Foster I
Cardiff University
School of Engineering

Spezi E
Cardiff University
School of Engineering

Wheeler P
Velindre NHS Trust
Velindre Cancer Centre, Cardiff

ENGINEERING FOR HEALTH

Inter-planner Variability in Expert Driven Pareto-guided Automated Planning Solutions

An intra-planner study has been carried out to compare planning choices among qualified professionals when utilising Pareto-guided automated planning (PGAP) navigation. PGAP was used to calibrate planning goal weights of a protocol-based automatic iterative optimisation automated planning system. Four qualified professionals (Participant A-D) navigated solutions for eight prostate seminal vesicle (PSV) patient cases using PGAP. Plans were based on an existing clinically approved planning protocol containing seven planning goals (PGs). Three PG weights were navigated per plan (rectum Dmean, bladder Dmean and PTV conformality) with all other weights held constant at a value assigned in the original clinically approved protocol. Statistically significant differences were observed between participants for all PG groups except bladder Dmean. However, dosimetrically the PGAP system mitigated the majority of discrepancies in deviations at the calibration stage with few statistically significant dose-volume metric differences observed, none of which were clinically significant.

Keywords:
Pareto optimisation, radiotherapy planning, prostate cancer, inter-observer variability.

Corresponding author:
FosterID@cardiff.ac.uk

I. Foster, E. Spezi, and P. Wheeler, 'Inter-planner Variability in Expert Driven Pareto-guided Automated Planning Solutions', *Proceedings of the Cardiff University Engineering Research Conference 2023*, Cardiff, UK, pp. 23-26.

doi.org/10.18573/conf1.g

INTRODUCTION

Automated planning (AP) is fast becoming the state-of-the-art in radiotherapy planning for intensity-modulated radiotherapy (IMRT) and volumetric-modulated radiotherapy (VMAT) [1, 2, 3] and can be classified into one of two categories: knowledge-based planning (KBP) or rules-based planning (RBP). KBP uses statistical techniques [4, 2, 5, 6] and is trained on historical clinical datasets to inform planning for novel cases through prediction of optimisation objectives [7], dose-volume histograms [8, 9] or voxel-level dose [2]. RBP employs logic to converge on a solution. For example, a lexicographic ordering that optimises planning goals (PGs) in strict sequential order [10, 11, 12] and protocol-based automatic iterative optimisation (PBAIO) that uses algorithms to automatically adapt planning parameters during optimisation.

Protocol based automatic iterative optimisation (PBAIO) uses common protocols for patients within a cancer site. This leads to improved quality compared to manual planning although may not result in individualised planning. The most clinically desirable plans are "Pareto optimal" meaning no dosimetric improvements can be made to anyone planning goal except at the detriment of another. "Pareto guided automated planning" (PGAP) refers to any branch of AP that incorporates Pareto navigation as part of its functionality and studies suggest such methods have may have congruence with oncological preference leading to clinically desirable solutions [13, 14, 15]. However, even in highly regulated fields such as radiotherapy planning where qualified practitioner adhere to strict local and universal practices, it is not unreasonable to expect some variance in performance.

This study aims to explore discrepancies in choices made by different qualified practitioners when using PGAP built on a PBAIO framework known as the Experience Driven plan Generation Engine by Velindre Cancer Centre or EdgeVcc. Given a PGAP system is used, it is hypothesised the interactive and intuitive nature of this approach enables observer-relative interpretations of oncological preference. Hence, findings of this work will help to determine the clinically relevant region of the Pareto front as defined by a range of qualified individuals.

MATERIALS AND METHODS

All sessions took place between 1 December 2019 and 28 February 2020 and 4 participants were selected to take part:

- medical physicist (Participant A)
- two oncologists (Participant B and D).
- professional planner (Participant C)

Participants were fully qualified, highly familiar with the prostate seminal vesicles (PSV) treatment site and had multiple years of experience. Plan generation with EdgeVcc is dependent upon a base site-specific "AutoPlan protocol" containing a set of prioritised planning goals (PGs). The PGs are prioritised according to assigned weight values defied using the Pareto navigation functionality in EdgeVcc. Figure 1 shows a CT scan of a typical patient including delineated regions used during planning.

The base AutoPlan protocol contains seven PGs and had been designed for clinical use and validated prior to this study. The weights for PGs 1-3 were navigated by participants in this study with all other PG weights held constant. PG 1-3 were as following: (1) rectum D_{mean}, (2) bladder D_{mean} and (3) PTV conformality. Navigated PGs were chosen for this study based on preliminary research of PGs showing the most significant trade-off relationships.

Eight PSV patients were randomly chosen for this retrospective study. This number of patients was chosen because Pareto plan generation is computationally expensive and in addition, navigation of large numbers of patients can become a time-consuming task for participants outside of their clinical duties. This set of patients was therefore chosen such that the number of cases was considered large enough to observe a sufficient range of anatomies but small enough not to become unnecessarily time-consuming for participants.

Observers completed the task in an environment fit for clinical planning and had access to the clinical goals defined by an oncologist and could interact with the TPS however they desired. However, given not all participants were familiar with the PGAP system, they were all required to complete a practice case before completing the eight study cases. The results of the practice case were not considered in this study. Resulting plans were compared for difference in terms of relative weighting factors and dosimetric features and were tested using ANOVA when appropriate and a Friedman test otherwise. Sørensen–Dice coefficients (DiceC) were also used to describe similarity between participant choices.

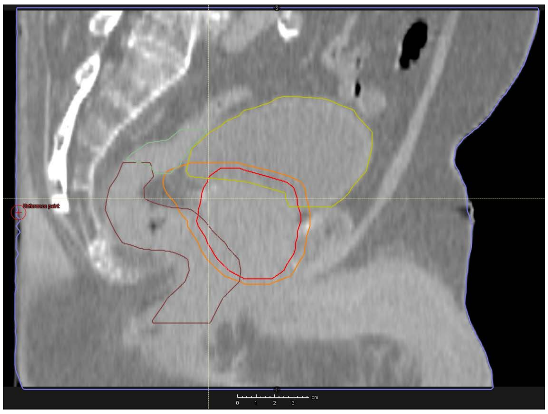

Fig. 1. An example sagital plane indicating delineated regions-of-interest used during planning. Shown are the external contour (purple), bladder (yellow), rectum (brown), high dose PTV (red), low dose PTV (orange).

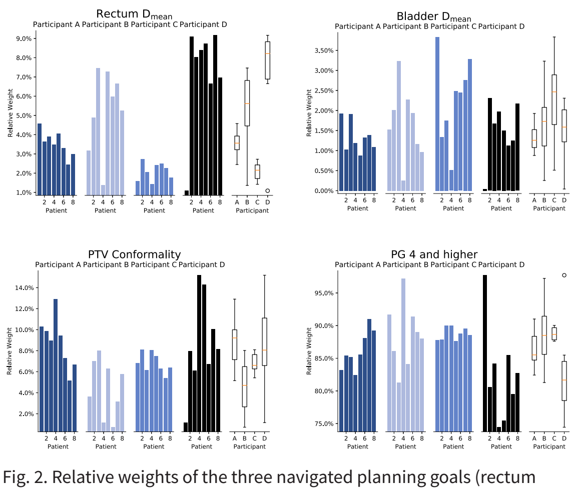

Fig. 2. Relative weights of the three navigated planning goals (rectum D_{mean}, bladder D_{mean} and PTV conformality) and non-navigated planning goals (PG 4 and higher).

RESULTS

Weights
Differences were observed between participants for all PG groups except bladder D_{mean} (Figure 2). For rectum D_{mean}, the highest degree of similarity was observed between B and D with a DiceC values 0.884. The PG with the highest degree of agreement between participants was bladder D_{mean}. The DiceC metric indicates A&C prioritised PTV conformality similarly with a values of 0.950. All DiceC values are higher than 0.99 for the PG 4 and higher group.

Dosimetry

	DVH Statistic	Participant A	Participant C	Participant B	Participant D
PTV60	**D_{98}% (Gy)**	57.6 ± 0.3	57.7 ± 0.2	57.6 ± 0.3	57.5 ± 0.3
	$D_{2\%}$ (Gy)	61.7 ± 0.1	61.6 ± 0.1	61.7 ± 0.1	61.7 ± 0.1
	CI	0.85 ± 0.0	0.82 ± 0.02	0.845 ± 0.01	0.84 ± 0.02
PTV48	D_{98}% (Gy)	46.2 ± 0.2	46.4 ± 0.5	46.4 ± 0.24	46 ± 0.64
	$D_{2\%}$ (Gy)	59.2 ± 0.2	59.4 ± 0.3	59.2 ± 0.1	59.3 ± 0.3
	CI	0.82 ± 0.02	0.72 ± 0.06	0.81 ± 0.02	0.77 ± 0.04
Rectum	$V_{24.3Gy}$ (%)	27.2% ± 5.8%	25.3% ± 5.2%	*28.6% ± 5.9%*	24.8% ± 4.7%
	$V_{56.8Gy}$ (%)	5.25% ± 1.4%	5.16% ± 1.5%	*5.32% ± 1.5%*	4.78% ± 1.1%
	D_{mean} (Gy)	17.9 ± 2.8	*17.0 ± 2.8*	*18.7 ± 2.8*	16.7 ± 2.3
Bladder	$V_{40.5Gy}$ (%)	12.6% ± 6.3%	12.9% ± 6.7%	12.5% ± 6.4%	12.8% ± 6.3%
	$V_{56.8Gy}$ (%)	4.8% ± 2.7%	5.1% ± 3.0%	4.8% ± 2.7%	4.9% ± 2.8%
	D_{mean} (Gy)	15.9 ± 6.4	15.2 ± 6.2	15.5 ± 6.7	15.7 ± 6.3

Table 1. Summary of key dose metrics. Values shown are mean ± 1 standard deviation and statistical difference at the 95% level of significance indicated in boldface.

As reported in Table 1, differences were observed between participant C&D for PTV60 $D_{98\%}$ (Gy). Higher doses were observed for C than D with a mean difference of 0.155 Gy. PTV48 differences were observed for $D_{50\%}$ (Gy) related to participants A&B only. Observed dose was lower for A than B with a mean difference of 0.832 Gy. For CI60 observed difference relate to participant A&B with observed indices low for A on average given a mean difference of 0.0315 units. CI48 saw participant B observe lower indices than all other participants with deviations of 0.0973, 0.0899 and 0.0546 units for A, C and D respectively. All observed difference were considered clinically small indicating differences is planning decisions may be clinically negligible with this PGAP system.

DISCUSSION

There still exists a gap in the literature for further inter-planner studies, but of those that do exist, there is evidence showing inconsistencies in participant choices [16, 17, 18, 19]. Given this expectation, the aim of applying PGAP to mitigate discrepancies was explored here with a view of observing clinically significant differences.
It was observed that oncologists (participant B and D) applied a higher priority to sparing the rectum than do planners or physicists. Following interviews with the participants, participant B stated a preference to push dose in the anterior direction to help spare the rectum even at some cost to conformality or even increasing dose to the bladder. This participant considered the rectum a notably higher clinical priority over the bladder and would increase sparing of the rectum given a suitable dose distribution to PTVs was still achieved and all clinical goals were being met.

A key difference between the oncologists in this study was the tendency for participant D to use a higher range of the navigation scale than participant B. The tendency of participant D to use the higher end of the scale resulted in a generally higher priority to the navigated PGs over the other PGs than is seen for any other participant. However, interviews with participant D revealed simple preferences. Participant D wanted to ensure the achievement of the clinical goals but had fewer concerns about the planning details than some of the other participants. Although traditional IMRT planning methods have been criticised for being tedious and lacking an intuitive approach that facilitates interaction of physicians [13], this work suggests clinical preference can at times be broad. The number of clinically applicable choices can be overwhelming for physicians even with the use of intuitive techniques such PGAP.

The physicist (participant A) and the planner (participant C) performed the most similarly by default with participant A in particular showing notably greater levels of consistency in planning choices between patients. Nevertheless, the PBAIO system was valuable in mitigated the majority of discrepancies in deviations at the calibration stage with few statistically significant dosimetric differences observed none of which were clinically significant.

CONCLUSION

There is evidence that expert-driven PGAP can be used to deliver consistent dosimetric planning with the clinically relevant region of the Pareto front defined comparably by any expert.

Conflicts of interest
The authors declare no conflict of interest.

REFERENCES

[1] M. Hussein, B.J.M. Heijmen, D. Verellen, and A. Nisbet, 'Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations', *British Journal of Radiology*, vol. 91, no. 1092, p. 20180270, 2018. doi.org/10.1259/bjr.20180270

[2] Y. Ge and Q.J. Wu, 'Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches', *Medical Physics*, vol. 46, no. 6, pp. 2760-2775, 2019. doi.org/10.1002/mp.13526

[3] C. Parkinson, C. Matthams, K. Foley, and E. Spezi, 'Artificial intelligence in radiation oncology: A review of its current status and potential application for the radiotherapy workforce', *Radiography*, vol. 27, pp. S63-S68, 2021. doi.org/10.1016/j.radi.2021.07.012

[4] S. Momin, Y. Fu, Y. Lei, J. Roper, J.D. Bradley, W.J. Curran, T. Liu, and X. Yang, 'Knowledge-based radiation treatment planning: A data-driven method survey', *Medical Physics*, vol. 22, no. 8, pp. 16-44, 2021. doi.org/10.1002/acm2.13337

[5] M. Zarepisheh, T. Long, N. Li, Z. Tian, H.E. Romeijn, X. Jia, and S.B. Jiang, 'A DVH-guided IMRT optimization algorithm for automatic treatment planning and adaptive radiotherapy replanning', *Medical Physics*, vol. 41, no. 6Part1, p. 061711, 2014. doi.org/10.1118/1.4875700

[6] L. Yuan, Y. Ge, W.R. Lee, F.F. Yin, J.P. Kirkpatrick, and Q.J. Wu, 'Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans', *Medical Physics*, vol. 39, no. 11, pp. 6868-78, 2012. doi.org/10.1118/1.4757927

[7] J.J. Boutilier, T. Lee, T. Craig, M.B. Sharpe, and T.C.Y. Chan, 'Models for predicting objective function weights in prostate cancer IMRT', *Medical Physics*, vol. 42, no. 4, pp. 1586-1595, 2015. doi.org/10.1118/1.4914140

[8] J. Ma, D. Nguyen, T. Bai, M. Folkerts, X. Jia, W. Lu, L. Zhou, and S. Jiang, 'A Feasibility Study on Deep Learning--Based Individualized 3D Dose Distribution Prediction', *Medical Physics*, vol. 48, no. 8, pp. 4438-4447, 2021. doi.org/10.1002/mp.15025

[9] E. Cagni, A. Botti, A. Chendi, M. Iori, and E. Spezi, 'Use of knowledge based DVH predictions to enhance automated re-planning strategies in head and neck adaptive radiotherapy', *Physics in Medicine & Biology*, vol. 66, no. 13, 2021. doi.org/10.1088/1361-6560/ac08b0

[10] A.C. Spalding, K.-W. Jee, K. Vineberg, M. Jablonowski, B.A. Fraass, C.C. Pan, T.S. Lawrence, R.K. Ten Haken, and E. Ben-Josef, 'Potential for dose-escalation and reduction of risk in pancreatic cancer using IMRT optimization with lexicographic ordering and gEUD-based cost functions', *Medical Physics*, vol. 34, no. 2, pp. 521-529, 2007. doi.org/10.1118/1.2426403

[11] S. Breedveld, P.R. Storchi, M. Keijzer, A.W. Heemink, and B. J. Heijmen, 'A novel approach to multi-criteria inverse planning for IMRT', *Physics in Medicine & Biology*, vol. 52, no. 20, p. 6339, 2007. doi.org/10.1088/0031-9155/52/20/016

[12] M.-C. Biston, M. Costea, F. Gassa, A.-A. Serre, P. Voet, R. Larson, and V. Gregoire, 'Evaluation of fully automated a priori MCO treatment planning in VMAT for head-and-neck cancer', *Physica Medica*, vol. 87, pp. 31-38, 2021. doi.org/10.1016/j.ejmp.2021.05.037

[13] D.L. Craft, T.S. Hong, H.A. Shih, and T.R. Bortfeld, 'Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy', *International Journal of Radiation Oncology-Biology-Physics*, vol. 82, no. 1, p. e83–e90, 2012. doi.org/10.1016/j.ijrobp.2010.12.007

[14] P.A. Wheeler, M. Chu, R. Holmes, O.W. Woodley, C.S. Jones, R. Maggs, J. Staffurth, N. Palaniappan, E. Spezi, D.G. Lewis, S. Campbell, J. Fitzgibbon, and A.E. Millin, 'Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer', *Radiotherapy & Oncology*, pp. 220-226, 2019. doi.org/10.1016/j.radonc.2019.08.001

[15] P.A. Wheeler, M. Chu, R. Holmes, M. Smyth, R. Maggs, E. Spezi, J. Staffurth, D.G. Lewis, and A.E. Millin, 'Utilisation of Pareto navigation techniques to calibrate a fully automated radiotherapy treatment planning solution', *Physics and Imaging in Radiation Oncology*, pp. 41-48, 2019. doi.org/10.1016/j.phro.2019.04.005

[16] I. Das, K. Cashon, K. Chopra, K. Khadivi, H. Malhotra, and C. Mayo, 'WE-E-T-617-06: Intra- and inter-planner dosimetric variations in inverse planning of IMRT', *Medical Physics*, vol. 32, no. 6Part20, pp. 2147-2147, 2005. doi.org/10.1118/1.1998607

[17] S. Srivastava, O. Nohadani, C. Medawar, C. Cheng and I. Das, 'Su-e-t-604: Inter planner dosimetric variations in imrt', *Medical Physics*, vol. 39, no. 6, Part 19, p. 3845, 2012. doi.org/10.1118/1.4735693

[18] Y.E. Erkal, A. Karabey, M.C. Sahin, A. Karayel, B. Tirpanci and H.Ş. Erkal, 'Intensity- modulated radiotherapy planning for prostate cancer: The evaluation of inter-observer variability and treatment delivery efficiency', *International Journal of Radiation Research*, vol. 20, no. 1, pp. 49-53, 2022. doi.org/10.52547/ijrr.20.1.8

[19] M. Esposito, G. Maggi, C. Marino, L. Bottalico, E. Cagni, C. Carbonini, M. Casale, S. Clemente, V. D'Alesio, D. Fedele, F. R. Giglioli, V. Landoni, and A. Martinotti, 'Multicentre treatment planning inter- comparison in a national context: the liver stereotactic ablative radiotherapy case', *Physica Medica*, vol. 32, no. 1, pp. 277-283, 2016. doi.org/10.1016/j.ejmp.2015.09.009

E. Spezi and M. Bray (eds.) 2024. *Proceedings of the Cardiff University Engineering Research Conference 2023.* Cardiff: Cardiff University Press.
doi.org/10.18573/conf1

Cardiff University Engineering Research Conference 2023 was organised by the School of Engineering and held from 12 to 14 July 2023 at Cardiff University.

The work presented in these proceedings has been peer reviewed and approved by the conference organisers and associated scientific committee to ensure high academic standards have been met.

First published 2024

Cardiff University Press
Cardiff University, PO Box 430
1st Floor, 30-36 Newport Road
Cardiff CF24 0DE

cardiffuniversitypress.org

Editorial design and layout by
Academic Visual Communication

ISBN: 978-1-9116-5349-3 (PDF)