Article Type: Research Article

A method for \textit{a priori} estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning

Saeed Ahmed

\textit{Department of Physics, University of South Florida, Tampa, FL, 33612}

Benjamin Nelms

\textit{Canis Lupus LLC, Merrimac, WI, 53561}

Dawn Gintz

\textit{Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, 33612}

Jimmy Caudell

\textit{Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, 33612}

Geoffrey Zhang

\textit{Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, 33612}

Eduardo G. Moros

\textit{Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, 33612}

Vladimir Feygelman

\textit{Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, 33612}

Vladimir.feygelman@moffitt.org

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.12500

This article is protected by copyright. All rights reserved.
ABSTRACT

Purpose

Despite improvements in optimization and automation algorithms, the quality of radiation treatment plans still varies dramatically. A tool that allows \textit{a priori} estimation of the best possible sparing (Feasibility DVH, or FDVH) of an organ at risk (OAR) in high-energy photon planning may help reduce plan quality variability by deriving patient-specific OAR goals prior to optimization. Such a tool may be useful for 1) meaningfully evaluating patient-specific plan quality and 2) supplying best theoretically achievable DVH goals, thus pushing the solution towards automatic Pareto-optimality. This work introduces such a tool and validates it for clinical Head and Neck (HN) datasets.

Methods

To compute FDVH, first the targets are assigned uniform prescription doses, with no reference to any particular beam arrangement. A benchmark 3D dose built outside the targets is estimated using a series of energy-specific dose-spread calculations reflecting observed properties of radiation distribution in media. For the patient, the calculation is performed on the heterogeneous dataset, taking into account the high- (penumbra-driven) and low- (PDD and scatter-driven) gradient dose spreading. The former is driven mostly by target dose and surface shape, while the latter adds the dependence on target volume. This benchmark dose is used to produce the “best possible sparing” FDVH for an OAR, and based on it, progressively more easily achievable FDVH curves can be estimated. Validation was performed using test cylindrical geometries as well as ten clinical HN datasets. For HN, VMAT plans were prepared with objectives of covering the primary and the secondary (bilateral elective neck) PTVs while

This article is protected by copyright. All rights reserved.
addressing only one OAR at a time, with the goal of maximum sparing. The OARs were each parotid, the larynx, and the inferior pharyngeal constrictor. The difference in mean OAR doses was computed for the achieved vs. FDVHs, and the shapes of those DVHs were compared by means of the Dice similarity coefficient (DSC).

Results

For all individually optimized HN OARs (N=38) the average DSC between the planned DVHs and the FDVHs was 0.961±0.018 (95% CI 0.955 – 0.967), with the corresponding average of mean OAR dose differences of 1.8 ± 5.8% (CI -0.1 – 3.6%). For realistic plans the achieved DVHs run no lower than the FDVHs, except when target coverage is compromised at the target/OAR interface.

Conclusions

For the validation VMAT plans, the OAR DVHs optimized one-at-a-time were similar in shape to and bound on the low side by the FDVHs, within the confines of planner’s ability to precisely cover the target(s) with the prescription dose(s). The method is best suited for the OARs close to the target. This approach is fundamentally different from “knowledge-based planning” because it is 1) independent of the treatment plan and prior experience, and 2) it approximates, from nearly first principles, the lowest possible boundary of the OAR DVH, but not necessarily its actual shape in the presence of competing OAR sparing and target dose homogeneity objectives.

Keywords: treatment planning, IMRT optimization, dose volume histograms, organ at risk dose, treatment plan quality
I. INTRODUCTION

Radiation treatment planning today is a combination of science and art, described in 2014 by Moore et al as “a very time-consuming task with great output variability.” An inter-institutional study by Nelms et al demonstrated that for the same CT and pre-segmented structure set, and using a fully transparent, objective multi-criteria scoring algorithm (to avoid any variability associated with differing treatment planning protocols), there was dramatic variability in plan quality. This variability could not be assigned to obvious technical factors such as delivery modality or the specific treatment planning system (TPS) employed, and therefore was largely attributed to a nebulous variable of “planner skill.”

Attempts to improve the quality of treatment planning on the whole would include increasing the average plan quality while decreasing variability across all plans. One logical and important approach is to improve the training of professional planners and incorporate objective metrics by which to benchmark performance and guide continual improvement. Another approach is to strive towards more effective, computer-aided automation. Plan automation is a popular topic today with several commercial strategies already available, such as (1) explicitly mimicking the steps taken by experienced users (AutoPlanning, AP), (2) referencing the database of previous similar plans (knowledge-based planning, KBP), and (3) developing plans based on multi-criteria optimization (MCO).

It could be argued that any automation strategy might benefit if the system had a priori expectations of the ideal achievable results based on each patient’s unique anatomy. Such knowledge could be used in two main ways: (1) during optimization, to provide superior inputs (as compared to standard tolerances used across all patients) that might allow the automated planning to exceed standard goals while avoiding chasing impossible ones, and (2) after
optimization, to help gauge plan quality by comparing achieved results to theoretical but patient-specific limits. The current commercial AP method does not generate patient-specific optimization goals but recently plan quality was shown to improve when a priori feasibility estimates are fed into the AP objectives instead of generic standard objectives. In published MCO approaches, some offer exploration of the de novo generated Pareto surfaces while others are used to generate a constraint list that performs well as a class solution for patients with the same tumor type. As for KBP, though the details of the specific methods differ, a common theme is to build a database derived from previously-designed plans for generally similar disease presentations, and from them predict the achievable results for any new patient dataset – a machine learning approach. This method was applied to quality control and, more recently, to the design of the plans. One challenge of KBP is that it requires careful selection of learning plans. To that end, a priori dosimetric feasibility data could serve as the basis for additional patient-specific benchmarks, complementing the more basic pass/fail inspections of population-based objectives. Such tools could enhance KBP, by helping to select the high quality plans for the learning database and providing an additional measure of quality control for the output plans. They could also be useful in other approaches to treatment planning automation as well as for evaluation of traditional manual plans.

Towards these aims, a tool called Feasibility DVH (FDVH) was introduced in the PlanIQ software (Sun Nuclear Corp., Melbourne, FL). The primary goal of this tool is to estimate the best case sparing (lowest possible) DVH for the OARs of any patient, given the full coverage target volumes by the prescribed doses. It is intended as a guide to help generate challenging patient-specific dose objectives, not as a “predicted DVH” engine. The secondary goal is to
allow scored plan metrics, components of a composite Plan Quality Metric (PQM) formalism described previously, \textsuperscript{2,5} to be adjusted for the challenges presented by patient-specific anatomy.

The goal of this paper is to describe the algorithm behind the FDVH as implemented in PlanIQ v. 2.1.2, and to validate its performance across a series of head and neck (HN) cancer patients.

II. METHODS

II.A. Feasibility calculation

II.A.1. Major steps

The process is broken into three steps: 1) specify target volumes, prescription doses, and calculation parameters, 2) build the benchmark dose grid, and 3) generate the FDVH curves. The method presented here does not require any user set up of radiation machine details or parameters. After the user performs step 1, the software executes the rest. The details of these steps are presented in the following sections.

II.A.2. Target doses and calculation parameters

Patient anatomy data are required as input, including at a minimum the contoured target volumes and OARs. To model the effects of the patient density, the CT dataset is necessary. The user selects the planning target volumes (PTVs) and their respective prescription dose levels. The user is prompted to enter a few calculation parameters, some of which are dose grid spatial resolution, beam energy, and an option to correct for CT-based density. If the latter option is declined, a water-equivalent volume is assumed.
II.A.3. Benchmark dose grid

The benchmark dose (BD) is a fictitious 3D dose grid that ensures 100% coverage of each of the target volumes with its respective prescription dose, and then estimates the minimal dose that any voxel outside the targets must receive given the specified target coverage. The BD is unachievable by design, and is ultimately used as the basis for the feasibility estimates. The calculation of the BD is a multi-phased process, described below and illustrated in Figure 1.

Figure 1. Summary of the FDVH process. A,C,E: Benchmark dose (BD) distributions resulting from assigning uniform doses to the targets (2cm cylinder, 10 cm cylinder, and realistic HN, respectively). B,D,F: Corresponding BD profiles (along the arrows) from the border of the target showing High Gradient Dose Spread (HGDS) and Low Dose Spread (LDS) components separately as solid lines, and the final combination as circles (see Eq. (6)).
In all equations, the coordinate system is the right-handed IEC couch coordinate system, with $Y$ increasing into the gantry and $Z$ upward normal to the couch surface.

For each target, in the specified layer overlap order, the dose grid points $[x, y, z]$ inside it are assigned the corresponding dose. During this process, target “edge voxels” are also tagged and stored for use in subsequent steps. Note that if the OAR overlaps the target, the latter “owns” the common voxels and that will be reflected in the OAR DVH. The output is a simple target coverage grid:

$$D_{\text{Coverage}}[x, y, z] = \begin{cases} D[x, y, z]_{Rx,\text{topmost}}, & \text{for voxels inside target(s)} \\ 0, & \text{for voxels outside target(s)} \end{cases}$$

(1)

Next, the algorithm starts assigning dose to non-target voxels. The first step uses energy-specific 3D high gradient dose spread ($HGDS$) function. Critical to this is a library of pre-calculated, energy-specific high-gradient dose spread functions that cover a range of media densities relative to water (see examples in Figure 1 B,D,F). The inputs to the HGDS function are nominal photon energy along with the physical and radiological distances from nearest target voxel. The HGDS is best conceptualized as capturing the general penumbra effects at a beam edge, tangential to the target surface. The HGDS function is really a series of multi-dimensional lookup tables built into the software, generated using a set of observed, reliable dose distributions. Specifically, a representative library of measurement-guided dose reconstruction (MGDR) dose grids were built for a number of static beams conforming to small- to medium-sized targets over a range of depths in water-equivalent media, and for a variety of linac models and nominal energies. In other words, building a library of observed penumbra gradients was done by extracting data from a strategic set of single-beam dose grids per energy. These source
grids needed to have high spatial resolution, because the lookup tables used for HGDS have resolution of 1 mm within 20 mm from the target. To achieve best possible accuracy, we used measurement-based (MGDR) rather than just TPS-calculated dose. From this library of 3D source dose grids, the dose falloff just outside a covered target surface were extracted by drawing dose profile lines perpendicular to the beam direction, saving the dose vs. distance data, and normalizing them to target dose. This was repeated in low (0.25) and high (2.0) density media relative to water to cover a range of ratios of radiological-to-physical distances. The practical HGDS function is thus a collection of multi-dimensional numerical lookup tables rather than a formulaic algorithm. To ensure 100% coverage of each target volume even under the condition of interpolated dose near the target surface, the HGDS return value is set to 1.0 for any distance up to half a voxel away from the target surface in all dimensions, which effectively extends the prescription dose out to a thin rind. This is a conservative approach, erring on the side of coverage for abutting or overlapping OARs. For each non-target voxel, the output is dose assignment equal to the maximum spread value searched over all neighboring target voxels $(D_{HGDS}[x, y, z])$:

$$D_{HGDS}[x, y, z] = \max_{all \ (x', y', z') \} \{D_{\text{Coverage}}[x', y', z'] \ HGDS[E, r, r_{\text{rad}}] \}$$

(2)

where $r$ and $r_{\text{rad}}$ are the physical and radiological distances, respectively, from the non-target point $[x, y, z]$ to each searched target point $[x', y', z']$, and $E$ is the nominal beam energy. The HGDS is not a convolution, it can only add dose to surrounding voxels but not remove or blur dose inside the assigned target volumes.

This article is protected by copyright. All rights reserved.
With the best-possible high-gradient portion of the dose spread now determined, the next step is to account for additional dose outside the targets that will occur due to a combination of factors, including dose attenuation along beams’ axes, i.e. the percent-depth-dose (PDD) effect, and also low dose outside the steep portion of the penumbra due to scatter out-of-field. This low-dose spread (LDS) calculations are highly influenced by the target size (compare Figure 1 B and D). Since the total amount of energy needed to raise the target to a certain dose is proportional to its volume (~r² for a cylinder) and the area available for fluence delivery is only the surface (~r), more energy has to cross each surface element along the beamlets’ axes for a larger target, increasing the dose spread outside the target along the beamlet direction; hence the “PDD effect”.

Two LDS dose operations are followed in sequence. The first covers “mid-range” distances away from targets (D_{LDS,Mid}) and the second covers “far-range” (D_{LDS,Far}). For 6 MV x-rays, mid-range has influence up to about 5 cm and far-range up to about 10 cm away from the target surfaces. Each operation starts with a 2D convolution in axial planes of a signal function with energy-dependent 2D kernels. This strategy works well for beam setups that are expected to be nearly coplanar and allows dose gradients at the most superior and inferior edges of targets to remain steep, i.e. not influenced by beam PDD effects. The 2D LDS kernels are radially symmetric and were derived by fitting the dose spread curves to experimentally achieved ones over a suite of test datasets optimized for VMAT or for IMRT with 7 to 9 coplanar beams. These kernels are pre-programmed, just like the HGDS lookup tables, and are not at this time exposed to the user to edit. The convolution signal function for D_{LDS,Mid} operation is D_{Coverage}, and the signal function for D_{LDS,Far} operation is D_{LDS,Mid}. The final operation reduces dose in very low density regions so that it tapers to zero in air.

This article is protected by copyright. All rights reserved.
\[ D_{LDS,Mid}[y][x,z] = CF[\rho_{rel}(x,y,z)] \sum_k \sum_i D_{Coverage}[y][i,k] LDS_{Mid}[x-i,z-k] \]

(3)

\[ D_{LDS,Far}[y][x,z] = CF[\rho_{rel}(x,y,z)] \sum_k \sum_i D_{Mid}[y][i,k] LDS_{Far}[x-i,z-k] \]

(4)

where the summation is done for all voxels in the dataset outside of the target(s) and

\[ CF[\rho_{rel}(x,y,z)] = \begin{cases} \frac{\rho_{rel}}{\rho_{rel,threshold}}, & \text{for relative density } < 0.05 \\ 1.0, & \text{everywhere else} \end{cases} \]

(5)

\( LDS_{Mid}[x,z] \) and \( LDS_{Far}[x,z] \) are the aforementioned energy-dependent, radially symmetric kernels, i.e. \( LDS[r] \).

After the parameterization of \( LDS_{Mid}[x,z] \) and \( LDS_{Far}[x,z] \), the final benchmark dose (\( BD_{final} \)) is the maximum value at each point \([x,y,z]\) from the three input grids (Figure 1 B,D and F). Simply put, there are now three equally sized and spaced 3D dose grids capturing different traits of potential dose-at-distance from target, and the benchmark dose at each voxel \([x,y,z]\) is assigned the highest value at that location from each grid:

\[ BD_{final}[x,y,z] = \max \{ D_{HGDS}[x,y,z], D_{LDS,Mid}[y][x,z], D_{LDS,Far}[y][x,z] \} \]

(6)

II.A.4. Feasibility DVH (FDVH)

Feasibility DVH (FDVH) can be calculated for any OAR, and is intended for OARs only, as target coverage is implied. The first step is simply to compute the OAR’s DVH based on the benchmark dose grid. This FDVH is assigned feasibility \((f)\) level of zero (FDVH(0)). The \( f \) value
ranges from zero (lower bound of achievable under best circumstances, referred to from here on as “unachievable”) to 1.0 (easily achievable).

With FDVH(0) known for an OAR, an arbitrary DVH coordinate defined by a specific dose (D) and cumulative percent volume (V) can be assessed for feasibility, using:

\[
F[(D, V)] = \begin{cases} 
0, & \text{for } (D, V) \text{ inside the FDVH(0) curve} \\
CLI[(D, V), FDVH(0)] & C2F[(D, V), D_{\text{max}}], \text{ elsewhere}
\end{cases}
\]

(7)

where

\[
CLI[(D, V), FDVH(0)] = \text{Normalized distance from point } (D, V) \text{ to } FDVH(0) \text{ curve}
\]

(8)

and

\[
C2F[(D, V), D_{\text{max}}] = \max \left\{ C2F_{\text{min}}, \left[ C2F_{max, Dose} \frac{D}{D_{max}} \right], \left[ C2F_{max, Vol} \frac{V}{100} \right] \right\}
\]

(9)

CLI is the “closeness index” equal to the distance of the \((D, V)\) coordinate from the FDVH(0) curve; the distance is unitless with \(V\) normalized to the OAR volume and \(D\) normalized to the highest target dose \((D_{max})\). \(C2F\) is a “closeness-to-feasibility” function that converts the closeness index to an \(F[(D, V)]\) value. The \(C2F\) has a minimum \((C2F_{min})\), but can increase up to value \(C2F_{max, Dose}\) as \(D\) approaches the maximum target dose or to \(C2F_{max, Vol}\) as \(V\) approaches 100%, respectively.

Using this function, coordinates for an FDVH curve for any specific feasibility value \(f\), i.e. \(FDVH(f)\), can be calculated. First, an array of volumes \([V_j: 0 \text{ to } 100\%]\) of sufficient resolution is defined. For each sampled volume \((V_j)\), increasing dose values \((D_i)\) are sampled, starting at intersection of \(V_j\) and the \(FDVH(0)\) curve. When \(F[(D_i, V_j)] \geq f\), the point \((D_i, V_j)\)
is added as a coordinate on the curve $FDVH(f)$ and the process repeats for the next volume point, $V_{j+1}$.

Several FDVH achievability zones are defined and separated by discrete $FDVH(f)$ curves. Area below the curve $FDVH(0)$ is labeled unachievable in the case of full target coverage. The region between $f=0$ and $f=0.1$ curves is qualitatively called “difficult” (but achievable). The $FDVH(0)$ and $FDVH(0.1)$ were previously examined in terms of predictive power across a population of OAR metrics for the highest-quality plans from multi-institutional plan studies for the four different disease sites: abdomen, head and neck, lung, and prostate. A statistical approach was used where the feasibility estimates (i.e. predicted achievability for each metric/objective) were compared to empirical observations from the highest quality plans pulled from the library of plans submitted by a large, international group of professional dosimetrists and physicists. Examining the highest scoring plans was critical, as those were the ones that achieved both excellent coverage of targets and sparing of critical organs. Quantified in this study were the following: Positive Predictive Value (PPV) equal to the ratio: No. of True Positives / (No. of True Positives + No. of False Positives), Negative Predictive Value (NPV) equal to the ratio: No. of True Negatives / (No. of True Negatives + No. of False Negatives), and Accuracy equal to the ratio: (No. of True Positives + No. of True Negatives) / (No. of Total Results). For each of these three metrics, the ideal value is 1.000. In the current context, True Positive means that plan does not achieve an OAR DVH below FDVH prediction, and the other three scenarios are easily derived from that. The results from these studies for PPV, NPV, and Accuracy were 1.000, 0.88 and 0.89, respectively, for $FDVH(0)$ and 0.81, 0.95, and 0.92 for $FDVH(0.1)$. Given this strong performance, this study is focused on examining the behavior of $FDVH(f)$ with $f$-values from 0 to 0.1.
II.A.5. Benchmark Dose Calculation Time

The time to calculate benchmark dose (3 mm resolution) and subsequent FDVH data (all OARs) for a typical HN case with multiple PTVs is under 3 minutes on a 2.7 GHz Windows laptop, including the time for loading patient CT images and structure contours. In addition to the number, size, and shape of the OARs, other parameters affecting calculation time are the user-selected options of dose grid resolution (1.0 to 3.0 mm) and inclusion of optional CT density correction.

II.B. Validation strategy

II.B.1. Geometrical test cases

First, we examined a collection of model geometries (Figure 2), as they are helpful for both illustrating the concepts and testing the basics of the FDVH algorithm. All targets and OARs were cylinders of equal length (5 cm) but varying diameters (2 - 10 cm). The centers of the targets were placed at the center of a water-equivalent 22 cm diameter cylindrical phantom with the OARs arranged as shown in Figure 2. In the axial view, sets A,B,D, and E in represent large (10 cm diameter) and small (2 cm) targets either adjacent to, or 1 cm away from a 3 cm diameter OAR. Set C consist of a 10 cm diameter target with a 2.3 cm diameter doughnut hole containing a concentric 2 cm diameter OAR.
Figure 2. Geometrical test structure sets. All dimensions are in mm. A,B: A large (100 mm diameter) cylindrical target with a 20 mm OAR. D,E: A small (20 mm) target with the same OAR. C: A doughnut target with an OAR in the middle. The surrounding 22 cm diameter phantom is not shown.

VMAT plans were generated for comparison to FDVH. Plans used 6MV beams from a TrueBeam linear accelerator with a 120-leaf Millennium MLC (Varian Medical Systems, Palo Alto, CA) using the collapsed cone convolution (CCC) algorithm in Pinnacle TPS (v. 9.8, Philips Radiation Oncology Systems, Fitchburg, WI). A single full arc was used for each plan with gantry angle spacing of 2° and isotropic 2 mm calculation voxel size. For this and subsequent tests it is important to underscore the challenge of creating a target coverage in the TPS that approximates the ideal uniform dose assigned to the targets in PlanIQ. For the OAR, the goal was to drive the maximum and mean (equivalent uniform dose, or EUD, in planning) doses as low as possible. It is impossible to have a truly homogeneous target dose in the real plan. The main objective for the target was to achieve the near-perfect conformity of the prescription dose to the portion of its border facing the OAR, while maintaining visually acceptable prescription isodose conformity elsewhere. Dose homogeneity within the PTV was maintained within ±10%. This 10% number is clearly higher than is normally clinically desirable, but it was chosen...
to minimize the OAR dose, since we were looking for the lowest achievable values, and the target dose homogeneity and OAR sparing exhibit exponential trade-off.\textsuperscript{37}

To demonstrate the effect of the realistic delivery constraints, we approximated idealized dose distributions. To that end, for the same geometrical test cases, optimization was run in Pinnacle for 100 iterations without the Convolution/Superposition dose calculation step. This resulted in idealized optimized dose clouds calculated with a Pencil Beam algorithm with no regard to physical delivery constraints.\textsuperscript{38} While full DICOM dose export is disabled in the TPS in this configuration, the DVHs could be extracted via a custom script and qualitatively compared to the full planning and FDVHs.

II.B.2. Clinical Head & Neck test cases

In the next step, for validation of FDVH predictions on realistic datasets, ten previously described advanced HN cases were used.\textsuperscript{5} Each patient had two targets: a primary PTV treated to 70 Gy in 35 fractions and an elective bilateral neck clinical target volume (CTV) simultaneously irradiated to 56 Gy. The OARs selected for this study were each parotid, larynx, and inferior pharyngeal constrictor (IPC). In two patients the entire larynx was a part of the target and therefore was excluded from analysis. The planning strategy with respect to the targets and the OAR was the same as described above for the geometrical test cases, except two full arcs were used instead of one in order to provide the optimizer with sufficient degrees of freedom. One OAR at-a-time was considered in each plan, thus avoiding competition with others and ensuring the best possible sparing.

This article is protected by copyright. All rights reserved.
Two metrics were used to compare the planned and FDVHs: 1) the difference in mean doses, $\Delta D_{\text{mean}}$, and 2) the Dice similarity coefficient (DSC),\(^{39}\) which allowed us to quantitate the similarity between the shapes of the DVHs. The DSC adapted for cumulative DVH curve comparison is defined as

$$DSC = 2 \frac{A_1 \cap A_2}{A_1 + A_2}$$

(10)

where $A_1$ and $A_2$ are areas under the planned and FDVH curves, respectively, and $A_1 \cap A_2$ is the intersection of those areas. The DSC can take values from 0.0 (no overlap between the curves) to 1.0 (identical curves).

The $f$-level for comparison FDVHs (sampled at 0.01 intervals) was chosen so that the lower boundary of the 95\% confidence interval of the average of the mean planned doses to all OARs across 10 patients was just above the corresponding average for FDVHs. The DSCs were then computed for that baseline feasibility level. The data were analyzed for all OARs in aggregate and for each organ type separately.

To illustrate how the presence of competing objectives changes the achieved DVHs in relation to the single-OAR FDVHs, the mean doses and DSCs for the same organs as before were calculated for the previously described Pinnacle AP test plans.$^5$ The AP software was shown to achieve lower OAR doses compared to the corresponding human-driven plans,$^4,5,34$ as it attempts to drive the OAR dose lower than specified by the user, if possible. Thus the AP plans were chosen for comparison. The realistic plans reflected all optimization objectives routinely employed in our clinic, which are more stringent than similar objectives from RTOG protocols.$^{40}$

Some pertinent examples are as follows. The gross tumor volume (GTV) must be entirely covered by the prescription isodose. PTV single voxel doses must be within ±5\% of the
prescription, with at least 95% of the PTV covered by the prescription dose. Parotid mean dose should not exceed 26 Gy. The mean IPC dose should not exceed 50 Gy, with the volume-dose percentage indices $V_{40\text{Gy}}<65\%$, $V_{50\text{Gy}}<47\%$, and $V_{60\text{Gy}}<11\%$. Similarly, the mean larynx dose should be ideally $<20\text{Gy}$ (dysphonia) or at least below 41 Gy (aspiration), with the volume-dose indices $V_{35\text{Gy}}<79\%$, $V_{45\text{Gy}}<45\%$, $V_{55\text{Gy}}<32\%$, and $V_{65\text{Gy}}<22\%$. The other structures included in optimization, as appropriate, are brain stem, spinal cord, submandibular glands, oral cavity and lips, mandible, superior and middle pharyngeal constrictors, optics and thyroid.

**II.B.3. Multi-institutional Head & Neck case with OAR overlapping target**

The case of an OAR overlapping (or abutting) a target is interesting considering that feasibility benchmark dose enforces 100% target coverage while real-world coverage goals are often less (e.g. 95%). To study this effect, we analyzed the top 10% quality plans from an international, multi-institutional plan study recently completed for the head and neck case from the MPPG 5.a  library of standard TPS commissioning cases. The controlled plan study followed the general methodology previously employed by Nelms et al. and used cloud-based data collection and real-time scoring system (ProKnow Systems, Sanford, FL). In this study, the ideal coverage for a large PTV_5600 volume was set at 95%, while the mean dose to the contralateral right parotid had a minimal requirement of 30 Gy with an ideal objective of 24 Gy. The total right parotid volume was 28.2 cc, with 3.4 cc (12.2%) inside the PTV_5600. The OAR DVHs for the top 10% total score plans (N=23) of the 238 submitted and analyzed plans were extracted and compared to the FDVH. The top 10% plans encompassed a range of x-ray treatment techniques (VMAT, IMRT, helical tomotherapy), TPS vendors, and linac models.

This article is protected by copyright. All rights reserved.
III. RESULTS

III.A. Geometrical test cases

Figure 3 and Figure 4 illustrate the most important general concepts embedded in the Feasibility algorithm. Figure 3 demonstrates how the conformal target coverage is paramount for the TPS and FDVH agreement in the low-dose region. The best sparing for the OAR immediately adjacent to the target (Figure 3 B and D) could be easily achieved by a vertical parallel-opposed beam pair. Only the high-gradient (penumbra) dose spread would affect the OAR DVH. However the target coverage would not be conformal.

![Figure 3](image)

Figure 3. Two VMAT plans for a 10 cm target with immediately adjacent 3 cm OAR with lower (A,B) and higher (C,D) level of dose conformality to the target. A: Feasibility DVH curves with $f=0$ ("unachievable") generated with and without the low dose spread component. C: The FDVHs at $f=0$ and 0.1 for a more conformal plan. The dotted lines in panels A and C are TPS DVHs. The lower panels (B,D) depict the axial dose distributions for the two plans. The "unachievable" area under the FDVH with $f=0$ is shaded.

To achieve tight target dose conformality, some fluence must enter and exit through the adjacent OAR, triggering the low (PDD) dose spread throughout it. In the left panels of Figure 3
the dose conformality is not strictly enforced, and one can see how the dose distribution near the target-OAR interface is shaped by the relatively high weight assigned to the vertical beamlets, partially mimicking a parallel-opposed beam arrangement near the target/OAR interface (Figure 3 B). As a result, the low-dose portion of the TPS DVH dips in the “unachievable” area. If the low dose spread in the benchmark dose calculation is turned off, the FDVH retracts below the achieved one (Figure 3 A). Conversely, if the target dose conformality is enforced (Figure 3 B,D), the achieved OAR DVH shifts to the right, towards the FDVH calculated in normal manner, with both low-and high-dose spread, reflecting additional fluence entering and exiting the OAR.

**Figure 4.** Comparison of FDVHs with TPS DVHs for realistic VMAT (“Full”) and idealized optimization (“Ideal”) dose distributions. Panels A-D correspond to Panels B-E in Figure 2, which contain the exact dimensions. Feasibility OAR DVHs for $f=0$ and 0.1 are presented along with the TPS-generated target and OAR DVHs for a full CCC algorithm calculations and idealized Pencil Beam dose at the end of optimization. The “unachievable” area under the FDVH with $f=0$ is shaded. The target/OAR arrangements from Figure 2 are illustrated.

Figure 4 demonstrates that by invoking parameterization on the geometrical targets the feasibility algorithm takes into account the realistic capabilities of the delivery system. For the same target coverage, the idealized optimized OAR DVHs dip into the “impossible” zone under
FDVHs with \( f = 0 \). The effect is present on every graph, but is more pronounced for the larger targets, and in particular with the doughnut geometry (Figure 4 B). This can be easily explained by the extra central dose in the full CCC calculation due to the MLC transmission.

### III.B. Head & Neck test cases

Dose-Volume Histogram metrics results for individually optimized OARS in HN plans are presented in Table 1. The feasibility level \( f \) with the lower limit of the 95% CI of the average \( \Delta D_{\text{mean}} \) closest to zero was 0.05. Thus Table 1 uses this \( f=0.05 \) value for all FDVH(\( f \)). A good agreement between carefully planned achieved DVHs and FDVHs with the \( f \)-value somewhere between 0.0 and 0.1 was expected from the preliminary data.\(^{36}\)

<table>
<thead>
<tr>
<th>OAR</th>
<th>N</th>
<th>Ave. DSC</th>
<th>DSC 95% CI</th>
<th>Ave. ( \Delta D_{\text{mean}} ) (%)</th>
<th>( \Delta D_{\text{mean}} ) 95% CI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>38</td>
<td>0.961±0.018</td>
<td>0.955 – 0.967</td>
<td>1.8 ± 5.8</td>
<td>-0.1 – 3.6</td>
</tr>
<tr>
<td>Parotids</td>
<td>20</td>
<td>0.954±0.021</td>
<td>0.944 – 0.964</td>
<td>-0.5±6.4</td>
<td>-3.5 – 2.5</td>
</tr>
<tr>
<td>Larynx</td>
<td>8</td>
<td>0.963±0.012</td>
<td>0.953 – 0.973</td>
<td>5.7±3.3</td>
<td>3.0 – 8.5</td>
</tr>
<tr>
<td>IPC</td>
<td>10</td>
<td>0.972±0.010</td>
<td>0.964 – 0.979</td>
<td>3.0±3.7</td>
<td>0.4 – 5.7</td>
</tr>
</tbody>
</table>

An example of the achieved DVHs for different individually optimized OARs on one of the datasets, superimposed on FDVHs with \( f \)-values from 0.0 to 0.1 is presented in Figure 5.
Figure 5. Example feasibility DVHs with $f$ from 0.0 to 0.1 and the achieved TPS DVHs for individual OARs from one dataset (#6). The “unachievable” area under the FDVH with $f=0$ is shaded.

III.C. Comparison of achieved and FDVHs in realistic plans

On average, as expected, in the presence of competing priorities the mean dose for every OAR is higher for clinically realistic plans compared to the single-organ optimization ones. The results are shown in Table 2 in the same fashion as in Table 1, as the average DSC values and a percentage difference in average $\Delta D_{\text{mean}}$ from the FDVH (0.05). The lower DSC values comport with the greater differences in the mean doses.

Table 2. Descriptive statistics comparing the achieved and FDVHs ($f=0.05$) for clinically realistic AP plans. Presented are the DSC and mean dose differences, with 1SD and ranges.

<table>
<thead>
<tr>
<th>OAR</th>
<th>Ave. DSC</th>
<th>DSC Range</th>
<th>Ave. $\Delta D_{\text{mean}}$ (%)</th>
<th>$\Delta D_{\text{mean}}$ Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parotids</td>
<td>0.882±0.087</td>
<td>0.616 – 0.979</td>
<td>22.4±26.7</td>
<td>-16.0 – 125.5</td>
</tr>
<tr>
<td>Larynx</td>
<td>0.809±0.143</td>
<td>0.564 – 0.994</td>
<td>54.7±47.9</td>
<td>1.2 – 154.2</td>
</tr>
<tr>
<td>IPC</td>
<td>0.844±0.090</td>
<td>0.668 – 0.944</td>
<td>37.4±27.6</td>
<td>11.9 – 98.5</td>
</tr>
</tbody>
</table>

Comparing Table 1 and Table 2 one can notice much larger ranges and standard deviations of values in the latter. This is a reflection of the fundamental difference between the single-OAR optimization plans and clinical ones with multiple competing objectives. The range of planning results is illustrated in Figure 6. Panel A depicts the rather prevalent one when an OAR cannot be spared as well as in the single-OAR plan without unacceptable sacrifices in other objective(s). Panel B shows an achieved parotid DVH that was close to the FDVH, indicating high priority placed on sparing that organ. Finally, as shown in Panel C, occasionally the achieved OAR DVH dips noticeably below the FDVH. Every case where that happened has been carefully examined and inevitably it was found that PTV coverage at the interface with the OAR in question was sacrificed. This violates the major assumption of the Feasibility algorithm, namely the complete coverage of the target(s) by the prescription dose(s).
As evident from Table 2, the realistic parotid DVHs are on average closer to the corresponding FDVHs than the IPC or larynx ones. This is understandable, as the parotids’ dose objectives are generally lower and the parotid mean dose being a high clinical priority. With the IPC and larynx, the low-dose portion of the DVH is of relatively less importance and is not being pushed down as hard (Figure 6 A). We also expect that with the larynx and IPC being central structures, as opposed to the laterally situated parotids, it is harder for the optimizer to find a solution where the fluence at least does not exit through the OAR, raising the low-dose portion of the DVH.

![Graphs A, B, C](image)

**Figure 6.** Examples of the changes in achieved DVH between the single-organ optimization plans and clinically realistic Auto-Plans (AP), in relation to the “impossible” FDVH (f=0). A: IPC #2. B: Lt. Parotid #6. C: Lt. Parotid #5.

The case of an overlapping OAR in the multi-institutional study where PTV coverage, as is common, was specified at 95% is shown in Figure 7 A,B. Here, the achieved right parotid DVHs for the top 10% quality plans are plotted along with the FDVH. The achieved DVH
curves venture inside the “impossible” FDVH curve in places, and this is because the PTV_5600 coverage goal was allowing 5% (37.2 cc) of the total target volume (744.4 cc) to fall under 56 Gy without penalty. The experience planners concluded that sacrificing some target coverage in favor of the parotid sparing in this case was rewarded by a higher overall score. A useful observation can be made that if an achieved DVH falls inside the FDV, it indicates that the target coverage was likely compromised. As shown in Figure 7, when the PTV coverage is enforced at 100% (Panel C vs. B), the OAR DVH is fairly close to the FDVH (Panel A).

**Figure 7.** A: Achieved right parotid DVHs for the top 10% quality plans from a controlled plan study, along with the FDVH and a DVH from a plan with 100% PTV_5600 coverage. B: PTV_5600 coverage by the 56 Gy isodose line for the best study plan. Note how PTV coverage is sacrificed. C: Isodose coverage corresponding to the dashed DVH curve, for the plan covering 100% of PTV_5600.
IV. DISCUSSION

IV.A. Study strengths and weaknesses

To the best of our knowledge, the approach described here is unique as a whole and has never been implemented before. However, while the method is different, the results are largely determined by the same physics principles, and certain elements of the described algorithm by necessity overlap with the previous body of work on KBP.\(^6\)\(^{-12}\),\(^{18}\)\(^{-33}\) In particular, the geometry of the OAR in relation to the target is the major driver of the achievable OAR DVH. To that effect, Petit et al\(^{29}\) used a subset of previous plans with less favorable PTV/OAR configuration to find the minimum achievable OAR dose for a new patient. Different mathematical frameworks were proposed to correlate minimum achievable dose to a voxel with the distance from that voxel to the target surface,\(^{22,25,30,31,33}\) sometimes coupled with the attempts to identify additional descriptors of the target/OAR geometrical relationship through principal component analysis.\(^{8,9,22,25}\)

In the entire body of KBP work, partially cited above, the dose/geometry relationship is always determined from the set of existing treatments plans. This is where our Feasibility approach fundamentally diverges from KBP. Feasibility is independent of the treatment planning technique and prior experience - no learning database of similar treatment plans is necessary. Hence one of the strengths of the method lies in its simplicity and minimal, if any, commissioning effort. Tailoring the parameters of the method to accommodate a specific machine type or technique could produce slight improvements in the feasibility analyses, but would also introduce variation and thus remove the ability to be used as a benchmark against which to compare achievements of a various TPS or delivery system.
By the same token this simplicity leads to limitations. Most importantly it approximates the lowest possible boundary of the OAR DVH, but not necessarily its actual shape in the presence of competing objectives of multiple OAR sparing, target dose coverage\textsuperscript{42,43} and homogeneity.\textsuperscript{3,37} In particular, Tol et al \textsuperscript{37} noted a strong (exponential) tradeoff between OAR sparing and target dose homogeneity. These tradeoffs are implicitly built in the KBP learning approach,\textsuperscript{26} assuming a uniform planning protocol.\textsuperscript{3} Such realistic delivery tradeoff information is not available to the Feasibility algorithm that assumes that the target is just uniformly painted with the prescription dose. Similarly, the algorithm effectively deals with one OAR at a time and is agnostic to fluence redistribution necessary to fulfill multiple competing objectives. This is why the Feasibility algorithm is a much stronger predictor of the unachievable (lower) DVH boundary than of the likely achievable DVH.

In addition, the Feasibility method is better suited for OARs close to the target. As the distance between the OAR and the target increases, specific beam arrangements, unknown to the algorithm, exert more influence on the OAR DVH. This can be understood by comparing for example the rectum with the femoral head in prostate cancer treatment. While the dose to the rectum at first approximation is largely independent of the IMRT beam angles, the femoral heads can be almost completely protected from both entrance and exit fluence, depending on the beam configuration. Thus the beam-agnostic Feasibility prediction becomes unreliable.

The DVH curves often differ most from the corresponding FDVHs at the point of transition from the high-to low-gradient portion of the curve (Figure 4 C). It can be attributed to a simplistic mathematical form of piecewise constructing the final dose spread from the high- and low-gradient portions (the higher of the two, see Eq. (6) and Figure 1). In the future a more sophisticated function could be used to describe the transition region.
Finally, the method is well-suited for approximating the general shape of the best-possible OAR DVH; however, because it enforces 100% coverage whereas real-world plans often sacrifice coverage near OARs, there can be deviations (actual vs. FDVH) in the curves as they approach the high dose region. This characteristic is easily internalized by the learned treatment planner, though, and the actual DVH dipping below the FDVH curve becomes a useful indicator of sacrificed coverage. This phenomenon is illustrated by the multi-institutional plan study where the top 10% scoring plans had OAR DVHs (for the right parotid, which overlapped the PTV_5600 target) that dipped inside the FDVH curve because target coverage was driven only to 95%, leaving 5% of the target to be under-dosed without penalty.

The benchmark dose algorithm is currently designed for coplanar (or nearly-coplanar) beams, stemming from the cylindrical form of integration (summation) in convolution Eqs. (3) and (4). Furthermore, it was honed and deployed for common photon delivery techniques technique (coplanar VMAT or IMRT with 7+ beams) and linac hardware. One disease site (HN) was explored here. It has a moderate amount of density heterogeneity and for the purposes of this study is comparable to most body sites, but dose to (or near) lung volumes should be evaluated separately.

IV.B. Current and potential applications

IV.B.1. Plan review (scoring) adjusted for feasibility

In the PlanIQ software, the user can set up plan quality algorithms that are relevant dosimetric metrics and corresponding scoring functions for each target and OAR. The output is a composite plan quality metric (PQM). If desired, the FDVH can be used in the current software to normalize any individual metric score based on the best possible sparing for the
specific patient, rather than the generic OAR objective. This way the plan quality score can be adjusted for any metric where the standard clinical objective is fundamentally unachievable. This essentially allows grading on a curve in the cases of difficult patient anatomy. In authors’ experience, occasionally it was demonstrated outright that the OAR dosimetric objectives could not be achieved without compromising PTV coverage and appropriate clinical adjustments were made. In addition to straightforward plan evaluation, this method could be useful in the assessment and commissioning of KBP and other auto-planning methods in the future, by helping grade existing plans in terms of quality, to determine which ones are candidates to be added to the knowledge base.

IV.B.2. Plan optimization

Another possible application is based on the multiple observations that feeding the optimizer challenging objectives specific to the patient anatomy improves plan quality.\textsuperscript{12,26,28,33} We hypothesize that FDVH data (curves in the range of $f=0$ to 0.1) derived prior to optimization would provide patient-specific planning objectives that are challenging yet reality-based. The optimization solution in this case should be automatically Pareto-optimal.\textsuperscript{44} That is, for a given level of target dose coverage and homogeneity, the OAR DVH curves will be driven as close as possible to, but never below, the corresponding FDVH. Hence no individual region of interest objective function can be further improved without sacrificing at least one other. Other groups are currently pursuing this application of the Feasibility algorithm to plan optimization, with promising preliminary results.\textsuperscript{16,45,46} This paper only endeavors to describe and validate the method of generating the FDVHs.
V. CONCLUSIONS

A novel method (Feasibility DVH, or FDVH) for \textit{a priori} estimation of the lower achievable boundary of the OAR DVHs in inverse treatment planning was introduced and validated on a series of HN VMAT plans. The algorithm does not require a database of prior plans but rather derives the FDVH from nearly first principles, assuming that the targets are uniformly covered with the prescription doses. It is easily parameterized based on a short list of model geometrical datasets. The method is agnostic to the planning technique and beam arrangement, requiring only the regions of interest, the energy, and, optionally, the CT dataset as inputs. But this simplicity is also the cause of the algorithm’s limitations. It is designed to approximate the lowest possible OAR DVH based on that OAR’s geometrical relationship to the target, but not the likely achievable one for the class of plans in the presence of realistic competing objectives. The method is best suited for the parallel OARs close to the target and is currently implemented for (nearly) coplanar beam arrangements. It is useful for plan quality assessment and can potentially be used to supplement and augment automated planning techniques.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from Sun Nuclear Corporation (SNC).

CONFLICT OF INTEREST

Ben Nelms is the original inventor of the method \textsuperscript{47} and a paid consultant to Sun Nuclear Corporation (SNC), to whom the patent is assigned. Vladimir Feygelman is a PI on research
grant from SNC and has received speaker fees. Saeed Ahmed is a student supported by the SNC grant.

REFERENCES


This article is protected by copyright. All rights reserved.


40. RTOG 1016. Phase III trial of radiotherapy plus cetuximab versus chemoradiotherapy in HPV-associated oropharynx cancer.


47. Nelms BE, Inventor; Sun Nuclear Corporation, assignee. Benchmark system for radiation therapy planning. US patent 94633362016.