ACTA ONCOLOGICA
2022, VOL. 61, NO. 8, 1012–1018
https://doi.org/10.1080/0284186X.2022.2093615

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

A practical method to improve the performance of knowledge-based VMAT planning for endometrial and cervical cancer

Zheng Kang[a,b]*, Lirong Fu[a]*, Jun Liu[a]*, Liwan Shi[a,c] and Yimin Li[a,b,c]

[a]Department of Radiation Oncology, The First Affiliated Hospital of Xiamen University, Xiamen, China; [b]Xiamen Key Laboratory of Radiation Oncology, Xiamen, China; [c]Teaching Hospital of Fujian Medical University, Xiamen, China

ABSTRACT

Purpose: The aim of this work was to demonstrate a practical and effective method to improve the performance of RapidPlan (RP) model.

Methods: 203 consecutive clinical VMAT plans (P_0) for cervical and endometrial cancer were used to train an RP model (M_0). The plans were then reoptimized by M_0 to generate 203 new plans (P_1). Compared with P_0, 150 plans with a lower mean dose (MD) of bladder, rectum and PBM were selected from P_1 to configure a new RP model (M_1). A final RP model (M_2) was trained using plans in M_1 and the remaining 53 plans from P_1 (excluding OARs with worse MD) and the corresponding plans from P_0 (only including OARs with better MD). The models were validated on the mentioned 53 plans (closed-loop set) and 46 patient cohorts outside the training library (open-loop set). $p < 0.05$ was considered statistically significant.

Results: For closed-loop validation, the difference of $D_{2\%}$, $D_{98\%}$ and $CI_{95\%}$ between groups was of no statistical significance, the homogeneity index (HI) was lower in the groups of RP models ($p < 0.05$). The MD of all OARs decreased monotonically in the sequence of the clinical group, group M_0, M_1 and M_2, except the MD of bowel in M_1 and MD of LFH in M_2. Similarly, for open-loop validation, there was no significant difference in $D_{2\%}$, $D_{98\%}$ and HI between groups, but $CI_{95\%}$ was larger in the clinical group ($p < 0.05$). The MD of all OARs decreased monotonically in the sequence of the clinical group, group M_0, M_1 and M_2, with the exception of bowel in M_1.

Conclusion: The practical method of incorporating plan data of better-sparing OARs from both the clinical VMAT plans and the re-optimized plans could further improve the performance of the RP model.

ARTICLE HISTORY
Received 1 March 2022
Accepted 20 June 2022

KEYWORDS
Knowledge-based planning; cervical cancer; endometrial cancer; RapidPlan; dosimetry; iterative model training

Background

Optimally utilized radiotherapy can improve the overall survival rate and local control benefits of patients with endometrial and cervical cancer [1,2]. Recently, the volumetric modulated radiotherapy (VMAT) technique has been increasingly implemented to treat endometrial and cervical cancer [3], for VMAT plan has been proved to achieve excellent dose distribution and favorable clinical outcomes [4]. The process of VMAT planning is a complex, labor-intensive and time-consuming workflow. To achieve the optimal dose distribution of the VMAT plan, a planner manually adjusts the dose objectives through an inverse-planning method and several iterations. The trial and error process highly depends on the skills and experience of planners, which may lead to the variation in plan quality [5,6]. Plenty of studies have shown that knowledge-based planning (KBP) is an effective approach to reducing the inter-planner varieties of plan quality and improving planning efficiency [7–15].

The KBP method utilizes prior high-quality treatment plans to configure dose volume histogram (DVH) estimation for future radiotherapy plans. Nowadays, a commercial KBP software, namely RapidPlan (Varian Medical Systems, Palo Alto, CA, USA), has been widely implemented in clinical practice. By dividing the OARs into four different regions (i.e., the out-of-field, in-field, leaf-transmission, and overlap regions), RapidPlan uses the principal component analysis (PCA) and stepwise regression analysis to build a DVH estimation model, which builds strong correlations between geometric features and dosimetric features by using the corresponding information derived from plans of the training set [7–10]. Since the predicted DVH automatically generated by RapidPlan is used to optimize the plan, it is easier to improve the plan consistency and the planning efficiency.

Although well-trained RP models can generate plans that are at least as good or better than conventional manual plans in terms of target coverage, OAR sparing, plan consistency and efficiency, there is a known limitation of the RP model [16], i.e., the plan quality for new patients strongly depends on the quality of plans in the training set. The performance of RP models relies on the consistency between

CONTACT Shi Liwan, Li Yimin kangzheng6@163.com The First Affiliated Hospital of Xiamen University, Xiamen, 361000, Fujian, China
*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/0284186X.2022.2093615.

the anatomic features of the new patients and the population used for training. Nan Li *et al.* reported on a strategy to identify high-quality plans for a refined training set of cervical cancer [11]. In their study, a refined model was constructed by selecting plans emphasizing PBM sparing and bowel sparing by comparing the DVHs of original plans with DVHs of plans generated by an unfiltered model. Fusella *et al.* proposed an adjusted Plan Quality Metric (APQM) scoring system to select the training set of RP model for prostate cancer patients [10]. They found that the APQM model showed target coverage and OARs sparing equal or superior to the Uncleaned and Cleaned model [10]. In the studies of RP models of glioblastoma and prostate cancer by Chatterjee *et al.* [8,17], in order to improve the goodness-of-fit of the model, an intermediate model was generated using the DVHs of best-spared OARs of the initial model, then a final model was trained by plans re-planned using the intermediate model and manual refinement. However, they didn't specify what plan was the best spared OARs plan. Previously, Fogliata *et al.* configured an RP model with clinical VMAT plans and then trained a refined RP model with plans re-optimized by the initial model on the same patients [12]. They concluded that the refined RP model could generate plans with improved quality, mostly for parallel organs at risk, and possibly also the intrinsic model quality.

Previously, Hussein *et al.* validated a clinically adopted RP model for the treatment of cervical cancer [7]. They didn't use special techniques to select patients for training. In recent research, Marisol Tinoco *et al.* randomly selected 46 cervical cancer patients for RP model training and achieved a model capable of generating treatable VMAT plans for cervical cancer patients [18]. The aim of the present work was to estimate the performance of a practical method for selecting the training set for the RP model, the method was based on the selection of plans with a lower mean dose of three OARs (bladder, rectum and PBM) in the plans generated by an initial RP model. Our approach to training the final model M_2 is the first to date for RP model configuration using the plan data from RP model re-optimization plans and the clinically accepted plans.

Material and methods

Patient sample

This is a quality improvement study. 203 consecutive clinical VMAT plans for endometrial and cervical cancer treatment in the First Affiliated Hospital of Xiamen University from December 2019 to August 2021 were included in the training set, and another forty-six clinical VMAT plans were set to the test dataset. Patients with metallic femoral head prostheses were not included in our study. There were some patients with the PTV of the lymph nodes extending to the skin included. Since this was a retrospective review of patient data and replanning, we did not require formal research approvals. These clinical VMAT plans were generated by experienced medical physicists and reviewed by another experienced physicist and the chief physician before treatment. A CT scan was acquired for each patient in head first and supine position at a 5 mm slice spacing. Clinical target volume (CTV) was delineated in accordance with internationally accepted guidelines [19,20]. PTV was created by adding an isotropic 6 mm margin to the CTV and cropped 5 mm inside the body contour. The plans were optimized with VMAT and 6 MV photon beams generated by Varian linacs (TrueBeam or Unique) in the department. The prescribed dose for PTV was 50.4 Gy or 45 Gy in 1.8 Gy daily fractions, or 50 Gy or 46 Gy in 2.0 Gy daily fractions. The planning goal was to deliver 95% of the prescribed dose to 100% of the PTV and limit overdosage to 110% of the prescribed dose.

Model building and refinement

The 203 VMAT plans (P_0) in the training set were used to create an initial RP model (M_0) in the Eclipse version 15.6 (Varian Medical Systems). For model M_0, only statistical outlier OARs with Cook's Distance of more than 30 were removed from the training set, rather than deleting the whole plan [7,10,14,21]. Subsequently, plans in the whole training set were reoptimized using M_0, yielding 203 new plans (P_1). Three OARs were used to select the better plans relative to P_0 in P_1, to avoid observer-dependent evaluation preferences especially when the DVH lines were crossed, the mean doses were compared and the plan with lower MD was considered better [13,22]. The three OARs were bladder, rectum and PBM. Then, a refined RP model (M_1) was constructed by identifying 150 plans in P_1 with simultaneously lower MD of bladder, rectum and PBM. Another new RP model (M_2) was configured with all plans in model M_1 and the remaining 53 plans in P_1 (removing the OARs with an MD higher than the corresponding plans from P_0 but retaining the whole plan) and the corresponding plans from P_0 (only including OARs with better MD).

Model assessment

To compare the clinical performance of the three models, open-loop validation was performed on another consecutive 46 clinical VMAT plans that were not included in any model. Besides, to test whether the MD of the three OARs (bladder, rectum and PBM) of the 'worse' plans could be further improved by the refined models, a closed-loop validation was also carried out on the mentioned 53 cases using model M_0, M_1 and M_2. To ensure a fair comparison, the same beam configuration as the original clinical VMAT plans was implemented. A single optimization without human intervention was carried out to assess the quality of the models.

The dose metrics based on the cumulative DVH of the OARs and PTVs were evaluated. The V_{30Gy}, V_{45Gy} and MD were evaluated for the bladder, bowel and rectum. For the PBM, V_{10Gy}, V_{20Gy}, V_{30Gy}, V_{40Gy} and MD were recorded. The MD, V_{30Gy} and D_{1cm^3} were reported for the left and right femoral head (LFH and RFH). As for the PTV, the near maximum dose ($D_{2\%}$), near minimum dose ($D_{98\%}$), homogeneity index (HI) and 95% isodose conformity index ($CI_{95\%}$) were evaluated. Since the prescribed dose and fractionations were different in the validation sets, therefore the MD, D_{1cm^3}, $D_{2\%}$

and $D_{98\%}$ were expressed as the dose relative to the prescribed dose. The HI was defined as $HI=(D_{2\%}-D_{98\%})/D_{50\%}$, and the $CI_{95\%}$ was calculated according to the equation [23] $CI_{95\%}=V_{95\%}/V_{PTV}$.

Statistical analysis

The statistical significance of any difference between dose values created by the three models and the clinical VMAT plans was assessed using analysis of variance (ANOVA). Post-hoc Turkey tests were used to further analyze the difference between any two groups of the three models and the clinical VMAT plans. A p-value <0.05 was considered statistically significant.

Results

Mean doses were reduced for all OARs of the 53 plans re-optimized by M_0, M_1 and M_2 compared to plans in the clinical group (P_0), moreover, most OARs had lower dose metrics, except for V_{30Gy} of the rectum, V_{10Gy}, V_{30Gy} and V_{40Gy} of PBM (Tables 1 and 2 and Figure S1). In addition, the magnitude of the mean dose reduction increased monotonically in the sequence of M_0, M_1 and M_2 groups, with the exception of the MD of bowel in M_1 and MD of LFH in M_2 (Table 2). The 53 clinical VMAT plans and the plans re-optimized by M_0, M_1 and M_2 yielded comparable target coverage. There were no significant differences in $D_{2\%}$, $D_{98\%}$ and $CI_{95\%}$ between groups. However, the HI was lower in the groups of the RP model ($p<0.05$). According to the results of the post-hoc Turkey tests, the differences between the clinical group and M_0 group in the MD, V_{30Gy} and D_{1cm^3} of LFH and RFH and HI of PTV were statistically significant; the differences between the clinical group and M_1 group in the MD, V_{30Gy} and D_{1cm^3} of LFH and RFH, V_{30Gy} of the bladder, V_{30Gy} of PBM and HI of PTV were statistically significant; the differences between the clinical group and M_2 group in the MD, V_{30Gy} and D_{1cm^3} of LFH and RFH, MD and V_{30Gy} of bladder, V_{30Gy} of PBM and V_{30Gy} of rectum were statistically significant; and the difference between the M_0 group and M_2 group in V_{30Gy} of bladder was statistically significant ($p<0.05$, Table 1).

Similarly, relative to the clinical VMAT plans, mean doses of all OARs of the 46 open-loop cases re-optimized by M_0, M_1 and M_2 were reduced. In addition, all of the dose metrics of M_0, M_1 and M_2 groups were lower than the clinical group except for the V_{10Gy} of PBM (Table 3, Figure S2), the differences were statistically significant ($p<0.05$) for all dose metrics, with the exception of dose metrics of the bowel, V_{45Gy} of bladder and V_{40Gy} of PBM. Furthermore, the magnitude of the mean dose reduction increased monotonically in the sequence of M_0, M_1 and M_2 groups, with the exception of the bowel (Table 4). The 46 clinical VMAT plans and the plans re-optimized by M_0, M_1 and M_2 yielded comparable target coverage. The differences of $D_{2\%}$, $D_{98\%}$ and HI were not statistically significant between groups, but the $CI_{95\%}$ was larger in the clinical group ($p<0.05$, Table 3), indicating that the conformity of clinical plans was worse than that of the three RP models. According to the results of the post-hoc Turkey tests, the differences between the clinical group and M_0 group, the clinical group and M_1 group, the clinical group and M_2 group in the MD, V_{30Gy} and D_{1cm^3} of LFH and RFH,

Table 2. Dose differences for the mean dose of the OARs of the closed-loop validation cases (uncertainty expressed as 1 standard deviation).

	M_0-Clinical	M_1-Clinical	M_2-Clinical
Bladder (%)	−1.5 ± 6.1	−2.2 ± 6.3	−4.0 ± 6.2
Bowel (%)	−1.7 ± 2.6	−1.6 ± 2.5	−1.8 ± 2.5
LFH (%)	−4.3 ± 3.0	−5.9 ± 3.1	−5.9 ± 3.1
RFH (%)	−3.8 ± 3.2	−5.1 ± 3.3	−5.2 ± 3.2
Rectum (%)	−0.2 ± 6.6	−1.4 ± 6.7	−2.8 ± 6.7
PBM (%)	−0.0 ± 2.6	−0.3 ± 2.6	−0.8 ± 2.6

Table 1. Summary of the dosimetric metrics comparing the three RapidPlan models with the clinical treatment plans for the 53 closed-loop validation cases (uncertainty expressed as 1 standard deviation).

	Parameter	Clinical (Mean ± SD)	M_0 (Mean ± SD)	M_1 (Mean ± SD)	M_2 (Mean ± SD)	p-Value
PTV	$D_{2\%}$ (%)	106.1 ± 1.0	105.2 ± 0.6	105.3 ± 0.8	105.4 ± 2.6	0.072
	$D_{98\%}$ (%)	98.5 ± 0.5	98.7 ± 0.6	98.6 ± 0.7	98.5 ± 2.2	0.996
	HI	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.073
	$CI_{95\%}$	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	0.736
Bladder	MD (%)	83.9 ± 6.6	82.3 ± 3.8	81.7 ± 3.9	79.9 ± 4.1	<0.001
	V_{30Gy} (%)	83.2 ± 14.3	78.9 ± 6.1	77.2 ± 6.4	73.2 ± 6.6	<0.001
	V_{45Gy} (%)	45.5 ± 12.1	43.7 ± 11.2	43.5 ± 11.1	42.7 ± 10.9	0.595
Bowel	MD (%)	55.8 ± 8.7	54.2 ± 8.3	54.2 ± 8.4	54.0 ± 8.3	0.673
	V_{30Gy} (%)	42.8 ± 9.2	42.0 ± 8.7	42.0 ± 8.9	41.7 ± 8.8	0.912
	V_{45Gy} (%)	20.1 ± 7.2	19.7 ± 6.4	19.6 ± 6.3	19.4 ± 6.2	0.952
LFH	MD (%)	29.0 ± 4.7	24.7 ± 3.5	23.2 ± 3.1	23.2 ± 3.1	<0.001
	V_{30Gy} (%)	7.1 ± 4.7	4.1 ± 3.2	2.9 ± 3.0	3.0 ± 3.0	<0.001
	D_{1cm^3} (%)	77.4 ± 6.7	71.5 ± 6.1	68.2 ± 6.6	68.6 ± 6.3	<0.001
RFH	MD (%)	28.3 ± 4.3	24.5 ± 2.5	23.2 ± 2.5	23.0 ± 2.3	<0.001
	V_{30Gy} (%)	7.2 ± 4.5	4.3 ± 3.6	3.4 ± 3.0	3.3 ± 2.7	<0.001
	D_{1cm^3} (%)	77.1 ± 6.8	72.2 ± 6.1	69.7 ± 6.1	70.0 ± 6.0	<0.001
PBM	MD (%)	58.6 ± 3.0	58.6 ± 1.6	58.3 ± 1.6	57.7 ± 1.7	0.113
	V_{10Gy} (%)	85.3 ± 2.6	86.5 ± 2.7	85.5 ± 2.5	86.5 ± 2.8	0.094
	V_{20Gy} (%)	72.7 ± 5.3	71.5 ± 2.0	70.6 ± 2.0	69.9 ± 2.3	<0.001
	V_{30Gy} (%)	48.4 ± 8.5	48.7 ± 6.0	48.3 ± 5.8	46.8 ± 5.9	0.591
	V_{40Gy} (%)	21.1 ± 6.0	21.5 ± 6.3	21.9 ± 6.4	21.0 ± 6.0	1.0
Rectum	MD (%)	84.6 ± 6.9	84.4 ± 4.7	83.1 ± 5.0	81.8 ± 5.4	0.049
	V_{30Gy} (%)	85.1 ± 11.7	87.5 ± 5.6	85.7 ± 6.1	83.1 ± 7.2	0.597
	V_{45Gy} (%)	51.3 ± 21.7	46.3 ± 17.7	43.6 ± 17.5	42.5 ± 17.1	0.073

The reported p-values refer to the statistical significance between clinical group and M_2 group.

Table 3. Summary of the dosimetric metrics comparing the three RapidPlan models with the clinical treatment plans for the 46 open-loop validation cases (uncertainty expressed as 1 standard deviation).

	Parameter	Clinical (Mean ± SD)	M_0 (Mean ± SD)	M_1 (Mean ± SD)	M_2 (Mean ± SD)	*p*-Value
PTV	$D_{2\%}$ (%)	105.5 ± 1.2	105.4 ± 0.9	105.5 ± 1.2	105.4 ± 0.9	0.976
	$D_{98\%}$ (%)	99.0 ± 0.5	98.9 ± 0.5	99.0 ± 0.5	98.9 ± 0.5	0.859
	HI	0.6 ± 0.0	0.6 ± 0.0	0.6 ± 0.0	0.6 ± 0.0	0.225
	$CI_{95\%}$	1.3 ± 0.1	1.2 ± 0.1	1.2 ± 0.0	1.2 ± 0.0	<0.001
Bladder	MD (%)	86.2 ± 6.5	83.2 ± 3.7	83.0 ± 4.2	81.3 ± 4.6	<0.001
	V_{30Gy} (%)	86.1 ± 13.0	80.3 ± 5.3	79.4 ± 6.1	75.5 ± 6.7	<0.001
	V_{45Gy} (%)	51.6 ± 14.7	47.3 ± 9.8	47.5 ± 9.7	46.6 ± 9.8	0.136
Bowel	MD (%)	56.2 ± 9.7	54.4 ± 9.4	54.5 ± 9.4	54.2 ± 9.4	0.742
	V_{30Gy} (%)	43.5 ± 9.4	42.6 ± 10.2	42.8 ± 10.4	42.2 ± 10.3	0.930
	V_{45Gy} (%)	22.4 ± 9.6	21.1 ± 7.6	21.1 ± 7.7	21.0 ± 7.5	0.846
LFH	MD (%)	30.6 ± 3.7	25.1 ± 2.4	23.5 ± 2.2	23.5 ± 2.3	<0.001
	V_{30Gy} (%)	10.1 ± 4.6	4.8 ± 2.7	3.4 ± 2.3	3.4 ± 2.2	<0.001
	D_{1cm^3} (%)	82.4 ± 6.6	73.5 ± 6.3	70.2 ± 6.5	69.9 ± 6.1	<0.001
RFH	MD (%)	30.0 ± 4.3	24.4 ± 2.6	23.1 ± 2.4	22.9 ± 2.3	<0.001
	V_{30Gy} (%)	9.3 ± 4.5	4.7 ± 2.3	3.5 ± 2.0	3.4 ± 1.8	<0.001
	D_{1cm^3} (%)	81.0 ± 7.6	73.4 ± 5.7	70.5 ± 5.8	70.3 ± 5.8	<0.001
PBM	MD (%)	60.7 ± 3.6	58.9 ± 1.9	58.8 ± 1.9	58.2 ± 1.8	<0.001
	V_{10Gy} (%)	84.3 ± 2.3	87.2 ± 2.6	86.8 ± 3.0	87.8 ± 3.2	<0.001
	V_{20Gy} (%)	74.2 ± 3.9	72.0 ± 1.9	71.1 ± 2.0	70.5 ± 2.0	<0.001
	V_{30Gy} (%)	54.8 ± 9.7	49.7 ± 5.1	49.6 ± 5.1	48.0 ± 4.9	<0.001
	V_{40Gy} (%)	25.2 ± 7.2	23.1 ± 5.8	23.3 ± 5.7	22.4 ± 5.6	0.117
Rectum	MD (%)	88.8 ± 6.1	83.1 ± 5.4	81.9 ± 5.9	80.3 ± 6.2	<0.001
	V_{30Gy} (%)	92.1 ± 7.3	87.2 ± 5.0	85.2 ± 6.1	82.6 ± 7.2	<0.001
	V_{45Gy} (%)	60.2 ± 24.5	45.3 ± 19.3	43.4 ± 18.8	41.4 ± 18.5	<0.001

The reported *p*-values refer to the statistical significance between clinical group and M_2 group.

Table 4. Dose differences for the mean dose of the OARs of the open-loop validation cases (uncertainty expressed as 1 standard deviation).

	M_0-Clinical	M_1-Clinical	M_2-Clinical
Bladder (%)	−3.0 ± 5.5	−3.1 ± 5.9	−4.9 ± 6.0
Bowel (%)	−2.0 ± 2.7	−1.7 ± 2.8	−2.0 ± 2.7
LFH (%)	−5.5 ± 3.1	−7.1 ± 3.1	−7.1 ± 3.1
RFH (%)	−5.7 ± 3.6	−6.9 ± 3.7	−7.1 ± 3.5
Rectum (%)	−5.8 ± 5.3	−6.9 ± 5.5	−8.5 ± 5.6
PBM (%)	−1.8 ± 3.2	−1.9 ± 3.2	−2.5 ± 3.2

V_{30Gy} of the bladder, V_{10Gy}, V_{20Gy}, V_{30Gy} of PBM, V_{30Gy}, V_{45Gy} and MD of the rectum were statistically significant; the difference between the M_0 group and M_1 group in MD of LFH was statistically significant; the differences between the M_0 group and M_2 group in the MD of LFH, V_{30Gy} of the bladder, V_{20Gy} of PBM, and V_{30Gy} of the rectum were statistically significant ($p < 0.05$, Table 1).

The mean dose reduction of all OARs in the open-loop cases was larger than that in the closed-loop cases (Tables 2 and 4), indicating that the RP models are applicable not only to the closed-loop cases but also to the open-loop cases.

Figure 1(a) plots the mean DVHs for the 53 closed-loop validation cases, while Figure 1(b) demonstrates the mean DVHs for the 46 open-loop validation cases. Both plots compare the DVHs of the original VMAT plans and plans optimized by models M_0, M_1 and M_2. According to the plots, model M_2 outperformed model M_1, M_0 and the clinical plans in DVH of almost all OARs. Furthermore, the DVHs of PTVs largely overlap, suggesting comparable target coverage.

Discussion

In the current study, we investigated the performance of different knowledge-based RP models in closed-loop and open-loop validation by comparing the dose metrics of clinical VMAT plans and plans re-optimized by these models in a single optimization. Our study illustrates an efficient and practical method to refine the model training library and provides a model (M_2) that on average outperforms M_1, M_0 and clinical plans in terms of dosimetric outcomes. In addition, we found in this research that the RapidPlan model (M_0) generated using only the library of clinically accepted VMAT plans was able to produce VMAT plans that were comparable or better than the conventional trail-and-error clinical VMAT plans, consistent with previous studies [7,11]. And the plan quality of the validation cases can be further improved by model M_1, which is in line with previous publications [8,12,13]. Our approach to training the final model M_2 is the first to date for RP model configuration using the plan data from RP model re-optimization plans and the clinically accepted plans. In this method, M_1 re-optimization plans with lower MD of bladder, rectum and PBM, as well as other M_1 re-optimization plans (herein we call them the defective re-optimization plan) but excluding OARs with larger MD, and clinical plans corresponding to the defective re-optimization plans but only including OARs with lower MD were selected.

It's fundamental that the RP model represents broad geometric diversity and tackles the interpatient OAR variability. To achieve this, it would be beneficial to enlarge the number of initial training cases with broad anatomical variability. Although it is clinically desirable to enroll as many cases as possible in the model library to expand geometric diversity, in practice, it's technically impossible to include all anatomical variability. On one hand, our training library contains more than two hundred cases, which is the bedrock of providing plenty of anatomical variety. On the other hand, our solution better-protected OARs both from clinical plans and the re-optimized plans to construct the training set, which may well preserve population diversity and variability in dose

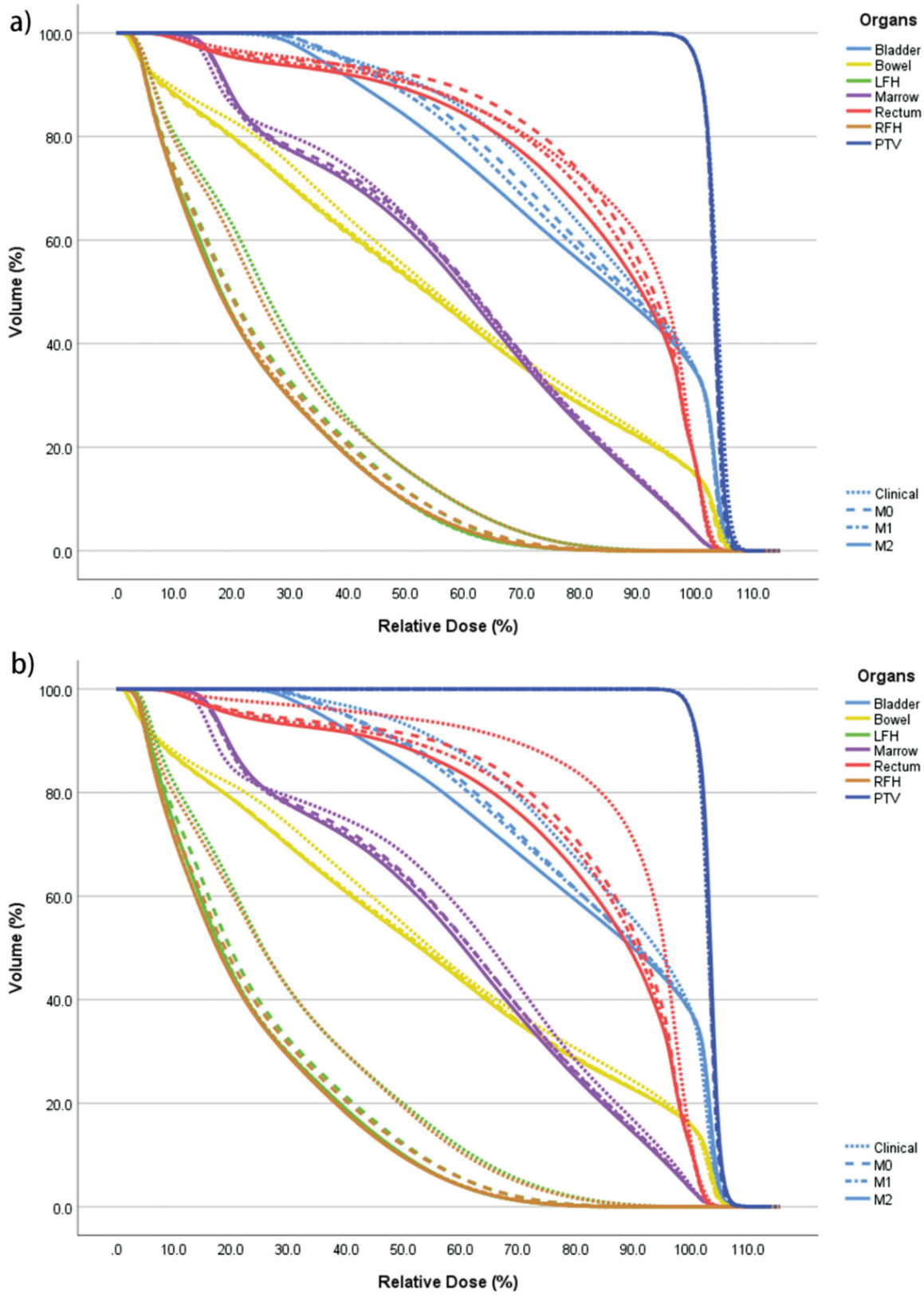

Figure 1. The average DVH plots of the closed-loop validation plans (a) and the open-loop validation plans (b). As shown in the legend, each OAR and PTV is represented by one color, and the dotted line is clinical plans, the dashed line is model M_0, the dot-dashed line is M_1 and the solid line is M_2.

distribution. Unlike other studies, in which poor plans were either removed or only qualified cases were selected [10,11].

In this study, model M_1 outperformed model M_0, suggesting that closed-loop re-optimization using the same training cases can improve suboptimal plans and eliminate inter-planner variation in training cases, even though the number of training cases of model M_1 is about three-quarters of that of model M_0. Consistent with the study by Fogliata *et al.*, they demonstrated

that an RP model could be improved by re-training it with a set of plans originated by itself in an iterative learning process [12], although the mean dose differences for the OARs they studied were modest in absolute terms. Recently, Chatterjee *et al.* replanned 48 cases in the training library with an intermediate model generated by DVH of the best-spared OARs of the initial model, and used these replan to configure the final model [8]. However, they didn't define the concept of best-spared OARs.

Ideally, planners try to pursue Pareto optimal plans, in which situation the additional protection of one OAR can only be achieved at the expense of other OARs in that plan. However, through the re-planning of model M_0, the mean dose of the three OARs can be reduced simultaneously in about three-quarters of the training cases. This indicates that, in everyday practice, the insufficient sparing of an OAR is not because of tradeoffs between OARs, but because of suboptimal optimization. According to the results (Tables 2 and 4), the RP models can reduce the mean doses of the OARs more in the open-loop validation set than in the closed-loop validation set, which is contrary to the results of Wang *et al.* [13]. It implies that the clinical plans in the closed-loop set are closer to the optimal status than those in the open-loop set, although, intuitively, the closed-loop cases that are part of the training set will result in better performance of the closed-loop validation set than the open-loop validation set. In fact, the closed-loop validation set was selected from cases where the mean doses of bladder, rectum and PBM could not be reduced simultaneously, making it difficult to further improve these plans.

In our study, we also investigated the influence of bladder volume on the improvement of the mean dose of the bladder in the three models. By performing linear regression analysis on the bladder volume and the improvement of the mean dose of the bladder in the three models, we found that there was no significant correlation between the bladder volume and the improvement of the mean dose of the bladder ($p = 0.242$ in the open-loop group, $p = 0.854$ in the closed-loop group). However, inconsistent with our results, Tang *et al.* [24] reported a linear correlation of D_{2cm^3} with the overlapping volume of bladder and PTV. They also found that the V_{40Gy} of the bladder was related to the overlapping volume of bladder and PTV, and the V_{40Gy} of the non-overlapping volume of bladder and PTV in the multiple linear regression analysis.

The first limitation of this study is that these RP models were configured for a single dose level prescription and a single treatment site. The practical techniques mentioned in our study need to be validated in future multiple dose level prescriptions and different treatment sites. A second limitation is that the training set in our research was selected directly from the clinically accepted plans, which resulted in inter-planner variabilities. In previous research, Wang *et al.* manually refined the clinical VMAT plans in the training library by experts to narrow the initial plan quality diversity [13]. The third limitation is that the training library, although containing more than two hundred cases, cannot exhaust all geometric diversity.

Conclusions

In summary, the RapidPlan models can generate appropriate DVH objectives for the optimization of VMAT plans. In patients with endometrial and cervical cancer, VMAT plans produced by the RP models through a single optimization without manual intervention had comparable target coverage and conformity and better OAR sparing to the clinically accepted plans. Our study also demonstrates that configuring RP models configured with plans generated by previous RP models can improve model performance. Finally, our practical approach of incorporating plan data of better sparing OARs from both the clinical group and the re-planned group can further improve the performance of the RP model.

Disclosure statement

The authors report no conflicts of interest.

ORCID

Zheng Kang http://orcid.org/0000-0002-1688-7799

Data availability statement

The data that support the findings of this study are openly available in [repository HARVARD Dataverse] at https://doi.org/10.7910/DVN/QZ8IEQ

Funding

This research did not receive any special grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Hanna TP, Shafiq J, Delaney GP, et al. The population benefit of radiotherapy for cervical cancer: local control and survival estimates for optimally utilized radiotherapy and chemoradiation [meta-analysis research support, Non-U.S. Gov't]. Radiother Oncol. 2015;114(3):389–394.

[2] Droge LH, von Sivers FF, Schirmer MA, et al. Conventional 3D conformal radiotherapy and volumetric modulated arc therapy for cervical cancer: comparison of clinical results with special consideration of the influence of patient- and treatment-related parameters [comparative study]. Strahlenther Onkol. 2021;197(6):520–527.

[3] Guo M, Huang E, Liu X, et al. Volumetric modulated arc therapy versus fixed-field intensity-modulated radiotherapy in radical irradiation for cervical cancer without lymphadenectasis: dosimetric and clinical results [comparative study]. Oncol Res Treat. 2018;41(3):105–109.

[4] Lin Y, Ouyang Y, Chen K, et al. Clinical outcomes of volumetric modulated arc therapy following intracavitary/interstitial brachytherapy in cervical cancer: a single institution retrospective experience. Front Oncol. 2019;9:760.

[5] Wang J, Hu W, Yang Z, et al. Is it possible for knowledge-based planning to improve intensity modulated radiation therapy plan quality for planners with different planning experiences in left-sided breast cancer patients? Radiat Oncol. 2017;12(1):85.

[6] Nelms BE, Robinson G, Markham J, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Pract Radiat Oncol. 2012;2(4):296–305.

[7] Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy [validation study. Radiother Oncol. 2016;120(3):473–479.
[8] Chatterjee A, Serban M, Faria S, et al. Novel knowledge-based treatment planning model for hypofractionated radiotherapy of prostate cancer patients. Phys Med. 2020;69:36–43.
[9] Alpuche Aviles JE, Cordero Marcos MI, Sasaki D, et al. Creation of knowledge-based planning models intended for large scale distribution: minimizing the effect of outlier plans. J Appl Clin Med Phys. 2018;19(3):215–226.
[10] Fusella M, Scaggion A, Pivato N, et al. Efficiently train and validate a RapidPlan model through APQM scoring [validation study]. Med Phys. 2018;45(6):2611–2619.
[11] Li N, Carmona R, Sirak I, et al. Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials [multicenter study validation study]. Int J Radiat Oncol Biol Phys. 2017;97(1):164–172.
[12] Fogliata A, Cozzi L, Reggiori G, et al. RapidPlan knowledge based planning: iterative learning process and model ability to steer planning strategies. Radiat Oncol. 2019;14(1):187.
[13] Wang M, Li S, Huang Y, et al. An interactive plan and model evolution method for knowledge-based pelvic VMAT planning. J Appl Clin Med Phys. 2018;19(5):491–498.
[14] Yu G, Li Y, Feng Z, et al. Knowledge-based IMRT planning for individual liver cancer patients using a novel specific model. Radiat Oncol. 2018;13(1):52.
[15] Yu S, Xu H, Sinclair A, et al. Dosimetric and planning efficiency comparison for lung SBRT: CyberKnife vs VMAT vs knowledge-based VMAT. Med Dosim. 2020;45(4):346–351. Winter
[16] Hussein M, Heijmen BJM, Verellen D, et al. Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations. Br J Radiol. 2018;91(1092):20180270.
[17] Chatterjee A, Serban M, Abdulkarim B, et al. Performance of knowledge-based radiation therapy planning for the glioblastoma disease site [validation study]. Int J Radiat Oncol Biol Phys. 2017;99(4):1021–1028.
[18] Tinoco M, Waga E, Tran K, et al. RapidPlan development of VMAT plans for cervical cancer patients in low- and middle-income countries [validation study]. Med Dosim. 2020;45(2):172–178.
[19] Small W Jr, Mell LK, Anderson P, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer [consensus development conference practice guideline]. Int J Radiat Oncol Biol Phys. 2008;71(2):428–434.
[20] Small W Jr, Bosch WR, Harkenrider MM, et al. NRG oncology/RTOG consensus guidelines for delineation of clinical target volume for intensity modulated pelvic radiation therapy in postoperative treatment of endometrial and cervical cancer: an update [research support, N.I.H., extramural]. Int J Radiat Oncol Biol Phys. 2021;109(2):413–424.
[21] Delaney AR, Tol JP, Dahele M, et al. Effect of dosimetric outliers on the performance of a commercial knowledge-based planning solution [research support, Non-U.S. Gov't]. Int J Radiat Oncol Biol Phys. 2016;94(3):469–477.
[22] Fan J, Wang J, Zhang Z, et al. Iterative dataset optimization in automated planning: implementation for breast and rectal cancer radiotherapy. Med Phys. 2017;44(6):2515–2531.
[23] Lomax NJ, Scheib SG. Quantifying the degree of conformity in radiosurgery treatment planning [comparative study]. Int J Radiat Oncol Biol Phys. 2003;55(5):1409–1419.
[24] Tang H, Chen Y, Jiang J, et al. Dose prediction models based on geometric and plan optimization parameter for adjuvant radiotherapy planning design in cervical cancer radiotherapy [research support, Non-U.S. Gov't]. J Healthcare Eng. 2021;2021:1–11.