

This is an 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:

Wheeler, Philip Anthony, Chu, Michael K., Holmes, Rosemary A., Woodley, Owain W., Jones, Ceri S., Maggs, Rhydian, Staffurth, John , Palaniappan, Nachi, Spezi, Emiliano , Lewis, David G., Campbell, Sue, Fitzgibbon, Jim and Millin, Anthony E 2019. Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer. *Radiotherapy and Oncology* 141 , pp. 220-226. 10.1016/j.radonc.2019.08.001

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.



# 1 Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer

## 2 Keywords

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

## 5 Abstract

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

10 *Materials and methods:* For 40 randomly selected patients (20 prostate and seminal vesicles (PSV) and 20 prostate and pelvic  
11 nodes (PPN)) automatically generated plans (VMAT<sub>Auto</sub>) were compared against plans created by expert dosimetrists under clinical  
12 conditions (VMAT<sub>Clinical</sub>) and no time pressures (VMAT<sub>Ideal</sub>). Plans were compared through quantitative comparison of dosimetric  
13 parameters and blind review by an oncologist.

14 *Results:* Upon blind review 39/40 and 33/40 VMAT<sub>Auto</sub> plans were considered preferable or equal to VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub>  
15 respectively, with all deemed clinically acceptable. Dosimetrically, VMAT<sub>Auto</sub>, VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub> were similar, with observed  
16 differences generally of low clinical significance. Compared to VMAT<sub>Clinical</sub>, VMAT<sub>Auto</sub> reduced hands-on planning time by 94% and  
17 79% for PSV and PPN respectively. Total planning time was significantly reduced from 22.2 mins to 14.0 mins for PSV, with no  
18 significant reduction observed for PPN.

19 *Conclusions:* A novel automated planning solution has been evaluated, whose Pareto navigation based calibration enabled clinical  
20 decision-making on trade-off balancing to be intuitively incorporated into automated protocols. It was successfully applied to two  
21 sites of differing complexity and robustly generated high quality plans in an efficient manner.

## 22 Introduction

23 Intensity modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) treatment plan generation is a complex  
24 process, traditionally performed manually by medical physicists or specialist dosimetrists. Manual methods can be time consuming  
25 and dependent on the treatment planner's experience [1]. A solution to this problem is automated planning, where high quality  
26 plans are generated autonomously with minimal operator interaction [2–9].

27 A key challenge in automated planning is incorporating treatment planners' or oncologists' clinical experience and decision-making  
28 within the autonomous process. A number of different methods have been employed: knowledge based planning (KBP) utilises  
29 databases of previous clinical plans to correlate the relationship between patient geometry and the resultant dose distribution,  
30 which then informs the optimisation of new patients [3,10–13]; sequential  $\epsilon$ -constraint planning ( $\epsilon$ c) optimises plans based on a  
31 list of clinically prioritised goals [2,7,8,14–16]; and protocol based automatic iterative optimisation (PB-AIO) adapts optimisation  
32 parameters during the planning process, tailoring the optimisation to the individual patient [4,17–19]. Whilst these techniques  
33 have been successfully applied to automated planning, a method for intuitive exploration of different 'trade-off' options during  
34 their calibration has not yet been demonstrated.

35 Recently we developed a fully automated treatment planning solution, which is uniquely calibrated using Pareto navigation  
36 principles. This novel calibration process allows differing trade-off options to be intuitively explored, ensuring clinical experience  
37 and decision-making can be effectively incorporated into the autonomous plan generation process. Utilisation of Pareto navigation  
38 techniques on a per patient basis has been shown to improve congruence between the oncologist's clinical preference and the  
39 final clinical plan [20], improve efficiency [20–22], and enable novice operators to generate high quality plans [21]. It is anticipated  
40 that utilising such an approach to inform and calibrate an automated solution would have similar benefits and provide significant  
41 advantages over current methods, which are reliant on trial and error, or calibration against historical datasets.

42 In a previous publication we presented in detail the algorithms behind our automated approach, demonstrated the calibration  
43 process for the tumour site of prostate and seminal vesicles (PSV), and presented results from a limited proof of principle pilot  
44 study on 10 patients [23]. The objective of this study was to additionally calibrate the solution for the complex site of prostate and  
45 pelvic nodes (PPN), and for both PSV and PPN perform a comprehensive clinical evaluation comparing this new automated  
46 technique with plans generated manually by expert dosimetrists. It is hypothesised that this novel approach to calibration will  
47 result in high quality plans that are closely aligned with oncologist clinical preferences.

## 48 Methods and Materials

49 *Patient Selection and Planning Protocol*

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

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

61 Treatments were prescribed for 20 fractions using a simultaneous integrated boost (SIB) technique, with the PTV's suffix denoting  
62 its prescribed dose in Gy. The local clinical planning goals, adapted from the UK clinical trial PIVOTAL [24], are detailed in the  
63 supplementary file S2. All plans in this study were generated within RayStation (v4.99, Raysearch Laboratories, Stockholm) using  
64 identical computer clients, treatment units (Elekta Agility, Elekta Ltd, Crawley) and VMAT arc configurations (6MV single 360° arc  
65 for PSV; 6MV dual 360° arc for PPN).

### 66 *Automated System Overview*

67 Automated planning was performed using EdgeVcc: an 'in-house' automated treatment planning solution, implemented within  
68 RayStation using its scripting functionality. This section provides an brief overview of the system, with full technical details  
69 provided by Wheeler *et al* [23].

70 Prior to automated plan generation a site-specific 'AutoPlan protocol' must be created and calibrated. The AutoPlan protocol  
71 specifies the treatment modality, beam arrangement and planning goals for a given tumour site. Planning goals are split into three  
72 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  
73 PPN are presented in the supplementary file S3.

74 Planning goals do not require any user defined optimisation weighting factors, instead weights are automatically assigned during  
75 plan generation through one of two processes. For  $P_1$  and  $P_2$  goals, where the handling of competing clinical trade-offs is explicitly  
76 defined (i.e. target coverage is compromised to maintain normal tissue goals), weights are derived from a set of hard-coded  
77 nominal weights, which are common to all tumour sites. When derived, weighting factors are scaled according to the volume of  
78 their corresponding region of interest to account for the observation that to obtain the same effect, small volumes require lower  
79 weighting factors than large volumes. For  $P_3$  goals, interaction between conflicting trade-offs is complex, site specific and requires  
80 careful balancing of competing clinical demands.  $P_3$  nominal weights are therefore derived through an intuitive Pareto navigation  
81 based calibration process, where the operator sequentially explores differently weighted options of each  $P_3$  goal using an  
82 interactive slider GUI, with DVHs and dose distributions updated in real-time to inform the decision-making. The calibration is  
83 initially performed on a single patient, with the resultant solution tested against the remaining patients in the calibration cohort  
84 to ensure robustness against the whole population. Where there are large inter-patient anatomical variations, repeat navigations  
85 over population outliers may be required to improve the robustness of the solution. In this situation the operator decides if the  
86 final weighting is based a particular patient, or averaged over multiple patients. Once calibrated, a single high quality treatment  
87 plan can be automatically generated for delineated patients within that tumour site.

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

### 93 *Automated Plan Generation*

94 Using the calibration patient dataset an AutoPlan protocol for PPN was created and calibrated. The final PPN protocol and  
95 previously calibrated PSV protocol [23] were used to generate a single automated plan (VMAT<sub>Auto</sub>) for each patient in the  
96 corresponding independent validation datasets. Plans were reviewed for clinical acceptability, with manual dose scaling  
97 performed if required. All work was performed by a single clinical scientist (PW).

### 98 *Study Design and Statistical Analysis*

99 To benchmark VMAT<sub>Auto</sub>, experienced IMRT/VMAT dosimetrists (CJ for PSV; OW for PPN) generated two manual treatment plans  
100 (VMAT<sub>Clinical</sub> & VMAT<sub>Ideal</sub>) for each patient in the validation dataset. VMAT<sub>Clinical</sub> was generated under simulated clinical conditions  
101 following standard protocols, which utilise an efficient template-based class-solution methodology. As per clinical practice the  
102 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<sub>Clinical</sub> as far as they deemed possible to produce an 'ideal' treatment plan, VMAT<sub>Ideal</sub>.

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<sub>Auto</sub> was compared to both VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub> 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<sub>1</sub> nominal weight for primary conformality goals was considered suboptimal and manually increased to match the weight for P<sub>1</sub> primary OAR goals. The post calibration nominal weights are presented in the supplementary file S4.

39/40 VMAT<sub>Auto</sub> 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<sub>Ideal</sub> led to small reductions in all OAR metrics when compared to VMAT<sub>Clinical</sub> and across all three techniques observed differences were generally considered of low clinical significance. For PSV VMAT<sub>Auto</sub>, the noteworthy statistically significant ( $p < 0.05$ ) differences with VMAT<sub>Ideal</sub> and VMAT<sub>Clinical</sub> were: reductions in rectum mean dose and V24.3Gy, increases in the majority of bladder metrics, improved (increased) CI compared to VMAT<sub>Clinical</sub>, and decreased CI compared to VMAT<sub>Ideal</sub>. For PPN VMAT<sub>Auto</sub> the noteworthy differences ( $p < 0.05$ ) were: reduction in bowel V36.5Gy; increased mean bladder dose; increased PTV48 CI; and when compared to VMAT<sub>Clinical</sub> only, decreased rectum V24.3Gy. For PSV, automation led to a moderate increase in plan MU of 7% and 9% compared to VMAT<sub>Ideal</sub> and VMAT<sub>Clinical</sub> 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<sub>Auto</sub> plans were considered preferable or equal to VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub> respectively. When compared to VMAT<sub>Clinical</sub>, 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<sub>1</sub> and P<sub>2</sub>, which were based on previous clinical experience, may need further refining, as evidenced by the requirement to increase the nominal weight for P<sub>1</sub> 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<sub>Auto</sub> was the superior technique both in terms of quality and efficiency. In addition, results indicate VMAT<sub>Auto</sub> is non-inferior to manual planning by expert dosimetrists under no time pressure. In general, dosimetric differences between VMAT<sub>Ideal</sub> and VMAT<sub>Auto</sub> were small, which

157 was considered supportive evidence that implementation of 'dynamic objectives' within the automated planning process were  
158 yielding plans which were, or were near to, Pareto optimal.

159 Interestingly clinical preference towards automation was stronger for the more complex site of PPN. It is hypothesised that for  
160 PPN the high degrees of freedom within the optimisation problem made the manual trial and error exploration of trade-offs  
161 difficult. In contrast, implementation of Pareto navigation techniques allowed intuitive exploration of these trade-offs and whilst  
162 calibration was challenging, this approach resulted in plans more closely aligned to the clinician's preference. Improved  
163 congruence with the clinician's clinical preference is a key benefit of Pareto navigation, which has been demonstrated on a per-  
164 patient level [20] and this work supports the hypothesis that similar benefits can be realised by applying this technique at a patient  
165 cohort level.

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

171 Compared to existing methods of calibrating automated solutions, Pareto navigation presents a clear alternative. For both  $\epsilon$ c and  
172 PB-AIO, automated solution calibrations are reliant on trial and error. It is envisaged that the methods presented in this study  
173 would enhance many of the existing  $\epsilon$ c and PB-AIO solutions and bring the advantages of intuitive trade-off exploration into the  
174 wider field of automated planning. When compared to KBP, the employed calibration methodology benefits from having no  
175 requirement for a database of reference treatment plans. Automated solutions are therefore not influenced by the quality or  
176 quantity of historical plans and new techniques can be developed without the time consuming manual creation of a training  
177 dataset. In addition, it is envisaged that due to flexibility in the calibration process this approach is ideal for successful  
178 implementation in radiotherapy centres with differing clinical protocols.

179 When comparing to previously published studies, for the tumour site of PSV a thorough summary has recently been presented by  
180 Heijmen et al [26]; with 12 studies identified as demonstrating small differences between automated and manual plans  
181 [2,10,18,27–35], and only their more recent multi-centre study showing the overall dosimetric superiority of automation through  
182 reduced rectum doses [26]. For PPN, to the authors' knowledge two studies have been published. The first being a methodological  
183 paper presenting results from a single patient [36], which will not be discussed further, and the second a 30 patient evaluative  
184 study comparing automated planning using  $\epsilon$ c with manual planning under no time pressures [8]. The study demonstrated a clear  
185 preference towards automated planning, with notable improvements in a wide range of dosimetric parameters. Direct comparison  
186 between these examples in the literature and results from the study presented in this manuscript is not possible or appropriate  
187 due to the wide range of confounding factors including: patient selection criteria, planner and institutional expertise, and clinical  
188 protocol complexity. However, what can be ascertained is that results from this study, which demonstrate that automated  
189 planning is non-inferior to expert manual planning, are consistent with existing literature and supportive of Pareto navigation  
190 guided automated planning. Furthermore, in a recent review on automated planning by Hussein et al [37] only two out of the 81  
191 identified evaluative studies were for complex pelvis treatments (SIB technique with nodal irradiation) [8,38]. Our work builds on  
192 this limited evidence base, providing further data in support of automation for even the most complex tumour sites.

## 193 **Conclusions**

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

## 197 **Acknowledgements**

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

## 199 **Conflicts of Interest**

200 The authors declare that they have no conflicts of interest.

## 201 **References**

- 202 [1] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality:  
203 An inter-institutional study of planners and planning systems. *Pract Radiat Oncol* 2012;2:296–305.  
204 doi:10.1016/j.prro.2011.11.012.
- 205 [2] Voet PWJ, Dirx MLP, Breedveld S, Al-Mamgani A, Incrocci L, Heijmen BJM. Fully automated volumetric modulated arc  
206 therapy plan generation for prostate cancer patients. *Int J Radiat Oncol Biol Phys* 2014;88:1175–9.  
207 doi:10.1016/j.ijrobp.2013.12.046.

- 208 [3] Wu B, McNutt T, Zahurak M, Simari P, Pang D, Taylor R, et al. Fully automated simultaneous integrated boosted-intensity  
209 modulated radiation therapy treatment planning is feasible for head-and-neck cancer: A prospective clinical study. *Int J*  
210 *Radiat Oncol Biol Phys* 2012;84:e647–53. doi:10.1016/j.ijrobp.2012.06.047.
- 211 [4] Song Y, Wang Q, Jiang X, Liu S, Zhang Y, Bai S. Fully automatic volumetric modulated arc therapy plan generation for rectal  
212 cancer. *Radiother Oncol* 2016;119:531–6. doi:10.1016/j.radonc.2016.04.010.
- 213 [5] Hansen CR, Bertelsen A, Hazell I, Zukauskaitė R, Gyldenkerne N, Johansen J, et al. Automatic treatment planning improves  
214 the clinical quality of head and neck cancer treatment plans. *Clin Transl Radiat Oncol* 2016;1:1–7.  
215 doi:10.1016/j.ctro.2016.08.001.
- 216 [6] Purdie TG, Dinniwell RE, Fyles A, Sharpe MB. Automation and Intensity Modulated Radiation Therapy for Individualized  
217 High-Quality Tangent Breast Treatment Plans. *Int J Radiat Oncol Biol Phys* 2014;90:688–95.  
218 doi:10.1016/j.ijrobp.2014.06.056.
- 219 [7] Sharfo AWM, Stieler F, Kupfer O, Heijmen BJM, Dirx MLP, Breedveld S, et al. Automated VMAT planning for postoperative  
220 adjuvant treatment of advanced gastric cancer. *Radiat Oncol* 2018;13:74. doi:10.1186/s13014-018-1032-z.
- 221 [8] Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Goldner G, Georg D, et al. Automated volumetric modulated  
222 arc therapy planning for whole pelvic prostate radiotherapy. *Strahlenther Onkol* 2017;333–42. doi:10.1007/s00066-017-  
223 1246-2.
- 224 [9] Quan EM, Chang JY, Liao Z, Xia T, Yuan Z, Liu H, et al. Automated volumetric modulated arc therapy treatment planning  
225 for stage III lung cancer: How does it compare with intensity-modulated radio therapy? *Int J Radiat Oncol Biol Phys*  
226 2012;84:e69–76. doi:10.1016/j.ijrobp.2012.02.017.
- 227 [10] Yang Y, Ford EC, Wu B, Pinkawa M, van Triest B, Campbell P, et al. An overlap-volume-histogram based method for rectal  
228 dose prediction and automated treatment planning in the external beam prostate radiotherapy following hydrogel  
229 injection. *Med Phys* 2013;40:011709. doi:10.1118/1.4769424.
- 230 [11] Wu B, Pang D, Simari P, Taylor R, Sanguineti G, McNutt T. Using overlap volume histogram and IMRT plan data to guide  
231 and automate VMAT planning: a head-and-neck case study. *Med Phys* 2013;40:021714. doi:10.1118/1.4788671.
- 232 [12] Mcintosh C, Welch M, Mcniven A. Fully automated treatment planning for head and neck radiotherapy using a voxel-based  
233 dose prediction and dose mimicking method 2017;62:5926–44.
- 234 [13] Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution  
235 predicted from deep learning technique. *Med Phys* 2018. doi:10.1002/mp.13271.
- 236 [14] Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. iCycle: Integrated, multicriterial beam angle, and profile optimization  
237 for generation of coplanar and noncoplanar IMRT plans. *Med Phys* 2012;39:951–63. doi:10.1118/1.3676689.
- 238 [15] Sharfo AWM, Voet PWJ, Breedveld S, Mens JWM, Hoogeman MS, Heijmen BJM. Comparison of VMAT and IMRT strategies  
239 for cervical cancer patients using automated planning. *Radiother Oncol* 2015;114:395–401.  
240 doi:10.1016/j.radonc.2015.02.006.
- 241 [16] Voet PWJ, Dirx MLP, Breedveld S, Franssen D, Levendag PC, Heijmen BJM. Toward fully automated multicriterial plan  
242 generation: A prospective clinical study. *Int J Radiat Oncol Biol Phys* 2013;85:866–72. doi:10.1016/j.ijrobp.2012.04.015.
- 243 [17] Zhang X, Li X, Quan EM, Pan X, Li Y. A methodology for automatic intensity-modulated radiation treatment planning for  
244 lung cancer. *Phys Med Biol* 2011;56:3873–93. doi:10.1088/0031-9155/56/13/009.
- 245 [18] Winkel D, Bol GH, van Asselen B, Hes J, Scholten V, Kerkmeijer LGW, et al. Development and clinical introduction of  
246 automated radiotherapy treatment planning for prostate cancer. *Phys Med Biol* 2016;61:8587–95. doi:10.1088/1361-  
247 6560/61/24/8587.
- 248 [19] Vanderstraeten B, Goddeeris B, Vandecasteele K, van Eijkeren M, De Wagter C, Lievens Y. Automated instead of manual  
249 treatment planning? A plan comparison based on dose-volume statistics and clinical preference. *Int J Radiat Oncol* 2018;In  
250 Press. doi:10.1016/j.ijrobp.2018.05.063.

- 251 [20] Craft DL, Hong TS, Shih HA, Bortfeld TR. Improved planning time and plan quality through multicriteria optimization for  
252 intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys* 2012;82:83–90. doi:10.1016/j.ijrobp.2010.12.007.
- 253 [21] Kierkels RG, Visser R, Bijl HP, Langendijk JA, van 't Veld AA, Steenbakkers RJ, et al. Multicriteria optimization enables less  
254 experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. *Radiat*  
255 *Oncol* 2015;10. doi:10.1186/s13014-015-0385-9.
- 256 [22] Liao X, Lang J, Li N, Wang P, Li J, Yang J, et al. Dosimetric comparisons of IMRT planning using MCO and DMPO techniques.  
257 *Technol. Heal. Care*, vol. 25, 2017, p. S107–14. doi:10.3233/THC-171312.
- 258 [23] Wheeler PA, Chu M, Holmes R, Smyth M, Maggs R, Spezi E, et al. Utilisation of Pareto navigation techniques to calibrate a  
259 fully automated radiotherapy treatment planning solution. *Phys Imag Radiat Oncol* 2019;10:41–8.  
260 doi:10.1016/j.phro.2019.04.005.
- 261 [24] Dearnaley D, Griffin C, Harris V, Lewis R, Mayles P, Scrase C, et al. OC-0155: First toxicity results of a phase II randomised  
262 trial of prostate and pelvis versus prostate alone radiotherapy. *Radiother Oncol* 2017;111:S59–60. doi:10.1016/S0167-  
263 8140(15)30260-7.
- 264 [25] Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. *J Neurosurg* 2000;93:219–22.
- 265 [26] Heijmen B, Voet P, Franssen D, Penninkhof J, Milder M, Akhlat H, et al. Fully automated, multi-criterial planning for  
266 Volumetric Modulated Arc Therapy – An international multi-center validation for prostate cancer. *Radiother Oncol*  
267 2018;128:343–8. doi:10.1016/j.radonc.2018.06.023.
- 268 [27] Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK. A knowledge-based approach to improving and homogenizing intensity  
269 modulated radiation therapy planning quality among treatment centers: An example application to prostate cancer  
270 planning. *Int J Radiat Oncol Biol Phys* 2013;87:176–81. doi:10.1016/j.ijrobp.2013.03.015.
- 271 [28] Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the pre-clinical validation of a commercial model-  
272 based optimisation engine : Application to volumetric modulated arc therapy for patients with lung or prostate cancer.  
273 *Radiother Oncol* 2014;113:385–91. doi:10.1016/j.radonc.2014.11.009.
- 274 [29] Nwankwo O, Mekdash H, Sihono DSK, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning  
275 versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. *Radiat Oncol*  
276 2015;10:111. doi:10.1186/s13014-015-0416-6.
- 277 [30] Nawa K, Haga A, Nomoto A, Sarmiento R a., Shiraishi K, Yamashita H, et al. Evaluation of a commercial automatic treatment  
278 planning system for prostate cancers. *Med Dosim* 2017;42:203–9. doi:10.1016/j.meddos.2017.03.004.
- 279 [31] Kubo K, Monzen H, Ishii K, Tamura M, Kawamorita R, Sumida I, et al. Dosimetric comparison of RapidPlan and manually  
280 optimized plans in volumetric modulated arc therapy for prostate cancer. *Phys Med* 2017;44:199–204.  
281 doi:10.1016/j.ejmp.2017.06.026.
- 282 [32] Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter validation of a knowledge based model  
283 for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan  
284 Consortium. *PLoS One* 2017;12:1–13. doi:10.1371/journal.pone.0178034.
- 285 [33] Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. *Med Phys*  
286 2011;38:2515–22. doi:10.1118/1.3574874.
- 287 [34] Hussein M, South CP, Barry M a., Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-  
288 based IMRT and VMAT treatment planning in pelvic anatomy. *Radiother Oncol* 2016;120:473–9.  
289 doi:10.1016/j.radonc.2016.06.022.
- 290 [35] Yang J, Zhang P, Zhang L, Shu H, Li B, Gui Z. Particle swarm optimizer for weighting factor selection in intensity-modulated  
291 radiation therapy optimization algorithms. *Phys Med* 2017;33:136–45. doi:10.1016/j.ejmp.2016.12.021.
- 292 [36] Khaferllari I, Wong E, Bzdusek K, Lock M, Chen J. Automated IMRT planning with regional optimization using planning  
293 scripts. *J Appl Clin Med Phys* 2013;14:4052. doi:10.1120/jacmp.v14i1.4052.

- 294 [37] Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—a  
 295 review of recent innovations. Br J Radiol 2018;91:20180270. doi:10.1259/bjr.20180270.
- 296 [38] Wu H, Jiang F, Yue H, Li S, Zhang Y. A dosimetric evaluation of knowledge-based VMAT planning with simultaneous  
 297 integrated boosting for rectal cancer patients. J Appl Clin Med Phys 2016;17:78–85. doi:10.1120/jacmp.v17i6.6410.

298

**Table 1**

Dosimetric comparison of VMAT<sub>Auto</sub>, VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub> for the treatment sites PSV and PPN (mean ± standard deviation)

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

Statistical significance: VMAT<sub>Clinical</sub> and VMAT<sub>Ideal</sub> dosimetric and timing data are presented in bold where statistically significant differences (p<0.05) with VMAT<sub>Auto</sub> are observed.

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

299

300

301 **Figure Legends**

302

303 **Fig. 1.** Comparison of rectum, bladder and bowel dosimetric plan parameters between automatically generated plans (VMAT<sub>Auto</sub>)  
304 and plans generated by expert dosimetrists under no time pressure (VMAT<sub>Ideal</sub>).

305 **Fig. 2.** DVH and dose distributions for patient 1 in the PPN and PSV validation cohort. (A) PPN VMAT<sub>Auto</sub> dose distribution. (B) PPN  
306 VMAT<sub>Ideal</sub> dose distribution. (C) PSV VMAT<sub>Auto</sub> dose distribution. (D) PSV VMAT<sub>Ideal</sub> dose distribution. (E) PPN DVH for VMAT<sub>Auto</sub>  
307 (solid line) and VMAT<sub>Ideal</sub> (dotted line). (F) PSV DVH for VMAT<sub>Auto</sub> (solid line) and VMAT<sub>Ideal</sub> (dotted line). Note: BowelBagRegion  
308 ROI omitted from dose distribution images to improve clarity.

309

310