HHS Public Access

HIGHLY EFFICIENT TRAINING, REFINEMENT, AND VALIDATION OF A KNOWLEDGE-BASED PLAN QUALITY CONTROL SYSTEM FOR RADIOTHERAPY CLINICAL TRIALS

Nan Li, Ph.D.[1], **Ruben Carmona, M.D.**[1], **Igor Sirak, M.D., Ph.D.**[2], **Linda Kasaova, Ph.D.**[2], **David Followill, Ph.D.**[3], **Jeff Michalski, M.D.**[4], **Walter Bosch, Ph.D.**[4], **William Straube, MS.**[4], **Loren K. Mell, M.D.**[1], and **Kevin L. Moore, Ph.D.**[1]

[1]Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA

[2]Department of Oncology and Radiotherapy, University Hospital, Hradec Kralove, Czech Republic

[3]Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, TX

[4]Department of Radiation Oncology, Washington University in St. Louis, St. Louis, MO

Abstract

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

Methods—We analyzed 86 stage IB-IVA cervical cancer patients treated with intensity modulated radiotherapy (IMRT) at two institutions according to XXXXX multi-institutional protocol standards. The protocol utilized a planning target volume (PTV) and two primary organs-at-risk (OARs): pelvic bone marrow (PBM) and bowel. Secondary OARs were rectum and bladder. Initial UNFILTERED dose-volume histogram (DVH) estimation models were trained using all 86 plans. REFINED training sets and DVH-estimation models were constructed by identifying 30/86 plans emphasizing PBM-sparing (comparing protocol-specified V_{10} and V_{20} with UNFILTERED predictions), and another 30/86 plans emphasizing bowel sparing (comparing V_{40} and V_{45}, 9 in common with PBM set). To obtain deliverable KBP plans, REFINED models must inform patient-specific optimization objectives/priorities (an auto-planning "routine"). Four candidate routines emphasizing different tradeoffs were composed and a script was developed to

---

Corresponding Author: Kevin L. Moore, Ph.D., Department of Radiation Medicine and Applied Sciences, 3960 Health Sciences Drive, MC0865, La Jolla, CA 92093, kevinmoore@ucsd.edu, (858) 822-6056.

automatically re-plan multiple patients with each routine. After selection of the routine best meeting protocol objectives in the 51-patient training sample (KBP_{FINAL}), protocol-specific DVH metrics and normal tissue complication probability (NTCP) were compared for original vs. KBP_{FINAL} plans across the 35-patient validation set. Paired t-tests tested differences between planning sets.

Results—KBP_{FINAL} plans outperformed manual planning across the validation set in all protocol-specific DVH cutpoints. The mean NTCP for GI toxicity was lower for KBP_{FINAL} vs. validation-set plans (48.7% vs. 53.8%, p<.001). Similarly, the estimated mean white blood cell count (WBC) nadir was higher (2.77 vs 2.49, 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 radiotherapy clinical trials with minimal effort. This patient-specific plan QC 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 higher clinical trial accruing centers was associated with longer overall survival in 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 utilized KBP on a large-scale multi-institutional clinical trial (RTOG 0126) to assess the frequency and clinical severity of sub-optimal treatment plans (6). The results indicated that patient-specific plan QC tools, e.g. KBP-driven DVH estimation, should be part of the quality apparatus of multi-institutional radiotherapy trials, particularly if the 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/priorities derived from these estimations. The patient-specific DVH estimations result from mathematical models that correlate patient anatomical 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 multi-variable plan quality problem posed by multiple OARs and planning target volumes (PTV(s))(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 compels 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 quality control system for clinical trials, we developed a highly efficient method to train, refine and automatically validate a KBP system on a cohort of patients for implementation in trials with a radiotherapy component. The aim is to demonstrate a streamlined process that could work on any radiotherapy 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," i.e. with no manual effort.

As a test bed for this method, we focused on the Phase II clinical trial X (NCTN identifier: XXXX), which was an international multi-institutional trial of bone marrow-sparing IMRT with a primary endpoint of reducing hematologic and gastrointestinal 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 II component would be ideal for implementation as a quality system on the critical Phase III.

METHODS AND MATERIALS

Patient Sample

We selected 86 stage IB-IVA cervical cancer patients treated with fixed-field IMRT according to trial X guidelines: 63 patients from XXXXX and 23 from YYYYY. 26 patients were participants in the clinical trial and 60 patients were treated off-trial (e.g., due to refusal), or on a competing clinical trial using the same treatment planning approach. All patients received 45.0–50.4 Gy in 1.8 Gy daily fractions over 5–5.5 weeks with concurrent chemotherapy. 73 patients were treated with pelvic IMRT fields only and 13 patients were treated with extended field (pelvic/para-aortic) IMRT fields. The clinical target volume (CTV) 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; 10 mm around vagina and parametria; 5–7 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 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. Initially, 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 dose-volume histogram information(19). Application of the RapidPlan modeling routine to all 86 plans resulted in an initial UNFILTERED model.

Since the results that the model produces are roughly dependent on the mean quality of the training plans(18), a strategy to identify high quality plans for a refined training set for each OAR is critical(3,17). We decided to select the top 30 (~1/3 of the entire sample) plans with an apparent emphasis on PBM or bowel sparing to generate refined training sets for these models, since these are 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} and V_{20} to the UNFILTERED model DVH estimations. Another 30 plans were selected in the same manner for bowel optimization by comparing V_{40} and V_{45}, nine of which were found to be in common with the PBM sample. As PBM and bowel were the two primary OARs, we used the 51-patient samples selected for these two 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 (see Figure 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 get the frontier estimates for both PBM and bowel sparing and then use the optimization routines to properly balance these multi-criterial objectives against each other and achieve acceptable plans (see next sub-section).

For rectum and bladder, a rough refinement was carried out using the model evaluation tools established in RapidPlan(22), i.e. only obvious outliers with clinical DVHs far in excess of their estimated DVHs were removed. In the end, 52 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/PTV criteria, and it is not at all obvious what the right blend of OAR objectives/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 four 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 51*4+35=239 treatment plans 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) to re-plan a large plurality of patients, automatically generating all desired N*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. In order 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 volume with 100% of the prescribed dose before plan comparison.

After generating the 51*4=204 re-plans, the ESAPI-driven system exported the resultant plan data one-by-one for each of the training set patients for aggregate analysis. Following the XXXX 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 (PI), 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 also used validated NTCP models (20,21) that map bowel and PBM DVHs to the probability of developing grade 2 gastrointestinal (GI) toxicity and hematologic toxicity, respectively. 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 .05) differences in DVH metrics and NTCP.

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

RESULTS

Auto-planning routines comparison

To determine the best routine of the four expressions of the REFINED DVH estimation models, the ESAPI-driven system applied the four routines to the 51 training patients and the resultant plans were compared to the original protocol plans. With identical PTV normalization, we focused on the four 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%}, BOWEL routine {53%, 53%, 61%, 88%}. Similarly, the average ± standard deviation DVH metric excess in the clinical sample could be assessed per routine {PBM $V10_{clinical}-V10_{routine}$, PBM $V20_{clinical}-V20_{routine}$, bowel $V40_{clinical}-V40_{routine}$, bowel $V45_{clinical}-V45_{routine}$}: 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%}, 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 PI, 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 (Figure 2). KBP_{FINAL} plans exhibited no significant difference in PTV homogeneity (Original: D1%–D99%=5.15±2.9 Gy, KBP_{FINAL}=5.15±3.1 Gy, p=0.98). On average KBP_{FINAL} lowered the <30 Gy region of the PBM DVH and the >30 Gy region of the bowel DVH compared to original plans.

Secondly, the protocol-specified DVH metrics were also used to do the comparison. Each pair of KBP_{FINAL} and original plans were plotted to show how the plan quality changes for each patient in training set and validation set (Figures 3). As indicated, the quality of most KBP_{FINAL} plans were 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(20), we observed that the mean NTCP for GI toxicity (based on bowel V_{45}) was lower for the 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(21), the estimated mean WBC nadir (based on PBM V_{20}) was higher across the validation set (2.8 vs. 2.5, p<. 001), indicating lower expected hematologic toxicity in the KBP_{FINAL} cohort compared to the original plans. Figure 4 shows the frequency histogram of the difference of NTCP and estimated WBC nadir between original and KBP_{FINAL} plans across the training set (grey) 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 (e.g. 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 Jim P. Tol *et al.* (26) showed the potential of using 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 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 either in normal clinical operations or on multi-institutional clinical trials.

This proposed QC system is currently being implemented in two multi-center randomized trials for patients with locoregionally advanced cervical cancer. The first is the Phase III component of trial X which sets positron emission tomography (PET) image-guided bone marrow-sparing IMRT (IG-IMRT) against non-bone marrow sparing radiation techniques. The second is trial Y (NCTN identifier: XXXX), a Phase II randomized trial of concurrent chemoradiotherapy with or without triapine. As using IMRT to limit radiation dose to OARs and reduce toxicity is a key objective in both trials, these will serve as excellent testbeds 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 pre-treatment KBP-driven quality report in the Supplementary Material. After the participating centers submitted their plans, a patient-specific reference plan was automatically generated for each of the patient using the KBP_{FINAL} routine presented in this manuscript. (This entire process requires only minutes to complete, perfect for the rapid turnaround required in pre-treatment 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 stand-alone script, study coordinators and participating centers can be assured that the plan submissions are not only benchmarked 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 VMAT plans, which may have narrowed the generality of the KBP model. Future work is ongoing to build robust VMAT-based models with active bone marrow sparing in larger samples, for applications in both 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 (SIB) 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 sub-sample, the relatively small statistical numbers 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 upon expansion to a larger patient sample. Addressing this statistical limitation will likely be possible upon application of these techniques on the trial X Phase III and trial Y.

In conclusion, we developed a highly efficient method to train, refine, and validate a KBP system automated planning system. Model refinement was accomplished according to clinical trial 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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. [PubMed: 24674168]

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 United States. 2008:300–7.

3. Appenzoller LM, Michalski JM, Thorstad WL, et al. Predicting dose-volume histograms for organs-at-risk in imrt planning. Med Phys. 2012; 39:7446–61. [PubMed: 23231294]

4. Moore KL, Mutic S, Brame RS, et al. Developing predictive dose-volume relationships for a radiotherapy treatment. Book Developing predictive dose-volume relationships for a radiotherapy treatment: Google Patents. 2012 Editor, editor^editors.

5. Moore KL, Brame RS, Low DA, et al. Quantitative metrics for assessing plan quality. Seminars in radiation oncology. 2012; 22:62–69. [PubMed: 22177879]

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; 92:228–35. [PubMed: 25847605]

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. Journal of the National Cancer Institute. 2016; 108:djw034. [PubMed: 27206636]

8. Wu B, Ricchetti F, Sanguineti G, et al. Patient geometry-driven information retrieval for imrt treatment plan quality control. Med Phys. 2009; 36:5497–505. [PubMed: 20095262]

9. Moore, KL.; Brame, RS.; Low, DA., et al. Int j radiat oncol biol phys. United States: Elsevier Inc; 2011. Experience-based quality control of clinical intensity-modulated radiotherapy planning; p. 545-51.2011

10. Wu, B.; Ricchetti, F.; Sanguineti, G., et al. Int j radiat oncol biol phys. United States: Elsevier Inc; 2011. Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning; p. 1241-7.2011

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:e647–53. [PubMed: 22867890]

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. [PubMed: 23387737]

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; 87:176–81. [PubMed: 23623460]

14. Zhu X, Ge Y, Li T, et al. A planning quality evaluation tool for prostate adaptive imrt based on machine learning. Medical physics. 2011; 38:719–726. [PubMed: 21452709]

15. Moore, K.; Appenzoller, L.; Tan, J., et al. Journal of Physics: Conference Series. IOP Publishing; 2014. Clinical implementation of dose-volume histogram predictions for organs-at-risk in imrt planning; p. 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. [PubMed: 24506619]

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. [PubMed: 25652503]

18. Varian medical systems. Eclipse photon and electron instructions for use. Palo Alto: CA; 2014. p. 183-199.

19. Varian medical systems. Eclipse photon and electron algorithms reference guide. Palo Alto: CA; 2014. p. 220-229.

20. Simpson, DR.; Song, WY.; Moiseenko, V., et al. Int j radiat oncol biol phys. United States: Elsevier Inc; 2012. Normal tissue complication probability analysis of acute gastrointestinal toxicity in cervical cancer patients undergoing intensity modulated radiation therapy and concurrent cisplatin; p. e81-6.2012

21. 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; 79:800–7. [PubMed: 20400238]

22. Varian medical systems. Eclipse photon and electron reference guide. Palo Alto: CA; 2014.

23. Tol JP, Delaney AR, Dahele M, et al. Evaluation of a knowledge-based planning solution for head and neck cancer. International Journal of Radiation Oncology* Biology* Physics. 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. Radiotherapy and Oncology. 2014; 113:385–391. [PubMed: 25465726]

25. Fogliata A, Wang P-M, Belosi F, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiation Oncology. 2014; 9:1. [PubMed: 24382205]

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? Radiation Oncology. 2015; 10:1. [PubMed: 25567003]
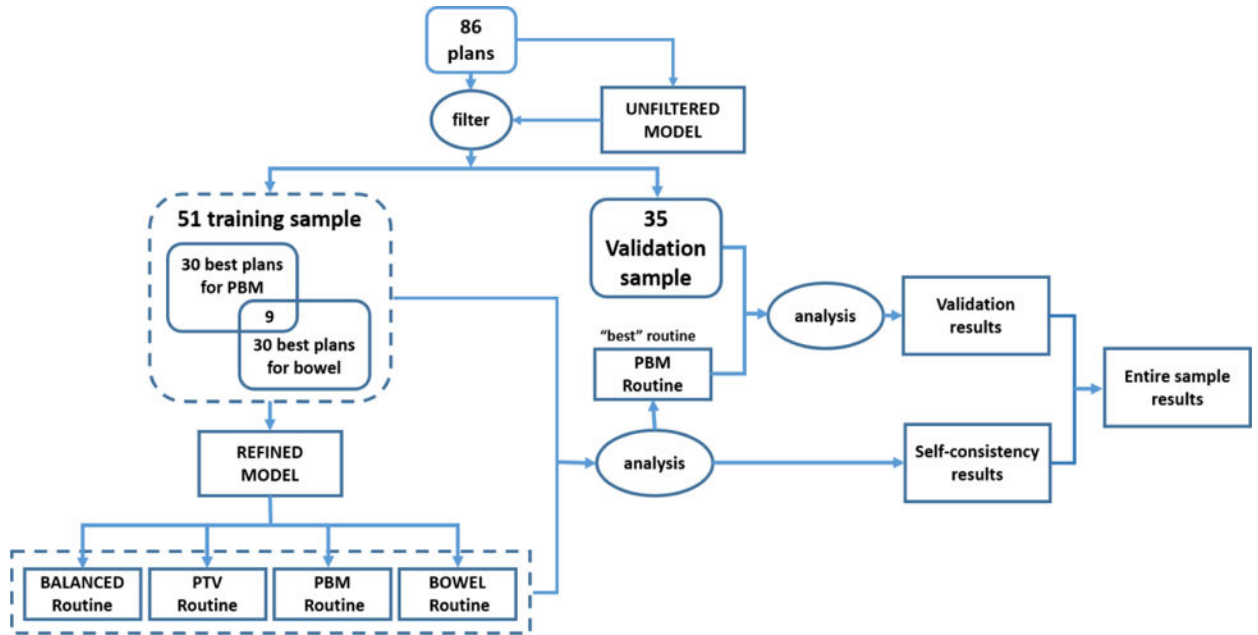
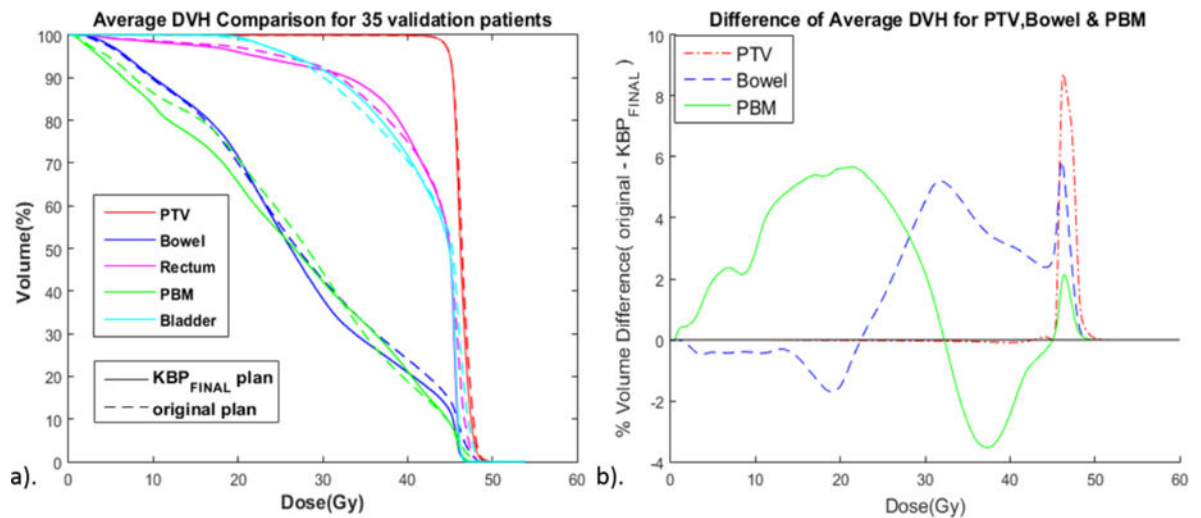**Figure 1.**
Flow Diagram of this study.

Figure 2.
Comparison of average DVHs of the original and KBP$_{FINAL}$ plans for 35 validation patients.
a). Average DVH comparison. PTV, planning target volume; PBM, pelvic bone marrow. b).
Difference of average DVH between original and KBP$_{FINAL}$. Note that the regions of
improvement are in the protocol-specified 10–20 Gy region of the PBM DVH and in the
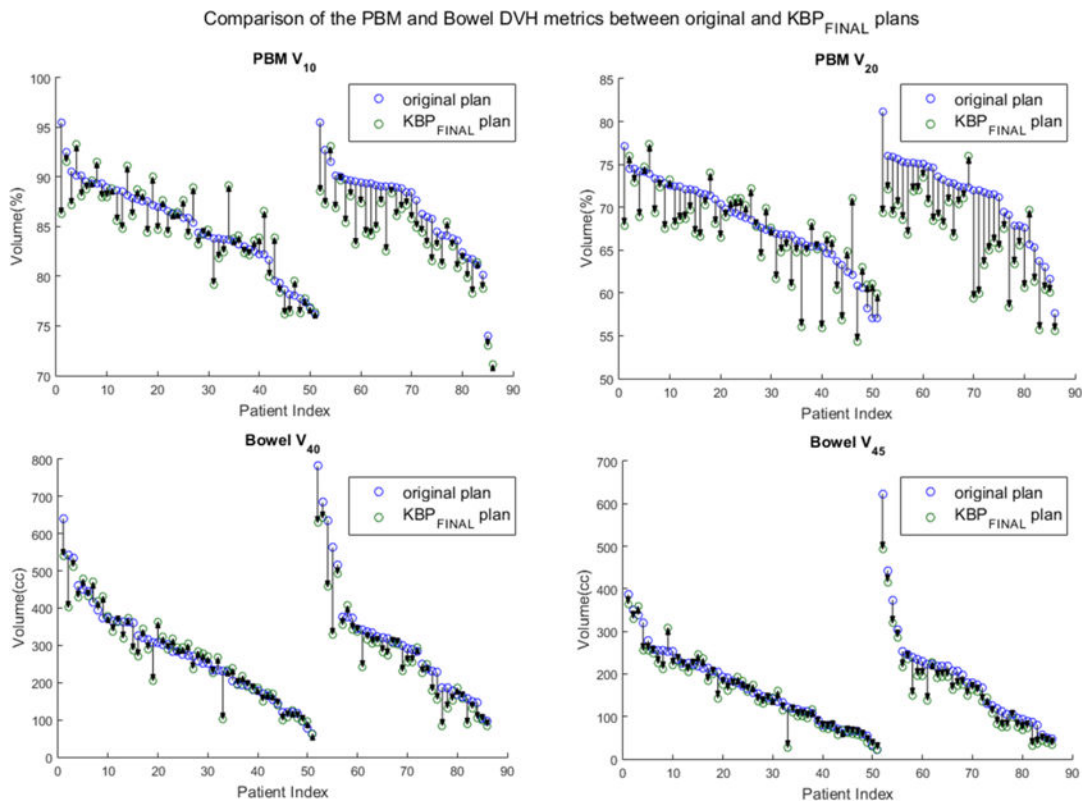high dose region >40 Gy region of the bowel DVH.

Figure 3.
Comparison of the PBM and Bowel DVH metrics between original and KBP$_{FINAL}$ in the training and validation sets. PBM a) V$_{10}$ and b) V$_{20}$ comparison between original plan and KBP$_{FINAL}$ for each patient. Bowel c) V$_{40}$ and d) V$_{45}$ comparison between original plan and KBP$_{FINAL}$ for each patient.
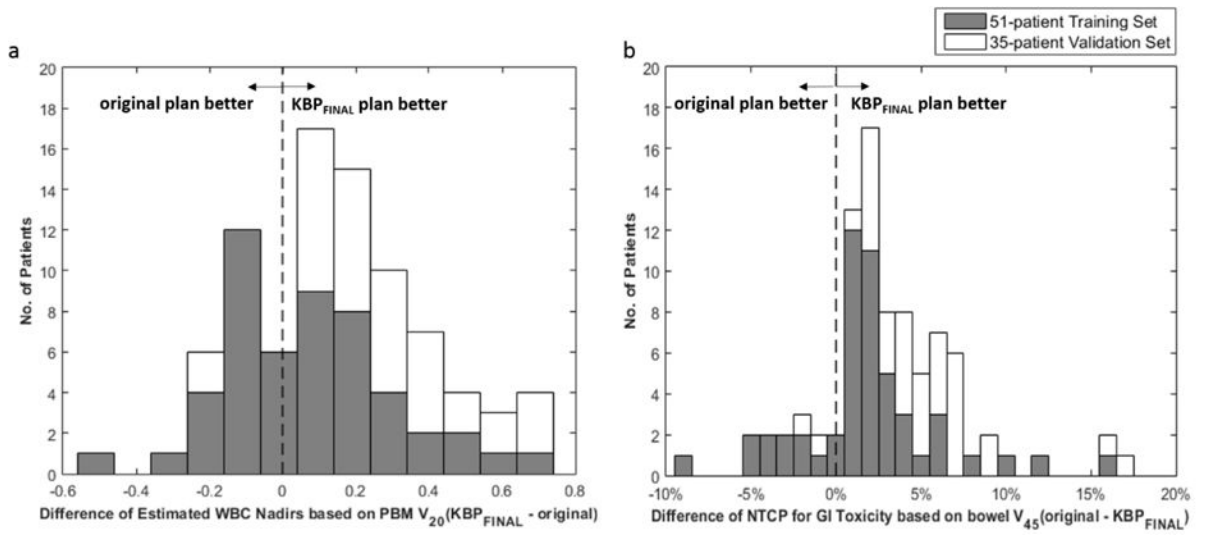
Figure 4.
Frequency histogram of the difference of NTCP and estimated WBC nadir between original and KBP$_{FINAL}$ plans across the training set (grey) and validation set (white).

Table 1

Optimization and priority settings Priorities

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

Structure	Objectives	Priorities				
		BALANCED routine	PTV routine	PBM routine	BOWEL routine	
Rectum	$D_{max} < Rx$	80	80	80	80	
	$V_{20} <$ Model-Generated	50	50	50	50	
	$V_{40} <$ Model-Generated	50	50	50	50	

PTV, planning target volume; PBM, pelvic bone marrow; Rx, prescription

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	0.98
	D$_{99\%}$ (Gy)	44.1±1.7	43.7±3.4	0.43
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$_{40}$ (cc)	268.0±139.1	311.4±159.7	<.001
	V$_{45}$ (cc)	158.3±103.9	190.2±116.3	<.001
Rectum	V$_{30}$ (%)	92.0±5.8	92.7±10.4	0.65
	V$_{45}$ (%)	51.7±20.4	52.1±20.1	0.79
	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	<0.01

IMRT, intensity modulated radiation therapy; DVH, dose-volume histogram; PTV, planning target volume; PBM, pelvic bone marrow.