Clinical implementation of dose-volume histogram predictions for organs-at-risk in IMRT planning

K L Moore, L M Appenzoller, J Tan, J M Michalski, W L Thorstad and S Mutic

1 Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, CA 92093, USA
2 Department of Radiation Oncology, Washington University in St. Louis, St. Louis, MO 63110, USA
E-mail: kevinmoore@ucsd.edu

Abstract. True quality control (QC) of the planning process requires quantitative assessments of treatment plan quality itself, and QC in IMRT has been stymied by intra-patient anatomical variability and inherently complex three-dimensional dose distributions. In this work we describe the development of an automated system to reduce clinical IMRT planning variability and improve plan quality using mathematical models that predict achievable OAR DVHs based on individual patient anatomy. These models rely on the correlation of expected dose to the minimum distance from a voxel to the PTV surface, whereby a three-parameter probability distribution function (PDF) was used to model iso-distance OAR subvolume dose distributions. DVH models were obtained by fitting the evolution of the PDF with distance. Initial validation on clinical cohorts of 40 prostate and 24 head-and-neck plans demonstrated highly accurate model-based predictions for achievable DVHs in rectum, bladder, and parotid glands. By quantifying the integrated difference between candidate DVHs and predicted DVHs, the models correctly identified plans with under-spared OARs, validated by replanning all cases and correlating any realized improvements against the predicted gains. Clinical implementation of these predictive models was demonstrated in the PINNACLE treatment planning system by use of existing margin expansion utilities and the scripting functionality inherent to the system. To maintain independence from specific planning software, a system was developed in MATLAB to directly process DICOM-RT data. Both model training and patient-specific analyses were demonstrated with significant computational accelerations from parallelization.

1. Introduction
Recent studies focused on IMRT treatment plan quality have quantified that the use of inverse-planned radiotherapy does not guarantee a truly optimal treatment plan for every patient. IMRT treatment plan quality assessments are currently compromised by their reliance on individual expertise, alertness, and subjective evaluation, with no effective quantitative measures to ensure high plan quality with respect to what is actually achievable. Plan quality variations can be so great that a significant proportion of patients who should have been at low risk of radiation-induced complications

3 To whom any correspondence should be addressed.
are at much higher risk for poor outcome.6
To address this problem, we have developed a methodology of model-based algorithms for IMRT treatment plan dose-volume histogram prediction (pDVH) which use patient-specific anatomical features to predict achievable organ at risk (OAR) DVHs. The pDVHs facilitate quantitative assessment of measured vs. expected values for OAR dosimetric parameters, and in this work we describe two separate clinical implementations of this quality control (QC) method.

2. Model-based predictions for OAR DVHs in IMRT
Our framework for model-based predictions of OAR DVHs was described in Ref. [1], and here we summarize the method and primary results. The initial assumption for this work involves the identification of a cohort of \( N \) site-similar plans that were developed using identical clinical goals and quality assessment criteria. The planning datasets are comprised of structure sets \( SS_{ij} \) \((i=1...N \) cases, \( j=1...M \) structures with \( j=1 \) representing the PTV and \( j=2...M \) representing \( M-1 \) OARs) and dose matrices \( D_i \) \( (x) \), where \( x \) is the 3-D position vector with arbitrary origin. Operations on \( D_i \), \( SS_{ij} \), and derivative structures result in differential and cumulative DVHs for volumes of interest.

Figure 1 depicts the pDVH process, whereby the boundary distance vector \( \bar{r}(x) \) forms the geometric information that will be correlated to the measured DVHs in the training sets. (The use of this quantity has also been explored by other investigators2,4.)

---

**Figure 1.** Outline of the pDVH method. (a) A cohort of \( N \) prior patients form a training set, where (b) the boundary distance vector field quantifies the geometric arrangement of OAR voxels around PTV(s). (c) OAR voxels are grouped into iso-distance shells, themselves structures with differential DVHs, or “sub-DVHs” of a full OAR. (d) A skew-normal probability function forms a three-parameter basis function that is used to fit the measured iso-distance sub-DVHs, and (e) the evolution of the skew-normal parameters as a function of distance results in (f) the ability to compute pDVHs for any structure set \( SS_{ij} \) on new or prior patients, i.e. in the absence of any \( D_i \).
Clinical cohorts of 20 prostate and 24 head-and-neck plans trained pDVH models for rectum, bladder, and parotids. Sums-of-residuals (SR) quantified the integrated difference between clinical DVHs and model-predicted DVHs. The predictive ability for new patients was initially evaluated on a cohort of 20 prostate plans. The models identified plans with large SR values, indicating clinical DVHs that significantly exceeded model-predicted DVHs. All plans were replanned, and the correct identification of outliers was assessed by quantifying the correlation between realized dosimetric gains (SR_{replan}-SR_{original}) and the model-predicted gains (SR_{original}). The OAR DVHs of the training cohorts were accurately modelled with small mean sums-of-residuals compared to standard deviations. The training and validation sets’ sums-of-residuals were statistically indistinguishable, confirming predictive power in new patients. The models exhibited excellent suboptimal plan detection with large correlation coefficients between predicted and realized gains: $r_{\text{rectum}}=0.92$, $r_{\text{bladder}}=0.88$, and $r_{\text{parotid}}=0.84$.

3. Clinical implementation in Philips PINNACLE treatment planning system

From the perspective of clinical implementation, it is advantageous to incorporate quality control elements directly into the treatment planning system as this reduces the need for extraneous data transfers during the planning process. The results of the prostate pDVH models were encoded into a PINNACLE script that can be executed by the planner for any new patient, and has been in use at Washington University in St. Louis for over a year. The script functions by successive use of PINNACLE’s contour expansion and Boolean operations to generate the iso-distance shells depicted in Figure 1c. While this is a very computationally inefficient method, it is mathematically equivalent to direct computation of the boundary distance field. The script then computes volume and sub-DVH information for the OAR shells, which can be used to compute a pDVH for the active plan. As shown in Figure 2, multiple pDVH models (derived from different training sets) can be presented as feedback to the user who can compare the current DVH to the pDVH(s). In addition to the full DVH curves, discrete DVH cut points can be encoded and displayed, as well as red flags for situations where the active plan exceeds expected variation. With a simple extension of the PINNACLE scripting routines it was also possible to convert a pDVH into IMRT optimization objectives, a facilitating step for automated treatment planning.

![Figure 2. Clinical pDVH module for prostate implemented with PINNACLE scripting. (a) AVERAGE models (trained with uncontrolled prior patient sample) and REFINED models (trained with high plan quality patient sample) for rectum and bladder allow reference pDVHs to be shown aside the current DVHs of an active PINNACLE plan. (b) If the sum-of-residuals analysis shows that the plan is more than one standard deviation outside of the norm, the clinical DVH is flagged in red. Report also shows mean and max dose, as well as $V_{40}$ and $V_{65}$, for comparison.](image)

4. A universal system for predictive DVH modeling and analysis

To support all treatment planning systems (TPS), we developed an efficient DICOM-based tool to create pDVH models for IMRT QC independent of the TPS. RT dose and structure files exported from
different TPSs including Varian’s Eclipse, Philips’ PINNACLE\(^3\), and Accuray’s Tomotherapy as input to the pDVH DICOM tool (Figure 3).

To date, models have been created for many different treatment sites (intact prostate, prostate bed, head-and-neck, endometrial, cervical, CNS, and lung). MATLAB was used to implement multiple techniques that accelerate the computation of a distance vector field for efficient model building. The program can run on GPUs or, if GPUs are not available, in parallel on multiple CPU cores. The inclusion of GPUs reduces computation time by 200 times, requiring <1 second to compute an OAR vector field and only a few seconds to build a pDVH model. The tool can run in batch mode with minimal user interaction to create a pDVH model for a cohort of patients with a specified treatment site. The tool automatically identifies outliers for treatment plan quality control and pDVH model refinement (Figure 3b).

**Figure 3.** pDVH module in MATLAB. (a) Program interface allows user to import patients and efficiently train new OAR pDVH models. (b) “Validate & refine” module of system allows for the elimination of suspected outliers via residuals analysis, accomplished by removing plans from the training of refined model if their sum-of-residuals score exceeds the standard deviation on any OAR.

5. Summary and conclusions

We have demonstrated that the pDVH methodology can serve as an efficient tool for QC of IMRT treatment plans, providing clinicians a novel and highly efficient plan quality control mechanism by predicting OAR DVHs based on local institutional best practices. This technology can be used to facilitate inter-institutional and inter-TPS quality comparisons based on aggregate clinical data instead of contrived benchmark planning sets, as well as facilitating automated planning routines.

6. References

[1] Appenzoller L M, Michalski J M, Thorstad W L, Mutic S and Moore K L 2012 *Med Phys* **39** (12), 7446
[2] Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Chuang M, Taylor R, Jacques R and McNutt T 2009 *Med Phys* **36** (12), 5497-5505
[3] Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Jacques R, Taylor R and McNutt T 2011 *Int J Radiat Oncol Biol Phys* **79** (4), 1241-1247
[4] Zhu X, Ge Y, Li T, Thongphiew D, Yin F F and Wu Q J 2011 *Med Phys* **38** (2), 719-726
[5] Siochi R A, Balter P, Bloch C D, Santanam L, Blodgett K, Curran B H, Engelsman M, Feng W, Mechalakos J, Pavord D, Simon T, Sutlieff S and Zhu X R 2011 *J Appl Clin Med Phys* **12** (1), 3479
[6] Moore K L, Brame R S, Low D A and Mutic S 2011 *Int J Radiat Oncol Biol Phys* **81** (2), 545-551
[7] Yang D and Moore K L 2012 *Med Phys* **39** (3), 1542-1551