Physica Medica 53 (2018) 86–93

Contents lists available at ScienceDirect

Physica Medica

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

Original paper

# Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution

Alessandro Scaggion[a,1], Marco Fusella[a,1,*], Antonella Roggio[a], Simonetta Bacco[a], Nicola Pivato[a], Marco Andrea Rossato[a], Lucia Mariel Arana Peña[a,b], Marta Paiusco[a]

[a] Medical Physics Department, Veneto Institute of Oncology IOV-IRCCS, via Gattamelata 64, 35128 Padova, Italy
[b] The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

ARTICLE INFO

*Keywords:*
VMAT
Knowledge-based planning
Planning quality
Consistency
Planning experience

ABSTRACT

*Purpose:* This study measured to which extent RapidPlan can drive a reduction of the human-caused variability in prostate cancer treatment planning.
*Methods:* Seventy clinical prostate plans were used to train a RapidPlan model. Seven planners, with different levels of planning experience, were asked to plan a VMAT treatment for fifteen prostate cancer patients with and without RapidPlan assistance. The plans were compared on the basis of target coverage, conformance and OAR sparing. Inter-planner and intra-planner variability were assessed on the basis of the Plan Quality Metric formalism. Differences in mean values and InterQuartile Ranges between patients and operators were assessed.
*Results:* RapidPlan-assisted plans matched manual planning in terms of target coverage, homogeneity, conformance and bladder sparing but outperformed it for rectum and femoral heads sparing. 8 out of 15 patients showed a statistically significant increase in overall quality. Inter-planner variability is reduced in RapidPlan-assisted planning for rectum and femoral heads while bladder variability was constant. The inter-planner variability of the overall plan quality, IQR of PQM%, was approximately halved for all patients. RapidPlan assistance induced a larger increase in plan quality for less experienced planners. At the same time, a reduction in intra-planner variability is measured with a significant overall reduction.
*Conclusions:* The assistance of RapidPlan during the optimization of treatments for prostate cancer induces a significant increase of plan quality and a contextual reduction of plan variability. RapidPlan is proven to be a valuable tool to leverage the planning skills of less experienced planners ensuring a better homogeneity of treatment plan quality.

## 1. Introduction

Large variations of radiotherapy treatment quality have been observed between institutions [1–3] or among planners [4–7], and many authors reported the need for a study focused on its accurate quantification [1,4,3,6,8].

The operator's experience has been indicated as the main cause of this variability [4–6] and the difficulty of the planner to a priori asses the attainable tradeoff between the PTV coverage and OAR sparing has been also shown to contribute [1,6,9,10]. Knowledge Based Planning (KBP) have been suggested as a solution to reduce this variation [1].

KBP systems has been developed as a machine learning process designed to assist the human planner in the effort to efficiently achieve an optimal dose distribution [11]. KBP have been also employed as plan quality assurance tool [12,13], to prevent the poor clinical outcomes correlated with sub-optimal plans [14,15], and as a knowledge sharing tool to facilitate planner learning curve [1,5,9,16,17].

The recent investigations about the capability of KBP systems to reduce the human-caused variability are affected by some limitations. Cross-institutional comparisons have been performed on large databases without a common cohort of patients [1,3,10], treatment of a single patient has been planned by many planners [4,5] or many plans have been administered to a single experienced operator [18,19].

This study present the attempt to overcome some of these

* Corresponding author.
*E-mail addresses:* alessandro.scaggion@iov.veneto.it (A. Scaggion), marco.fusella@iov.veneto.it (M. Fusella), antonella.rogggio@iov.veneto.it (A. Roggio), simonetta.bacco@iov.veneto.it (S. Bacco), nicola.pivato@iov.veneto.it (N. Pivato), marcoandrea.rossato@iov.veneto.it (M.A. Rossato), larana_p@ictp.it (L.M.A. Peña), marta.paiusco@iov.veneto.it (M. Paiusco).
[1] Alessandro Scaggion and Marco Fusella equally contributed to this work and should be considered as co-first authors.

https://doi.org/10.1016/j.ejmp.2018.08.016
Received 29 December 2017; Received in revised form 1 August 2018; Accepted 15 August 2018

limitations using a robust and systematic approach to correctly quantify the impact of the RapidPlan KBP system (Varian Medical Systems, Palo Alto, CA) on inter- and intra-planner variability. Seven planners with different levels of clinical planning experience were asked to plan a VMAT treatment for the same cohort of prostate patients with and without the support of RapidPlan. To address the problem more effectively, instead of the troublesome and clinically questionable analysis of the average DVHs, the PQM formalism has been employed. This novel measure is gaining attention in the community and allowed to assess whether RapidPlan assistance affected differently the performances of planners with different degrees of experience.

## 2. Materials and methods

### 2.1. Planners

Six planners in our department consisting of physicists and dosimetrists with different degrees of expertise in RT planning were involved in the study. They were ranked by the total number of planned VMAT treatments (from 100 to 700). An internship student, without prior experience in IMRT or VMAT planning, was also involved to fully investigate the benefits of RapidPlan in an educational pathway.

### 2.2. Patient selection

For this study we have chosen eighty-five patients treated for radical prostate cancer, between 2016 and 2017 at our institute. PTV was obtained adding to a GTV, the prostate gland, a posterior margin of 5 mm and 7 mm margin in all the other directions. Rectum, bladder and femoral heads were delineated as OARs. The contouring procedure was undertaken by two dedicated radiation oncologists.

All the patients were treated with Volumetric Modulated Arc Therapy using 1 or 2 full arcs and 6-MV photons delivered with a Millennium 120 MLC based on Varian Unique linac. The treatments were planned with Eclipse and Progressive Resolution Optimizer (PRO) v. 11 to deliver 78 Gy or 70 Gy (PTV) over 39 or 28 fractions [20,21].

The planning goals were to fully cover the PTV with 95% of the prescribed dose limiting the overdosage to 110% of the prescribed dose. All plans were optimized according to our department prostate radical treatment protocol which is based on RTOG 0126 (see Table 2).

### 2.3. Model configuration and validation

Data from seventy patients were imported in Eclipse v.13.7 and used to train and validate the RapidPlan model. The model was configured following the recommendation of the Varian operator's manual and suggestions from the literature [22–25]. Any outlier identified by RapidPlan was carefully re-checked and eventually replanned. The model was validated through a closed- and open-loop process as proposed in the literature [23–26]. The details of the process are given in a previous publication form the same group [27].

The RapidPlan model was configured with the list of objectives given in Table 1.

### 2.4. Planning protocol

The remaining fifteen patients from the initial group, all treated with a prescription of 78 Gy, were used to conduct the prospective study in two subsequent phases. First, during routine clinical activity, data from each patient were copied, renamed and distributed to every planner as a clinical treatment to be optimized following a standard manual approach (manual planning). After the introduction of RapidPlan, the entire patient sample was anonymized with univocal IDs and administered to the planners to be optimized with the assistance of RapidPlan (RapidPlan assisted planning). This strategy was adopted to minimize possible bias due to planners' memory.

**Table 1**
Summary of the optimization objective used in RapidPlan-assisted planning. The *gen.* indicates those values generated by RapidPlan on the basis of the prostate model. $D_{presc}$ indicates the prescription dose.

| ROI | Optimization Objective | | | |
|---|---|---|---|---|
| | Objective Type | D [Gy] | V [%] | Weight |
| PTV | Lower | 77.22 | 100 | 130 |
| | Upper | 79.56 | 0 | 120 |
| Rectum | Upper | *gen.* | 0, 10, 30, 50, 80 | *gen.* |
| Bladder | Upper | *gen.* | 0, 10, 30, 60 | *gen.* |
| Femur L | Upper | *gen.* | 0, 50 | *gen.* |
| Femur R | Upper | *gen.* | 0, 50 | *gen.* |
| Body | Normal Tissue Objective | DistanceFromTargetBorder = 0.2 cm<br>StartDose = 100<br>EndDose = 50<br>FallOff = 0.2 $cm^{-1}$ | | 100 |

All plans were created to be delivered with the same Linac respecting the original clinical set-up. All the planners were forced to maintain the isocenters identified during the CT-simulation. In both the planning procedures planners were left free to set the treatment geometry: one or two full arcs and an arbitrary collimator angle.

During the manual planning phase operators were free to set DV optimization constraints and draw ghost structures for dose containment. Conversely, during the RapidPlan assisted planning, they were provided with the DVH predictions given by RapidPlan and were limited to use and modify, but not delete, the set of predefined optimization objectives generated by the RapidPlan model. In addition, planners were not allowed to draw ghost structures to support the optimization. This method allows to make full use of RapidPlan capability which inherently takes into account the relative geometrical relationship when predicting the DVH curves. All plans were normalized to cover the 100% of the PTV with 76.44 Gy (95% of the prescription dose) in 39 fractions.

### 2.5. Plan evaluation

The dosimetric features of *manual* and *RapidPlan assisted* plans were compared on the basis of DVH metrics based on RTOG 0126 and complementary low-dose DVH points. In detail: 1. The near minimum dose ($D_{98\%}$), the near maximum dose ($D_{2\%}$), the Homogeneity index $[(D_{2\%} - D_{98\%})/78\,Gy]$ and conformity index $[V_{100\%}/V_{PTV}]$ for the PTV; 2. $V_{30Gy}$, $V_{40Gy}$, $V_{50Gy}$, $V_{60Gy}$, $V_{65Gy}$, $V_{70Gy}$ and Mean Dose for the rectum and bladder; 3. the Mean Dose and the $D_{1cc}$ for the femoral heads.

Together with the standard DVH metric used in clinical practice, to simplify the overall scoring of plans and to limit the subjectivity of judgment, the Plan Quality Metric (PQM) was adopted as a global measure of quality. PQM was first introduced by Nelms [6] and is now implemented in PlanIQ software (v2.1.1, Sun Nuclear Corp., Melbourne, FL).

PQM is a user-defined metric intended to compare the quality of treatment plans. It gathers into a single number the judgment of quality expressed by a clinical team on the basis of its knowledge and experience. It is built through a list of submetrics, e.g. DVH metrics, which should schematically represent the peculiar goals of the treatment (Table 2). To each metric, the user associates a numerical scoring function to model as accurately as possible the judgment criteria of the clinicians (Fig. 1). The PQM is the sum of the score obtained by each submetric and measures how much the plan adheres to the list of indentified goals. The percentage PQM (PQM%), i.e. the ratio of the achieved score to the maximum achievable, thus represents a relative measure of plan goodness.

For the purpose of this work, PQM offered a prompt and objective method to compare the quality of different plans pertaining to the same

**Table 2**
List of metrics, definitions and PQM value ranges used to form the PQM algorithm used in this article.

| Structure | Metric | Definition | PQM value range | |
|---|---|---|---|---|
| | | | Min | Max |
| PTV | $V_{0.98Dpresc}$ [%] | Percent of PTV volume ≥ 98% of the prescription dose | 0 | 15 |
| PTV | $D_{0.03\ cc}$ [Gy] | Dose [Gy] covering highest 0.03 cc of PTV | 0 | 10 |
| CTV | $V_{Dpresc}$ [%] | Percent of CTV volume ≥ prescription dose | 0 | 10 |
| PTV | Conformity index | (PTV V95% [cc])$^2$/(PTV total volume [cc] * 0.98Dpresc isosurface volume [cc]) | 0 | 5 |
| Body - PTV | $V_{Dpresc}$ [%] | Volume [cc] of tissue outside PTV ≥ Dpresc | 0 | 10 |
| Rectum | $V_{40Gy}$ [%] | Percent of rectum volume ≥ 40 Gy | 0 | 10 |
| Rectum | $V_{65Gy}$ [%] | Percent of rectum volume ≥ 65 Gy | 0 | 10 |
| Rectum | $V_{Dpresc}$ [cc] | Volume [cc] of rectum ≥ Dpresc | 0 | 10 |
| Rectum | Serial rectum | Number of axial planes with all rectum voxels exceeding 34 Gy | −10 | 0 |
| Bladder | $V_{40Gy}$ [%] | Percent of bladder volume ≥ 40 Gy | 0 | 3 |
| Bladder | $V_{65Gy}$ [%] | Percent of bladder volume ≥ 65 Gy | 0 | 7 |
| Femur R | $D_{1\ cc}$ [Gy] | Dose [Gy] covering highest 1 cc of right femour | 0 | 5 |
| Femur L | $D_{1\ cc}$ [Gy] | Dose [Gy] covering highest 1 cc of left femour | 0 | 5 |
| | Global maximum location | Anatomic location of global maximum: CTV, PTV or elsewhere | 0 | 5 |
| Total | | | −10 | 105 |

patient or even different patients.

### 2.6. Clinical impact

To quantify the impact of RapidPlan in the clinical routine and to also show that the sample of plans considered in this work is large enough to support the conclusions we simulated a clinical workflow performing a bootstrap analysis on the collected PQM% data. We considered an inflow of 90 prostate patients per year, each of which is planned by one of the six resident planners of our department, all considered as randomly and equally available. The patient population was simulated selecting randomly 90 times one of the 15 patients considered in this study. To each of these 90 random patients one of the six planners participating the study was randomly assigned. This is equivalent to randomly select 90 couples patient-planner from the 90 available in this study (6 planners times 15 patients) allowing for replacement, i.e. bootstrapping. The difference between the mean PQM% score obtained during the manual planning and the RapidPlan assisted planning procedure of the 90 randomly sampled couples has been computed together with the percentage of patients with higher PQM% between the two planning techniques. To associate statistic significance to these quantities the whole procedure has been repeated 10.000 times as usually done in bootstrap techniques.

### 2.7. Statistical analysis

Statistical analyses were performed to compare *manual* and *RapidPlan assisted* plans, in terms of quality, *inter-planner* and *intra-planner* variability.

To measure the *inter-planner* variability, the interquartile range (IQR) of each DVH metric and of PQM% was computed on the whole set of planners for each patient. IQR, the distance between the first and the third quartiles of a distribution, was chosen because of its inherent statistical robustness when dealing with skewed populations and/or data with outliers. To assess the *intra-planner* variability, the IQR of PQM was computed on the whole set of patients for each fixed planner.

DVH metrics were compared through a two-sided *t*-test with a significance level of 0.05. IQR and PQM% values were compared through a Wilcoxon signed rank test with a significance level of 0.05.

## 3. Results

### 3.1. Global comparison

A total of 210 treatment plans were created for the study: 7 planners, 15 patients, 2 plans per patient (105 *manual* and 105 *RapidPlan assisted* plans). The plans were all judged acceptable by a clinician, although in a few cases the protocol criteria were not fully satisfied.

Despite the freedom left to the planners when setting the treatment geometry each planner kept the same setup for any given patient. As a result, only a limited degree of variability is present in the number of arcs and collimator angles in the whole database while no differences can be noted between Manual and RapidPlan assisted plans once patient and planner are fixed. Moreover the same number of arcs was chosen by all planners to treat a given patient, 12 out of 15 patients were treated with two full arcs. The planners generally showed different approaches to set the collimator angle, but all the treatments were planned with angles between 10° and 30° and complementary angles were set by all planners in the case of 2 arcs.

Table 3 shows the percentage of plans, of the two planning procedures, that met the protocol criteria.

Table 4 confronts the PTV coverage and homogeneity. No significant differences emerged between the two planning procedures. Table 4 summarizes also the descriptive statistics for the principal metric characterizing OAR sparing along with the p-value of a paired *t*-test. A complete version of this table can be found in the Supplementary Material. In general, *RapidPlan assisted* plans outperform *manual* plans for the sparing of rectum and femoral heads. For the bladder, a weak dose reduction is observed, although the differences are not significant. Looking in more detail, 46% (48 plans) of the *RapidPlan assisted* plans showed a consistent better sparing of both rectum and bladder than the related *manual* plan. Moreover, 19% (20 plans) showed a consistent better sparing for only one structure and 11% (12 plans) were inferior with a lesser sparing for both structures, even if clinically acceptable. For the remaining 24% (25 plans), the DVH curves crossed one another preventing a quantitative unequivocal evaluation.

### 3.2. Inter-planner variability comparison

PQM% values were computed and compared between Manual and Rapid Plan assisted plans on a per patient basis (Fig. 2). We noted an overall increase in plan PQM% values when RapidPlan assisted planning was compared to manual planning, in particular: 14 out of 15 patients show higher median PQM% values. Only 1 patient out of 15 (#14) has a lower median PQM%, but the difference is not statistically significant. A Wilcoxon signed-rank test states that the plan quality increase is significant for 8 out of 15 patients.

The overall increase in plan quality induced by RapidPlan is accompanied by a general reduction in its variability. In Table 5, the IQR computed on each patient and for the principal DVH metrics is reported along with the number of patients for which the RapidPlan assisted plans showed a reduced IQR. A complete version of this table can be

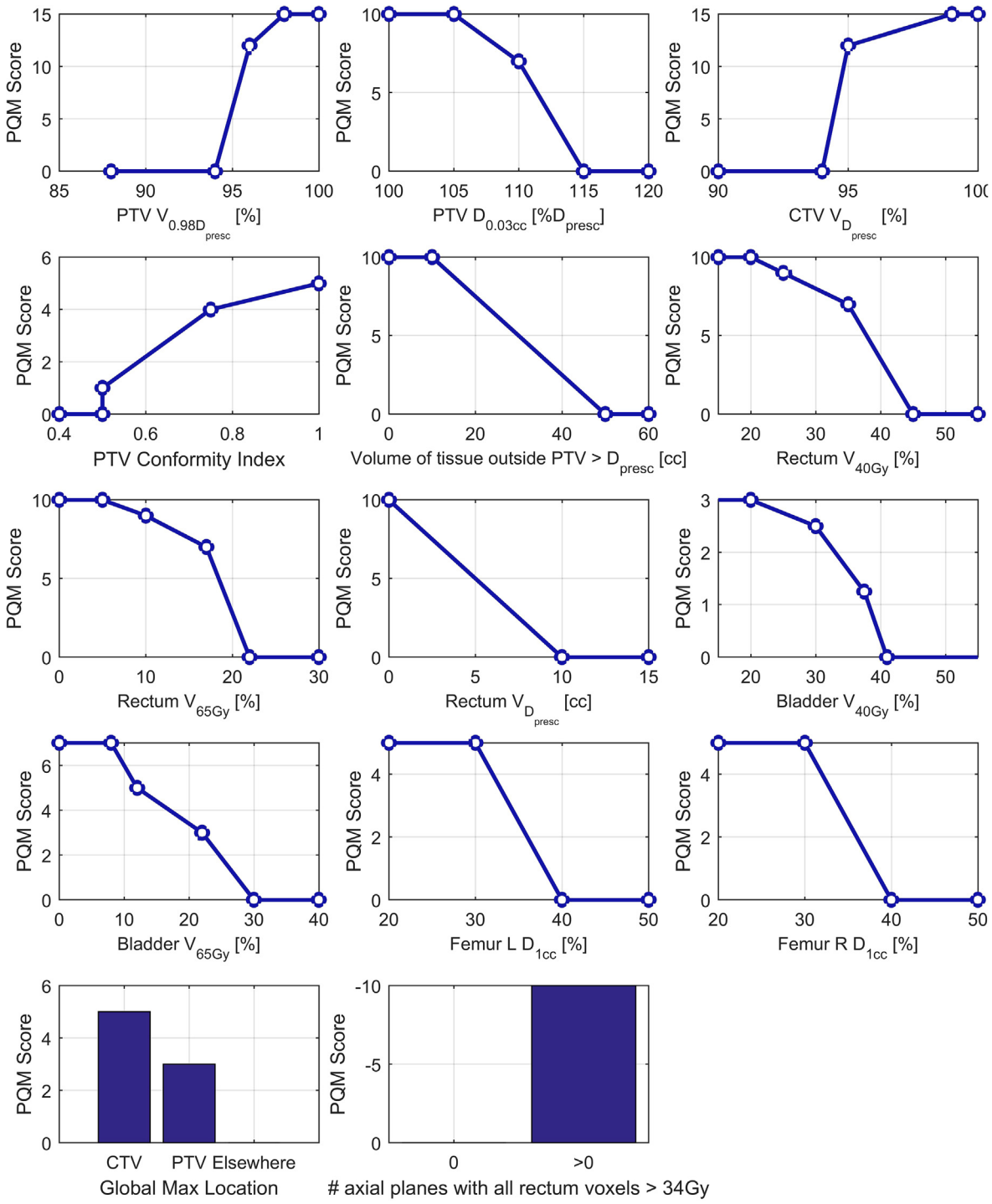


**Fig. 1.** Metrics related scoring functions that compose the PQM algorithm used in this study.

found in the Supplementary Material. The narrowing of the variability range of DVH metrics is significant and larger than 30% for rectum and femoral heads. For bladder, the IQR reduction lies generally between 10% and 20% but is not statistically significant. The inter-planner variability decrease is also perceived in the overall plan quality as shown in Fig. 1: 12 out 15 (80%) patients show a reduced IQR. For 8 patients out of 15 the reduction is statistically significant.

A Wilcoxon signed-rank test, performed on the whole sample of patients, showed that the overall variability reduction is statistically significant (p-value = 0.0046). PQM% IQR reduces from 8.32 ± 4.19 to 4.73 ± 3.79, approximately 40%.

### 3.3. Intra-planner variability comparison

To evaluate if RapidPlan had a different impact on the planning performance of planners with different experience, an *intra-planner* comparison of PQM% was performed (Fig. 3). A Wilcoxon signed-rank test was performed and statistically significant differences are marked. The IQR of PQM% showed a reduction for all planners but #6. A Wilcoxon signed-rank test, performed on the whole sample of planners, showed that the overall reduction is statistically significant (p-value = 0.033).

**Table 3**
Dose-volume objectives and meeting percentage.

| Structures | Endpoints | Meeting Percentage | | | |
|---|---|---|---|---|---|
| | | Manual | | RapidPlan | |
| PTV | $V_{78Gy}$ > = 95% | 100% | (n = 105) | 100% | (n = 105) |
| | $D_{98\%}$ > 95% | 100% | (n = 105) | 100% | (n = 105) |
| | $D_{2\%}$ < 107% | 99% | (n = 104) | 98% | (n = 103) |
| Bladder | $V_{40Gy}$ < = 40% | 92% | (n = 97) | 97% | (n = 102) |
| | $V_{65Gy}$ < = 25% | 100% | (n = 105) | 100% | (n = 105) |
| | $V_{75Gy}$ < = 10 cc | 26% | (n = 27) | 28% | (n = 29) |
| Rectum | $V_{40Gy}$ < = 45% | 92% | (n = 97) | 97% | (n = 102) |
| | $V_{60Gy}$ < = 25% | 98% | (n = 103) | 100% | (n = 105) |
| | $V_{75Gy}$ < = 10 cc | 100% | (n = 105) | 100% | (n = 105) |
| Femoral Head R | $D_{1cc}$ < = 45 Gy | 100% | (n = 105) | 100% | (n = 105) |
| Femoral Head L | $D_{1cc}$ < = 45 Gy | 100% | (n = 105) | 100% | (n = 105) |

### 3.4. Clinical impact

The bootstrapping procedure stated that, given the result of the present study, the introduction of RapidPlan in our clinical routine would result in an average 5.35% increase of PQM% (C.I. = [4.78%, 5.91%]) with a percentage of obtaining a better overall plan for every new patient of 75.1% (C.I. = [72.3%, 77.8%]).

## 4. Discussion

This study was designed to measure to what extent RapidPlan can reduce the human-caused variability in the planning of VMAT treatments thus increasing the homogeneity of planning performance of a planners cohort. At the same time, this study investigated the use of RapidPlan as an instrument to speed up the learning curve of VMAT planning. Manual planning and RapidPlan assisted planning of prostate treatments were compared in terms of quality.

In order to measure if RapidPlan can drive an improvement in plan quality, the DVH metrics reported in RTOG0126 were complemented by supplementary metrics at low doses for rectum and bladder together with mean doses to OARs. No statistically relevant differences were noted for target coverage and homogeneity and bladder sparing, while a net increase in rectum and femoral heads sparing (Tables 3 and 4) was measured for RapidPlan assisted planning. The equivalence of the planning methods in target coverage was probably due to the fact that target optimization and priority is largely standardized in our department and RapidPlan was configured to follow this standard. Conversely, the equivalence in bladder sparing was probably due to the small space left for optimization in well-prepared patients, for whom bladder sparing is largely determined by the relative geometric relationship with the target. Even if not significant, it is worth noting that bladder metrics were consistently smaller for RapidPlan planning for both mean and maximum values. Comparing plans in their entirety, 46% of RapidPlan-assisted plans were of superior or equivalent quality when compared with manual plans. RapidPlan-assisted plans had consistently inferior dose sparing in 11% of cases.

The main result of this work is that, besides the general increase in plan quality, RapidPlan was capable of driving a general reduction of human-caused variability as measured by the IQR of DVH metrics (Table 5). RapidPlan was found to be effective in reducing inter-planner variability of rectum and femoral heads, this is not true for bladder. Smaller room for optimization turns into lower differences among possible solutions and thus, less difference between different planners. Nonetheless, RapidPlan assisted planning outperformed the manual planning procedure in every considered DVH metric.

It is well known that the different shapes of DVH curve limit the possibility to univocally define if one plan is better than another. This difficulty can be partially overcome using a global, albeit subjective, metric of overall plan quality. To this end, the PQM score evaluation was included in the study and allowed a systematic comparison of the variability between patients (inter-planner) or planners (intra-planner). At the moment no systematic study has been presented about the robustness of PQM score. Limited to the scope of the presented study (the comparison of concurrent plans), the results are robust against small changes of the score lists used to define PQM%.

Considering Fig. 2, RapidPlan was capable of driving a larger or equivalent quality for the entire cohort of patients with a significant PQM% improvement in eight out of fifteen. For 12 out 15 patients the IQR is reduced, and on average a 40% reduction is seen (from 8.32 to 4.73) with a statistical significance of 0.0046 from a Wilcoxon signed-rank test. PQM analysis also confirmed the inter-planner variability reduction: IQR of PQM% scoring was significantly reduced for the entire sample of patients.

Fig. 3 depicts the capability of RapidPlan to enhance the plan quality achieved by the single planners. It is remarkable that the use of RapidPlan improves the planning skills of an inexperienced user (PL #7) up to the level of an average experienced one (PL #4). On the other hand, PL#5 produced the worst RapidPlan-assisted plans of the group despite his experience. Such result indicates that the skills of the single operator still significantly impact the planning process, despite the KBP prediction. The number of faced VMAT could be a measure of experience but it is not directly usable as planning skill indication.

**Table 4**
Comparison of target coverage, target conformation and OAR sparing between manual and RapidPlan assisted planning. Statistically significant differences are marked.

| Metric | | | Manual | | RapidPlan assisted | | |
|---|---|---|---|---|---|---|---|
| | | Endpoint | mean ± std | [min;max] | mean ± std | [min;max] | p-value |
| PTV | $D_{98\%}$ | | 98.5 ± 0.3 | [97.5;98.9] | 98.6 ± 0.3 | [96.9;99.3] | 0.441 |
| | $D_{2\%}$ | | 105.6 ± 0.9 | [103.7;109.4] | 105.8 ± 1.0 | [103.6;109.6] | 0.618 |
| | HI | | 7.12 ± 1.16 | [5.03;11.80] | 7.21 ± 1.25 | [4.46;12.68] | 0.837 |
| | CI | | 1.00 ± 0.02 | [0.97; 1.10] | 0.99 ± 0.02 | [0.97;1.06] | 0.089 |
| Rectum | $V_{40Gy}$ [%] | < =45 | 30.43 ± 10.13 | [11.39;58.07] | 26.38 ± 8.60 | [9.42;50.12] | < 0.001* |
| | $V_{60Gy}$ [%] | < =25 | 13.94 ± 6.33 | [3.13;34.43] | 11.31 ± 4.98 | [3.42;25.69] | < 0.001* |
| | $V_{75Gy}$ [cc] | < =10 | 2.38 ± 1.84 | [0.03;10.09] | 2.24 ± 1.62 | [0.04;7.80] | 0.059 |
| | mean dose [Gy] | | 31.03 ± 6.35 | [18.90;46.07] | 29.16 ± 5.38 | [19.18;41.29] | < 0.001* |
| Bladder | $V_{40Gy}$ [%] | < =40 | 24.17 ± 10.38 | [7.38;52.22] | 23.41 ± 9.69 | [7.49;44.23] | 0.091 |
| | $V_{65Gy}$ [%] | < =25 | 10.75 ± 5.19 | [3.75;23.50] | 10.37 ± 4.92 | [3.65;22.19] | 0.118 |
| | $V_{75Gy}$ [cc] | < =10 | 13.19 ± 6.01 | [4.55;32.99] | 12.83 ± 5.80 | [4.78;33.99] | 0.0136 |
| | mean dose [Gy] | | 24.43 ± 8.50 | [8.35;41.59] | 24.04 ± 8.32 | [8.99;39.24] | 0.149 |
| Femoral Head R | $D_{1cc}$ [Gy] | < =45 | 29.67 ± 8.50 | [20.41;43.59] | 26.68 ± 3.35 | [8.99;39.24] | < 0.001* |
| | mean dose [Gy] | | 14.00 ± 4.68 | [7.28;21.12] | 13.01 ± 2.57 | [8.17;18.62] | < 0.001* |
| Femoral Head L | $D_{1cc}$ [Gy] | < =45 | 30.29 ± 5.23 | [18.79;45.18] | 27.95 ± 3.68 | [18.29;37.03] | < 0.001* |
| | Mean dose [Gy] | | 14.58 ± 3.14 | [6.70;23.24] | 13.68 ± 2.68 | [7.06;21.36] | < 0.001* |

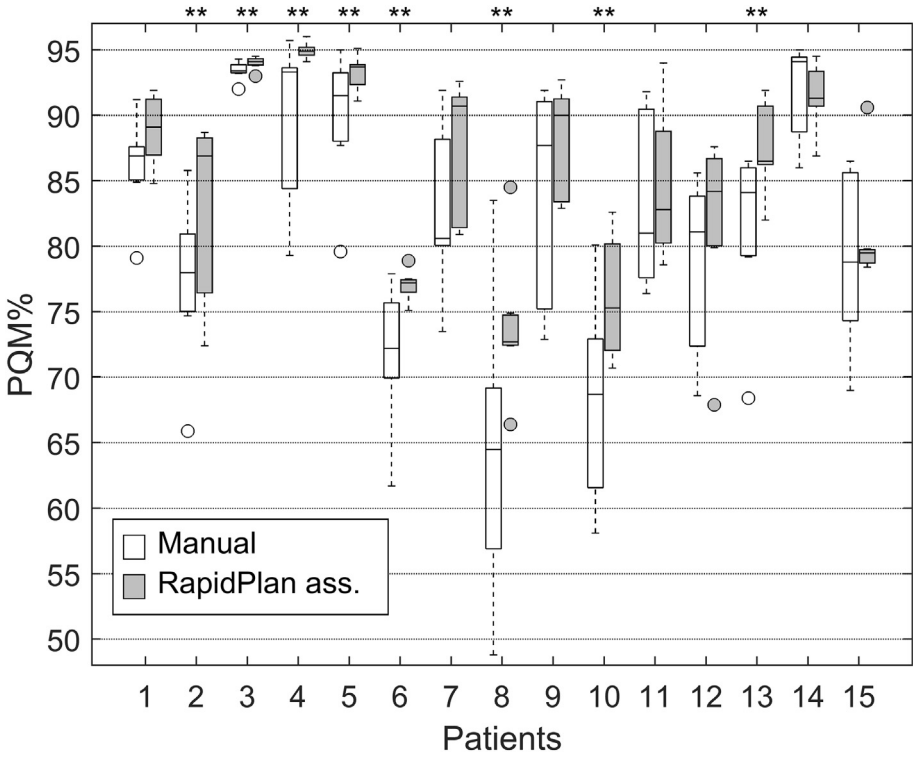


**Fig. 2.** Whiskers box plot of PQM%, inter-planner variability comparison. For each patient, manual and RapidPlan assisted planning are compared. The central line marks the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the adjacent values, which are the most extreme data value that are not outliers, and the circles represent the outliers. The extent of the boxes represents the Inter-Quartile Range of PQM%, herein used as measure of the inter-planner variability. Statistically significant differences are marked.

**Table 5**
Summary of InterQuartile Range reduction of DVH metrics due to RapidPlan assistance. For each DVH metric the IQR values are reported along with the percentage of patients showing a reduction. Statistical significant differences are marked.

| Metric | | IQR | | IQR reduction | |
|---|---|---|---|---|---|
| | | Manual | RapidPlan | # of cases | p-value |
| Rectum | V40 [%] | 11.69 ± 3.71 | 6.45 ± 3.34 | 14 (93%) | < 0.001* |
| | V60 [%] | 4.26 ± 2.77 | 2.17 ± 1.02 | 13 (86%) | 0.002* |
| | V75 [cc] | 0.81 ± 0.74 | 0.42 ± 0.34 | 13 (86%) | 0.001* |
| | mean dose [Gy] | 6.57 ± 1.90 | 3.82 ± 1.89 | 15 (100%) | < 0.001* |
| Bladder | V40 [%] | 4.62 ± 3.75 | 3.54 ± 2.93 | 11 (73%) | 0.107 |
| | V65 [%] | 1.42 ± 1.19 | 1.14 ± 0.89 | 10 (66%) | 0.389 |
| | V75 [cc] | 1.26 ± 0.69 | 1.12 ± 0.67 | 9 (60%) | 0.330 |
| | mean dose [Gy] | 2.35 ± 1.65 | 1.95 ± 1.45 | 10 (66%) | 0.277 |
| Femur R | D1cc [Gy] | 4.77 ± 2.65 | 2.49 ± 1.04 | 12 (80%) | 0.008* |
| | mean dose [Gy] | 2.09 ± 0.87 | 1.45 ± 0.78 | 10 (66%) | 0.043* |
| Femur L | D1cc [Gy] | 5.66 ± 2.33 | 2.76 ± 1.57 | 13 (86%) | 0.001* |
| | mean dose [Gy] | 3.02 ± 1.05 | 1.56 ± 0.97 | 14 (93%) | 0.001* |

Our results compare well with [19] where the authors presented a reduced standard deviation of DVH metrics as demonstration of the enhanced consistency of plan quality driven by RapidPlan. Our results extend those of Wu et al. as we have shown that RapidPlan predictions can increase the quality consistency even on a cohort of planners with different degrees of experience and skills.

Previous studies reported poor or no correlation between the differences in plan quality and the amount of planning experience; in our study, a moderate difference in plan quality was observed among the planners. Fig. 3 shows that for manual planning the more experienced planners generally outperform the less experienced ones. The introduction of RapidPlan predictions seems to reduce the gap. Nonetheless, the perceivable trend in Fig. 3 does not hold if each patient is considered alone. In other words, RapidPlan planning leads to better planning results only generally speaking and not for the individual patient case.

Limited to the quite simple treatment site here studied (prostate without nodes) these findings prove the effectiveness of RapidPlan predictions to enhance the consistency of plan quality as a result of the improved homogeneity among planners. This can be attributed to the capability of RapidPlan to simplify the planner's approach to the ubiquitous problem of reachable OAR sparing, reducing the experience gap between highly and poorly skilled planners. A reduction of the degree of human-caused variability related to the lack of planner's experience, is here reported together with its usefulness in clinical practice.

The increase of plan quality and its consistency did not come at the cost of a fewer time efficiency. Even if precise measurement were not performed, all the planners reported that with RapidPlan assistance the planning time was never longer than standard planning.

Part of the human-caused variability measured in this study might be due to the difference in arc arrangements (number of arcs and collimator angle) which was left to the free choice of the planner. This amount of variability could not be diminished by the use of RapidPlan which, at the moment, did not help the user in the choice of treatment geometry. Despite a detailed evaluation of this variability is hardly attainable, the limited differences of arc arrangements and the simple site studied should result in minimal dosimetric differences as reported by early literature [28,29]. On the other hand, the different arc arrangements did not threatens the result here presented because the same setup was kept between manual and RapidPlan assisted plans for any given patient and planner. This a posteriori evidence matches de facto the study design proposed in [19] where energy and geometry were kept fixed when comparing manual and RapidPlan assisted planning.

This study is affected by the following limitations: only fifteen patients and a single and simple treatment site were considered. Moreover it was conducted on planners of a single department. These limits do not threaten the validity of the proposed methodology, but it hinders, at the moment, the possibility to extend the results to different treatment sites. Future works should also asses the improvements in planning efficiency driven by KBP algorithm or the possible trade-off between

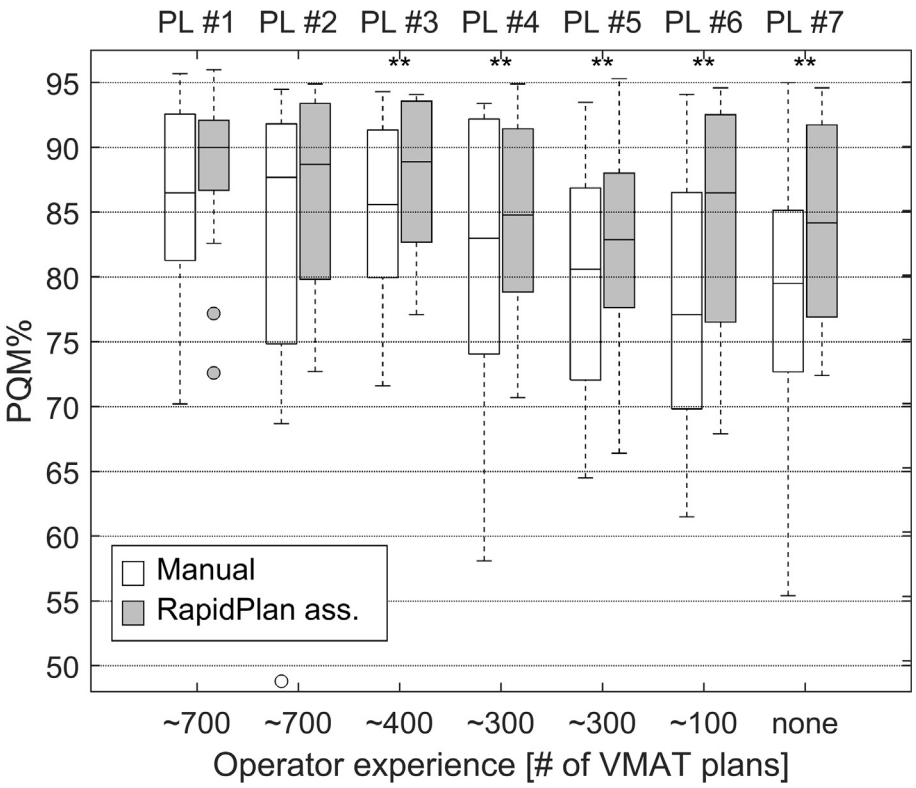


**Fig. 3.** Whiskers box plot of PQM%, intra-planner variability comparison. For each planner, manual and RapidPlan assisted planning are compared. The operators are sorted by experience in terms of total number of planned clinical VMAT. The central line marks the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the adjacent values, which are the most extreme data value that are not outliers, and the circles represent the outliers. The extent of the boxes represents the Inter-Quartile Range of PQM%, herein used as measure of the intra-planner variability. Statistically significant differences are marked.

efficiency and planning quality.

The results here presented support the conclusion that KBP systems can improve the mean plan quality of a single institution. At the same time, they can reduce the dependence of plan quality on planner skills thus increasing the robustness and homogeneity of the radiotherapy process. They can also be regarded as powerful tools for knowledge sharing and early education. Lastly the use of a PQM score here proposed simplifies the quality comparison in a large cohort of plans.

## 5. Conclusion

The RapidPlan knowledge-based planning engine can be trained to develop suitable models to improve the quality and consistency of treatment plans even when generated by planners with different education and expertise. Thus, the KBP approach can be used to homogenize plan quality by transferring planning expertise among operators at the same or different institutions or among different ones.

Exploiting the study design proposed herein, further work can be done to confirm that these results have a more general validity, to investigate more complex treatment sites and to extend the study with a multicentre participation.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ejmp.2018.08.016.

## References

[1] Good D, Lo J, Lee R, Wu J, Yin FF, Das S. 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 Biol Phys 2013;87:176–81.
[2] Williams MJ, Bailey MJ, Forstner D, Metcalfe PE. Multicentre quality assurance of intensity-modulated radiation therapy plans: A precursor to clinical trials. J Med Imaging Radiation Oncol 2007;51(5):472–9.
[3] 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(2):349–55.
[4] Batumalai V, Jameson MG, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. Practical Radiation Oncol 2013;3(3):e99–106.
[5] Berry SL, Boczkowski A, Ma R, Mechalakos J, Hunt M. Interobserver variability in radiation therapy plan output: results of a single-institution study. Practical Radiation Oncol 2016;6(6):442–9.
[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. Practical Radiation Oncol 2012;2(4):296–305.
[7] Moore KL, Brame RS, Low DA, Mutic S. Experience-based quality control of clinical intensity-modulated radiotherapy planning. International Journal of Radiation Oncology* Biology*. Physics 2011;81(2):545–51.
[8] Ueda Y, Fukunaga JI, Kamima T, Adachi Y, Nakamatsu K, Monzen H. Evaluation of multiple institutions' models for knowledge-based planning of volumetric modulated arc therapy (VMAT) for prostate cancer. Radiation Oncol 2018;13:46.
[9] Nwankwo O, Mekdash H, Sihono DS, 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(1):111.
[10] Tol JP, Doornaert P, Witte BI, Dahele M, Slotman BJ, Verbakel WF. A longitudinal evaluation of improvements in radiotherapy treatment plan quality for head and neck cancer patients. Radiother Oncol 2016;119(2):337–43.
[11] Hunt MA, Hsiung CY, Spirou SV, et al. Evaluation of concave dose distributions created using an inverse planning system. Int J Radiat Oncol Biol Phys 2002;54:953–62.
[12] Tol JP, Dahele M, Delaney AR, et al. Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans? Radiat Oncol 2015;10(1):234.
[13] 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. Int J Radiat Oncol Biol Phys 2017;97(1):164–72.
[14] Moore KL, Schmidt R, Moiseenko V, Olsen LA, Tan J, Xiao Y, et al. Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126. Int J Radiat Oncol Biol Phys 2015 Jun 1;92(2):228–35.
[15] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. J Clin Oncol 2010;28(18):2996–3001.
[16] Schubert C, Waletzko O, Weiss C, 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):e0178034.
[17] Fogliata A, Nicolini G, Clivio A, et al. A broad scope knowledge based model for optimization of VMAT in esophageal cancer: validation and assessment of plan quality among different treatment centers. Radiat Oncol 2015;10(1):220.
[18] 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. Physica Med 2017;44:199–204.
[19] Wu H, Jiang F, Yue H, Li S, Zhang Y. A dosimetric evaluation of knowledge-based VMAT planning with simultaneous integrated boosting for rectal cancer patients. J Appl Clin Med Phys 2016;17(6):78–85.
[20] Dal Pra A, Souhami L. Prostate cancer radiation therapy: a physician's perspective. Physica Med 2016;32(3):438–45.
[21] McVicar N, Popescu IA, Heath E. Techniques for adaptive prostate radiotherapy. Physica Med 2016;32(3):492–8.
[22] Eclipse Photon and Electron Reference Guide v13.7, Varian Medical System, June 2015.
[23] Fogliata A, Wang PM, Belosi F, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiat Oncol 2014;9(1):236.
[24] Tol JP, Delaney AR, Dahele M, Slotman BJ. Verbakel WF. Evaluation of a knowledge-based planning solution for head and neck cancer. International Journal of Radiation Oncology* Biology*. Physics 2015;91(3):612–20.
[25] Delaney AR, Tol JP, Dahele M, et al. Effect of dosimetric outliers on the performance of a commercial Knowledge-Based Planning solution. Int J Radiat Oncol Biol Phys 2016;94(3):496–1477.
[26] Cagni E, Botti A, Micera R, Galeandro M, Sghedoni R, Orlandi M, et al. Knowledge-based treatment planning: An inter-technique and inter-system feasibility study for prostate cancer. Physica Med 2017;36:38–45.
[27] Fusella M, Scaggion A, Pivato N, Rossato MA, Zorz A, Paiusco M. Efficiently train and validate a RapidPlan model through APQM scoring. Med Phys 2018.
[28] Guckenberger M, Richter A, Krieger T, Wilbert J, Baier K, Flentje M. Is a single arc sufficient in volumetric-modulated arc therapy (VMAT) for complex-shaped target volumes? Radiother Oncol 2009;93(2):259–65.
[29] Wolff D, Stieler F, Welzel G, Lorenz F, Abo-Madyan Y, Mai S, et al. Volumetric modulated arc therapy (VMAT) vs. serial tomotherapy, step-and-shoot IMRT and 3D-conformal RT for treatment of prostate cancer. Radiother Oncol 2009;93(2):226–33.