Physica Medica 57 (2019) 231–237

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Comprehensive Intra-Institution stepping validation of knowledge-based models for automatic plan optimization

R. Castriconi[a,*], C. Fiorino[a], S. Broggi[a], C. Cozzarini[b], N. Di Muzio[b], R. Calandrino[a], G.M. Cattaneo[a]

[a] Medical Physics, San Raffaele Scientific Institute, Milano, Italy
[b] Radiotherapy, San Raffaele Scientific Institute, Milano, Italy

ARTICLE INFO

Keywords:
Knowledge-based
Automatic planning
Machine learning
Prostate radiotherapy

ABSTRACT

Purpose: To develop and apply a stepping approach for the validation of Knowledge-based (KB) models for planning optimization: the method was applied to the case of concomitant irradiation of pelvic nodes and prostate + seminal − vesicles bed irradiation in post-prostatectomy patients.

Methods: The clinical VMAT plans of 52 patients optimized by two reference planners were selected to generate a KB-model (RapidPlan, v.13.5 Varian). A stepping-validation approach was followed by comparing KB-generated plans (with and without planner-interaction, *RP* and *only-RP* respectively) against delivered clinical plans (*RA*). The validation followed three steps, gradually extending its generalization: 20 patients used to develop the model (closed-loop); 20 new patients, same planners (open-loop); 20 new patients, different planners (wide-loop). All plans were compared, in terms of relevant dose-volume parameters and generalized equivalent uniform dose (*gEUD*).

Results: KB-plans were generally better than or equivalent to clinical plans. For *RPvsRA*, PTVs coverage was comparable, for OARs *RP* was always better. Comparing *only-RPvsRA*, PTVs coverage was always better; bowel \bladder V_{50Gy} and $D_{1\%}$, bowel\bladder\rectum D_{mean}, femoral heads V_{40Gy} and penile bulb V_{50Gy} were significantly improved. For *RPvsRA gEUD* reduction > 1 Gy was seen in 80% of plans for rectum, bladder and bowel; for *only-RPvsRA,* this was found in 50% for rectum/bladder and in 70% for bowel.

Conclusion: An extensive stepping validation approach of KB-model for planning optimization showed better or equal performances of automatically generated KB-plan compared to clinical plans. The interaction of a planner further improved planning performances.

1. Introduction

External beam radiotherapy treatments require individually optimized planning. Despite the use of specific protocols, the final result of the optimization process is strongly planner dependent. Moreover, plan optimization is time consuming and needs several iterations before achieving a clinically acceptable plan. In order to reduce the inter-operator variability, spare the time for planning and, possibly, improve the quality of treatment plans, automatic planning systems were largely investigated and developed during recent years [1–5].

The application of machine-learning techniques to planning optimization recently lead to the development of the so called Knowledge-based (KB) optimization approach. The basic concept is quite simple: given a certain delivery/planning technique and treatment site, existing clinical treatment plans may be modelled to individually estimate the most likely dosimetric features expected in new patients, taking into account the individual anatomical and morphological characteristics of each individual patient. For this reason, differently from “automatic planning”, the performances of any KB approach is highly depending on the quality of the available plans managed during the modelling phase, being them representative of the center’s clinical experience. Different KB approach were developed and investigated in literature: Good et al. assessed the capability of a mutual information method based on a dataset of treated patient to predict the DVH [6]. Other studies investigated the feasibility of a principle component analysis (PCA) based model to estimate the expected organs-at-risk (OAR) sparing [7,8]. Nwankwo et al. developed a predicted model based on the geometric position relative to the target [9] while Wu et al. tested the potentiality of the overlap volume histogram (OVH) approach to estimate the best agreement between target coverage and OAR sparing [10].

* Corresponding author.
E-mail address: castriconi.roberta@hsr.it (R. Castriconi).

https://doi.org/10.1016/j.ejmp.2018.12.002
Received 16 July 2018; Received in revised form 23 November 2018; Accepted 5 December 2018
Available online 10 December 2018

Hence, once implemented and validated, a KB model predicts what may be obtained based on the previous experience. The same model may be used to create an optimization template potentially able of largely automatizing the planning phase and consequently reducing the time necessary to obtain a clinically acceptable plan. Moving forward, any KB-model may also be tuned to guide the optimization to obtain "better" plans with respect to "average" plans, also resulting in a net improvement in terms of reduction of inter-operator variability and plan quality.

RapidPlan is a commercially available KB planning tool (Varian Inc.), implemented into the Eclipse planning system as an optional module. Combining PCA and advanced regression techniques the tool generates an estimated DVH range suggesting where the DVH of a structure will most likely land; the plan may be automatically optimized building a template based on the KB individually optimized constraints. RapidPlan has quite widely been investigated in literature and validated in several clinical scenarios [11–14].

This investigation had two major aims:

1) To develop and validate a KB-model with an associated automatic planning template in the case of concomitant treatment of lymph nodes, seminal vesicles and prostatic bed with different dose levels: despite the feasibility of the Rapid Plan implementation to prostate treatment has already been shown [13,15,16], to our knowledge, it has never been reported for the current case.
2) We suggested (and applied to this case) a validation stepping procedure that implicitly assume reference planners. The methodology has the advantage to accomplish an overall Intra-Institution validation (internal and external) and to split the contribution of "other planners" against "new patients" on the performance of the model during automatic plan optimization.

2. Material and methods

2.1. Volumes, dose and planning

At our institute, pelvic nodes are usually treated for post-operative prostate patients in presence of clinical risk factors [17–20]. Our protocol involves 3 PTVs (prostatic bed PTV_2, seminal vesicles bed PTV_1 and pelvic lymph nodes PTV_{LN}) which are treated in two sequential phases: the first with a prescribed dose of 60 Gy to PTV_2 + PTV_1 concomitantly delivering 52.5 Gy to PTV_{LN} in 30 fractions; the second phase delivers 10/14 Gy to PTV_2 in 5/7 fractions (with adjuvant/salvage intent respectively).

CTV_2 included the prostate bed and encompassed the area behind the lower bladder, the anastomosis and the peri-urethral area, taking clip positioning and pathological specimen into account; CTV_1 included the seminal vesicles bed. The CTV referring to nodes (CTV_{LN}) included obturator, hypogastric, external and internal iliac and pre-sacral LN, anterior to the first sacral segment. PTVs were defined by adding a margin of 8/8/10 mm (lateral/superior-anterior/cranial-caudal) for CTV_1 and CTV_2; concerning CTV_{LN} a 7 mm margin was applied in all directions.

The organs at risk (OARs) were rectum, bladder, bowel (i.e.: defined as the intestinal cavity out of PTVs [18]), femoral heads and penile bulb; planning strategies in our clinical practice were previously discussed [17–20]. All plans were optimized by the Eclipse inverse planning (Varian Inc, v. 13.5) and delivered with 6MV X-Rays with a *CLINAC-IX 2300* equipped with a Millenium-MLC system, with the Rapid Arc (VMAT) technique. All VMAT plans consisted of two fully reverse arcs with a collimator inclination angle of 15° and normalized to the mean dose of the PTV_2.

2.2. KB-model

In order to reduce the inter-operator variability, all clinical plans optimized by the two planners responsible of the protocol at the time of the generation of the model were considered; in total, 52 patients (half plans for each planner) were available for training the model, a number generally considered as largely sufficient to provide a robust KB model in similar contexts [13,15]. We focused on the more demanding first phase, using the RapidPlan tool implemented in the Eclipse system (v.13.5). The training-model process is well explained in literature [21,22]: the RapidPlan performed the data extraction and the model training for each set of treatment plans. The result was a set of coefficients for each OARs providing an evaluation of the principal component scores of the DVH based on the geometric and anatomic characteristics of OARs and PTVs. The tuning of the model was performed by using the statistical tool of the system that permits to evaluate and eliminate potential outliers: in total from 2 to 10 patients were skipped for each OARs, considering outliers the OARs that exceeded by > 2SD the principal components fitted with the model.

2.3. KB-based optimization

The resulting DVH prediction model was set on the lower DVH population quartile for each OARs. Selected DVH constraints may be extracted from the KB-prediction model to generate an individually optimized template for plan optimization. Based on our experience, and after accomplishing several fine tuning tests by planning sample patients, the KB-based template for planning optimization, shown in Table 1, was finally generated and used for automatic optimization: according to our practice, we asked the system to obtain a highly homogenous dose distribution within PTVs; then, in order to control the maximum dose and to avoid hot spots in the bladder and the rectum we used an upper 'soft' constraint for these OARs equal to the prescribed dose. The Eclipse treatment planning system is able to manage this template with a good coverage of the PTVs and reduction of the hot spots both for PTV and OARs. In other words, once arcs are positioned, the operator may launch an automatic optimization, based on this template, individually tailored to the patient's features or decide to use

Table 1
Objective template as defined in the KB-model for automatic optimization.

Organs	Objectives	Volume (%)	Dose (Gy)	Priority
PTV_2	Upper	0	60	150
	Lower	100	60	150
PTV_1	Upper	0	60	150
	Lower	100	60	150
PTV_{LN}	Upper	2	53.5	130
	Lower	100	52.5	150
Bowel	Upper	Generated	50	120
	Upper	Generated	40	120
	Upper	Generated	30	100
	Upper	Generated	15	100
	Line	Generated	Generated	Generated
Bladder	Upper	Generated	60	120
	Upper	Generated	45	120
	Upper	Generated	30	120
	Upper	Generated	20	100
	Line	Generated	Generated	Generated
Rectum	Upper	Generated	60	130
	Upper	Generated	55	130
	Upper	Generated	40	120
	Upper	Generated	30	120
	Upper	Generated	20	100
	Line	Generated	Generated	Generated
Femoral heads	Upper	Generated	50	100
	Upper	Generated	40	80
	Line	Generated	Generated	Generated
Penile bulb	Upper	0	Generated	100
	Upper	Generated	30	80

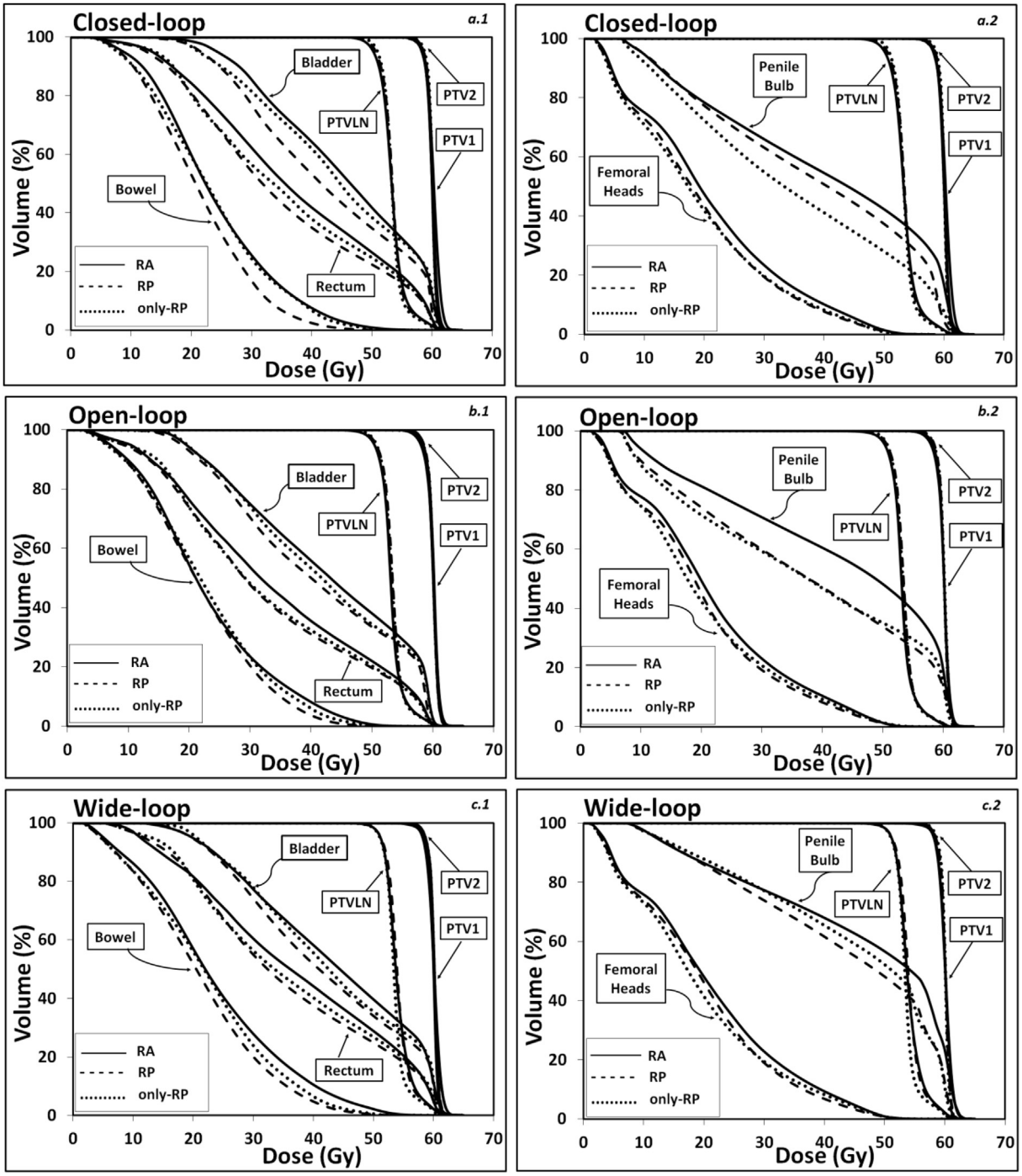

Fig. 1. Comparison of the mean-DVH of the original clinical plans (RA), re-optimized RapidPlan plus operator intervention plans (RP) and re-optimized fully automatic RapidPlan plans (only-RP) for each steps: closed-loop (a.1–2), open-loop (b.1–2) and wide-loop (c.1–2).

this template to start optimization and interactively act during the optimization process.

2.4. Stepping validation

After completing the model configuration, a step validation involving different patient sets was performed:

- Closed-loop validation: KB-replanning 20 randomly chosen plans used to generate the model; this step aimed to understand the robustness of KB-model to reproduce the same quality of the plans used to build it and then represents an internal validation step.
- Open-loop validation: KB-replanning 20 plans not used for training the model but previously planned by the same two operators (half plans per each planner); this external validation step aimed to test the capability of KB-model to produce high quality plans also for new patient cases.
- Wide-loop validation; KB-replanning 20 patients not used for the training model and planned by four different planners (5 plans each); aim of this external validation step was to understand the capability of KB-model to generalize the prediction to new patients as well as to produce high quality plans also for treatments previously planned by different operators.

The PTV volumes involved in the three validation-steps were similar without any statistically significant differences. The average values for each patient sets were: $V_{PTVLN} = 1148\,cm^3$, $V_{PTV1} = 129\,cm^3$, $V_{PTV2} = 120\,cm^3$ for closed-loop; $V_{PTVLN} = 1107\,cm^3$, $V_{PTV1} = 125\,cm^3$, $V_{PTV2} = 176\,cm^3$ for open-loop; $V_{PTVLN} = 1100\,cm^3$, $V_{PTV1} = 142\,cm^3$, $V_{PTV2} = 143\,cm^3$ for wide-loop.

KB-optimized plans were generated following two modalities:

- RapidPlan automatic optimization followed by additional operator intervention (*RP*): a single operator, different from all the considered operators performing the clinical plans, performed the optimization;
- fully automatic RapidPlan (*only-RP*), without interaction of the planner.

All KB-plans were compared, in terms of relevant dose-volume parameters, against the original clinical plans (*RA*). All *RA* plans were originally optimized by senior planners with similar (> 15 year) working experience.

2.5. Analysis

Average DVHs of PTVs and OARs for *RP* and only-*RP* were compared against the clinical plans (*RA*). Selected relevant dose statistics and dose-volume parameters were considered: two-tails paired t-tests were performed to assess statistically significant differences ($p < 0.05$). Concerning PTVs, $V_{95\%}$, D_{mean} and $D_{1\%}$ were considered; regarding OARs V_{40Gy}, V_{50Gy}, D_{mean} and $D_{1\%}$ were considered. Moreover, we also evaluated OARs differences in terms of generalized equivalent uniform dose (gEUD). Values of a (describing the volume effect) were chosen based on AAPM TG 166 [23] and recent reviews [20,24]: for rectum we set $a = 1$ (faecal incontinence) and $a = 8$ (bleeding); for bladder $a = 8$ and $a = 40$ (severe urinary toxicity, urinary incontinence); for bowel $a = 10$ (acute diarrhea); for femoral heads $a = 12$ (fracture).

The differences of the performances in the three steps were also compared by paired t-tests: in addition, the type of validation (internal vs external: closed vs open + wide) and the type of operator independently on the type of validation (closed + open vs wide) were also compared. The aim of these comparisons was to assess if the consideration of new patients or new planners has more impact on the performances of the KB-model compared to the clinical RA plans.

3. Results

3.1. DVH-based validation

In general, all KB-based plans were clinically acceptable in terms of PTVs coverage and OAR sparing. The PTV/OARs average values of relevant dose-volume parameters of the clinical plans included in the model are shown in Table s.1 (supplementary material). KB-based plans were generally better than or equivalent to clinical plans in all three steps.

Fig. 1 showed the mean-DVH of the original clinical plans (*RA*), re-optimized RapidPlan plus operator intervention plans (*RP*) and re-optimized fully automatic RapidPlan plans (*only-RP*) for each steps: closed-loop (a.1–2), open-loop (b.1–2) and wide-loop (c.1–2). Similar trends could be appreciated without major differences between the three steps. In Table 2 the difference values for PTVs and OARs parameters for the three-steps were reported. When comparing *RPvsRA*, PTVs coverage was comparable and the average differences were not statistically significant except for PTV_1 and PTV_{LN} $D_{1\%}$ parameters where *RP* was better for the closed-loop and the wide-loop. Concerning OARs, *RP* was always better with most of the improvements statistically significant. When comparing *only-RP* vs *RA*, PTVs coverage was always better with $V_{95\%}$ for PTV_{LN} statistically significant for the closed-loop and the open-loop. Moreover, a statistically significant improvement was found for PTVs $D_{1\%}$ for the closed-loop and the wide-loop. Concerning OARs, *only-RP* was always better, in particular bowel\bladder V_{50Gy} and $D_{1\%}$; bowel\bladder\rectum D_{mean}; femoral heads V_{40Gy} and penile bulb V_{50Gy} were significantly improved. Moreover, a reduction of the planning time was obtained: for only-RP the computation time was of about 15 min; for RP the planning time was of about 30 min (15 min for automatic optimization plus 15 min for planner interaction). Unfortunately the individual timing for planning was not previously recorded for RA: however, based on the experience, the time needed for clinically acceptable plan may be estimated to be on average around 90 min.

3.2. gEUD-based validation

The *gEUD* difference values were reported in Table 3 for the three steps. The average values for both KB-modalities were always better than *RA*, with most of the improvements statistically significant for *RP*.

The *gEUD* analysis for bowel ($a = 10$ acute diarrhea) was shown in Fig. 2: we reported the histogram distribution in terms of *gEUD* differences for the closed-loop (Fig. 2a.2), the open-loop (Fig. 2a.3) and the wide-loop (Fig. 2a.4). For *RP,* a > 1 Gy *gEUD* reduction was found in 80% of the cases with a maximum reduction up to 9 Gy. For *only-RP,* a > 1 Gy reduction was found in 70% of the cases; in particular for the wide-loop the *gEUD* average value was significantly improved (Table 3) with a reduction up to 9 Gy.

As well as for bowel toxicity, the same *gEUD* analysis is also shown for bladder ($a = 8$ severe urinary toxicity) and rectum ($a = 8$ bleeding) in Figs. 3 and 4 respectively. In both case, for *RP* a *gEUD* reduction was found in 80% of the cases with most of the improvements statistically significant. For *only-RPvs.RA,* the average differences in *gEUD* were significant only in few cases (Table 3): despite this, an improvement was found in 50% of patients for both rectum and bladder gEUD values.

Importantly, a > 1 Gy *gEUD* increase was seen in only 5–15% of the patients depending on the organ and method (*RP* or *only-RP*), with a maximum increase of 4 Gy. Moreover, the 'negative-difference' cases between RA and onlyRP never resulted in plan unacceptability. Of note, the worst-values for each OAR were close to mean values of the clinical plans (Table s.1).

3.3. Loop comparison

No significant differences in terms of comparison between *only-RP* plans and clinical *RA* plans were found, apart a significant impact of the planners (closed + open vs wide) for bowel sparing both in terms of DVHs and gEUD, with a significantly larger improvement of KB plans for the new planners with respect to the two original planners ($p < 0.05$). In fact, the results confirmed the observed bowel sparing in the wide-loop for *only-RPvsRA* (Fig. 1), as well as the gEUD reduction up to 3 Gy (Table 2). The results suggested a slightly larger impact of the planner compared to the type of validation, although only for bowel sparing.

4. Discussions

Scope of the investigation was to develop and apply a stepping validation procedure for KB-based optimization models, focusing on the complex case of the concomitant treatment of pelvic nodes and prostatic + seminal vesicles bed in post-prostatectomy patients. The complexity of geometric irradiation and the different dose level for the three PTVs makes this planning situation challenging, time consuming and prone to high inter-operator variability, despite the application of specific protocols. The study aimed to assess the capability of the KB-model to handle the geometric and dosimetric variations in pelvis configuration, following a step-validation process that implicitly assume reference planners (in current case, two planners responsible of

Table 2
Difference values, Δ*(RA-RP)* and Δ *(RA-onlyRP)*, for PTV and OAR parameters for the three-steps; in bold the values with statistically significant differences (*p*-value < 0.05).

Organs	Features	Closed-loop		Open-loop		Wide-loop	
		Δ*(RA-RP)*	Δ*(RA-onlyRP)*	Δ*(RA-RP)*	Δ*(RA-onlyRP)*	Δ*(RA-RP)*	Δ*(RA-onlyRP)*
PTV_2	$V_{95\%}$ (%)	−0.1 ± 1.1	−0.3 ± 1.2	−0.6 ± 1.5	−0.3 ± 1.3	−0.5 ± 1.8	−0.7 ± 1.6
	D_{mean} (Gy)	–0.01 ± 0.05	–0.05 ± 0.11	0.01 ± 0.16	0.01 ± 0.22	–0.02 ± 0.26	–0.03 ± 0.27
	$D_{1\%}$ (Gy)	0.1 ± 0.5	**0.2 ± 0.5**	0.1 ± 0.4	–0.09 ± 0.51	**0.6 ± 0.5**	**0.8 ± 0.6**
PTV_1	$V_{95\%}$ (%)	0.01 ± 0.5	−0.5 ± 1.2	**−1.1 ± 2.2**	−1.1 ± 2.3	−0.03 ± 0.62	−0.1 ± 0.6
	D_{mean} (Gy)	0.1 ± 0.3	**0.1 ± 0.2**	–0.03 ± 0.27	–0.05 ± 0.41	0.1 ± 0.2	0.1 ± 0.3
	$D_{1\%}$ (Gy)	**0.8 ± 0.7**	**0.9 ± 0.5**	0.1 ± 0.5	0.06 ± 0.62	**0.8 ± 0.4**	**0.8 ± 0.7**
PTV_{LN}	$V_{95\%}$ (%)	−1.3 ± 2.7	**−2.3 ± 3.3**	−0.8 ± 1.9	−1.2 ± 2.7	−0.08 ± 1.2	−0.1 ± 1.2
	D_{mean} (Gy)	–0.1 ± 0.5	–0.03 ± 0.38	–0.3 ± 0.6	–0.2 ± 0.5	–0.08 ± 0.71	**0.5 ± 0.6**
	$D_{1\%}$ (Gy)	0.1 ± 0.7	0.3 ± 0.7	–0.2 ± 0.7	0.07 ± 0.71	0.1 ± 1.1	0.2 ± 0.9
Bowel	$V_{40Gy\ (\%)}$	**5.1 ± 3.4**	0.8 ± 3.4	**4.4 ± 3.1**	**2.1 ± 2.3**	**5.8 ± 5.7**	**4.1 ± 6.4**
	$V_{50Gy\ (\%)}$	**0.8 ± 0.8**	**0.4 ± 0.7**	**0.5 ± 0.7**	0.2 ± 0.6	**1.8 ± 3.2**	**1.7 ± 3.4**
	$D_{mean\ (Gy)}$	**2.4 ± 1.3**	–0.7 ± 5.4	.1 ± 3.1	–0.9 ± 4.2	**2.5 ± 2.3**	**1.5 ± 2.4**
	$D_{1\%\ (Gy)}$	**5.1 ± 4.4**	0.8 ± 3.4	0 **3.7 ± 2.8**	**1.7 ± 2.3**	**4.7 ± 3.2**	**3.2 ± 4.6**
Bladder	$V_{40Gy\ (\%)}$	**10.9 ± 7.8**	2.3 ± 8.4	5.8 ± 9.6	2.5 ± 10	**6.5 ± 9.9**	2.6 ± 9.9
	$V_{50Gy\ (\%)}$	**6.9 ± 5.7**	**3.3 ± 5.8**	4.6 ± 7.6	3.7 ± 8.2	**5.7 ± 7.2**	**4.2 ± 8.1**
	$D_{mean\ (Gy)}$	**2.8 ± 2.2**	1.1 ± 2.1	1.7 ± 2.8	0.7 ± 3.7	**2.1 ± 3.6**	**1.1 ± 3.2**
	$D_{1\%\ (Gy)}$	**0.8 ± 0.3**	**0.4 ± 0.5**	**0.5 ± 0.6**	0.04 ± 0.51	**0.5 ± 0.8**	0.8 ± 1.7
Rectum	$V_{40Gy\ (\%)}$	**6.8 ± 7.1**	**3.8 ± 6.1**	**3.3 ± 7.8**	3.1 ± 8.2	**5.3 ± 5.6**	**3.9 ± 6.1**
	$V_{50Gy\ (\%)}$	**4.3 ± 4.2**	1.7 ± 4.3	**2.1 ± 5.2**	1.5 ± 4.4	**3.7 ± 3.1**	2.1 ± 4.5
	$D_{mean\ (Gy)}$	**2.9 ± 2.8**	**1.8 ± 3.4**	**1.5 ± 2.1**	1.2 ± 3.1	**1.9 ± 2.7**	1.2 ± 3.1
	$D_{1\%\ (Gy)}$	0.1 ± 0.5	0.1 ± 0.5	**0.2 ± 03**	**0.3 ± 0.4**	0.1 ± 0.9	0.1 ± 1.1
Femoral Heads	$V_{40Gy\ (\%)}$	**2.4 ± 2.2**	**1.8 ± 1.9**	**2.2 ± 3.4**	1.2 ± 2.6	**2.5 ± 2.1**	**1.1 ± 1.3**
	$V_{50Gy\ (\%)}$	0.3 ± 1.2	**0.6 ± 1.1**	0.1 ± 1.2	0.1 ± 1.1	0.01 ± 1.04	0.03 ± 0.92
	$D_{mean\ (Gy)}$	**1.3 ± 1.4**	**1.4 ± 1.6**	**1.3 ± 1.8**	**–1.4 ± 1.7**	**1.2 ± 1.7**	**1.6 ± 1.3**
	$D_{1\%\ (Gy)}$	0.3 ± 2.2	0.6 ± 1.5	0.6 ± 2.3	–0.5 ± 2.3	0.6 ± 2.3	0.03 ± 1.8
Penile Bulb	$V_{40Gy\ (\%)}$	**3.8 ± 6.1**	6.2 ± 6.9	**13 ± 10**	**11 ± 10**	6.3 ± 9.9	12 ± 10
	$V_{50Gy\ (\%)}$	**4.9 ± 5.1**	**7.5 ± 7.1**	**14 ± 10**	**11 ± 10**	**8.5 ± 10**	14 ± 10
	$D_{mean\ (Gy)}$	**1.7 ± 3.2**	2.4 ± 3.1	**5.4 ± 5.3**	6.1 ± 6.8	2.4 ± 5.1	2.6 ± 4.9
	$D_{1\%\ (Gy)}$	**1.1 ± 1.7**	0.8 ± 1.8	2.2 ± 4.4	4.1 ± 6.8	**0.7 ± 0.9**	1.2 ± 1.2

Table 3
Difference values, Δ *(RA-RP)* and Δ *(RA-onlyRP),* in terms of gEUD values, for the three steps; in bold the values with statistically significant differences (*p*-value < 0.05).

Organs	Features	Closed-loop		Open-loop		Wide-loop	
		Δ*(RA-RP) (Gy)*	Δ*(RA-onlyRP) (Gy)*	Δ*(RA-RP) (Gy)*	Δ*(RA-onlyRP) (Gy)*	Δ*(RA-RP) (Gy)*	Δ*(RA-onlyRP) (Gy)*
Bowel	*a = 10*	**3.8 ± 2.7**	0.2 ± 2.4	**2.6 ± 1.9**	**1.1 ± 1.5**	**3.7 ± 2.5**	**2.7 ± 3.3**
Bladder	*a = 8*	**1.2 ± 0.7**	0.1 ± 0.8	**0.7 ± 1.1**	0.2 ± 1.1	**1.1 ± 1.2**	**0.7 ± 1.4**
	a = 40	**0.5 ± 0.4**	0.1 ± 0.5	**0.3 ± 0.5**	0.02 ± 0.42	**0.4 ± 0.5**	0.3 ± 0.6
Femoral Heads	*a = 12*	**0.9 ± 1.7**	**0.8 ± 1.2**	0.8 ± 1.7	0.02 ± 1.4	**0.9 ± 1.9**	0.3 ± 1.4
Rectum	*a = 1*	**2.9 ± 2.8**	**1.8 ± 3.4**	**1.4 ± 2.1**	1.1 ± 2.9	**1.9 ± 2.7**	1.3 ± 3.1
	a = 8	**1.1 ± 0.8**	0.3 ± 1.1	**1.1 ± 1.9**	0.2 ± 1.7	**1.2 ± 1.1**	**0.7 ± 1.5**

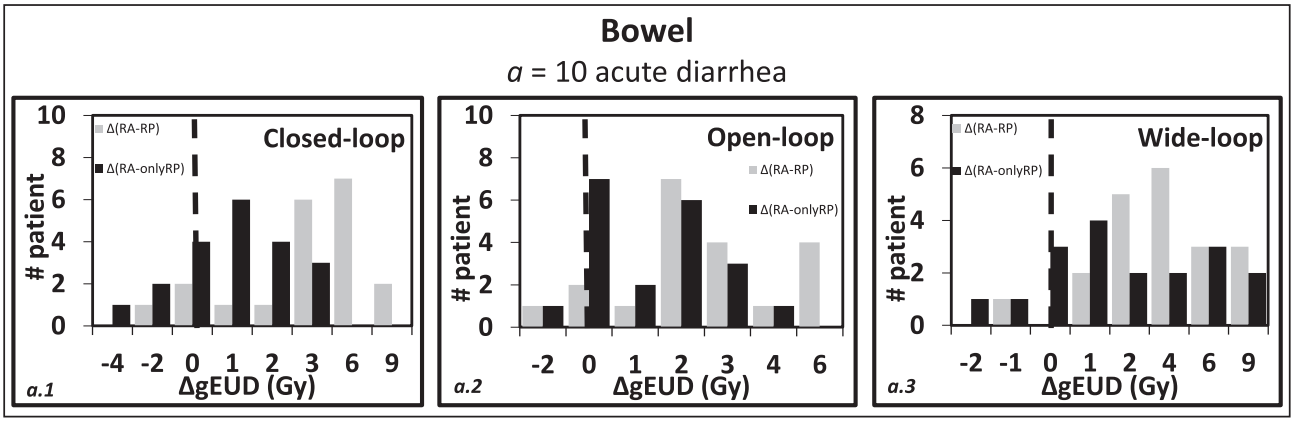

Fig. 2. The *gEUD* analysis for bowel (*a* = 10 acute diarrhea): the histogram distribution in terms of *gEUD* differences for closed-loop (*a.1*), open-loop (*a.2*) and wide-loop (*a.3*).

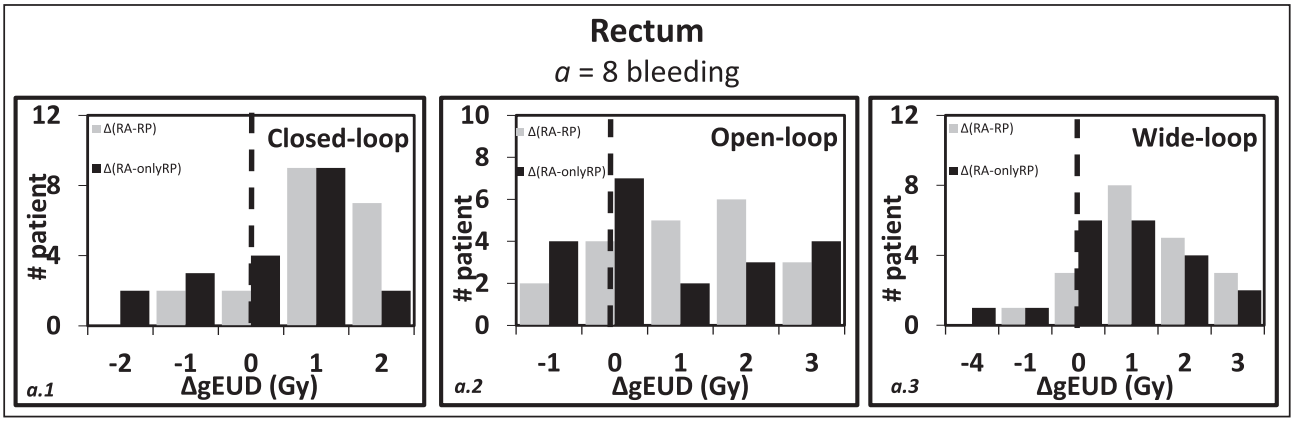

Fig. 3. The *gEUD* analysis for bladder (a = 8 severe urinary toxicity): the histogram distribution in terms of *gEUD* differences for closed-loop (*a.1*), open-loop (*a.2*) and wide-loop (*a.3*).

the clinical protocol) and additional planners inside the same Institute. The methodology has the advantage to give an overall validation (internal and external) and to split the contribution of "new patients" (open-loop) against "new planners" (wide-loop) on the performance of the model in automatic plan optimization.

In our investigation, the KB-model was found to be able to generate high-quality plans, at least comparable to previously optimized clinical plans, without any interaction with the planner (*only-RP* plans). Furthermore, an interaction of a planner showed additional potentialities to further improve planning performances (*RP* plans). Similar trends were found without major differences between the three steps (closed, open and wide loops). This step process assessed the ability of KB-model to generate high quality plans for new patient cases (open-loop validation) indirectly confirming the possibility of improving plan homogeneity between operators. Interestingly, the KB-model was found to produce high quality plans also for plans previously optimized by operators external to the model (wide-loop validation). In order to quantify the performances of the KB-model on new patients and/or new planners, the resulting plans were compared against the clinical ones grouping them according to the type of validation (internal vs external: closed vs open + wide) and to the type of operator (closed + open vs wide). Only the bowel dataset showed a statistically significant sparing, both in terms of *DVHs* and *gEUD*, between the closed + open and wide-loop. The results suggested a slightly larger impact of the planner compared to the type of validation, confirming the importance of a robust validation and the use of a dataset as heterogeneous as possible.

Scaggion et al [25] demonstrated the reduction of intra and inter planner variability in prostate cancer treatment planning by the assistance of RapidPlan during the optimization, consistent with our results in terms of *only-RP* quality, also for plans previously optimized by operators external to the model (wide-loop validation). The impact of the simultaneous delivery of multiple dose levels was investigated for head and neck treatment [11,12] and for prostate site [13], demonstrating the effectiveness of RapidPlan tool as well. Moreover, Hussein et al [13] mentioned the importance of the fine-tuning during model training in particular to manually set the optimization objectives. In agreement with our practice, in our KB-model the resulting DVH prediction model was set on the lower DVH population quartile for each OARs. Fogliata et al [11] mentioned the importance of a wide validation for each KB-model in an open-loop setting, in order to cover a large variety of patients, more heterogeneous than the data used for training the model.

To our knowledge, although different ways for validating KB-model have been reported and applied, a formalized step-validation approach as the one here suggested has never been reported in literature; the Intra-Institutional step-validation process is here suggested for the first time also aiming to gradually extend the validation of any KB-model to include new patients (outside the ones used for the model) and new planners and gradually increasing the confidence in using KB-based optimization. The method wished to contribute in defining a rational approach to Intra-Institution validation. Of course, we are aware that the novelty of the suggested approach is limited as several groups reported validation approaches that have some similarity with ours; however, we believe that the clear standardization of the followed phases may help users, especially in keeping attention to the inter-planner impact (and its minimization) in generating KB-models.

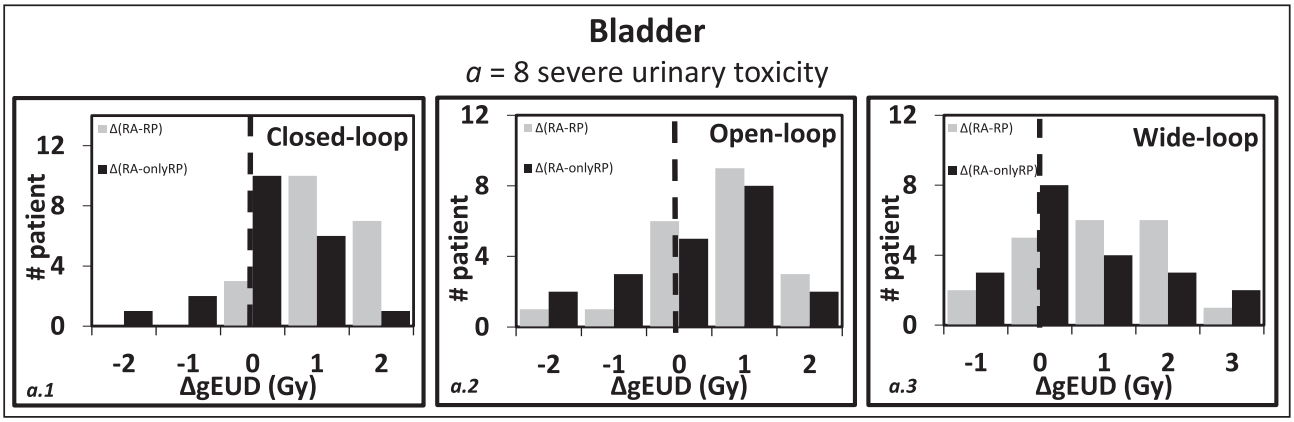

Fig. 4. The *gEUD* analysis for rectum (a = 8 bleeding): the histogram distribution in terms of *gEUD* differences for closed-loop (*a.1*), open-loop (*a.2*) and wide-loop (*a.3*).

A relevant and debatable point that deserves to be discussed is the extension of institutional KB models to other Institutions, as well as the extension of KB models to other treatment modality. Cagni et al [26] demonstrated the feasibility of training the RapidPlan model with plans of a different treatment modality. On the other hand, Berry et al [27] investigated the possibility of using a knowledge-based planning in order to identify systematic variation in esophagus IMRT plans between multiple campuses of a single institution. Their results showed KB approach can highlight differences in planning practices and can be used to increase consistency and quality among the campus. As well as, Schubert et al [15] demonstrated the capability of the RapidPlan tool in a multicentric cooperative initiative for high risk prostate cancer patients; in their case the pelvic volume (including the prostate, seminal vesicles and partially the pelvic nodes) was treated with a homogenous dose level. Their results demonstrated that the use of a KB-model allowed, on average, an improvement in D_{mean} and $D_{1\%}$ of the organs at risk compared to the routine clinical practice although very large variations were found between different Institutions. Anyway, the multicentric validation demonstrated the possibility of sharing models among different institutes highlighting the importance of an accurate validation for KB-models. Another important point regarded the impact of the dimension of the training set on the robustness of the KB-model prediction. In fact, the number of patients used for training is a critical issue for the resulting quality of KB-models. Cagni et al [28] demonstrated the dependence of the DVH-prediction performance on the size of the training patients in the model. This study showed a more accurate DVH-prediction with training sets ≥ 45 plans. This result was in agreement with our KB-model, consisting of 52-training plans. In addition, the choice to select plans optimized by only two (reference) planners should have further improved the quality and robustness of the model.

The implementation of a model outside the Institution(s) where it has been generated requires a careful evaluation, often adaptation to the local situation, and extensive validation; also in this context, the quality of the Institutional validation of a KB model, as the one we implemented, may help in increasing the confidence in the model itself and facilitating its acceptance outside the Institution. However, this point is controversial and deserves more attention in future studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2018.12.002.

References

[1] Zhang X, Li X, Quan E, Pan X, Li Y. A methodology for automatic intensity-modulated radiation treatment planning for lung cancer. Phys Med Biol 2011;56:3873–93.

[2] Quan E, Chang J, Liao Z, Xia T, Yuan Z, Liu H, et al. Automated volumetric modulated arc therapy treatment planning for stage III lung cancer: how does it compare with intensity-modulated radio therapy? Int J Radiat Oncol Biol Phys 2012;84:e69–76.

[3] Voet PWJ, Dirkx MLP, Breedveld S, Al-Mamgani A, Incrocci L, et al. Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. Int J Radiat Oncol Biol Phys 2014;88:1175–9.

[4] Sharfo AWM, Breedveld S, Voet PWJ, Heijkoop TS, Mens JWM, Hoogeman MS, et al. Validation of fully automated VMAT plan generation for library-based plan-of-the-day cervical cancer radiotherapy. PLoS One 2016. https://doi.org/10.1371/journal.pone.0169202.

[5] Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Georg GGD, Breedveld S, et al. Automated volumetric modulated arc therapy planning for whole pelvic prostate radiotherapy. Strahlenther Onkol 2017. https://doi.org/10.1007/s00066-017-1246-2.

[6] Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: an example application to prostate cancer planning. Int J Radiat Oncol 2013;87:176–81.

[7] Zhu X, Ge Y, Li T, Thongphiew D, Yin F-F, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys 2011;38(2):719–26.

[8] Yuan L, Ge Y, Lee WR, Yin FF, Kirkpatrick JP, Wu QJ. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. Med Phys 2012;39(11):6868–78.

[9] Nwankwo O, Mekdash H, Sihono DSK, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. Radiat Oncol 2015;10:111.

[10] Wu B, Ricchetti F, Sanguinetti G, Kazhdan M, SImari P, Jacques R, et al. Data-Driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning. Int J Radiat Oncol Biol Phys 2011;79:1241–7.

[11] Fogliata A, Wang PM, Belosi F, Clivio A, Nicolini G, Vanetti E. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiat Oncol 2014;9:236.

[12] Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol 2015;91:612–20.

[13] Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiat Oncol 2016;120:473–9.

[14] Fogliata A, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, et al. RapidPlan head and neck model: the objectives and possible clinical benefit. Radiother Oncol 2017;12:73.

[15] Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. PLoS One 2017;12(5). https://doi.org/10.1371/journal.pone.0178034.

[16] Kubo K, Monzen H, Ishii K, Tamura M, Kawamorita R, Sumida I, et al. Dosimetric comparison of RapidPlan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. Phys Med 2017;199:204–44.

[17] Alongi F, Fiorino C, Cozzarini C, Broggi S, Perna L, Cattaneo GM, et al. IMRT significantly reduces acute toxicity of whole-pelvis irradiation in patients treated with post-operative adjuvant or salvage radiotherapy after radical prostatectomy. Radiother Oncol 2009;93:207–12.

[18] Fiorino C, Alongi F, Perna L, et al. Dose-volume relationship for acute bowel toxicity in patients treated to pelvic nodal irradiation for prostate cancer. Int J Radiat Oncol Biol Phys 2009;75:29–35.

[19] Sini C, Chiorda BN, Gabriele P, Sanguineti G, Morlino S, Badenchini F, et al. Patient-reported intestinal toxicity from whole pelvis intensity-modulated radiotherapy: first quantification of bowel dose–volume effects. Radiother Oncol 2017;124:296–301.

[20] Fiorino C, Valdagni R, Rancati T, Sanguineti G. Dose–volume effects for normal tissues in external radiotherapy: pelvis. Radiother Oncol 2009;93:153–67.

[21] Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the pre-clinical validation of a commercial model-based optimisation engine: application to volumetric modulated arc therapy for patients with lung or prostate cancer. Radiother Oncol 2014;113:385–91.

[22] Eclipse Photon and Electron reference guide v.13.7, Varian Medical System, June 2015, https://www.myvarian.com.

[23] Li XA, Albert M, Deasy JO, Jackson A, Jee KWK, Marks LB, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM(a). Med Phys 2012;39:1386–409.

[24] Landoni V, Fiorino C, Cozzarini C, Sanguinetti G, Valdagni R, Rancati T. Predicting toxicity in radiotherapy for prostate cancer. Phys Med 2016;32:521–32.

[25] Scaggion A, Fusella M, Roggio A, Bacco S, Pivato N, Rossato MA, Paiusco M. Reducing inter and intra planner variability in radiotherpay plan output with a commerical knoweldge-based planning solution. Phys Med 2018;53:86–93.

[26] Cagni E, Botti A, Micera R, Galeandro M, Sghedoni R, Orlandi M. Knowledge-based tratment planning: an inter-technique and inter-system feasibility study for prostate canecer. Phys Med 2017;36:38–45.

[27] Berry SL, Ma R, Boczkowski A, Jackson A, Zhang P, Hunt M. Evaluating inter-campus plan consistency using a knowledge based planning model. Radiother Oncol 2016;120:349–55.

[28] Cagni E, Botti A, Wang Y, Iori M, Petit SF, Heijmen BJM. Pareto-optimal plans as gound truth for validation of a commercial system for knowledge-based DVH-prediction. Phys Med 2018;55:98–106.