:10.1016/j.ijrobp.2010.11.039.

[6] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02. J Clin Oncol 2010;28:2996–3001. doi:10.1200/JCO.2009.27.4498.

[7] Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, et al. Quantifying Unnecessary Normal Tissue Complication Risks due to Suboptimal Planning : A Secondary Study of RTOG 0126. Radiat Oncol Biol 2015;92:228–35. doi:10.1016/j.ijrobp.2015.01.046.

[8] Tol JP, Dahele M, Gregoire V, Overgaard J, Slotman BJ, Verbakel WFAR. Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance. Radiother Oncol 2019;130:75–81. doi:https://dx.doi.org/10.1016/j.radonc.2018.10.005.

[9] Wheeler PA, Chu M, Holmes R, Smyth M, Maggs R, Spezi E, et al. Utilisation of Pareto navigation techniques to calibrate a fully automated radiotherapy treatment planning solution. Phys Imaging Radiat Oncol 2019;10:41–8. doi:10.1016/j.phro.2019.04.005.

[10] Wheeler PA, Chu M, Holmes R, Woodley OW, Jones CS, Maggs R, et al. Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer. Radiother Oncol 2019;141:220–6. doi:http://dx.doi.org/10.1016/j.radonc.2019.08.001.

[11] Sharfo AWM, Dirkx MLP, Bijman RG, Schillemans W, Breedveld S, Aluwini S, et al. Late toxicity in the randomized multicenter HYPRO trial for prostate cancer analyzed with automated treatment planning. Radiother Oncol 2018;128:349–56. doi:10.1016/j.radonc.2018.05.028.

[12] Tol JP, Dahele M, Gregoire V, Overgaard J, Slotman BJ, Verbakel WF a. R. Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance. Radiother Oncol 2019;130:75–81. doi:10.1016/j.radonc.2018.10.005.

[13] Thompson MK, Poortmans P, Chalmers AJ, Faivre-Finn C, Hall E, Huddart RA, et al. Practice-changing radiation therapy trials for the treatment of cancer: where are we 150 years after the birth of Marie Curie? Br J Cancer 2018;119:389–407. doi:10.1038/s41416-018-0201-z.

[14] Gwynne S, Higgins E, Poon King A, Radhakrishna G, Wills L, Mukherjee S, et al. Driving developments in UK oesophageal radiotherapy through the SCOPE trials. Radiat Oncol 2019;14:1–8. doi:10.1186/s13014-019-1225-0.

[15] Venables K, Tsang Y, Ciurlionis L, Coles CE, Yarnold JR. Does Participation in Clinical Trials Influence the Implementation of New Techniques? A Look at Changing Techniques in Breast Radiotherapy in the UK. Clin Oncol 2012;24:e100–5. doi:10.1016/j.clon.2012.06.010.

8 Innovation Appendix: Estimated Resource Requirements and Example Work Plan

Staffing (3 years)				
Staff member	**Role**	**Institution**	**fte**	**Cost**
Project Director	Project Governance	Velindre	0.3	£ 74,090
Senior Clinical Scientist	Integration of EdgeVcc into Eclipse		0.2	£ 49,394
Scientist			1	£ 176,255
PPI rep	PPI engagement		0.1	£ 13,292
HnN Oncologist	Clinical Governance		0.1	£ 41,592
Prostate Oncologist			0.1	£ 41,592
Senior Clinical Scientist	Intergration of ASSURE-RT into RTTQA	RTTQA	0.2	£ 41,042
Scientist			1	£ 176,255
Post doc	Development of X-NAT interface	Academic Partner	0.5	£ 201,872

Other Costs		
UKIO Conference x1	£	700
ESTRO Conference x2	£	3,000
Open access publications x2	£	5,000
Nominal PPI costs	£	7,500
Travel & Subsistance	£	5,000

Summary		
Staff costs	£	815,384
Other costs	£	21,200
Total	**£**	**836,584**

Table 1: Estimated resource requirements (including staff on costs) to develop and evaluate ASSURE-RT for prostate and head and neck

ID	Task Mode	Task Name	Duration	Start
1		**WP1: Development of ASSURE-RT platform**	**547 days**	**Sat 01/01/22**
2		Develop X-Nat interface and Data Analytics Module	365 days	Sat 01/01/22
3		Integrate EdgeVcc algorithms into RTTQA's Eclipse TPS	456 days	Sat 01/01/22
4		Integrate all components to establish ASSURE-RT platform	182 days	Sun 01/01/23
5		**WP2: Train and validated ASSURE-RT AI models**	**638 days**	**Sat 01/01/22**
6		Curate Trial Datasets	456 days	Sat 01/01/22
7		Train AI models on calibration cohort	91 days	Sun 02/04/23
8		Validate AI models on validation cohorts	91 days	Sun 02/07/23
9		**WP3: Develop standard operating procedures (SoPs)**	**91 days**	**Sun 02/07/23**
10		Develop ASSURE-RT SoPs	91 days	Sun 02/07/23
11		**WP4: Clinical introduction and evaluation**	**456 days**	**Sun 01/10/23**
12		Implement ASSURE-RT in prostate and H&N trials	365 days	Sun 01/10/23
13		Analysis and write-up	91 days	Mon 30/09/24

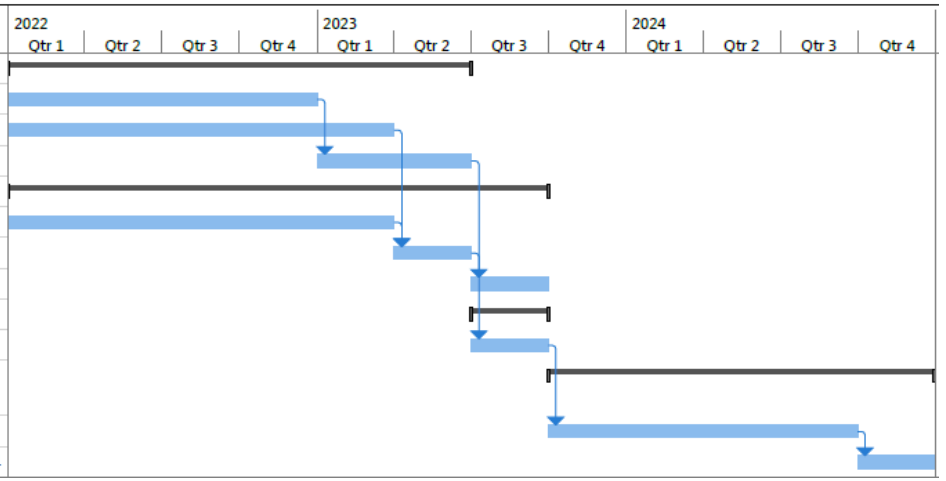


Figure 2: Example work plan to develop and evaluate ASSURE-RT for prostate and head and neck

Appendix E – Summary of Additional Modules Undertaken as Part of the Doctorate in Clinical Science

Alliance Manchester Business School – A Units		
Unit title	**Credits**	**Assignment Word Count**
A1: Professionalism and professional development in the healthcare environment	30	A1 – assignment 1 – 2500 words Group work/presentation – 10 minutes (10%) A1 – assignment 2 – 3000 words
A2: Theoretical foundations of leadership	20	A2 – assignment 1 – 3000 words A2 – assignment 2 – 3000 words
A3: Personal and professional development to enhance performance	30	A3 – assignment 1 – 1500 words A3 – assignment 2 – 4000 words
A4: Leadership and quality improvement in the clinical and scientific environment	20	A4 – assignment 1 – 3000 words A4 – assignment 2 – 3000 words
A5: Research and innovation in health and social care	20	A5 – Group work/presentation – 15 minutes (25%) A5 – assignment – 4000 words
Medical Physics – B Units		
B1: Medical Equipment Management	10	Group presentation 1500 word assignment
B2: Clinical and Scientific Computing	10	Group presentation 1500 word assignment
B3: Dosimetry	10	Group presentation 1500 word assignment
B4: Optimisation in Radiotherapy and Imaging	10	Group presentation 1500 word assignment
B6: Medical statistics in medical physics	10	3000 word assignment

B8: Health technology assessment	10	3000 word assignment
B9: Clinical applications of medical imaging technologies in radiotherapy physics	20	Group presentation 2000 word assignment
B10a: Advanced Radiobiology	10	Virtual experiment/1500 word report
B10c: Novel & External Beam Therapy	10	1500 word report
B10f: Radiation Protection Advice	10	1500 word report/piece of evidence for portfolio
Generic B Units		
B5: Contemporary issues in healthcare science	20	1500 word assignment + creative project
B7: Teaching Learning Assessment	20	20 minute group presentation