This is a n Open Access document downloaded from ORCA, Cardiff University's institutional repository: https://orca.cardiff.ac.uk/id/eprint/125381/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:


Publishers page: http://dx.doi.org/10.1016/j.radonc.2019.08.001

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer

Keywords
Intensity Modulated Radiotherapy; VMAT; Treatment Planning; Automation; Pareto Navigation; Prostate Cancer; Multicriteria optimization

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\textsubscript{Auto}) were compared against plans created by expert dosimetrists under clinical conditions (VMAT\textsubscript{Clinical}) and no time pressures (VMAT\textsubscript{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\textsubscript{Auto} plans were considered preferable or equal to VMAT\textsubscript{Clinical} and VMAT\textsubscript{Ideal} respectively, with all deemed clinically acceptable. Dosimetrically, VMAT\textsubscript{Auto}, VMAT\textsubscript{Clinical} and VMAT\textsubscript{Ideal} were similar, with observed differences generally of low clinical significance. Compared to VMAT\textsubscript{Clinical}, VMAT\textsubscript{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.

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 [1]. A solution to this problem is automated planning, where high quality plans are generated autonomously with minimal operator interaction [2–9].

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 [3,10–13]; sequential ε-constraint planning (εc) optimises plans based on a list of clinically prioritised goals [2,7,8,14–16]; and protocol based automatic iterative optimisation (PB-AIO) adapts optimisation parameters during the planning process, tailoring the optimisation to the individual patient [4,17–19]. 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 [20], improve efficiency [20–22], and enable novice operators to generate high quality plans [21]. 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 [23]. 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.

Methods and Materials

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 [24], 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).

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 [23].

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 (P1), target goals (P2) and trade-off goals (P3). 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 P1 and P2 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 P3 goals, interaction between conflicting trade-offs is complex, site specific and requires careful balancing of competing clinical demands. P1 nominal weights are therefore derived through an intuitive Pareto navigation based calibration process, where the operator sequentially explores differently weighted options of each P3 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 P1 related objectives are dynamically adjusted during the optimisation, such that the plan is tailored to the individual patient. Implementation of ‘dynamic objectives’ ensures P3 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.

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 [23] were used to generate a single automated plan (VMATAuto) 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).

Study Design and Statistical Analysis

To benchmark VMATAuto, experienced IMRT/VMAT dosimetrists (CI for PSV; OW for PPN) generated two manual treatment plans (VMATClinical & VMATIdeal) for each patient in the validation dataset. VMATClinical 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
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\textsubscript{Auto} was compared to both VMAT\textsubscript{Clinical} and VMAT\textsubscript{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) [25] 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.

**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\textsubscript{1} nominal weight for primary conformality goals was considered suboptimal and manually increased to match the weight for P\textsubscript{2} primary OAR goals. The post calibration nominal weights are presented in the supplementary file S4.

39/40 VMAT\textsubscript{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 1 and Fig. 1, with example dose distributions presented in Fig. 2. For both PPN and PSV, VMAT\textsubscript{Ideal} led to small reductions in all OAR metrics when compared to VMAT\textsubscript{Clinical} and across all three techniques observed differences were generally considered of low clinical significance. For PSV VMAT\textsubscript{Auto}, the noteworthy statistically significant (p<0.05) differences with VMAT\textsubscript{Ideal} and VMAT\textsubscript{Clinical} were: reductions in rectum mean dose and V24.3Gy, increases in the majority of bladder metrics, improved (increased) CI compared to VMAT\textsubscript{Clinical}, and decreased CI compared to VMAT\textsubscript{Ideal}. For PPN VMAT\textsubscript{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\textsubscript{Clinical} only, decreased rectum V24.3Gy. For PSV, automation led to a moderate increase in plan MU of 7% and 9% compared to VMAT\textsubscript{Ideal} and VMAT\textsubscript{Clinical} respectively, which may be indicative of increased modulation.

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\textsubscript{Auto} plans were considered preferable or equal to VMAT\textsubscript{Clinical} and VMAT\textsubscript{Ideal} respectively. When compared to VMAT\textsubscript{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.

**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 [20–22]. 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 [23].

During the calibration process several potential improvements in the implemented methodology were identified. Firstly, the hard-coded objective weights for P\textsubscript{1} and P\textsubscript{2}, which were based on previous clinical experience, may need further refining, as evidenced by the requirement to increase the nominal weight for P\textsubscript{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\textsubscript{Auto} was the superior technique both in terms of quality and efficiency. In addition, results indicate VMAT\textsubscript{Auto} is non-inferior to manual planning by expert dosimetrists under no time pressure. In general, dosimetric differences between VMAT\textsubscript{Ideal} and VMAT\textsubscript{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 patient level [20] 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 [26]; with 12 studies identified as demonstrating small differences between automated and manual plans [2,10,18,27–35], and only their more recent multi-centre study showing the overall dosimetric superiority of automation through reduced rectum doses [26]. For PPN, to the authors’ knowledge two studies have been published. The first being a methodological paper presenting results from a single patient [36], 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 [8]. 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 [37] only two out of the 81 identified evaluative studies were for complex pelvis treatments (SIB technique with nodal irradiation) [8,38]. Our work builds on this limited evidence base, providing further data in support of automation for even the most complex tumour sites.

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.

Acknowledgements

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


Table 1
Dosimetric comparison of VMAT_{Auto}, VMAT_{Clinical}, and VMAT_{ideal} for the treatment sites PSV and PPN (mean ± standard deviation)

<table>
<thead>
<tr>
<th>Metric</th>
<th>VMAT_{Auto}</th>
<th>VMAT_{Clinical}</th>
<th>VMAT_{ideal}</th>
<th>VMAT_{Auto}</th>
<th>VMAT_{Clinical}</th>
<th>VMAT_{ideal}</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTV60</td>
<td>D98% (Gy)</td>
<td>57.9 ± 0.1</td>
<td>57.8 ± 0.2</td>
<td>57.7 ± 0.1</td>
<td>57.8 ± 0.1</td>
<td>58.0 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>D2% (Gy)</td>
<td>61.6 ± 0.1</td>
<td>61.7 ± 0.2</td>
<td>61.7 ± 0.2</td>
<td>61.7 ± 0.1</td>
<td>61.9 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.86 ± 0.01</td>
<td>0.84 ± 0.03</td>
<td>0.88 ± 0.02</td>
<td>0.82 ± 0.02</td>
<td>0.81 ± 0.03</td>
</tr>
<tr>
<td>PTV50</td>
<td>D98% (Gy)</td>
<td>48.2 ± 0.2</td>
<td>48.0 ± 0.3</td>
<td>47.9 ± 0.2</td>
<td>52.3 ± 1.7</td>
<td>52.0 ± 1.0</td>
</tr>
<tr>
<td></td>
<td>D2% (Gy)</td>
<td>58.9 ± 0.2</td>
<td>59.0 ± 0.3</td>
<td>58.6 ± 0.3</td>
<td>59.5 ± 0.3</td>
<td>59.6 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.41 ± 0.05</td>
<td>0.41 ± 0.07</td>
<td>0.42 ± 0.07</td>
<td>0.65 ± 0.05</td>
<td>0.59 ± 0.07</td>
</tr>
<tr>
<td>PTV48</td>
<td>D98% (Gy)</td>
<td>46.8 ± 0.5</td>
<td>46.8 ± 0.4</td>
<td>46.5 ± 0.3</td>
<td>46.6 ± 0.6</td>
<td>46.7 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>D2% (Gy)</td>
<td>58.9 ± 0.2</td>
<td>59.0 ± 0.3</td>
<td>58.6 ± 0.3</td>
<td>59.5 ± 0.3</td>
<td>59.6 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.85 ± 0.01</td>
<td>0.82 ± 0.01</td>
<td>0.87 ± 0.01</td>
<td>0.65 ± 0.05</td>
<td>0.59 ± 0.07</td>
</tr>
<tr>
<td>PTV44</td>
<td>D98% (Gy)</td>
<td>42.3 ± 0.1</td>
<td>42.4 ± 0.1</td>
<td>42.4 ± 0.1</td>
<td>58.9 ± 0.3</td>
<td>59.6 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>D2% (Gy)</td>
<td>47.4 ± 1.6</td>
<td>47.8 ± 1.7</td>
<td>47.7 ± 1.8</td>
<td>61.7 ± 0.2</td>
<td>61.9 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.82 ± 0.02</td>
<td>0.81 ± 0.02</td>
<td>0.81 ± 0.02</td>
<td>0.81 ± 0.02</td>
<td>0.81 ± 0.02</td>
</tr>
<tr>
<td>Rectum</td>
<td>V24.3Gy (%)</td>
<td>36.7 ± 10.1</td>
<td>40.8 ± 11.1</td>
<td>38.0 ± 9.3</td>
<td>53.3 ± 9.3</td>
<td>59.3 ± 7.3</td>
</tr>
<tr>
<td></td>
<td>V40.5Gy (%)</td>
<td>20.4 ± 7.2</td>
<td>20.4 ± 7.4</td>
<td>20.0 ± 7.2</td>
<td>24.0 ± 6.1</td>
<td>23.8 ± 6.4</td>
</tr>
<tr>
<td></td>
<td>V52.7Gy (%)</td>
<td>8.5 ± 3.7</td>
<td>8.1 ± 3.6</td>
<td>8.0 ± 3.5</td>
<td>10.5 ± 3.0</td>
<td>10.0 ± 2.9</td>
</tr>
<tr>
<td></td>
<td>V60.8Gy (%)</td>
<td>0.0 ± 0.1</td>
<td>0.0 ± 0.1</td>
<td>0.0 ± 0.1</td>
<td>0.0 ± 0.1</td>
<td>0.0 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>DMean (Gy)</td>
<td>22.7 ± 3.9</td>
<td>25.1 ± 3.5</td>
<td>23.4 ± 3.5</td>
<td>29.5 ± 2.7</td>
<td>30.4 ± 2.6</td>
</tr>
<tr>
<td>Bladder</td>
<td>V40.5Gy (%)</td>
<td>19.2 ± 10.7</td>
<td>18.3 ± 9.6</td>
<td>17.4 ± 9.5</td>
<td>24.7 ± 10.4</td>
<td>23.7 ± 8.5</td>
</tr>
<tr>
<td></td>
<td>V52.7Gy (%)</td>
<td>8.8 ± 5.9</td>
<td>8.3 ± 5.2</td>
<td>7.9 ± 5.2</td>
<td>7.4 ± 4.8</td>
<td>7.6 ± 4.8</td>
</tr>
<tr>
<td></td>
<td>V56.8Gy (%)</td>
<td>6.1 ± 4.2</td>
<td>5.7 ± 3.8</td>
<td>5.6 ± 3.9</td>
<td>4.9 ± 3.1</td>
<td>5.3 ± 3.5</td>
</tr>
<tr>
<td></td>
<td>DMean (Gy)</td>
<td>23.0 ± 9.1</td>
<td>22.2 ± 8.6</td>
<td>21.6 ± 8.6</td>
<td>33.0 ± 3.9</td>
<td>31.3 ± 3.5</td>
</tr>
<tr>
<td>Bowel</td>
<td>V36.5Gy (cc)</td>
<td>0.9 ± 2.0</td>
<td>0.9 ± 1.9</td>
<td>0.7 ± 1.6</td>
<td>48.6 ± 35.9</td>
<td>53.9 ± 38.7</td>
</tr>
<tr>
<td></td>
<td>V44.6Gy (cc)</td>
<td>0.3 ± 0.7</td>
<td>0.3 ± 0.8</td>
<td>0.3 ± 0.8</td>
<td>3.6 ± 6.5</td>
<td>3.5 ± 6.0</td>
</tr>
<tr>
<td></td>
<td>V52.7Gy (cc)</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td></td>
<td>DMean (Gy)</td>
<td>8.6 ± 4.7</td>
<td>8.4 ± 4.7</td>
<td>7.7 ± 4.2</td>
<td>18.7 ± 2.6</td>
<td>19.6 ± 2.6</td>
</tr>
<tr>
<td>Patient Outline</td>
<td>D1.8cm³ (Gy)</td>
<td>61.6 ± 0.1</td>
<td>61.7 ± 0.2</td>
<td>61.7 ± 0.2</td>
<td>61.7 ± 0.1</td>
<td>61.9 ± 0.3</td>
</tr>
<tr>
<td>Plan MU</td>
<td>MU</td>
<td>616 ± 43</td>
<td>563 ± 58</td>
<td>575 ± 57</td>
<td>714 ± 60</td>
<td>695 ± 69</td>
</tr>
<tr>
<td>Planning Time</td>
<td>Hands on time (mins)</td>
<td>1.3 ± 0.3</td>
<td>22.2 ± 5.3</td>
<td>85.4 ± 39.9</td>
<td>4.4 ± 0.5</td>
<td>20.6 ± 6.3</td>
</tr>
<tr>
<td>Total time (mins)</td>
<td>14.0 ± 1.4</td>
<td>22.2 ± 5.3</td>
<td>85.4 ± 39.9</td>
<td>36.4 ± 3.1</td>
<td>41.8 ± 11.4</td>
<td>200.4 ± 53.1</td>
</tr>
<tr>
<td>Plan Quality</td>
<td>Score</td>
<td>5.0 ± 0.2</td>
<td>4.6 ± 0.5</td>
<td>4.9 ± 0.3</td>
<td>5.0 ± 0.2</td>
<td>4.8 ± 0.4</td>
</tr>
<tr>
<td>Plan Ranking vs VMAT_{Auto}</td>
<td>Plans Superior (%)</td>
<td>5%</td>
<td>15%</td>
<td>0%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plans Equivalent (%)</td>
<td>35%</td>
<td>55%</td>
<td>35%</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plans Inferior (%)</td>
<td>60%</td>
<td>30%</td>
<td>65%</td>
<td>65%</td>
</tr>
</tbody>
</table>

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.
Figure Legends

**Fig. 1.** Comparison of rectum, bladder and bowel dosimetric plan parameters between automatically generated plans (VMAT\textsubscript{Auto}) and plans generated by expert dosimetrists under no time pressure (VMAT\textsubscript{Ideal}).

**Fig. 2.** DVH and dose distributions for patient 1 in the PPN and PSV validation cohort. (A) PPN VMAT\textsubscript{Auto} dose distribution. (B) PPN VMAT\textsubscript{Ideal} dose distribution. (C) PSV VMAT\textsubscript{Auto} dose distribution. (D) PSV VMAT\textsubscript{Ideal} dose distribution. (E) PPN DVH for VMAT\textsubscript{Auto} (solid line) and VMAT\textsubscript{Ideal} (dotted line). (F) PSV DVH for VMAT\textsubscript{Auto} (solid line) and VMAT\textsubscript{Ideal} (dotted line). Note: BowelBagRegion ROI omitted from dose distribution images to improve clarity.