Title

Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials

Journal

International Journal of Radiation Oncology • Biology • Physics, 97(1)

ISSN

0360-3016

Authors

Li, Nan
Carmona, Ruben
Sirak, Igor
et al.

Publication Date

2017

DOI

10.1016/j.ijrobp.2016.10.005

Peer reviewed

International Journal of Radiation Oncology biology • physics

www.redjournal.org

Physics Contribution

Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials

Nan Li, PhD,* Ruben Carmona, MD,* Igor Sirak, MD, PhD,† Linda Kasaova, PhD,† David Followill, PhD,‡ Jeff Michalski, MD,§ Walter Bosch, PhD,§ William Straube, MS,§ Loren K. Mell, MD,* and Kevin L. Moore, PhD*

**Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California; †Department of Oncology and Radiotherapy, University Hospital, Hradec Kralove, Czech Republic; ‡Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, Texas; and §Department of Radiation Oncology, Washington University in St Louis, St Louis, Missouri*

Received Apr 20, 2016, and in revised form Sep 9, 2016. Accepted for publication Oct 4, 2016.

Summary

A highly efficient method is proposed to train and validate knowledge-based planning for use on radiation therapy clinical trials. Using plans from 86 patients who underwent planning according to a multi-institutional cervical cancer trial, we used

Purpose: To demonstrate an efficient method for training and validation of a knowledge-based planning (KBP) system as a radiation therapy clinical trial plan quality-control system.

Methods and Materials: We analyzed 86 patients with stage IB through IVA cervical cancer treated with intensity modulated radiation therapy at 2 institutions according to the standards of the INTERTECC (International Evaluation of Radiotherapy Technology Effectiveness in Cervical Cancer, National Clinical Trials Network identifier: 01554397) protocol. The protocol used a planning target volume and 2 primary organs at risk: pelvic bone marrow (PBM) and bowel. Secondary organs at risk were rectum and bladder. Initial unfiltered dose-volume histogram (DVH) estimation models were trained using all 86 plans. Refined training sets were created by removing sub-optimal

Reprint requests to: Kevin L. Moore, PhD, Department of Radiation Medicine and Applied Sciences, University of California San Diego, 3960 Health Sciences Dr, MC0865, La Jolla, CA 92093. Tel: (858) 822-6056; E-mail: kevinmoore@ucsd.edu

This study is supported by grants 1R21CA162718-01, KL2 RR031978-01, and LRP L30 CA135746-01 from the National Institutes of Health and grants from Varian Medical Systems.

Conflict of interest: W.B. reports receiving personal fees from Augmenix and support from the American Society for Radiation Oncology and the American Association of Physicists in Medicine outside the submitted work. W.S. reports receiving grants from the National Institutes of Health (NIH) during the conduct of the study. L.K.M. reports receiving grants from the NIH and grants and personal fees from Varian Medical Systems during the conduct of the study. K.L.M. reports receiving grants from the NIH and grants, personal fees, and nonfinancial support from Varian Medical Systems during the conduct of the study. In addition, K.L.M. has a patent licensed to Varian Medical Systems (US Patent 20120310615A1), as well as a patent pending (Three-Dimensional Radiotherapy Dose Distribution Prediction).

Supplementary material for this article can be found at www.redjournal.org.

Int J Radiation Oncol Biol Phys, Vol. 97, No. 1, pp. 164–172, 2017
0360-3016/$ - see front matter
http://dx.doi.org/10.1016/j.ijrobp.2016.10.005

protocol-specific dosimetric cutpoints to refine dose-volume histogram estimation models and implemented a program that automatically re-plans several candidate auto-planning routines across multiple patients. After identification of the knowledge-based planning routine that best exemplified trial dosimetric aims, improved normal tissue sparing was observed across the sample.

plans from the unfiltered sample, and DVH estimation models… and DVH estimation models were constructed by identifying 30 of 86 plans emphasizing PBM sparing (comparing protocol-specified dosimetric cutpoints V_{10} (percentage volume of PBM receiving at least 10 Gy dose) and V_{20} (percentage volume of PBM receiving at least 20 Gy dose) with unfiltered predictions) and another 30 of 86 plans emphasizing bowel sparing (comparing V_{40} (absolute volume of bowel receiving at least 40 Gy dose) and V_{45} (absolute volume of bowel receiving at least 45 Gy dose), 9 in common with the PBM set). To obtain deliverable KBP plans, refined models must inform patient-specific optimization objectives and/or priorities (an auto-planning "routine"). Four candidate routines emphasizing different tradeoffs were composed, and a script was developed to automatically re-plan multiple patients with each routine. After selection of the routine that best met protocol objectives in the 51-patient training sample (KBP_{FINAL}), protocol-specific DVH metrics and normal tissue complication probability were compared for original versus KBP_{FINAL} plans across the 35-patient validation set. Paired *t* tests were used to test differences between planning sets.

Results: KBP_{FINAL} plans outperformed manual planning across the validation set in all protocol-specific DVH cutpoints. The mean normal tissue complication probability for gastrointestinal toxicity was lower for KBP_{FINAL} versus validation-set plans (48.7% vs 53.8%, $P<.001$). Similarly, the estimated mean white blood cell count nadir was higher (2.77 vs 2.49 k/mL, $P<.001$) with KBP_{FINAL} plans, indicating lowered probability of hematologic toxicity.

Conclusions: This work demonstrates that a KBP system can be efficiently trained and refined for use in radiation therapy clinical trials with minimal effort. This patient-specific plan quality control resulted in improvements on protocol-specific dosimetric endpoints.

Introduction

Variations in treatment plan quality and the lack of systematic quality control (QC) are significant problems with current treatment planning techniques (1-5). These issues compromise the gains that can be realized with advanced technologies and can complicate inter-institution collaboration and multi-institutional clinical trials (6). There is even recently published evidence that treatment at centers with higher clinical trial accrual volume was associated with longer overall survival in patients with locally advanced non−small cell lung cancer (7). Multiple studies have shown that knowledge-based planning (KBP) can improve both the consistency of treatment plan quality and planning efficiency (8-13). A common theme in KBP methods is the incorporation of prior treatment plan experience into the treatment planning of future patients. Recent work has focused on developing models trained from previous patients to predict achievable dose-volume histograms (DVHs) for individual patients based on their unique anatomy (3, 14-17).

Although KBP is a promising method to improve quality and efficiency in clinical practice, its implementation as a multi-institutional clinical trial QC system has not been established. A recent retrospective study used KBP in a large-scale multi-institutional clinical trial (Radiation Therapy Oncology Group 0126) to assess the frequency and clinical severity of suboptimal treatment plans (6). The results indicated that patient-specific plan QC tools, for example, KBP-driven DVH estimation, should be part of the quality apparatus of multi-institutional radiation therapy trials, particularly if a trial has specified endpoints that hinge on normal tissue sparing.

Recently, commercial KBP software (RapidPlan; Varian Medical Systems, Palo Alto, CA) has become clinically available. RapidPlan allows clinicians to optimize new plans using model-based DVH estimations and optimization objectives and/or priorities derived from these estimations. The patient-specific DVH estimations result from mathematical models that correlate patient anatomic geometry to resultant DVHs, fed by previous patient plans that train individual organ-at-risk (OAR) estimation models. However, all knowledge-based systems require proper training and extensive validation for the multivariable plan quality problem posed by multiple OARs and planning target volumes (PTVs) (18). This process can be tremendously time-consuming, particularly the validation testing whereby the KBP system must be applied to and analyzed across a large plurality of patients. Any changes to the DVH estimation models or the optimization system compel another round of re-planning across the validation set, meaning the upfront costs can be a substantial barrier to implementation of KBP systems in resource-constrained environments.

In an attempt to lower the effort level required to develop and implement a KBP-driven QC system for clinical trials, we developed a highly efficient method to train, refine, and automatically validate a KBP system in a cohort of patients for implementation in trials with a radiation therapy component. The aim is to demonstrate a streamlined process

that could work in any radiation therapy trial irrespective of specific dosimetric constraints. The claim of this method as "highly efficient" follows not just from the inherent efficiency of an auto-planning system but in the process by which the parameter space of possible KBP-driven auto-planning routines is explored with a scripted program that can generate several hundred re-plans across many patients "at the click of a button," that is, with no manual effort.

We focused on the phase 2 clinical trial INTERTECC (International Evaluation of Radiotherapy Technology Effectiveness in Cervical Cancer, National Clinical Trials Network identifier: 01554397) as a test bed for this method; this was an international multi-institutional trial of bone marrow–sparing intensity modulated radiation therapy (IMRT) with a primary endpoint of reducing hematologic and gastrointestinal (GI) toxicity. Because the trial's main hypothesis hinged on normal tissue sparing in IMRT, this is an ideal test bed for a KBP system in clinical trial QC. A system validated on this phase 2 component would be ideal for implementation as a quality system on the critical phase 3.

Methods and Materials

Patient sample

We selected 86 patients with stage IB through IVA cervical cancer treated with fixed-field IMRT according to INTERTECC guidelines: 63 patients from University of California San Diego and 23 from University Hospital Hradec Králové. Of the patients, 26 were participants in the clinical trial and 60 were treated off trial (eg, because of refusal) or treated in a competing clinical trial using the same treatment planning approach. All patients received 45.0 to 50.4 Gy in 1.8-Gy daily fractions over a period of 5 to 5.5 weeks with concurrent chemotherapy. Seventy-three patients were treated with pelvic IMRT fields only, and 13 patients were treated with extended (pelvic or para-aortic) IMRT fields. The clinical target volume was defined as the gross tumor plus areas containing potential microscopic disease, including the cervix and uterus (if present), the superior third of the vagina (or half of the vagina, if clinically involved), the parametria, and the regional lymph nodes. Planning margins were 15 mm around the cervix and uterus and 10 mm around the vagina and parametria, with a 5- to 7-mm margin around nodal regions. Critical normal tissues for IMRT optimization consisted of bowel, bladder, rectum, and pelvic bone marrow (PBM). Only bowel and PBM were used as primary avoidance structures.

Terminology: Models and routines

We have used the term *model* to refer to the set of DVH estimation models that can predict achievable OAR DVHs for an individual patient and the term *routine* to refer to the use of these DVH estimation models to generate a set of patient-specific DVH-based inverse optimization objectives and priorities. This distinction is critical because the DVH estimates from a model are not the same as the DVHs from a plan created with a routine. As an example of this distinction, a single model can be used in multiple routines that convert the same model-predicted DVHs to distinct inverse planning objective sets that emphasize differing clinical priorities in the plan optimization stage.

Model building, training, and refinement

To create the RapidPlan DVH estimation models, we included one PTV and 4 OARs: PBM, bowel, rectum, and bladder. The entire library of 86 plans was included in the initial training set. The DVH estimation modeling engine processes each patient dataset, correlating the underlying OAR geometry to the DVH information (19). Application of the RapidPlan modeling routine to all 86 plans resulted in an initial unfiltered model.

Because the results that the model produces are roughly dependent on the mean quality of the training plans (18), it is critical to identify high-quality plans for a refined training set, i.e. a sample with suspected sub-optimal plan eliminated, for each OAR (3, 17). We decided to select the top 30 plans (approximately one-third of the entire sample) with an apparent emphasis on PBM or bowel sparing to generate refined training sets for these models, because these are the principal OARs for the protocol. To identify these high-quality plans, original plan DVHs were compared with the unfiltered model DVH estimations based on protocol-specified DVH metrics for each structure. More specifically, 30 original plans were identified for PBM optimization by comparing original DVHs for V_{10} (percentage volume of PBM receiving at least 10 Gy dose) and V_{20} (percentage volume of PBM receiving at least 20 Gy dose) with the unfiltered model DVH estimations. Another 30 plans were selected in the same manner for bowel optimization by comparing V_{40} (percentage volume of bowel receiving at least 40 Gy dose) and V_{45} (percentage volume of bowel receiving at least 45 Gy dose), 9 of which were found to be in common with the PBM sample. As PBM and bowel were the 2 primary OARs, we used the 51-patient samples selected for these 2 structures based on their protocol-specified DVH metrics as our training set for our entire study. The remaining 35-patient samples were used as the validation set (Fig. 1).

One concern with this approach is that, by treating the OAR DVH modeling separately in the selection of training plans, we might be predicting impossible-to-achieve DVH estimations when combining these model predictions if, as is likely, there is a tradeoff in the sparing of one OAR against the other. However, by using the plans that spare an individual OAR while still meeting protocol acceptance criteria, it is possible to obtain the frontier estimates for both PBM and bowel sparing and then use the optimization routines to properly balance these multi-criterion objectives against each other and achieve acceptable plans (as discussed in the next subsection).

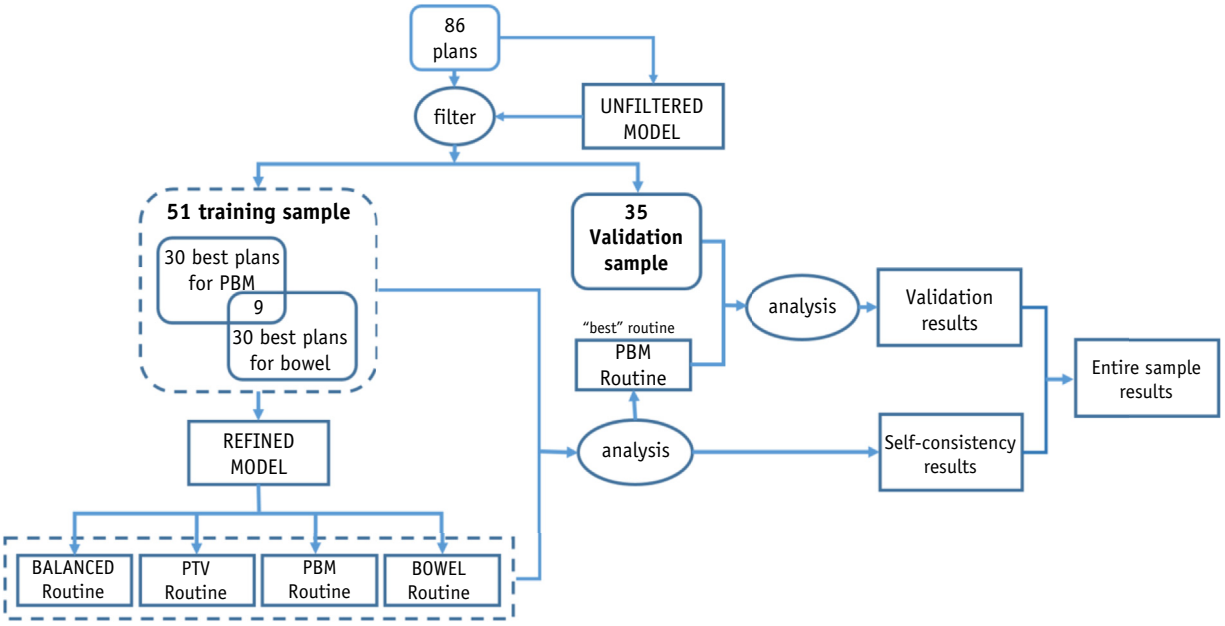

Fig. 1. Flow diagram of study. *Abbreviations:* PBM = pelvic bone marrow; PTV = planning target volume.

For rectum and bladder, a rough refinement was carried out using the model evaluation tools established in RapidPlan (20), that is, only obvious outliers with clinical DVHs far in excess of their estimated DVHs were removed. In the end, 52 plans and 63 plans were included in the training sets for rectum and bladder, respectively. As a result, a refined RapidPlan model that included DVH estimations for all structures was trained by the filtered training set for each OAR.

Model to routines

To truly validate the KBP quality system, the DVH estimation model must be converted into auto-planning routines that generate deliverable treatment plans. The final plans will necessarily have to balance multiple OAR and/or PTV criteria, and it is not at all obvious what the right blend of OAR objectives and priorities is. Four routines with different clinical emphases were created: balanced routine, PTV routine, PBM routine, and bowel routine. Table 1 lists the optimization objectives and priorities that we were using for each routine.

Automated routine analysis and validation

Patient-by-patient re-planning is extremely time-consuming, especially when there is more than one variant at play. For this work, we seek to quantify the performance of the 4 different automated planning routines across the 51-patient training set in a systematic way and validate the best routine using the remaining 35 samples. In total, this represents 239 treatment plans ($51 \times 4 + 35 = 239$) that need to be generated, a tremendously tedious re-planning task even with RapidPlan. To systematically analyze and validate the proposed planning routines with the minimum of manual effort, we developed a program using the Eclipse Scripting Application Programming Interface (ESAPI; Varian Medical Systems) to re-plan a large plurality of patients, automatically generating all desired $N \times M$ permutations for N patients with M routines with no human intervention. For each patient, the plan generation process includes beam configuration (using information from original plan), plan optimization with the specific planning routine, dose calculation, and plan normalization. To make equivalent comparisons across the routines, automated KBP plans were generated using the same modality, prescription dose, and beam setup as in the original plan. Following the protocol coverage requirement, all plans were normalized to cover 95% of PTV with 100% of the prescribed dose before plan comparison.

After generating the 204 re-plans ($51 \times 4 = 204$), the ESAPI-driven system exported the resultant plan data one by one for each of the training-set patients for aggregate analysis. Following the INTERTECC protocol, PBM V_{10}, PBM V_{20}, bowel V_{40}, and bowel V_{45} were used as criteria for DVH comparison. The performance of each routine on these DVH metrics, as well as review of the aggregate statistics with the protocol's principal investigator, identified the routine that best exemplified the clinical goals of the protocol (KBP_{FINAL}).

The validation-set evaluation was also accomplished by comparing PTV and OAR DVHs between resultant KBP_{FINAL} and original plans. In addition to the protocol DVH cutpoints, we used validated normal tissue complication probability (NTCP) models (21, 22) that map bowel and PBM DVHs to the probability of developing grade 2 or greater GI toxicity and hematologic toxicity, respectively.

Table 1 Optimization and priority settings

Structure	Objectives	Priorities			
		Balanced routine	PTV routine	PBM routine	Bowel routine
PTV	D_{max} < 105% of Rx	100	200	100	100
	$D_{10\%}$ < 103% of Rx	90	90	90	90
	$D_{99\%}$ > Rx	100	200	100	100
	D_{min} > 98% of Rx	100	200	100	100
PBM	D_{max} < Rx	90	90	90	90
	V_{10} < model generated	100	100	200	100
	V_{20} < model generated	100	100	200	100
	V_{30} < model generated	80	80	80	80
	V_{40} < model generated	80	80	80	80
Bowel	D_{max} < Rx	100	100	100	200
	V_{10} < model generated	70	70	70	70
	V_{20} < model generated	70	70	70	70
	V_{30} < model generated	90	90	90	90
	V_{40} < model generated	100	100	100	200
Bladder	D_{max} < Rx	80	80	80	80
	V_{20} < model generated	50	50	50	50
	V_{40} < model generated	50	50	50	50
Rectum	D_{max} < Rx	80	80	80	80
	V_{20} < model generated	50	50	50	50
	V_{40} < model generated	50	50	50	50

Abbreviations: PBM = pelvic bone marrow; PTV = planning target volume; Rx = prescription.

The NTCP estimations come with substantial uncertainties but were meant to provide a representation of the percentage-wise gains between manual and KBP planning rather than precise probability estimations. Paired, 2-sided Student *t* tests were used to identify significant ($P \leq .05$) differences in DVH metrics and NTCP.

Figure 1 is a flow diagram that describes the entire study.

Results

Comparison of auto-planning routines

To determine the best routine of the 4 expressions of the refined DVH estimation models, the ESAPI-driven system applied the 4 routines to the 51 training patients and the resultant plans were compared with the original protocol plans. With identical PTV normalization, we focused on the 4 key OAR DVH metrics for the protocol (PBM V_{10}, PBM V_{20}, bowel V_{40}, bowel V_{45}). The percentage of KBP plans for which the DVH metric was better than or equal to the original plans in each routine was as follows: balanced routine (49%, 51%, 47%, 76%), PTV routine (45%, 49%, 29%, 67%), PBM routine (51%, 59%, 49%, 78%), and bowel routine (53%, 53%, 61%, 88%). Similarly, the average ± standard deviation DVH metric excess in the clinical sample could be assessed per routine (PBM $V_{10\ clinical} - V_{10\ routine}$, PBM $V_{20\ clinical} - V_{20\ routine}$, bowel $V_{40\ clinical} - V_{40\ routine}$, bowel $V_{45\ clinical} - V_{45\ routine}$): balanced routine (0.1% ± 2.5%, 0.3% ± 3.9%, 0.6% ± 3.1%, 0.8% ± 1.9%), PTV routine (−0.2% ± 2.5%, 0.4% ± 3.9%, −0.5% ± 2.7%, 0.4% ± 1.8%), PBM routine (0.3% ± 2.5%, 1.3% ± 3.9%, 0.5% ± 3.2%, 0.8% ± 1.8%), and bowel routine (0.1% ± 2.5%, 0.4% ± 4.0%, 1.1% ± 3.1%, 1.4% ± 1.8%). These results are consistent with the PBM and bowel routines being generally superior with respect to the primary aims of the trial. In consultation with the protocol's principal investigator, the PBM routine was elected as the final routine (KBP_{FINAL}).

KBP quality comparison

We compared the average DVH and average DVH difference between original and KBP_{FINAL} plans for 35 validation samples (Fig. 2). Using $D_{x\%}$ to quantify the dose to the hottest x% volume, KBP_{FINAL} plans exhibited no significant difference in PTV homogeneity ($D_{1\%} - D_{99\%}$ = 5.15 ± 2.9 Gy for original and 5.15 ± 3.1 Gy for KBP_{FINAL}, $P = .98$). On average, KBP_{FINAL} lowered the <30-Gy region of the PBM DVH and the >30-Gy region of the bowel DVH compared with original plans.

Then, the protocol-specified DVH metrics were also used to perform the comparison. Each pair of KBP_{FINAL} and original plans was plotted to show how the plan quality changes for each patient in the training set and validation set (Fig. 3). As indicated, the quality of most KBP_{FINAL} plans was comparable to or better than the original plans in terms of DVH metrics for PBM and bowel for validation patients (Table 2). Using the NTCP model for GI toxicity (21), we observed that the mean NTCP for GI toxicity (based on bowel V_{45}) was lower for KBP_{FINAL} compared with original plans across the validation set (48.7% vs 53.8%, $P < .001$). Similarly, with the NTCP model for hematologic toxicity

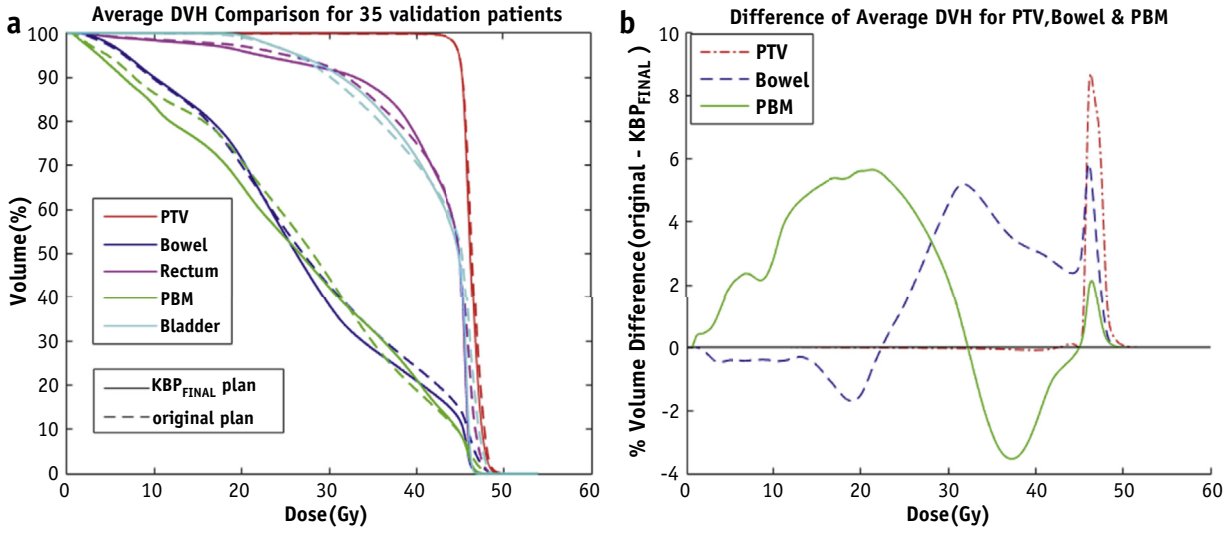

Fig. 2. Comparison of average dose-volume histograms (DVHs) of original and KBPFINAL plans for 35 validation patients. (a) Average DVH comparison. (b) Difference in average DVH between original and KBP_FINAL. One should note that the regions of improvement are in the protocol-specified 10- to 20-Gy region of the PBM DVH and in the high-dose region (>40-Gy region) of the bowel DVH. Abbreviations: PBM = pelvic bone marrow; PTV = planning target volume; KBP_{FINAL} = final KBP plan generated using the best routine; V_n = percentage or absolute volume receiving at least n Gy dose.

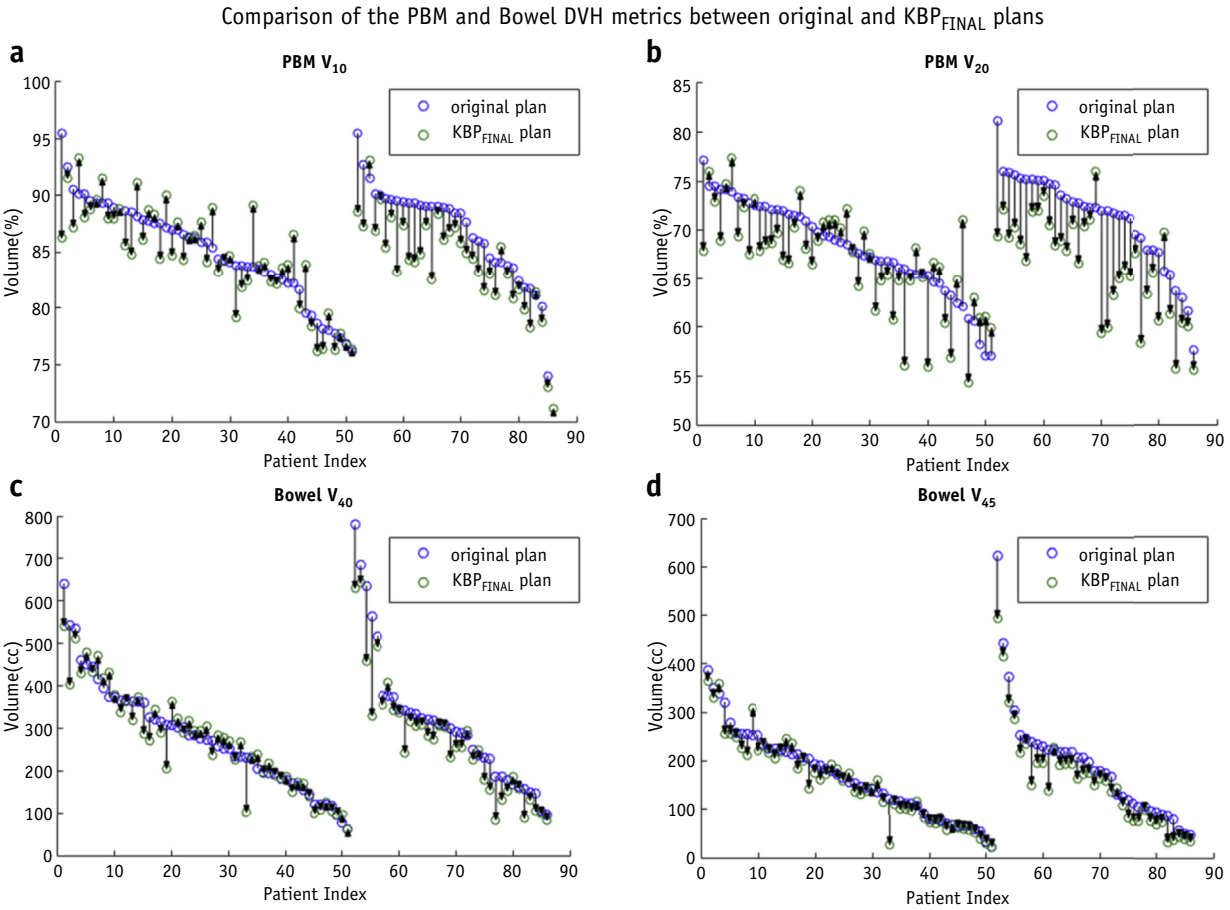

Fig. 3. Comparison of pelvic bone marrow (PBM) and bowel dose-volume histogram (DVH) metrics between original plan and KBPFINAL plan in training and validation sets: PBM V_{10} (a) and V_{20} (b) comparisons and bowel V_{40} (c) and V_{45} (d) comparisons for each patient. Abbreviations: KBP_{FINAL} = final KBP plan generated using the best routine; V_n = percentage or absolute volume receiving at least n Gy dose.

Table 2 Comparison of IMRT plans across 35 validation samples based on DVH metrics

Metrics	KBP_{FINAL} plan	Original plan	*P*
PTV			
$D_{1\%}$, Gy	48.9 ± 1.6	48.9 ± 1.6	.98
$D_{99\%}$, Gy	44.1 ± 1.7	43.7 ± 3.4	.43
PBM			
V_{10}, %	84.0 ± 4.4	86.5 ± 4.9	<.001
V_{20}, %	66.3 ± 5.2	71.2 ± 4.8	<.001
Bowel			
V_{40}, cm^3	268.0 ± 139.1	311.4 ± 159.7	<.001
V_{45}, cm^3	158.3 ± 103.9	190.2 ± 116.3	<.001
Rectum			
V_{30}, %	92.0 ± 5.8	92.7 ± 10.4	.65
V_{45}, %	51.7 ± 20.4	52.1 ± 20.1	.79
D_{max}, Gy	47.2 ± 1.6	48.1 ± 1.7	<.001
Bladder			
V_{45}, %	50.3 ± 15.5	54.9 ± 17.5	<.001
D_{max}, Gy	47.4 ± 1.6	48.4 ± 1.8	<.01

Abbreviations: D_{max} = maximum dose received; $D_{n\%}$ = dose received by n% of the volume; DVH = dose-volume histogram; IMRT = intensity modulated radiation therapy; KBP_{FINAL} = final KBP plan generated using the best routine; PBM = pelvic bone marrow; PTV = planning target volume; V_m = percentage or absolute volume receiving at least m Gy dose.

(22), the estimated mean white blood cell count nadir (based on PBM V_{20}) was higher across the validation set (2.77 vs. 2.49 k/mL, $P<.001$), indicating lower expected hematologic toxicity in the KBP_{FINAL} cohort compared with the original plans. Figure 4 shows the frequency histograms of the difference in NTCP and estimated white blood cell count nadir between original and KBP_{FINAL} plans across the training set (gray) and validation set (white). We can conclude that overall plan quality could be improved with a KBP-driven QC system, with significant clinical gains for individual plans that are significant outliers (eg, 3 plans observed with >15% excess risk of GI toxicity).

Discussion

This study demonstrated an efficient method to refine the model training library, explore the space of possible routines, and validate the routines with auto-scripting, which are important steps before KBP can be implemented as a QC tool for clinical trials. Several publications have reported the feasibility of using RapidPlan to improve OAR sparing (23-25), and Tol et al (26) showed the potential of using a RapidPlan DVH estimation model directly as a plan QC tool for head and neck cases. However, these studies built models directly on arbitrary training plans with no special selection criteria, which could influence the performance of the RapidPlan DVH estimation models (3, 17). Moreover, none of these works have systematically investigated the influence of using a different mix of optimization objectives and priorities for multiple OARs and PTVs. The ESAPI-driven system described in this work provides a way to explore and validate different routines automatically and efficiently before implementing them eithr in normal clinical operations or in multi-institutional clinical trials.

This proposed QC system is currently being implemented in 2 multicenter randomized trials for patients with locoregionally advanced cervical cancer. The first is the phase 3 component of INTERTECC which sets positron emission tomography image guided bone marrow−sparing IMRT against non−bone marrow−sparing radiation techniques. The second is NRG-GY006 (National Clinical Trials

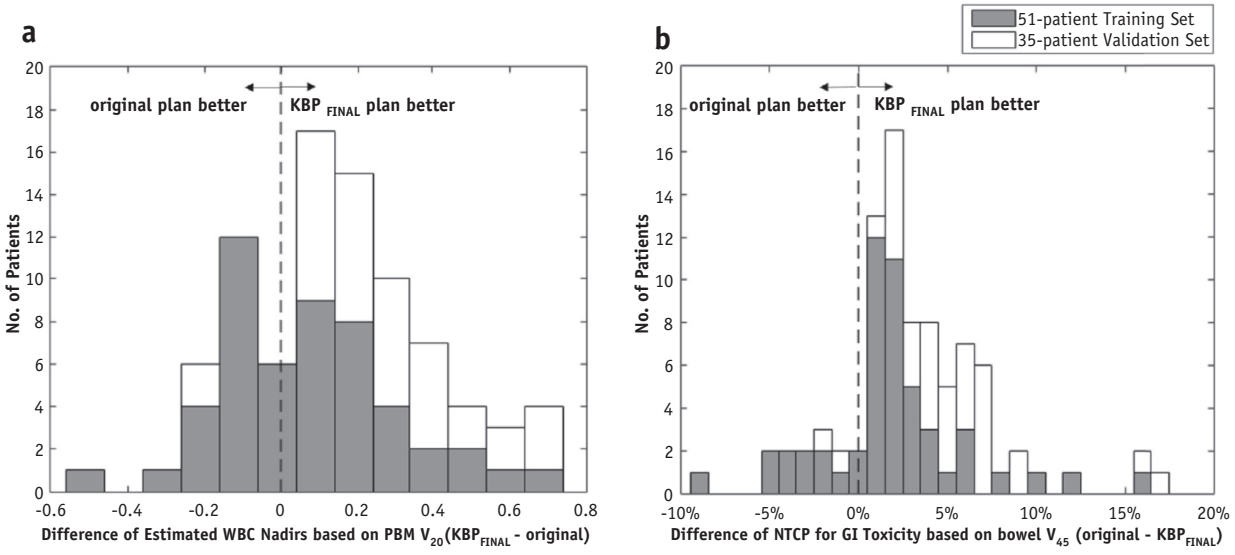

Fig. 4. Frequency histograms of difference in estimated white blood cell count (WBC, measured in k/mL) nadir (a) and normal tissue complication probability (NTCP) (b) between original and KBP_{FINAL} plans across training set (gray) and validation set (white). Abbreviations: GI = gastrointestinal; PBM = pelvic bone marrow; KBP_{FINAL} = final KBP plan generated using the best routine; V_n = percentage or absolute volume receiving at least n Gy dose.

Network identifier 02466971), a phase 2 randomized trial of concurrent chemoradiation therapy with or without **triapine (3-aminopyridine-2-carboxaldehyde thiosemicarbazone, NSC #663249)** As using IMRT to limit radiation dose to OARs and reduce toxicity is a key objective in both trials, these will serve as excellent test beds for prospective evaluation of KBP as a QC system.

To demonstrate what the KBP feedback might look like in practice, we have included an example pretreatment KBP-driven quality report in Appendix E1 (available online at www.redjournal.org). After the participating centers submitted their plans, a patient-specific reference plan was automatically generated for each of the patients using the KBP_{FINAL} routine presented in this article. (This entire process requires only minutes to complete, perfect for the rapid turnaround required in pretreatment quality review.) Subsequently, the submitted plan and the KBP_{FINAL} plan are directly evaluated against each other and against the protocol criteria. Using this kind of report, which is generated automatically by running another ESAPI standalone script, study coordinators and participating centers can be assured that the plan submissions are benchmarked not only against the population-based dosimetric objectives of the protocol but against knowledge-based, patient-specific standards as well.

A limitation of our study was that our training models used only fixed-field IMRT plans rather than volumetric modulated arc therapy plans, which may have narrowed the generality of the KBP model. Work is ongoing to build robust volumetric modulated arc therapy−based models with active bone marrow sparing in larger samples, for applications in both the clinic and clinical trials. While our initial studies have shown that using the refined DVH predictions for bone marrow and bowel from this work has some predictive power for active bone marrow sparing and simultaneous integrated boost cases, further investigation is required to demonstrate this conclusively.

One further limitation is that the sample was necessarily geometrically heterogeneous from the wide stage range of the trial inclusion criteria, notably the 13 patients treated with extended fields. While we did not observe any noticeably aberrant behavior of the models and routines in this subsample, the relatively small statistical number of extended-field patients was not sufficient to ascertain whether the wide anatomic span across the clinical sample had important geometric and/or dosimetric differences that might manifest themselves on expansion to a larger patient sample. Addressing this statistical limitation will likely be possible on application of these techniques in INTERTECC phase 3 and NRG-GY006 trials.

In conclusion, we developed a highly efficient method to train, refine, and validate a KBP automated planning system. Model refinement was accomplished according to clinical trial dosimetric objectives, and the best of several auto-planning routines was automatically identified using a program that automatically re-planned multiple prior patients without human intervention. The final KBP routine showed improved normal tissue sparing across both training and validation samples and will be incorporated as an automated QC system in future cervical cancer clinical trials.

References

1. 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;2:296-305.
2. Das IJ, Cheng CW, Chopra KL, et al. Intensity-modulated radiation therapy dose prescription, recording, and delivery: Patterns of variability among institutions and treatment planning systems. *J Natl Cancer Inst* 2008;100:300-307.
3. Appenzoller LM, Michalski JM, Thorstad WL, et al. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys* 2012;39:7446-7461.
4. Moore KL, Mutic S, Brame RS, et al. Developing predictive dose-volume relationships for a radiotherapy treatment. US patent application US20120310615A1. December 6, 2012.
5. Moore KL, Brame RS, Low DA, et al. Quantitative metrics for assessing plan quality. *Semin Radiat Oncol* 2012;22:62-69.
6. Moore KL, Schmidt R, Moiseenko V, 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;92:228-235.
7. Eaton BR, Pugh SL, Bradley JD, et al. Institutional enrollment and survival among NSCLC patients receiving chemoradiation: NRG Oncology Radiation Therapy Oncology Group (RTOG) 0617. *J Natl Cancer Inst* 2016;108:djw034.
8. Wu B, Ricchetti F, Sanguineti G, et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. *Med Phys* 2009;36:5497-5505.
9. Moore KL, Brame RS, Low DA, et al. Experience-based quality control of clinical intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys* 2011;81:545-551.
10. Wu B, Ricchetti F, Sanguineti G, et al. Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys* 2011;79:1241-1247.
11. Wu B, McNutt T, Zahurak M, et al. Fully automated simultaneous integrated boosted-intensity modulated radiation therapy treatment planning is feasible for head-and-neck cancer: A prospective clinical study. *Int J Radiat Oncol Biol Phys* 2012;84:e647-e653.
12. Wu B, Pang D, Simari P, et al. Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: A head-and-neck case study. *Med Phys* 2013;40:021714.
13. Good D, Lo J, Lee WR, et al. 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;87:176-181.
14. Zhu X, Ge Y, Li T, et al. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys* 2011;38:719-726.
15. Moore K, Appenzoller L, Tan J, et al. Clinical implementation of dose-volume histogram predictions for organs-at-risk in IMRT planning. *J Phys Conf Ser* 2014;489:012055.
16. Yuan L, Wu QJ, Yin FF, et al. Incorporating single-side sparing in models for predicting parotid dose sparing in head and neck IMRT. *Med Phys* 2014;41:021728.

17. Shiraishi S, Tan J, Olsen LA, et al. Knowledge-based prediction of plan quality metrics in intracranial stereotactic radiosurgery. *Med Phys* 2015;42:908.
18. Varian Medical Systems. Eclipse Photon and Electron Instructions for Use, Chapter 11 DVH Estimation Models, 2014;183-199.
19. Varian Medical Systems. Eclipse Photon and Electron Algorithms Reference Guide, Chapter 10 DVH Estimation Algorithm, 2014; 220-229.
20. Varian Medical Systems. Eclipse Photon and Electron Reference Guide, Chapter 11 Configuring DVH Estimation Models, 2014;263-329.
21. Simpson DR, Song WY, Moiseenko V, et al. Normal tissue complication probability analysis of acute gastrointestinal toxicity in cervical cancer patients undergoing intensity modulated radiation therapy and concurrent cisplatin. *Int J Radiat Oncol Biol Phys* 2012;83:e81-e86.
22. Rose BS, Aydogan B, Liang Y, et al. Normal tissue complication probability modeling of acute hematologic toxicity in cervical cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2011;79:800-807.
23. Tol JP, Delaney AR, Dahele M, et al. Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol Biol Phys* 2015;91:612-620.
24. Fogliata A, Belosi F, Clivio A, 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;113:385-391.
25. 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; 9:236.
26. 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;10:234.