

OPEN

# Template-based automation of treatment planning in advanced radiotherapy: a comprehensive dosimetric and clinical evaluation

Savino Cilla<sup>1\*</sup>, Anna Ianaro<sup>1</sup>, Carmela Romano<sup>1</sup>, Francesco Deodato<sup>2</sup>, Gabriella Macchia<sup>2</sup>, Milly Buwenge<sup>3</sup>, Nicola Dinapoli<sup>4</sup>, Luca Boldrini<sup>4</sup>, Alessio G. Morganti<sup>3</sup> & Vincenzo Valentini<sup>2,4</sup>

Despite the recent advanced developments in radiation therapy planning, treatment planning for head-neck and pelvic cancers remains challenging due to large concave target volumes, multiple dose prescriptions and numerous organs at risk close to targets. Inter-institutional studies highlighted that plan quality strongly depends on planner experience and skills. Automated optimization of planning procedure may improve plan quality and best practice. We performed a comprehensive dosimetric and clinical evaluation of the Pinnacle<sup>3</sup> AutoPlanning engine, comparing automatically generated plans (AP) with the historically clinically accepted manually-generated ones (MP). Thirty-six patients (12 for each of the following anatomical sites: head-neck, high-risk prostate and endometrial cancer) were re-planned with the AutoPlanning engine. Planning and optimization workflow was developed to automatically generate “dual-arc” VMAT plans with simultaneously integrated boost. Various dose and dose-volume parameters were used to build three metrics able to supply a global Plan Quality Index evaluation in terms of dose conformity indexes, targets coverage and sparing of critical organs. All plans were scored in a blinded clinical evaluation by two senior radiation oncologists. Dose accuracy was validated using the PTW Octavius-4D phantom together with the 1500 2D-array. AutoPlanning was able to produce high-quality clinically acceptable plans in all cases. The main benefit of AutoPlanning strategy was the improvement of overall treatment quality due to significant increased dose conformity and reduction of integral dose by 6–10%, keeping similar targets coverage. Overall planning time was reduced to 60–80 minutes, about a third of time needed for manual planning. In 94% of clinical evaluations, the AP plans scored equal or better to MP plans. Despite the increased fluence modulation, dose measurements reported an optimal agreement with dose calculations with a  $\gamma$ -pass-rate greater than 95% for 3%(global)-2 mm criteria. AutoPlanning engine is an effective device enabling the generation of VMAT high quality treatment plans according to institutional specific planning protocols.

In the past two decades, significant advances in planning and delivery of radiotherapy have been accomplished. In particular, the developments of new optimization algorithms in treatment planning systems (TPS) and the new capabilities of linear accelerators led to the implementation of intensity-modulated techniques (IMRT), able to improve dose distributions and conformity, reducing the irradiation to normal tissues and thereby minimising the risk of toxicity. Volumetric modulated arc therapy (VMAT), a complex technique able to deliver radiation beams in continuous sweeping arc around the patient, had a widespread diffusion in the last years<sup>1</sup>. This technology provided similar plan quality with respect to fixed-field IMRT in terms of target dose conformity and

<sup>1</sup>Medical Physics Unit, Fondazione di Ricerca e Cura Giovanni Paolo II - Università Cattolica del Sacro Cuore, Campobasso, Italy. <sup>2</sup>Radiation Oncology Unit, Fondazione di Ricerca e Cura Giovanni Paolo II - Università Cattolica del Sacro Cuore, Campobasso, Italy. <sup>3</sup>Radiation Oncology Department, DIMES Università di Bologna - Ospedale S. Orsola Malpighi, Bologna, Italy. <sup>4</sup>Dipartimento di Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, UOC Radioterapia Oncologica, Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Roma, Italy. \*email: [savino.cilla@fgps.it](mailto:savino.cilla@fgps.it)

minimal exposure to surrounding healthy tissues but with a large reduction in treatment time, potentially reducing the risk of intrafraction motion and improving patient comfort.

Despite the advanced developments in radiation therapy planning over the past years, complex treatment planning with VMAT or IMRT still remains challenging in order to achieve clinically acceptable plans. The manual planning process needed for these advanced techniques usually requires several trial-and-error optimization processes in which the continuous manual tuning of dosimetric objectives and weights translates in overall quality of plans strongly depending on planner experience<sup>2</sup>.

Various algorithms have been proposed for an automated optimization of the planning procedure and the search for the optimal patient plan in order to improve best practice. In particular, fully automated IMRT and VMAT plans have been successfully generated for clinical application using new commercial solutions. Knowledge-based planning (KBP) algorithms use a mathematical model, built and trained from a large number of previous optimal plans, to estimate the dose distribution and dose-volume histograms (DVH) for any new patient. This strategy has been implemented in the Varian Eclipse treatment planning system (TPS) as Rapidplan module (Varian Medical Systems, Palo Alto, Ca) and has been recently tested in different anatomical sites reporting an overall improving of treatment plans quality<sup>3–8</sup> with respect to conventional manual planning. Multi-criteria optimization (MCO) provides an alternative optimization workflow by generating a set of Pareto-optimal treatment plans regarding user-specified priorities and objectives (eg. plans for which no objectives can be improved without impairing another one)<sup>9</sup>. The planner can then navigate on the so-called Pareto surface from one plan to another to balance between clinical trade-offs. This strategy has been implemented in the Raystation TPS (Raysearch, Stockholm, Sweden) and in the Erasmus-Icycle algorithm developed at Erasmus MC-Cancer Institute in Rotterdam, reporting successful application in specific sites<sup>10–16</sup>. Another strategy for planning automation is the template-based planning optimization process. This algorithm, implemented in the Autoplanning module of Pinnacle<sup>3</sup> TPS (Philips Medical Systems, Fitchburg, WI) does not require any prior database of successful plans nor model training but uses an iterative approach of progressive optimization that mimic all the steps of experienced and skilled planners<sup>17</sup>. Autoplanning automates the inverse planning process using a so-called “Technique”, i.e. a template of parameters including beam setup, dose prescriptions and planning objectives that can be customized for each treatment protocol and tumor site. Then, the Autoplanning engine applies the Technique to iteratively optimize planning parameters to best meet the desired planning goals; different kinds of dummy structures are automatically generated and new objectives are added to the planning goals list in order to achieve better organs-at-risk (OARs) sparing and target uniformity and conformity by reducing cold/hot spots and managing the dose fall-off outside the targets. Autoplanning has been recently tested in a few publications for VMAT treatments of prostate cases<sup>18</sup>, head-neck tumours<sup>19</sup>, esophageal cancer<sup>20</sup> and for the extracranial stereotactic treatment of liver<sup>21</sup> and lung<sup>22</sup> metastasis, reporting in all cases promising results. All the aforementioned studies provided useful investigations of Autoplanning but were limited to single cancer sites.

The aim of the present study was to provide a more comprehensive evaluation of the Autoplanning potential including several challenging sites, normally treated in the clinical routine, as bilateral head-neck tumours, high-risk prostate cancer and endometrial cancer. In all these sites, several large irregular-shaped targets volumes, multiple dose prescription levels and the large number of complex anatomical organs at risk close or overlapped to targets represent a major challenge for the generation of optimal quality plans. VMAT plans for these cancer cases should then require highly conformal dose distributions and steep dose-gradient between targets and the many critical organs-at-risk.

Based on the recent literature data, in this study we hypothesized that automated radiotherapy treatment planning has the potential to increase consistency, improve plan quality and reduce workload for all routinely challenging treatments. The automated plans were then compared with the clinically accepted VMAT treatment plans manually generated by experienced medical physicists.

## Material and Methods

**Ethics statement.** This study was approved by the Institutional Review Board of the Fondazione di Ricerca e Cura Giovanni Paolo II in Campobasso, Italy. Because this was not a treatment-based but a retrospective dosimetric planning study, our institutional review board waived the need for written informed consent from the participants. The patient information was anonymized and de-identified to protect patient confidentiality. All the methods described here were performed in accordance with the relevant guidelines and regulations.

**Patient population and volumes definition.** Three different complex anatomic sites are presented to demonstrate the potential and challenges of automated template-based VMAT planning with respect to conventional manual planning. A total of 36 patients previously treated with SIB-VMAT were randomly selected, 12 patients for each of the following pathologies: bilateral head-neck, high-risk prostate and endometrial cancer. All patients underwent a simulation computed tomography (CT) scan (3-mm slice thickness); MRI-imaging for endometrial and prostate cases and PET-CT imaging for head-neck cases were then co-registered with the simulation CT for accurate volume delineation.

**Head-and-neck cancer.** The primary tumour and nodes showing metabolic activity at 18F PET-CT scan were defined as clinical target volume 1 (CTV1). CTV2 and CTV3 were defined as lymph nodes with high-risk and low-risk of occult metastases, respectively. Lymph nodal regions were outlined according to Grégoire *et al.* guidelines<sup>23</sup>. Corresponding planning target volumes (PTVs) were obtained by adding an isotropic 4-mm margin to CTVs. All PTVs were simultaneously irradiated over 30 daily fractions according to the SIB technique. Doses of 67.5 Gy (2.25 Gy/fraction), 60.0 Gy (2.0 Gy/fraction), and 55.5 Gy (1.85 Gy/fraction) were prescribed to the PTV1, PTV2, and PTV3, respectively. A planning organ-at-risk volume (PRV) was defined for serial OARs

ROI	Goal type	Dose	Volume
Head-neck cancer			
PTV1, PTV2 and PTV3	DVH	95%	98%
	DVH	98%	95%
	DVH	107%	5%
Parotids	Mean dose	25 Gy	
Spinal cord	Max dose	45 Gy	
PRV Spinal cord	Max dose	50 Gy	
Brainstem	Max dose	50 Gy	
PRV Brainstem	Max dose	54 Gy	
Optic chiasm	Max dose	50 Gy	
PRV Optic chiasm	Max dose	55 Gy	
Retina	Max dose	40 Gy	
Lens	Max dose	5 Gy	
Optic nerves	Max dose	55 Gy	
Pharyngeal constr. muscles	Mean dose	55 Gy	
Cochleas	Mean dose	45 Gy	
Mandible	DVH	68 Gy	2%
High-risk prostate and endometrial cancer			
PTV1 and PTV2	DVH	95%	98%
	DVH	98%	95%
	DVH	107%	5%
Rectum	DVH	65 Gy	25%
	DVH	60 Gy	35%
	DVH	50 Gy	50%
Bladder	DVH	70 Gy	35%
	DVH	65 Gy	50%
Femurs	Max dose	50 Gy	
Small Bowel	DVH	15 Gy	120cc

**Table 1.** Clinical objectives for treatment planning at Fondazione di Ricerca e Cura “Giovanni Paolo II”. For each PTV the doses are expressed as percentage of dose prescription.

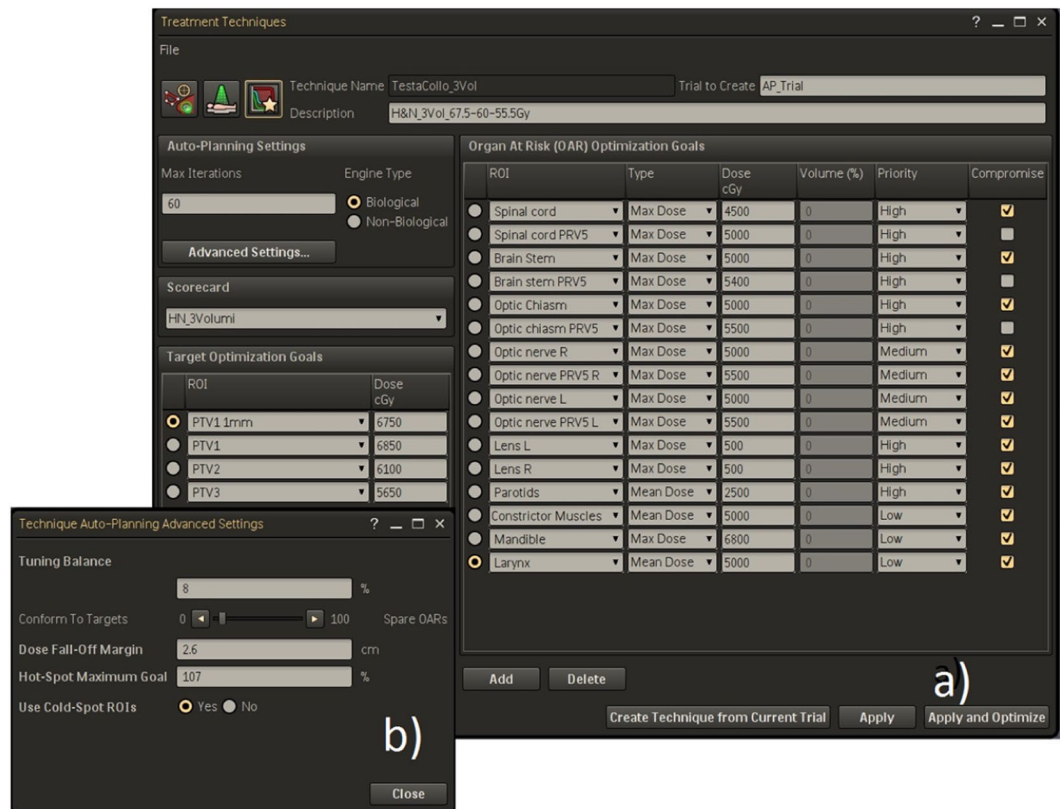
(spinal cord, brainstem, optic chiasm and optic nerves) isotropically expanding the corresponding OAR by 5 mm. For all serial PRV\_OARs, the dose to 0.035 cc was considered as maximal dose.

**High-risk prostate cancer.** The prostate plus 5-mm periprostate tissue and 5 to 20 mm of caudal seminal vesicles based on risk category<sup>24</sup> were defined as CTV1. CTV2 included the obturator, internal and external iliac, and presacral lymph nodes. PTV2 was obtained by adding an isotropic 8-mm margin to the CTV2; PTV1 was obtained by adding a 8-mm margin in all directions to the CTV1, except posteriorly where a 6-mm margin was given. The PTVs were simultaneously irradiated over 25 daily fractions prescribing 65.0 Gy (2.6 Gy/fraction) and 45.0 Gy (1.8 Gy/fraction) to PTV1 and PTV2, respectively. Main OARs were considered the rectum, the bladder, the small bowel and the femurs.

**Endometrial cancer.** The PTV1 consisted of the upper two thirds of vagina plus resection lines in the parametria (CTV1) with a 8 mm margin. CTV2 included the obturator, external iliac, internal iliac and presacral nodes. Corresponding PTV2 was obtained adding 8 mm margin. The PTVs were simultaneously irradiated over 25 daily fractions with a dose of 60.0 Gy (2.4 Gy/fraction) and 45.0 Gy (1.8 Gy/fraction) to PTV1 and PTV2, respectively. Main OARs were the rectum, the bladder, the small bowel and the femurs.

**Manual and auto-VMAT planning.** Manual VMAT plans were generated with “dual-arc” feature using the inverse optimization process previously described in more details<sup>25</sup> for coplanar 6–10 MV photon beams of an Elekta VersaHD linac (Elekta Ltd., Crawley, UK). A full gantry rotation was described by a sequence of 90 control points, i.e. one every 4°. Collimator was set at 10° to minimize the tongue-and-groove cumulative effect. For the three anatomical sites, treatment planning was performed accounting to the clinical objectives reported in Table 1.

Automated plans were created for each patient using the Autoplanning module implemented in the version 16.0 of Pinnacle<sup>3</sup> TPS. For each anatomical site, the Technique was created based on the same beam parameters, dose prescription and clinical objectives adopted for manual plans. During the optimization process, the AP engine automatically generates several dummy structures including: (a) rings around the PTVs to manage the dose fall-off, (b) residual targets structures where overlaps between no-compromised OARs are removed, (c) residual OARs structures where overlaps between targets are removed, (d) body structures used to control body dose and (e) hot-spot and cold-spot structures to manage target dose uniformity. New objectives are then



**Figure 1.** (a) AP setup template for head-neck cases; (b) advanced settings template.

automatically added to the aforementioned structures in order to achieve better OARs sparing and target uniformity and conformity. This process is iteratively performed during multiple optimization loops by adjusting the optimization parameters in order to continually spare the OARs without compromising the target coverage, i.e. mimicking what a manual experienced planner would usually do. Five patients for each site, not included in the present series, have been used to create and tweak the initial Techniques in order to generate plans fulfilling the clinical objectives. For both MP and AP plans, dose calculations were performed using the Pinnacle<sup>3</sup> collapsed cone convolution dose calculation algorithm with a dose grid resolution of 2 mm.

An example of Technique used for head-neck cases optimization is presented in Fig. 1. The objectives for the PTVs are only defined by numbers close to prescription doses (in our experience we chose as target goals the prescription doses plus 1 Gy, so as to avoid possible under dosage in PTVs boundary). The OARs objectives include maximum dose, mean dose and dose-volume histogram points; they can have three different priority levels (high, medium and low) and can be set compromised or uncompromised. These last choice is applicable when there is an overlap between PTVs and OARs; in the case of compromise option, the PTVs owns the overlapping voxels for the benefit of target coverage. Only for serial OARs, as spinal cord and brainstem, we chose to have higher priority than target volumes. In the advanced settings template (Fig. 1b), the planner can set up: (a) the tuning balance (i.e. the balance between target dose conformity and OARs sparing), (b) the dose fall-off margin (i.e. the distance across which the dose should decrease from 80% to 20% in an automatically generated tuning ring structure around the PTVs) and (c) the Cold-Spot ROI (i.e. the identification of cold regions inside the PTVs and the automatic creation of new tuning volumes and relative dose objectives to increase dose in the last optimization loops).

For head-neck cases, a manual fine tuning of 15 minutes at the end of the automated optimization process is usually needed in order to further lower the dose to serial OARs (eg. spinal cord).

Based on the Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) guidelines for normal tissue sparing<sup>26</sup>, similar templates were created for high-risk prostate and endometrial cancer cases.

**Plan evaluation and analysis.** Following the suggestions of Leung *et al.* paper for plan comparison<sup>27</sup>, several dosimetric parameters were used to build three metrics able to supply a global evaluation<sup>1</sup>: a healthy tissue conformity index (H) to describe the overall plan conformity<sup>2</sup>, a merit function (M) to describe the targets coverage and<sup>3</sup> a penalty function (P) to evaluate the sparing of critical organs.

**Healthy tissue conformity index (H).** This index was defined as

$$H = \frac{1}{r} \cdot \sum_{i=1}^n \left( \frac{TV_{RI,i}}{V_{RI,i}} \right)$$

where  $r$  is the number of PTVs of different prescription dose,  $TV_{RI}$  is the target volume covered by the reference isodose and  $V_{RI}$  is the volume of the reference isodose cloud. Reference isodoses were set at 95% of each prescription dose level.

For example, for the two PTVs in high-risk prostate case, the equation is expanded as:

$$H = \frac{1}{2} \cdot (H_1 + H_2) = \frac{1}{2} \cdot \left[ \left( \frac{TV_{95\%,PTV1}}{V_{95\%,PTV1}} \right) + \left( \frac{TV_{95\%,PTV2}}{V_{95\%,PTV2}} \right) \right]$$

**Target coverage index (M).** This index was defined as

$$M = \frac{1}{r} \cdot \sum_{j=1}^r \left[ \frac{\sum_{i=1}^p \left( \frac{V_{Tj,Di}}{V_{Tj,RDi}} \right) + \sum_{i=1}^q \left( 1 - \frac{V_{Tj,Di}}{V_{Tj,ADi}} \right)}{\sum_{i=1}^p \left( \frac{100}{V_{Tj,RDi}} \right) + q} \right]$$

where  $r$  is the number of targets of different prescription dose,  $p$  is the number of cold spot checks,  $q$  is the number of hot spot checks,  $V_{Tj,Di}$  is the volume of the  $j$ th target in % receiving a dose of at least the  $i$ th dose level,  $V_{Tj,RDi}$  is the minimum volume of the  $j$ th target in % receiving at least the  $i$ th dose level and  $V_{Tj,ADi}$  is the allowable volume of the  $j$ th target in % receiving at least the  $i$ th dose level.

For example, for the two PTVs objectives in high-risk prostate case as reported in Table 1, the equation is expanded as:

$$M = \frac{1}{2} \cdot (M_1 + M_2) = \frac{1}{2} \cdot \left[ \frac{\left[ \frac{V_{PTV1,95}}{98} + \frac{V_{PTV1,98}}{95} \right] + \left[ 1 - \frac{V_{PTV1,107}}{5} \right]}{\left[ \frac{100}{98} + \frac{100}{95} \right] + 1} + \frac{\left[ \frac{V_{PTV2,95}}{98} + \frac{V_{PTV2,98}}{95} \right] + \left[ 1 - \frac{V_{PTV2,107}}{5} \right]}{\left[ \frac{100}{98} + \frac{100}{95} \right] + 1} \right]$$

Note that the denominators represents the maximum possible score.

**Normal tissue sparing index (P).** This index was defined as:

$$P = \frac{1}{n} \cdot \sum_{j=1}^n \left[ \frac{1}{m} \cdot \sum_{i=1}^m \left( 1 - \frac{V_{Oj,Di}}{V_{Oj,ADi}} \right) \right]$$

where  $n$  is the number of critical organs to be monitored,  $m$  is number of check points used for the  $j$ th critical organ,  $V_{Oj,Di}$  is the volume of the  $j$ th critical organ in % receiving a dose of at least the  $i$ th dose level and  $V_{Oj,ADi}$  is the allowable volume of that organ in % receiving at least the  $i$ th dose level.

For example, in the high-risk prostate case,  $P$  becomes

$$P = \frac{1}{4} [P_{rectum} + P_{bladder} + P_{bowel} + P_{femurs}]$$

where, for the rectum QUANTEC objectives reported in Table 1, the equation is expanded as:

$$P_{rectum} = \frac{1}{3} \cdot \left[ \left( 1 - \frac{V50}{50} \right) + \left( 1 - \frac{V60}{35} \right) + \left( 1 - \frac{V65}{25} \right) \right]$$

and similarly for the other critical structures.

Main relevant OARs for the determination of  $P$  index were the rectum, the bladder, the small bowel and the femurs for prostate and endometrial cases, and PRV\_spinal cord, PRV\_brainstem, PRV\_optic chiasm, parotid glands, lens and pharyngeal constrictor muscles for head-neck cases.

**Plan quality index (PQI).** A comprehensive Plan Quality Index (PQI) can be formulated consolidating the three different metrics  $H$ ,  $M$  and  $P$  into a single figure obtained using the following Euclidean distance between the points  $(H,M,P)$  and  $(1,1,1)$ .

$$PQI = \sqrt{(1-H)^2 + (1-M)^2 + (1-P)^2}$$

This index represents the overall quality of a treatment plan. This choice was made because it was considered the most appropriate to represent the plan quality deviation from the ideal case. i.e. the point  $(0,0,0)$ . This must be interpreted as how far a plan is away from perfection, i.e.  $H=1$ ,  $M=1$ ,  $P=1$  or  $(1,1,1)$ . So, for an ideal case,  $PQI=0$  while for the worst scenario is  $PQI=\sqrt{3}$ .

In order to compare previous values with more common indices we calculated the conformation numbers (CNs) for each target volume as suggested by the Van't Riet *et al.*<sup>28</sup>:

$$CN = \frac{TV_{RI}}{TV} \times \frac{TV_{RI}}{V_{RI}}$$

where  $TV_{RI}$  was the target volume covered by the reference isodose,  $TV$  was the target volume, and  $V_{RI}$  was the volume of the reference isodose. The first part of this equation defines the quality of target coverage and the second part defines the volume of healthy tissues receiving a dose greater than or equal to the prescribed dose. CN ranges from 0 (complete PTV geographic miss) to the ideal value 1 (perfect conformity of the reference isodose to the PTV). Reference isodose was selected as 95% of the prescribed dose. Note that the second part of this equation represents the H value for each target.

Last, the integral dose (ID) received by non-tumour tissues was calculated as the product between mean dose and non-tumour tissue volume ( $Gy \cdot cc$ ).

Differences between manual and automated plans were quantified using the Wilcoxon matched-pair signed rank with a statistical significance at  $p < 0.05$ .

**Plans variability.** To evaluate the variability between MP and AP plans, we first calculated the coefficient of quartile variation (CQV) of the aforementioned metrics (CNs, H, M, P and PQI) for each of the three anatomical sites. CQV was defined as the ratio between the difference and the sum of first and third quartiles and was adopted because of its statistical robustness when dealing with data with outliers and/or skewed distributions<sup>29</sup>.

Then, in order to quantify the differences in standard deviations (SDs) of CNs, H, M, P and PQI metrics we performed the Levene's test for homogeneity of SDs when data comes from non normal distributions, with statistical significance at  $p < 0.05$ .

**Planning and treatment efficiency.** For all patients, the total planning time (human inputs, optimization loops and dose calculation times) and the total number of monitor units were recorded for both MP and AP plans; all optimization processes were performed on a local server (HP Z800 workstation, 2.80 GHz).

**Dosimetric verification.** Dose distributions were measured utilizing the 1500 2D ion-chamber array together with the Octavius-4D phantom<sup>30</sup> both developed by PTW (PTW, Freiburg, Germany). The 1500 2D-array consists of a matrix of 1405 ion chambers with a size of  $4.4\text{ mm} \times 4.4\text{ mm} \times 3.0\text{ mm}$ . This array is inserted into the Octavius-4D motorized cylindrical polystyrene phantom, capable to rotate synchronously with the gantry, so that the beam always hits the array in a perpendicular way, then allowing the possibility of 3-dimensional dose reconstruction and comparison. Measured and calculated dose distributions were compared by means of the gamma evaluation, based on the theoretical concept introduced by Low *et al.*<sup>31</sup>. Following the recent suggestions of the AAPM report No. 218<sup>32</sup>, dosimetric verification was considered optimal if the percentage of points fulfilling gamma index criteria exceeded 95% using 3% for dose criterion (global) and 2 mm for the distance to agreement criterion.

**Physician's plan scoring.** Two senior radiation oncologists independently performed a blind clinical evaluation of all AP and MP plans, based on the dose distributions, DVHs for all structures and a summary table reporting the most important parameters. The radiation oncologists rated the plans at first judging the clinical acceptability of each plan (pass or not pass) and secondary expressing their preferences using a clinical judgement based on a three-score scale (AP better than MP, MP better than AP and no preference). Cohen's kappa coefficient  $k$  was calculated to assess the inter-clinicians agreement<sup>33</sup>, with score defined as excellent ( $k > 0.81$ ), good ( $0.61 < k < 0.80$ ), moderate ( $0.41 < k < 0.60$ ), fair ( $0.21 < k < 0.40$ ) and poor ( $k < 0.20$ ).

## Results

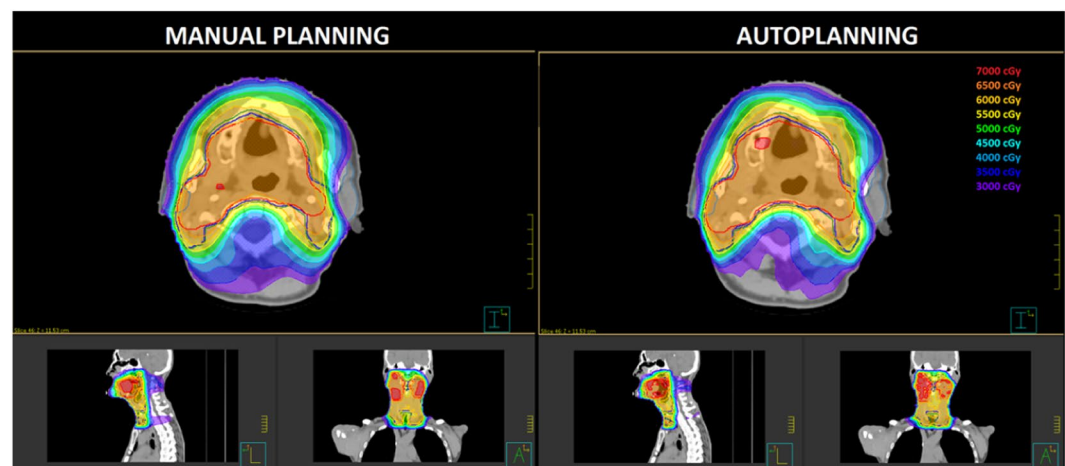
**Head-neck cancer cases.** A summary of H, M and P indexes for MP and AP plans is reported in Table 2. Global PQI scores for MP and AP plans were found to be  $0.796 \pm 0.059$  and  $0.722 \pm 0.056$ , respectively. No significant differences were found for the target coverage M index for nodal target volumes (PTV2 and PTV3) but MP plans showed larger hot-spot regions inside the PTV1 providing a worse M value. H index values show a higher capability of AP plans to better conform the doses to target volumes, especially to the prophylactic volumes. Regarding P index values for organs at risk sparing, AP plans provided a major sparing of parotid glands and a reduction of mean dose of 10% (3.7 Gy,  $p = 0.022$ ). Wilcoxon test also showed that the integral dose delivered to the patient body was significantly lower for AP plans than for the MP plans, with a reduction of 6.6% ( $p = 0.003$ ). Figure 2 shows the dose distributions comparison for a representative patient.

**High-risk prostate cancer cases.** A summary of H, M and P indexes for MP and AP plans is reported in Table 3. Global PQI scores for MP and AP plans were found to be  $0.429 \pm 0.053$  and  $0.400 \pm 0.049$ , respectively. No significant differences were found for the target coverage M index alone. H conformity index values was significantly better for AP plans ( $p = 0.003$ ) suggesting a higher capability of dose conformation to the large concave shaped nodal target volume, as shown graphically in Fig. 3 for a representative patient. This ability also translated into a significant reduction of integral dose to non-tumour volumes of 7.2%. Regarding P values for OARs sparing, no significant differences were found for all relevant OARs between the two techniques ( $p = 0.477$ ).

**Endometrial cancer cases.** A summary of H, M and P indexes for MP and AP plans is reported in Table 4. Global PQI scores for MP and AP plans were found to be  $0.532 \pm 0.095$  and  $0.472 \pm 0.081$ , respectively. As for prostate cases, while no significant differences were found for the target coverage M index, H values showed a higher capability of AP plans to better conform the dose distribution to target volumes, especially to the large concave nodal volumes. This capability is highlighted in Fig. 4 showing the dose distribution for a representative patient and it is also evidenced by a significant reduction in the integral dose to non-tumor tissues of 10% ( $p = 0.010$ ). Regarding P values for OARs sparing, no significant differences were found for all relevant OARs between the two techniques ( $p = 0.953$ ).

	Manual Planning (MP)		Automated Planning (AP)		P Wilcoxon
	Mean	Range (min-max)	Mean	Range (min-max)	
<b>Conformation numbers</b>					
CN 1	0.565	0.297–0.756	0.633	0.381–0.783	0.006
CN 2	0.572	0.234–0.722	0.602	0.305–0.713	0.012
CN 3	0.645	0.555–0.712	0.695	0.656–0.737	0.006
<b>Conformity Index of Healthy Tissue (H)</b>					
H <sub>PTV1</sub>	0.576	0.300–0.787	0.643	0.376–0.801	0.006
H <sub>PTV2</sub>	0.587	0.236–0.756	0.623	0.308–0.777	0.004
H <sub>PTV3</sub>	0.677	0.576–0.733	0.719	0.661–0.791	0.004
H	0.614	0.480–0.739	0.662	0.579–0.756	0.003
<b>Target Coverage (M)</b>					
M <sub>PTV1</sub>	0.815	0.634–0.927	0.946	0.909–0.997	0.003
M <sub>PTV2</sub>	0.954	0.922–0.983	0.956	0.924–0.981	0.790
M <sub>PTV3</sub>	0.934	0.871–0.970	0.939	0.885–0.983	0.477
M	0.901	0.854–0.937	0.947	0.912–0.968	0.004
<b>Normal Tissue Sparing (P)</b>					
P <sub>parotids</sub>	-0.161	-0.752–0.144	-0.039	-0.683–0.141	0.008
P <sub>lens</sub>	0.554	0.329–0.692	0.573	0.347–0.766	0.824
P <sub>PRV spine</sub>	0.216	0.146–0.328	0.242	0.194–0.324	0.026
P <sub>PRV brainstem</sub>	0.329	0.231–0.478	0.422	0.261–0.709	0.013
P <sub>PRV optic chiasm</sub>	0.917	0.829–0.965	0.921	0.866–0.959	0.477
P <sub>pharyngeal const. muscles</sub>	0.045	-0.038–0.184	0.048	-0.021–0.182	0.721
P	0.316	0.157–0.373	0.368	0.183–0.452	0.003
Plan Quality Index (PQI)	0.796	0.739–0.909	0.722	0.640–0.858	0.003
Integral Dose (Gy*cc*10 <sup>3</sup> )	1.079	0.615–1.528	1.008	0.586–1.374	0.003

**Table 2.** Comparison of scoring metrics between manual and automated planning for head-neck tumor cases.

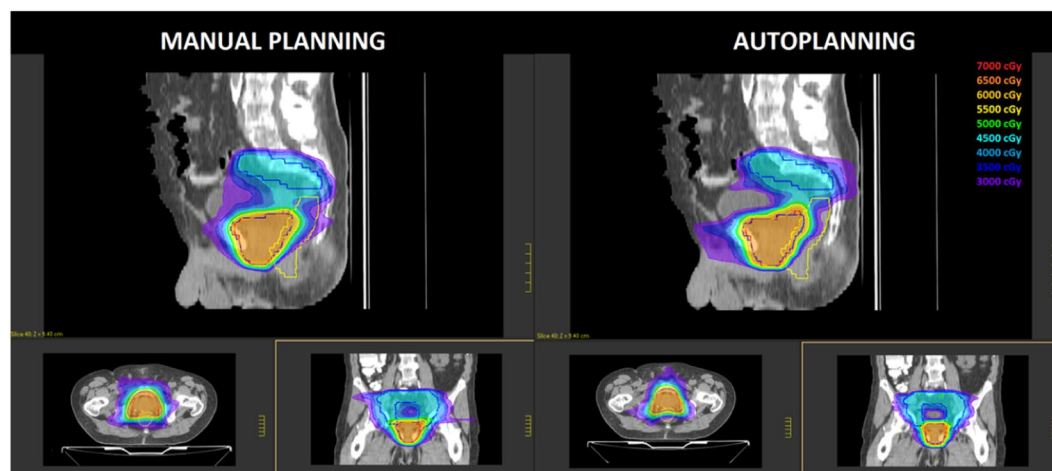


**Figure 2.** Comparison of dose distribution in axial, sagittal and coronal planes for a representative patient. Isodose curves are shown from 30 Gy to 70 Gy in 5 Gy steps. The PTV1, PTV2 and PTV3 target volumes are shown in red, blue and green contours, respectively.

**Plans variability comparison.** Figure 5 shows the whiskers box-plots of CNs, H, M, P and global PQI for the three anatomical sites. In particular, 1 out of 36 patients (an high-risk prostate patient) reported a worse PQI value for AP plans. The figure also shows a qualitative reduction of CQV values for plans optimized with AP technique for almost all metrics. Table 5 reports the CQV and SD values calculated for the principal metrics used for plan comparison. AP plans reported a narrowing of the variability range of CQV for the global PQI values from 21% for endometrial cases to 37% for head neck cases. For each metric, the results of Levene's test for homogeneity of SD between MP and AP plans are reported. In particular, AP plans reported a significant decrease in plans variability for the conformation numbers related to dose conformity to the large concave and irregular lymph-nodal volumes (CN2 and CN3).

	Manual Planning (MP)		Automated Planning (AP)		P Wilcoxon
	Mean	Range (min-max)	Mean	Range (min-max)	
<b>Conformation numbers</b>					
CN 1	0.820	0.754–0.866	0.821	0.767–0.864	0.722
CN 2	0.604	0.516–0.712	0.677	0.652–0.742	0.003
<b>Conformity Index of Healthy Tissue (H)</b>					
H <sub>PTV1</sub>	0.835	0.759–0.884	0.850	0.745–0.940	0.062
H <sub>PTV2</sub>	0.618	0.585–0.666	0.685	0.637–0.748	0.003
H	0.726	0.672–0.754	0.767	0.699–0.817	0.008
<b>Target Coverage (M)</b>					
M <sub>PTV1</sub>	0.959	0.843–0.986	0.968	0.913–0.995	0.878
M <sub>PTV2</sub>	0.943	0.925–0.964	0.941	0.913–0.970	0.721
M	0.951	0.900–0.975	0.954	0.913–0.981	0.878
<b>Normal Tissue Sparing (P)</b>					
P <sub>rectum</sub>	0.495	0.222–0.855	0.501	0.275–0.834	0.594
P <sub>bladder</sub>	0.641	0.250–0.893	0.651	0.267–0.997	0.859
P <sub>femurs</sub>	1.000	1.000–1.000	1.000	1.000–1.000	1.000
P <sub>small bowel</sub>	0.557	0.288–0.997	0.559	0.292–0.998	0.424
P	0.673	0.449–0.863	0.678	0.458–0.858	0.477
Plan Quality Index (PQI)	0.434	0.301–0.639	0.406	0.272–0.595	0.042
Integral Dose (Gy*cc) 10 <sup>5</sup>	2.671	2.103–3.340	2.452	1.950–3.019	0.003

**Table 3.** Comparison of scoring metrics between manual and automated planning for high-risk prostate cancer cases.



**Figure 3.** Comparison of dose distribution in axial, sagittal and coronal planes for a representative patient. Isodose curves are shown from 30 Gy to 60 Gy in 5 Gy steps. The PTV1 and PTV2 target volumes are shown in red and blue contours, respectively.

**Treatment efficiency and dosimetric verification.** Table 6 reports a summary of the treatment planning efficiency and the delivery metrics. Average total treatment planning time was about 60 minutes for AP plans in high-risk prostate and endometrial cases and 80 minutes for head-neck cases. Compared to MP plans, a significant larger number of monitor units was observed for AP plans, especially for head-neck cases, reflecting an increased level of fluence modulation.

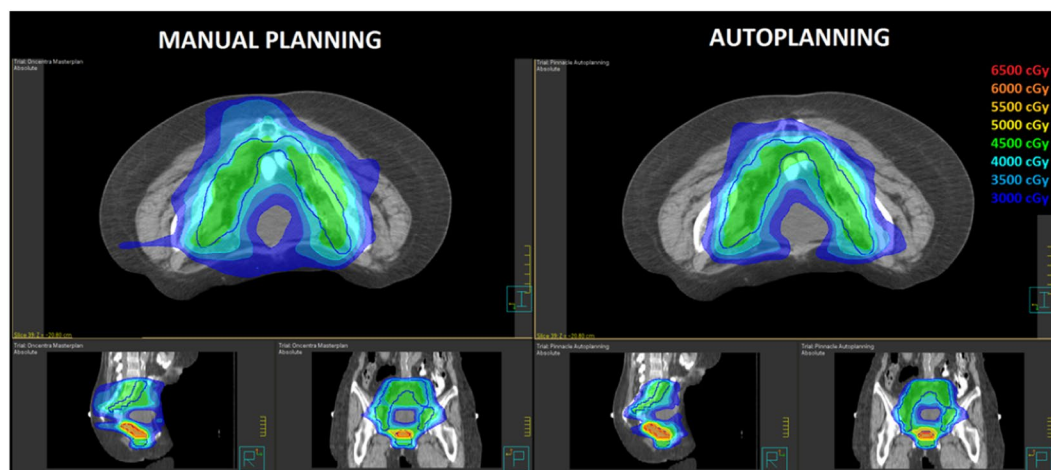
Pre-treatment verification was performed for all plans. With criteria equal to 3% (global) –2 mm for  $\gamma$ -index, the average pass-rate was  $98.2 \pm 1.4\%$  for MP plans and  $98.1 \pm 1.4\%$  for AP plans ( $p = 0.882$ ).

**Physician's plan scoring.** Cohen's kappa coefficient resulted in an intra-observer variability equal to 0.83 indicating an almost perfect agreement. Regarding the assessment of all plans by the two radiation oncologists, the clinical score of AP plans was equal or better than MP plans in 97% and 94% of cases for both clinicians, respectively. The clinical evaluation of plan quality was favourable to AP plans for both radiation oncologists.



	Manual Planning (MP)		Automated Planning (AP)		P Wilcoxon
	Mean	Range (min-max)	Mean	Range (min-max)	
<b>Conformation numbers</b>					
CN 1	0.602	0.472–0.681	0.647	0.543–0.724	0.074
CN 2	0.604	0.538–0.704	0.702	0.665–0.743	0.005
<b>Conformity Index of Healthy Tissue (H)</b>					
H <sub>PTV1</sub>	0.604	0.472–0.683	0.645	0.543–0.724	0.139
H <sub>PTV2</sub>	0.602	0.559–0.691	0.725	0.690–0.781	0.008
H	0.603	0.540–0.687	0.685	0.617–0.744	0.011
<b>Target Coverage (M)</b>					
M <sub>PTV1</sub>	0.979	0.917–0.999	0.985	0.958–0.998	0.374
M <sub>PTV2</sub>	0.961	0.927–0.988	0.960	0.928–0.984	0.515
M	0.970	0.943–0.994	0.972	0.961–0.988	0.515
<b>Normal Tissue Sparing (P)</b>					
P <sub>rectum</sub>	0.737	0.625–0.884	0.739	0.657–0.874	0.859
P <sub>bladder</sub>	0.891	0.767–0.981	0.895	0.760–0.958	0.767
P <sub>femurs</sub>	0.894	0.870–1.000	1.000	1.000–1.000	0.180
P <sub>small bowel</sub>	0.490	0.125–0.746	0.484	0.104–0.723	0.374
P	0.647	0.446–0.781	0.650	0.441–0.766	0.953
Plan Quality Index (PQI)	0.532	0.425–0.668	0.472	0.385–0.651	0.015
Integral Dose (Gy*cc*10 <sup>5</sup> )	3.196	2.392–4.468	2.881	2.111–3.999	0.010

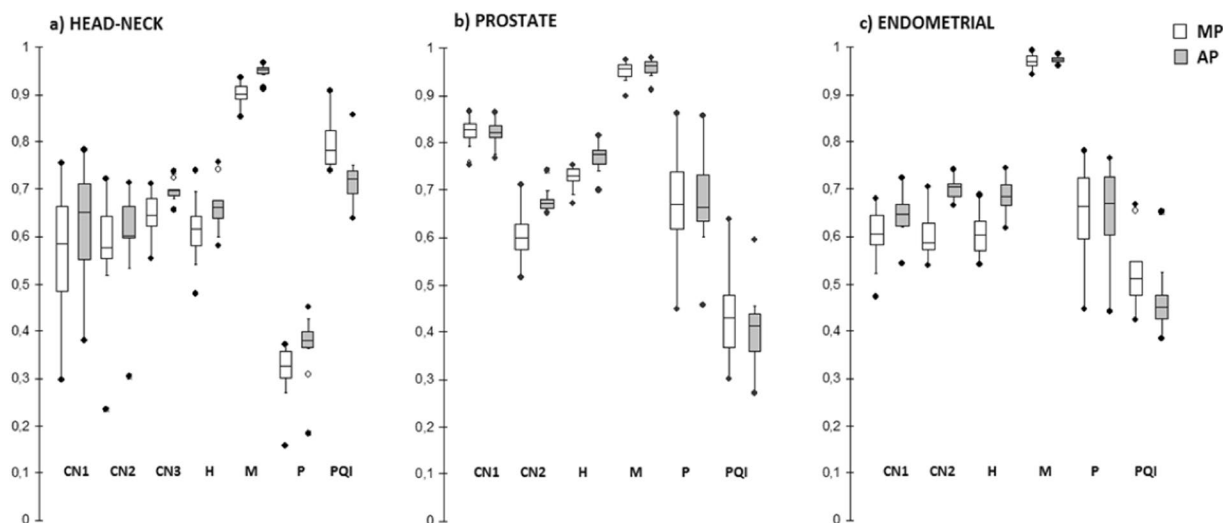
**Table 4.** Comparison of scoring metrics between manual and automated planning for endometrial cancer cases.



**Figure 4.** Comparison of dose distribution in axial, sagittal and coronal planes for a representative patient. Isodose curves are shown from 30 Gy to 60 Gy in 5 Gy steps. The PTV1 and PTV2 target volumes are shown in red and blue contours.

## Discussion

In the present study we explored the potential of a fully template-based automated VMAT planning engine implemented in Pinnacle TPS for challenging treatments executed in clinical routine. Head-neck, high-risk prostate and endometrial cancer sites were chosen because they involve large concave-shaped target volumes, multiple dose prescription, use of simultaneous integrated boost strategy and a large number of OARs adjacent or partially overlapping targets, then presenting the most complex and challenging problem for the plan optimization algorithms. The resulting plans were then compared with clinically accepted VMAT treatment plans generated by experienced medical physicists. The selection of optimal plans from different competing techniques or planning strategies has always been a daunting task, relying on dose volume histogram metrics and visual inspection of isodose distributions, often providing ambiguous evaluations. For this reason, a few quantitative indexes have been introduced to quantitatively describe the quality of a given plan<sup>27,34</sup>. In particular, Leung *et al.* proposed a new dose-volume based index for intensity-modulated plans called Plan Quality Index, able to simultaneously describe the overall plan conformity, the target coverage and the doses to critical organs<sup>27</sup>. The authors reported that this index improved the plan quality discerning power with respect to conventional comparison strategies. Following the suggestions of Leung *et al.* paper, we adopted the proposed PQI as fundamental metric



**Figure 5.** Whiskers box-plots of CNs, H, M, P and global PQI for the three anatomical sites for both MP and AP plans. The central line marks the median, the edge of the box are the 25<sup>th</sup> and 75<sup>th</sup> percentiles, black circles represent the extreme values. The whiskers extend to the adjacent values. The extent of the boxes represent the Inter Quartile Range (IQR).

Metric	MP		AP		p levene
	CQV	SD	CQV	SD	
Head-neck cases					
CN1	0.156	0.148	0.126	0.123	0.537
CN2	0.074	0.129	0.053	0.112	0.813
CN3	0.045	0.047	0.007	0.021	0.034
H	0.052	0.076	0.029	0.056	0.496
M	0.015	0.027	0.006	0.018	0.299
P	0.084	0.075	0.043	0.066	0.902
PQI	0.044	0.059	0.032	0.056	0.604
Prostate cases					
CN1	0.017	0.036	0.015	0.030	0.546
CN2	0.047	0.061	0.014	0.025	0.042
H	0.015	0.031	0.019	0.025	0.567
M	0.012	0.020	0.011	0.019	0.832
P	0.089	0.118	0.071	0.106	0.672
PQI	0.132	0.098	0.099	0.074	0.511
Endometrial cases					
CN1	0.053	0.067	0.035	0.053	0.510
CN2	0.048	0.053	0.018	0.022	0.024
H	0.053	0.050	0.031	0.040	0.600
M	0.011	0.017	0.004	0.008	0.167
P	0.098	0.116	0.091	0.107	0.801
PQI	0.068	0.095	0.056	0.081	0.682

**Table 5.** Summary of Coefficients of Quartile Variations (CQV) and standard deviations (SD) of CNs, H, M, P and global PQI for the three anatomical sites. For each metric, the results of Levene’s test for homogeneity of SD between MP and AP plans are reported.

to determine plan quality and for plan comparison purposes. Our results show that for all the anatomic sites, AP plans were able to provide similar, and in some cases better, plan quality of MP plans. AP plans significantly improved dose conformity, especially to large concave nodal target volumes, in all anatomical sites, but no statistically significant differences were found in terms of targets coverage. Similarly, no statistically significant difference were found for all relevant OARs dose sparing in high-risk prostate and endometrial cancer cases. However, AP plans showed an improvement in OARs sparing in head-neck cancer cases, i.e. in cases with the most complex anatomic scenario. In this case, maximal dose to PRV brainstem was reduced on average by 4.3 Gy and parotids mean dose was reduced on average by 3.7 Gy, that may be clinically relevant to reduce xerostomia.

	MP		AP		p
	Mean	STD	Mean	STD	
Head-neck cancer cases					
MUs	508	44	586	39	0.003
Planning time (minutes)	188	44	82	6	0.000
Beam-on-time (minutes)	2.3	0.2	2.4	0.2	0.133
$\gamma$ pass-rate (%)	97.9	1.4	97.8	1.4	0.749
High-risk prostate cancer cases					
MUs	528	58	572	45	0.003
Planning time (minutes)	180	40	61	4	0.000
Beam-on-time (minutes)	2.2	0.2	2.3	0.2	0.349
$\gamma$ pass-rate (%)	98.3	1.4	98.3	1.5	0.811
Endometrial cancer cases					
MUs	540	122	583	78	0.015
Planning time (minutes)	175	41	60	4	0.000
Beam-on-time (minutes)	2.1	0.2	2.2	0.2	0.286
$\gamma$ pass-rate (%)	98.5	1.4	98.3	1.4	0.983

**Table 6.** Overview of treatment delivery metrics.

These dosimetric findings were confirmed by the two radiation oncologist in the blind clinical evaluation session who considered AP plans better or equal to MP plans in more than 90% of cases.

A potential bias of this kind of planning comparison studies is that the quality of MP plans should be as high as possible (poor quality of MP plans obviously would favour AP plans). In our case, all clinically MP plans were created by two medical physicist with 10 years experience in VMAT planning, with the aim to obtain not only high quality plans but also a reduction of interplanner variability. As reported in Table 6, AP plans achieved a reduction of variability expressed by the CQV metrics for almost all dosimetric metrics for the three anatomical sites, with statistical significance for dose conformity. AP engine not only significantly improved dose conformity for complicated target geometry (including nodal involvement) but it has also the potential to drive a reduction of human-caused variability in VMAT planning for conformal coverage and dose distributions.

It must be highlighted that manual planning for complex cases is a challenging task, based mainly on the planner experience. Planners, although very experienced, never “a priori” know how much a plan can be optimized nor they can ensure that all dosimetric constraints on all OARs have been tightened as much as possible. This result has also been observed in other recent studies<sup>5,15</sup> focused on the optimization of prostate treatment with different automated algorithms as Rapidplan and Erasmus-Icycle. From this point of view, automated planning could allow more consistent outcomes in treatment planning studies and clinical trials thanks to their greater ability to reduce the inter- and intra-planner variations.

Regarding planning and treatment efficiency, AP plans resulted in 8–15% increase of MUs, a result in agreement with other experiences with Autoplanning engine<sup>22</sup>, and suggesting an increase of plan complexity and fluence modulation. However, the MUs increase did not translate in a lower pass rate during pre-treatment verification on the Octavius-4D phantom, which resulted in strong agreement with MP plans pass rate ( $p = 0.882$ ). Moreover, unlike expected, the increase of MUs number did not increase the integral dose to the patient, which was lower by 6.6–10.1%, theoretically reducing the risk of secondary malignancies<sup>35</sup>.

Mean overall planning time including human inputs, optimization loop processes and calculation times was 60 minutes and 80 minutes for pelvic and head-neck AP plans, respectively (about a third of time needed for manual planning).

Perhaps the most important feature of the Autoplanning module is its ability, according to the vendor, to push the OARs dose sparing beyond the constraints specified in the Technique, towards physical achievable limits. This feature is unique and represents a significant change compared to other dosimetric planning engines or to the natural human planning strategy, in which the primary goal is the achievement of objectives judged to be clinically effective. In the present paper, this ability was reported in the treatments of the head-neck district, where AP plans showed a significant reduction in the average dose to the parotid glands of about 10% and of the maximum dose to brainstem and spinal cord, well beyond the objectives that had been assigned in the Technique. Clearly, to definitively prove the aforementioned claim it would be necessary to demonstrate that AP plans are Pareto optimal, i.e. one or more objectives (as OARs sparing) cannot be improved without worsening at least one other (as target coverage). This demonstration is a challenging mathematical task<sup>36</sup> and is beyond the scope of the present paper.

An advantage of Autoplanning with respect to other strategies for automatic planning based on KBP knowledge-based approach is that it does not rely on a database of prior patients. This database must usually be filled with a large number of high quality plans for each protocol and disease site, whose clinical implementation translated in a labor-intensive process. Any changes in contouring protocol or dose prescription or planning techniques could require the generation of a new database. On contrary, in our experience only a small set of training patients for each anatomical site (five patients in our experience) was necessary as starting point for the implementation of the Techniques in Autoplanning by an expert team of medical physicists.

However, in the case of head-neck cancer site, we faced the problem of high point doses in serial OARs as the spinal cord or brainstem when they lie very close or partially inside the PTV. These cases have been solved with a further manual tuning of dose objectives in the post optimization so as to decrease the dose to these serial OARs below the acceptable values. This re-optimization step does not require more than 15 minutes of dose calculation time. From this point of view, AP strategy can always be considered a high quality starting point for further plan optimization, that is a tool able to increase the overall quality of planning, rather than a tool that could completely remove the need of manual optimization.

A potential limitation of the PQI evaluation method is that different combinations of H, M and P values can provide identical PQI values when comparing two plans (i.e. one plan may have a better M while another plan may have a better P). In this scenario, clinicians would inevitably decide which one would benefit the patient most, focusing attention on the coverage of the targets rather than on dose conformity or OARs sparing. Moreover, the actual H, M and P indexes are defined so that each dosimetric parameter has the same weight, while, in some specific clinical cases, clinicians may prefer that an objective for an OAR or a target have more weight than another.

Two additional potential benefits of Autoplanning template-based module are currently under investigation. The first one is the feasibility of rapid and easy knowledge-sharing between different institutions. In our experience, for high-risk prostate and endometrial treatments, Autoplanning normally created optimal plans in a “one-button click” procedure without any planner intervention for manual tuning. Techniques for specific anatomical sites can be successfully shared and implemented across multiple centres with simple adaptations to local protocols, allowing each centre to obtain optimal plans with the same quality<sup>37</sup>. The second benefit concerns the use of Autoplanning in the adaptive radiotherapy setting. In this case, the goal to correct for daily tumour and normal tissue variations through modification of original plan is hampered by the time-consuming re-planning process, representing nowadays the major obstacle for large scale implementation of this strategy. Improvement in Autoplanning, therefore, has the potential to make routine online adaptive radiotherapy a possibility<sup>38</sup>.

## Conclusion

We evaluated the Pinnacle Autoplanning engine to be a robust clinical tool, reporting significant increase of dose conformity with respect to manual planning. The blinded clinical scoring confirmed the dosimetric results, showing that in more than 90% of the evaluations AP plans were judged of equal or better quality with respect to MP plans. The reductions of plans variability and overall treatment time suggest the use of Autoplanning as a valuable tool to standardize high plan quality and improve clinic efficiency. Owing to dosimetric and clinical advantages, Autoplanning engine is an effective device enabling the generation of VMAT high quality treatment plans according to institutional specific planning protocols.

Future studies are needed to expand Autoplanning to other treatment techniques such as extracranial stereotactic radiotherapy.

Received: 17 July 2019; Accepted: 24 November 2019;

Published online: 16 January 2020

## References

- Otto, K. Volumetric Modulated Arc Therapy: IMRT in a single arc. *Med. Phys.* **35**, 310–7 (2008).
- Nelms, B. E. *et al.* Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. *Pract Radiat Oncol.* **2**(4), 296–305 (2012).
- Kubo, K. *et al.* Dosimetric comparison of Rapidplan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. *Phys Med.* **44**, 199–204 (2017).
- Li, N. *et al.* Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials. *Int J Radiat Oncol Biol Phys* **97**, 164–172 (2017).
- Scaggion, A. *et al.* Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution. *Phys Med.* **53**, 86–93 (2018).
- Hussein, M. *et al.* Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiother Oncol.* **120**, 473–9 (2016).
- Tol, J. P. *et al.* Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol Biol Phys.* **91**, 612–20 (2015).
- Chang, A. T. Y., Hung, A. W. M., Cheung, F. W. K., Cheng, Y. T. & Ng, W. T. Comparison of planning quality and efficiency between conventional and knowledge-based algorithms in nasopharyngeal cancer patients using intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys.* **95**, 981–90 (2016).
- Craft, D. L. *et al.* Improved planning time and plan quality through multicriteria optimization for intensity modulated radiotherapy. *Int J Radiat Oncol Biol Phys.* **82**, e83–90 (2012).
- Buschmann, M. *et al.* Advanced optimization methods for whole pelvic and local prostate external beam therapy. *Phys. Med.* **32**, 465–73 (2016).
- Young, M. R. *et al.* Volumetric-modulated arc therapy using multicriteria optimization for body and extremity sarcoma. *J. Appl. Clin. Med. Phys.* **17**, 283–91 (2016).
- Ghandour, S., Matzinger, O. & Pachoud, M. Volumetric-modulated arc therapy planning using multicriteria optimization for localized prostate cancer. *J. Appl. Clin. Med. Phys.* **16**, 258–69 (2015).
- Kierkels, R. G. J., Visser, R. & Bijl, H. P. Multicriteria optimization enables less experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. *Radiat Oncol.* **87**(10) (2015).
- Voet, P. W. J. *et al.* Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. *Int J Radiat Oncol Biol Phys.* **88**, 1175–1179 (2014).
- Buschmann, M. *et al.* Automated volumetric modulated arc therapy planning for whole pelvic prostate radiotherapy. *Strahlenther Onkol.* **194**, 333–342 (2018).
- Dg, G. *et al.* Fully automated VMAT treatment planning for advanced-stage NSCLC patients. *Strahlenther Onkol.* **193**, 402–409 (2017).
- Xia, P. & Murray, E. 3D treatment planning system – Pinnacle system. *Med Dos.* **43**(2), 118–128 (2018).

18. Nawa, K. *et al.* Evaluation of a commercial automatic treatment planning system for prostate cancer. *Med Dos.* **42**(3), 203–209 (2017).
19. Gintz, D. *et al.* Initial evaluation of automated treatment planning software. *J Appl Clin Med Phys.* **17**, 331–346 (2016).
20. Li, X. *et al.* Dosimetric benefits of automation in the treatment of lower thoracic esophageal cancer: is manual planning still an alternative option? *Med Dosim.* **42**(4), 289–295 (2017).
21. Gallio, E. *et al.* Evaluation of a commercial automatic treatment planning system for liver stereotactic body radiation therapy treatments. *Phys Med.* **46**, 153–159 (2018).
22. Vanderstraeten, B. *et al.* Automated instead of manual treatment planning? A plan comparison based on dose-volume statistics and clinical preference. *Int J Radiat Oncol Biol Phys.* **102**, 443–450 (2018).
23. Grégoire, V. *et al.* Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother. Oncol.* **110**(1), 172–81 (2014).
24. Boehmer, D. *et al.* Guidelines for primary radiotherapy of patients with prostate cancer. *Radiother. Oncol.* **79**(3), 259–69 (2006).
25. Cilla, S. *et al.* Assessing the feasibility of volumetric-modulated arc therapy using simultaneous integrated boost (SIB-VMAT): an analysis for complex headneck, prostate and rectal cancer cases. *Med Dosim* **39**, 108–16 (2014).
26. Bentzen, S. M. *et al.* Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys.* **76**(3 Suppl), S3–9 (2010).
27. Leung, L. H. *et al.* A new dose-volume based Plan Quality Index for IMRT plan comparison. *Radiother Oncol* **85**, 407–417 (2007).
28. van't Riet, A. *et al.* A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate. *Int J Radiat Oncol Biol Phys* **37**, 731–6 (1997).
29. Zwillinger, D. & Kokoska, S. *Standard Probability and Statistical Tables and Formula.* Chapman & Hall, Boca Raton (2000).
30. Stelljes, T. S. *et al.* Dosimetric characteristics of the novel 2D ionization chamber array OCTAVIUS Detector 1500. *Med Phys.* **42**(4), 1528–37 (2015).
31. Low, D. A. *et al.* A technique for the quantitative evaluation of dose distributions. *Med Phys.* **25**(5), 656–61 (1998).
32. Miften, M. *et al.* Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys.* **45**(4), e53–e83 (2018).
33. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics.* **33**, 159–174 (1977).
34. Feuvret, L. *et al.* Conformity index: a review. *Int J Radiat Oncol Biol Phys* **64**, 333e42 (2006).
35. Newhauser, W. D. *et al.* A review of radiotherapy induced late effects research after advanced technology treatments. *Front Oncol.* **6**, 13 (2016).
36. Janssen, T. *et al.* Pareto fronts in clinical practice for Pinnacle. *Int J Radiat Oncol Biol Phys.* **85**(3), 873–80 (2013).
37. Roach, D. *et al.* Adapting automated treatment planning configurations across international centres for prostate radiotherapy. *Physics and Imaging in Radiation Oncology* **10**, 7–13 (2019).
38. Sonke, J. J., Aznar, M. & Rasch, C. Adaptive Radiotherapy for Anatomical Changes. *Semin Radiat Oncol.* **29**(3), 245–257 (2019).

## Acknowledgements

The authors thank Dr. Davide Raspanti from Tema Sinergie, Faenza, Italy for his excellent technical assistance and precious suggestions.

## Author contributions

S.C. conceived and designed the study and draft the manuscript, A.I. and C.R. planned the treatments and collected the data, F.D., G.M., M.B., N.D. and L.B. analyzed the data and wrote the manuscript, A.G.M. and V.V. made intellectual contribution and critical review of the manuscript. All authors revised the draft manuscript and approved the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020