Advances in Radiation Oncology (2021) 7, 100841

advances in radiation oncology

www.advancesradonc.org

Scientific Article

Data-Driven Dose-Volume Histogram Prediction

Mitchell Polizzi, MS,[a,b] Robert W. Watkins,[a,c] and William T. Watkins, PhD[a,d,*]

[a]*Radiation Oncology, University of Virginia, Charlottesville, Virginia;* [b]*Radiation Oncology, Virginia Commonwealth University, Richmond, Virginia;* [c]*Virginia Polytechnic Institute and State University, Blacksburg, Virginia;* [d]*City of Hope National Medical Center, Duarte, California*

Received May 10, 2021; revised October 1, 2021; accepted October 19, 2021

Abstract

Purpose: To evaluate dose-volume histogram (DVH) prediction from prior radiation therapy data.

Methods and Materials: An Oncospace radiation therapy database was constructed including images, structures, and dose distributions for patients with advanced lung cancer. DVH data was queried for total lungs, esophagus, heart, and external body contours. Each query returned DVH data for the N-most similar organs at risk (OARs) based on OAR-to-planning-target-volume (PTV) geometry via the overlap volume histogram (OVH). The DVHs for 5, 20, and 50 of the most similar OVHs were returned for each OAR for each patient. The OVH(0cm) is the relative volume of the OAR overlapping with the PTV, and the OVH(2cm) is the relative volume of the OAR 2 cm away from the PTV. The OVH(cm) and DVH(%) queried from the database were separated into interquartile ranges (IQRs), nonoutlier ranges (NORs) (equal to 3 × IQR), and the average database DVH (DVH-DB) computed from the NOR data. The ability to predict the clinically delivered DVH was evaluated based on percentiles and differences between the DVH-DB and the clinical DVH (DVH-CL) for a varying number of returned patient DVHs for a subset of patients.

Results: The ability to predict the clinically delivered DVH was excellent in the lungs and body; the IQR and NOR were <4% and <16%, respectively, in the lungs and <1% and <5%, respectively, in the body at all distances less than 2 cm from the PTV. For 21/23 patients considered, the differences in lung DVH-DB and DVH-CL were <4.6% and in 14/23 cases, <3%. In esophagus and heart, the ability to predict DVH-CL was weaker, with mean DVH differences >10% for 12/23 esophagi and 10/23 hearts. In esophagus and heart queries, the NOR was often 10% to 100% volume in dose ranges between 0% and 50% of prescription, independent of the number of patients queried.

Conclusions: Using prior data to predict clinical dosimetry is increasingly of interest, but model- and data-driven methods have limitations if based on limited data sets. This study's results showed that prediction may be reasonable in organs containing tumors with known overlap, but for nonoverlapped OARs, planning preference and plan design may dominate the clinical dose.

Introduction

Radiation therapy (RT) treatment planning is evolving from a subjective process that relies on planner and physician expertise[1] toward automation.[2] The planning process uses sets of population- or model-derived DVH objectives and plan quality may be improved through manual effort and the planner's level of experience.[3] Recently, quantitative approaches to predict dose and dose-volume histograms (DVHs) based on models,[4,5] previous patient

Sources of support: This study was supported by the Ivy Foundation, University of Virginia (2018-2019).

Disclosures: none.

Data sharing statement: Research data are stored in an institutional repository and can possibly be shared upon request to the corresponding author with specific agreements.

*Corresponding author: William T. Watkins, PhD; E-mail: watkins.medical.physics@gmail.com

https://doi.org/10.1016/j.adro.2021.100841

data,[2,6-10] geometric relationships,[11-13] and other factors are being used to simplify (or automate) the planning process. There is significant ongoing effort in treatment planning automation based on prior data and big data[2,6]; however, dose or DVH prediction may not show achievable dosimetry. As an example, consider a dataset trained on arc-based delivery; a very low dose (or 0-dose) in an organ at risk (OAR) may not be achievable. However, a fixed-beam approach can conformally avoid this OAR and trivially achieve0-dose. A tangible example is heart dose in modern breast RT. A model trained on arc delivery will not show the achievable minimum heart dose. Similarly, anterior posterior (AP-PA) 3-dimensional (3D) conformal RT often maximizes the 0 dose-volume (and minimizes the irradiated volume) in photon therapy. If AP-PA plans are not included in data- or model-driven prediction, cognitive bias is introduced. The concept of standardizing treatment plans via predicted dosimetry has the potential to introduce cognitive bias in clinical plan selection and may blind clinicians to achievable dosimetry if models are constructed based on limited data. This work shows DVH prediction in an array of plan techniques for advanced lung cancer.

Wu et al[14] used a database of geometric and dosimetric data to estimate achievable DVH objectives for new patients, and in previous work introduced the overlap volume histogram (OVH).[12] OVH defines the spatial relationship between OARs and the planning target volume (PTV) in a method analogous to the DVH and was shown to be a powerful predictor of achievable dose.[12] Yuan et al[11] used an array of patient anatomic features and their relation to asses OAR dose sparing in intensity modulated RT (IMRT) using the distance-to-target histogram (DTH). DTH was previously introduced by Zhu et al[15] and is similar to OVH; it is the fractional volume of an OAR as a function of distance from the PTV. If an OAR voxel is within the PTV, the distance is negative. The DTH and OVH are equivalent when the Euclidean form of the distance function is used. Yuan et al[11] found that the median distance between the OAR and PTV, the portion of the OAR volume within distance-range, and other volumetric factors were correlated to dose.

Model-based predictions[1,5] are derived from assumptions about achievable dose or from prior data. Several studies have examined RapidPlan Knowledge-Based Planning (Varian Medical Systems, Palo Alto, California) for varying numbers of patients,[2,9,16] with a general consensus that 30 to 60 patients are sufficient to construct a reliable model. However, the ability to predict probable dose should be differentiated from achievable dose used for quality assurance and plan optimization. The probable dose is a function of the number of input patients with data quartiles defined by the consistency of input (eg, all patients in the model are treated via arc therapy). The achievable dose can be shown with as few as 2 patients (eg, AP-PA achieves a dose of 0 in an OAR, or arc provides the most conformality). Labeling prior data as similar or dissimilar in DVH prediction will introduce cognitive bias in individual patients and across patient populations if those data are used in plan optimization. It is not surprising that 30 to 60 patients labeled as similar generate a reliable model, but the inherent bias in that model is unknown. Prior-data and model-based methods of DVH prediction present powerful objective ways to assess the quality of treatment plans through prior RT.[1,12,15,17,18] However, the underlying variables that determine plan quality and optimality are unknown for data-based methods. This study evaluated the accuracy of DVH prediction based on anatomic similarity in a database of patients with advanced lung cancer.

Methods and Materials

Oncospace construction and OVH

A database of 130 patients with advanced-stage lung cancer were analyzed using an Oncospace database. Oncospace is a big-data platform that organizes RT data including segmentations, dose distributions, and clinical data.[19,20] The database has been shown through various studies to help assess plan quality and even predict RT-induced complications.[13-17] We constructed a database for advanced lung cancer, including various delivery types, prescription doses, and adaptive strategies as performed clinically from 2010 to 2018 at the University of Virginia Health System Department of Radiation Oncology (Charlottesville, VA). Delivery types included AP-PA/oblique 3D plans, multibeam 3D conformal plans, multibeam IMRT treatment plans, TomoTherapy helical plans (Accuray Inc, Sunnyvale, CA), and arc delivery on Varian linear accelerators (Varian Medical Systems). Prescription doses ranged from 60 to 72 Gy, delivered at the conventional 2 Gy per fraction. Approximately 30% of the patients were adaptively replanned owing to tumor regression; in these cases, the composite dose was evaluated on the final computed tomography scan. The database also included clinically used segmentations of the PTV, the external body contour, the lungs, esophagus, and heart for all patients.

To evaluate the predictive power of our database for a variety of input parameters, we queried and compared data for a 23-patient subset. These patients were chosen to represent an array of tumor location and delivery techniques. For each of the patient OAR and PTV segmentations, an input OVH was computed. The most similar database OVH was identified via a similarity score between the current patient and all patients in the database. The similarity score was the sum of the Euclidean distance between the input OVH and the database OVH at 2 distances: 0 cm and 2 cm. The OVH at 0 cm defined the volume overlap between the PTV and the OAR; the

OVH at 2 cm defined the relative volume of the OAR 2 cm away from the PTV. A similarity score of 0 meant an exact match between the current patient and the database patient, and a similarity score of 2 meant there was 100% difference in the OVH at 0 cm and at 2 cm for the current patient and the database patient. For each input OVH, the N-most similar ($N_{similar}$) organ and PTV OVHs and the corresponding delivered DVHs for those structures were returned from the database.

Validating the OVH query

To validate the 2-point OVH query method, we returned a varying number of the most similar organs and reported the OVH variations between the input OVH (the queried patient's organ OVH) and the database OVH (the returned OVHs of all similar organs). The results are shown as the OVH variation as a function of $N_{similar}$ = 5, 20, or 50 organs and the distance from the PTV, from −2 cm (overlapping the PTV) to 10 cm away from the PTV. Of note, the similarity was computed for organs, not for patients, so the 20 most similar hearts were not necessarily the same patients as the 20 most similar lungs.

The 2-point OVH query was derived based on simple radiation dosimetry concepts. At many radiation treatment sites, in advanced lung cancer, the prescription dose to the target is uniform. Therefore, the relative OAR volume at 100% dose, or DVH(100%), should be equal to OVH(0 cm). Because the target dose was uniform, OVH (<0 cm) did not provide additional information about the OAR-DVH. The OVH(2 cm) was the relative volume 2 cm away from the PTV. This value should be predictive of low dose-volume levels, eg, DVH(30%) to DVH(70%), based on varying target volume size and other factors. We anticipated that the OVH(2 cm) would be heavily influenced by planning parameters including beam arrangements and preferential sparing of OARs.

The similarity metric used to define the most similar patients in the database was the sum of Euclidean distances at 0 cm and 2 cm with equal weighting. Variations in OVH as a function of varying $N_{similar}$ for each of the 4 OARs are reported in the results.

Predicting DVH

The ability to predict the DVH from the database based on OVH similarity is also reported in the results. The clinically planned DVH (DVH-CL) and the predicted database DVH (DVH-DB) based on a query of the N-most similar organs were compared. An algorithm was developed to simplify the DVH-DB from the N-most similar organs to interpret the potentially large amount of data compared with a single clinical plan. For a return of 20 patients, there were thousands of DVH points with variable prescription dose to evaluate for each OAR. To address variable prescription doses in the database (ranging from 60-72 Gy with various PTV levels), all DVHs were converted to relative dose via normalization at PTV-D50 = 100%. The D50 was chosen owing to stability across patient populations. Other metrics such as D95 or D5 can vary substantially owing to minor changes in the DVH shoulder or tail. To further simplify the data and create a meaningful and clinically useful presentation, the N-most similar DVHs were combined using box-and-whisker plots on fixed dose-bin sizes. The use of big data and of quartiles to create weighted experience scores is discussed in Mayo et al.[6]

Combining prior DVH data to create predicted DVH

To create a predicted DVH by combining prior DVH data, each DVH was first normalized based on the PTV-D50. Ten dose bins were created on the relative dose interval 0% to 120%. For each bin, the 25th and 75th percentile data and the interquartile range (IQR) of all DVH data were computed (IQR = 75th − 25th percentile). The upper and lower whiskers were defined as 1.5 times the IQR above the 25th percentile and below the 75th percentile. Outliers, defined as all data points above or below the whiskers, were removed. The average DVH was computed from the nonoutlier data. The 10 box plots and the average DVH on the relative volume versus the relative dose DVH plot were displayed.

This algorithm allowed for simplification and an intuitive understanding of the current DVH compared with data from Oncospace including statistical analysis. Compared with clinical treatment plans, this method allows for visual comparison of the current plan compared to prior data in terms of quartiles.

The numerical difference between the DVH-CL and DVH-DB at the intervals of the 10 dose bins is given in the results. We evaluated the ability to predict the DVH-CL based on quartile data and the error between the DVH-CL and DVH-DB.

Results

Variations in OVH

The OVH variation as a function of increasing $N_{similar}$ patient orgnas and distance from the PTV for the 4 OARs is shown in Figure 1. Intuitively, the variation of the database OVH should increase as the distance from the PTV increases and the number of patients queried increases. Using an $N_{similar}$ of 5 or 20 resulted in OVH differences of less than 10% for distances up to 2 cm for all OARs considered. However, increasing the $N_{similar}$ to 50 (about 40% of the database) resulted in deviations of more than 10% in the OVH at 2 cm from the PTV, and in some

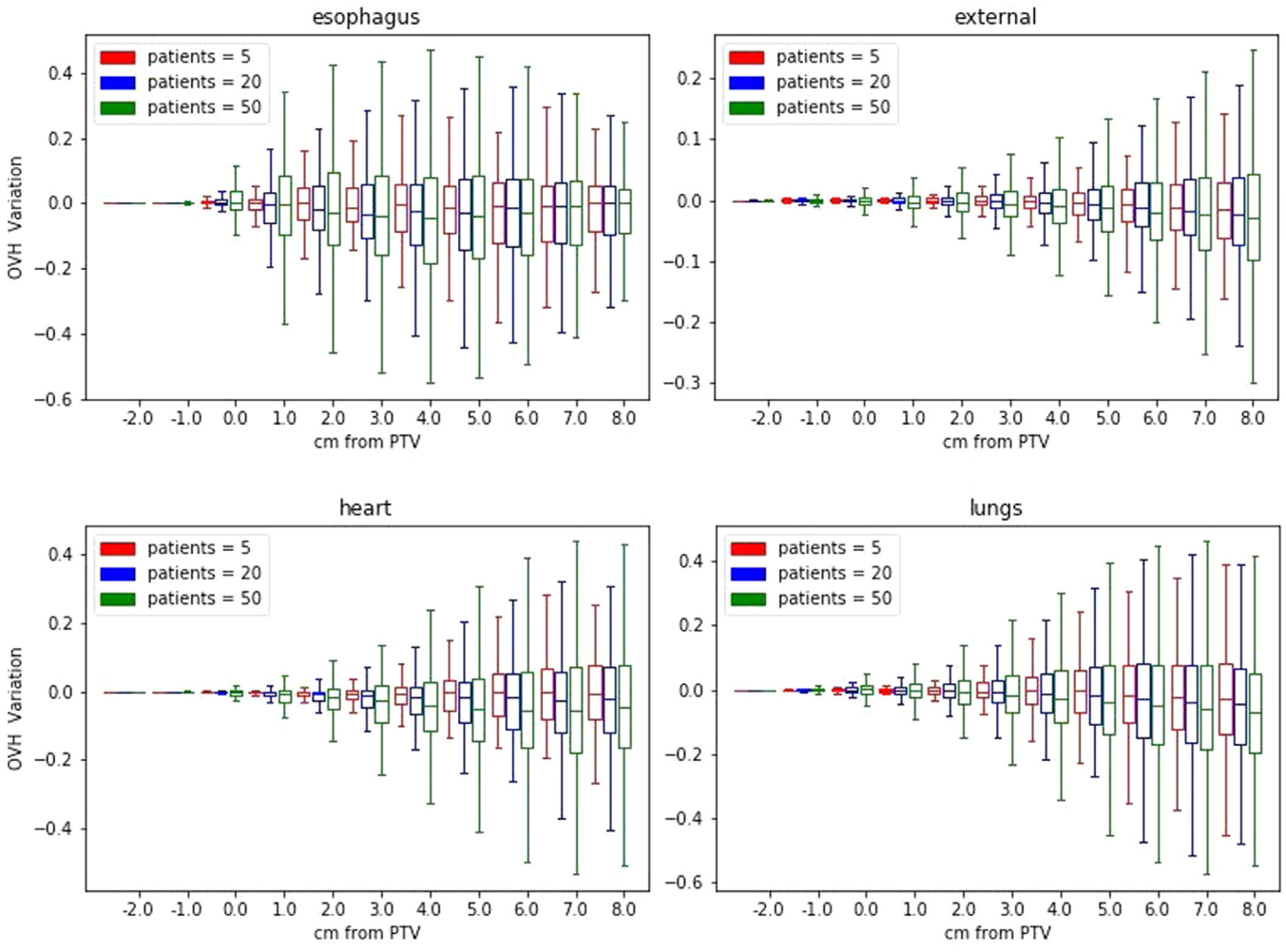

Fig. 1 The variation in overlap volume histogram (OVH) returned from the Oncospace database, as a function of increasing number of similar patients returned. The esophagus OVH (top left) had the most variation, with interquartile ranges of 25% to 75% exceeding 5% volume at distances greater than 1 cm from the planning target volume (PTV). The external OVH is the most consistent, which reflects similar patient and PTV sizes in the database.

cases, variations of more than 10% in the OVH at 0 cm to 1 cm from the PTV. In our database of 130 patients, using 5 to 20 of the most similar organs returned a reasonably small OVH variation and therefore should result in a reasonable DVH estimation.

The variation in the esophagus was the largest among all organs considered, but the OVH IQR remained less than 20% for an $N_{similar}$ of 20 or fewer organs at all distances. The esophagus OVH nonoutlier range (NOR) increased to more than 40% at 1.0 cm from the PTV and remained greater than 40% for the remainder of the distance for all values of $N_{similar}$ considered. The OVH IQR at a distance of 0.0 cm increased from less than 1% for 5 organs returned to 1% for 20 organs returned and to 5% for 50 organs returned. The IQR at a distance of 2.0 cm increased from 5% for 5 patients to 8% for 20 patients and to 25% for 50 patients. Beyond 2.0 cm, the IQR for 5 patients increased to approximately 10%; for 20 patients, to 20%; and for 50 patients, to approximately 35%. These results imply that a query of 5 to 20 of the most similar organs in this database gives a reasonable OVH similarity at distances between 0 to 2 cm from the PTV with values in the range of 10% to 20%.

Heart and lung OVH variations were similar and correlated with the increasing $N_{similar}$ and distance from the PTV, but the IQR was less than 20% for distances ranging from −2 cm to 10 cm when $N_{similar}$ was 20 or less, and the NOR was less than 40% at 2 cm with an $N_{similar}$ of 20 or fewer patients. The IQR of the OVH variation at distances of 0 to 2 cm remained less than 5% for an $N_{similar}$ of 5 patients or 20 patients. Like the results for the esophagus, these data support returning 5 to 20 of the most similar hearts and lungs in the database to possibly predict DVH accurately.

The external body OVH variation was small, independent of the number of patients queried. In external

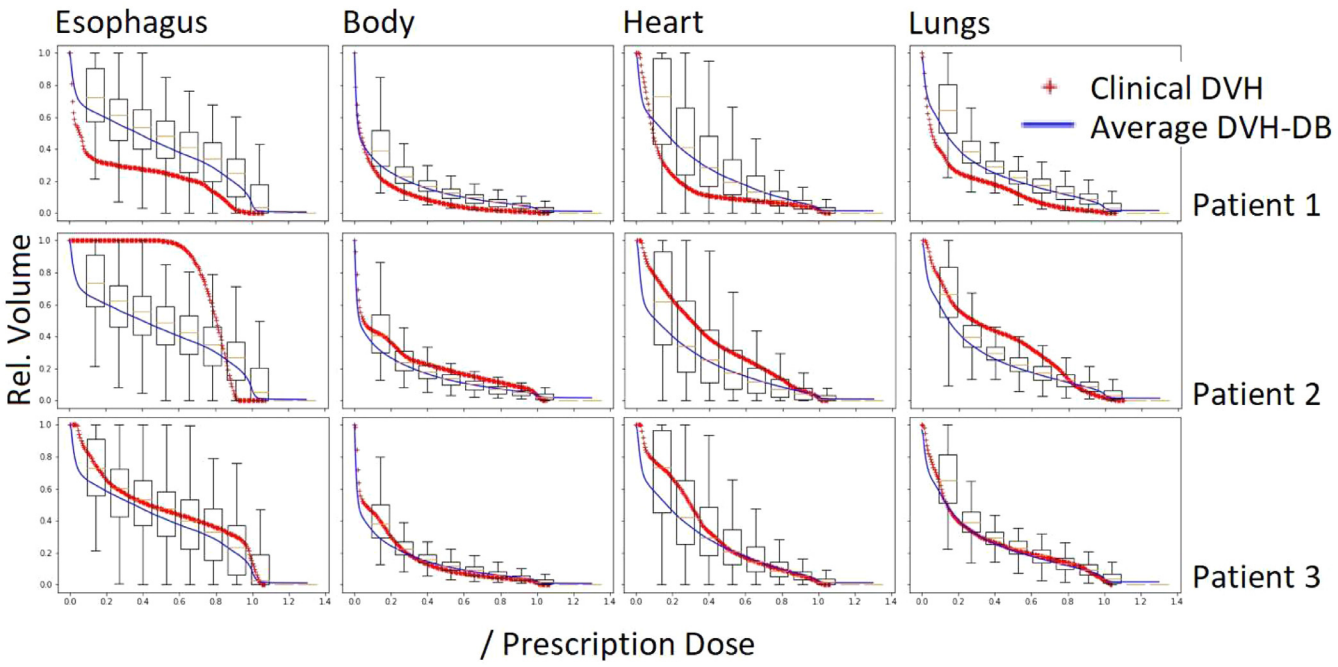

Fig. 2 The clinically delivered dose-volume histogram (DVH) (red) is compared with the predicted database DVH (DVH-DB) for 3 patients (1 per row) and 4 organs at risk per patient. The top row shows a clinical plan superior to the average DVH-DB, the middle row shows a clinical plan inferior to the average DVH-DB, and the bottom row shows a clinical DVH approximately equivalent to the DVH-DB. However, when considering the nonoutlier range of the data from the database, it is clear that a significant reduction in the heart and esophagus dose may be possible.

contours, the NOR was less than 5% and less than 20% for distances up to 5 cm from the PTV. The IQR at a distance of 0.0 cm remained less than 1% for 5 patients and for 20 patients and increased to just 2% for 50 patients.

Prediction of DVH

In many cases, there were large discrepancies between the clinically delivered DVH-CL and DVH-DB in the esophagus and heart, but lung and external DVHs were predicted robustly in all cases. The average difference between DVH-CL and DVH-DB in external contour was less than 7% for all cases, and less than 4.6% in 21 of 23. The average error in the predicted DVH for the lungs was less than 3% for 14 of 23 cases. In the esophagus and heart, the average error was more than 10% for 12 of 23 esophagi and 10 of 23 hearts. The results indicate a strong predictive accuracy for DVH prediction in organs that contain the PTV, such as the lungs and the external contour, but not for organs that do not necessarily overlap the PTV.

Three patient cases are shown in Figure 2 (1 in each row), each comparing the DVH-CL and DVH-DB queried with an $N_{similar}$ of 20 organs. The quartiles for the predicted DVH are shown via box plots; the mean predicted DVH (blue) and clinical DVH (red) are shown. Figure 2 shows the ability to identify superior (top row), inferior (middle row), and well-predicted (bottom row) DVHs for 3 different patients compared with the average DVH-DB. The top row shows a DVH-CL less than the average DVH-DB for all OARs; the second row shows a DVH-CL greater than the average DVH-DB for all OARs (indicating inferiority); and the third row shows minimal difference between the DVH-CL and the average DVH-DB. However, the IQR and NOR in the predictions for the esophagus and heart indicate there is an ability to reduce the DVH toward 0-dose (or avoid the OAR entirely). This is not the case for the external contour and the lungs which encompass the PTV. The difference between the achievable dose in OARs that contain the PTV (lungs and external contour) and those that do not (heart and esophagus) is apparent from Figure 2.

The underlying variables that determine the clinical DVH may not be related to geometric similarity in OARs that do not contain the PTV, as shown in Figure 2. In the case of the heart, the OVH variation was very small (<5% IQR for distances up to 5 cm from the PTV), but our ability to predict a useful DVH range was very limited at doses less than 50% of prescription.

The clinical and predicted DVHs for 1 patient across variable $N_{similar}$ OARs pulled from Oncospace are shown in Figure 3. Despite the large variation that was shown in OVH for 50 patients, the difference displayed between the 5 most similar DVHs versus the 50 most similar DVHs (including 4% vs 40% of the entire database) was minimal, with the IQRs and NORs almost indistinguishable in this case.

With an $N_{similar}$ of 5, the heart clinical DVH at a 50% dose was in the 83rd percentile, and this increased to the

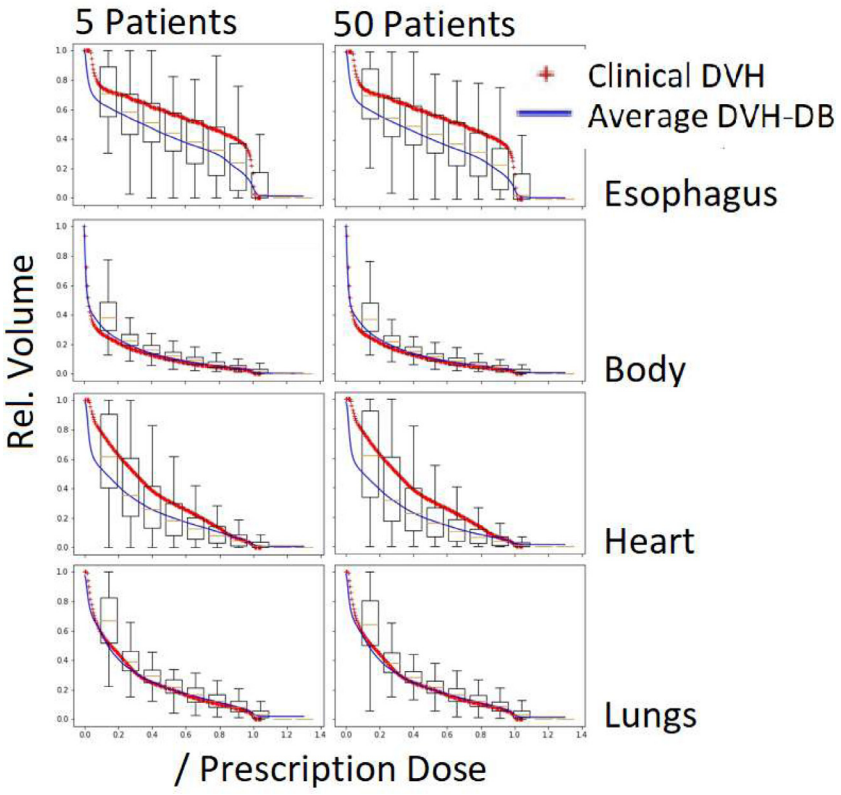

Fig. 3 The clinical dose-volume histogram (DVH) is compared with the database-derived DVH for a single patient. The columns increase the number of patients included in the query of similar patient data from 5 (left column) to 50 (right column). In this case, the clinical DVH is worse (or higher) than the average database DVH in the heart and esophagus.

90th percentile when the $N_{similar}$ was 50. This means that 83% of the DVH-DB datapoints were better than the clinical plan at that dose level using 5 patients, and increasing the amount of data tenfold increased the confidence that potentially better plans may exist. Comparing the left column of Figure 3 to the right column (5 patients vs 50 patients), visually, there was little difference in the IQR and NOR at all dose levels for all OARs. For this patient, whether 4% or 40% of the database was queried, Figure 3 shows that the clinical plan (or the prospective clinical plan, in the case of a new patient) was inferior to the database averages.

A similar result is shown in Figure 4, but in this case, the clinical DVH was superior to most of the database DVH data, independent of the number of patients included. In this case, the DVH-CL remained within the IQR for all doses and all OARs. The DVH-CL of the esophagus, with an $N_{similar}$ of 5 and at 50% of the prescription dose, was within the 18th percentile, indicating a good clinical plan. For an $N_{similar}$ of 50, the percentile of the DVH-CL increased slightly to the 20th percentile. In the case of the heart, which has the largest IQR and NOR, the variation in the percentile was greater between $N_{similar}$ cases, with the percentile initially at 21% (at an $N_{similar}$ of 5) but then increasing to 26% (at an $N_{similar}$ of 50) compared with the distribution of similar plans.

These 2 cases stand out to show the ease of identifying clinical plans that were superior to database averages (Fig 3) and inferior to database averages (Fig 4). In prospective treatment planning and in the design of optimization objectives, these cases may support clinical decision making. In both patients, the heart had the largest initial IQR, but this was reduced as the percentage of the prescription dose increased. The esophagus IQR was relatively large at a low dose but remained constant until the dose approached the D_{max}. The large IQR for both the heart and the esophagus shows that geometric similarity between organs may be a weak surrogate for dose prediction. Large variations were shown in the OVH for the esophagus but not for the heart, so OVH variation did not predict large deviation in the DVH. The lungs and external contour had a larger IQR at low doses, and this is indicative of planning technique. Simple 3D-beam arrangements minimize low dose spread compared with arc or helical delivery.

Discussion

Achievable dosimetry may not be directly related to the geometric relationship between the PTV and OARs. Instead, the dominant factor in OAR dosimetry may be

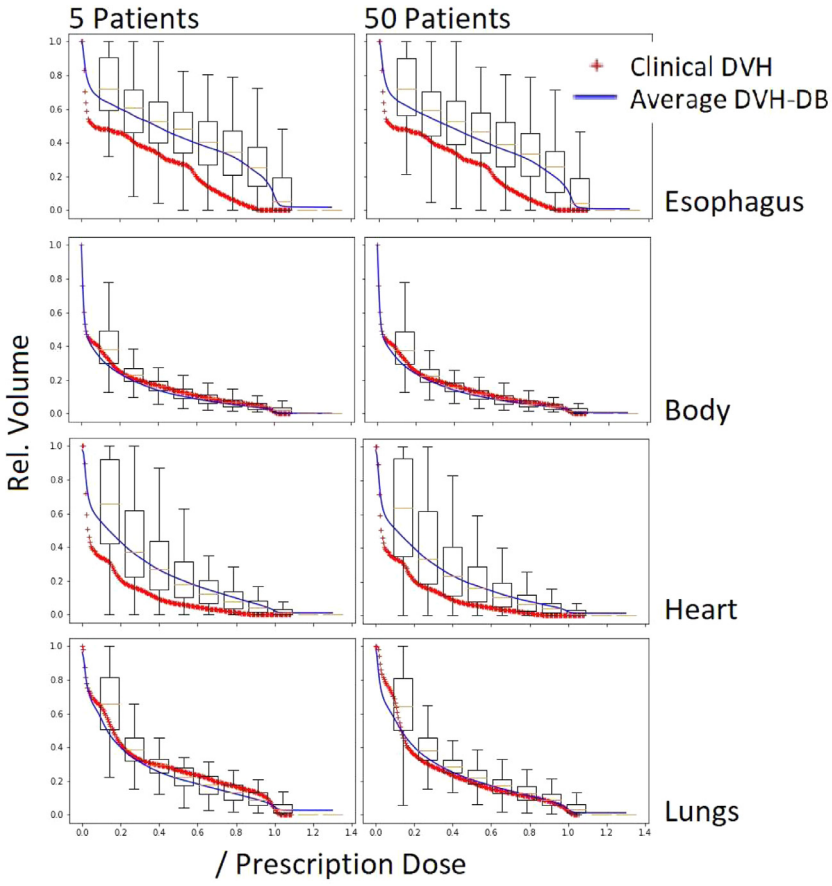

Fig. 4 The clinical dose-volume histogram (DVH) is compared with the database derived DVH for a single patient. The columns increase the number of patients included in the query of similar patient data from 5 (left column) to 50 (right column). In this case, the clinical DVH is better (or lower) than the average database DVH in the heart and esophagus.

planning parameters, including optimization preference and delivery techniques. The OVH variations found in this study indicate that consistent interpatient OAR-PTV geometry can be identified with 5 to 20 patients in a relatively small database. However, the ability to predict DVH does not appear to be predicated on geometric similarity measured in the OVH. Whether 5 or 50 patients were included in the DVH prediction, we found an ability to predict dose in OARs that encompassed the PTV accurately (in this case, lungs and external contours were consistently predicted to within approximately 5%). For organs that did not necessarily overlap the target, there existed significant variations in achievable dosimetry at low and intermediate dose levels in the patients considered.

This study tried to correlate geometric similarity with the ability to predict DVH. Previous approaches have developed predictions based on models or prior data with or without specific strategies to address anatomic similarities in patients.[2,4,8-10,13,15-17,21] However, none of these methods (including those in the current study) have accounted for clinical decision making including delivery technique and optimization preferences among multiple OARs. Multicriteria optimization considers the trade-offs in OAR dosimetry inherently; therefore, a range in the achievable dose is estimated based on variations in importance weighting.[22−28] Trade-offs are inherently built into current treatment planning and plan-selection approaches, although these trade-offs are limited by delivery technique and patient-specific dose limits.[29] Ultimately, the optimal plan is defined by the patient's radiation oncologist with a complete depiction of the patient condition; current algorithms do not incorporate all of the relevant clinical variables in defining optimality. This is a critical limitation of the current study. We used OAR-PTV geometry for DVH prediction without simultaneously assessing the geometry of all OARs. Clinical variables such as patient age and lung function were completely ignored in our DVH lookup. In future work, combining the geometry of multiple OARs with prior data may reveal clinical preference between different OARs.

This study demonstrated that data-driven prediction of the DVH is reasonable for OARs that encompass the PTV. This ability could be due to physical limits in dose delivery and conservation of the integral dose around the

target. Reese et al[30] showed that independent of the delivery technique and organ preference, there is conservation of the integral dose in concentric shells around a target. In practice, however, significant integral dose variations are observed in different modalities.[31-33] The integral dose can be represented by the external DVH, and this study's data showed large variations in the low-dose DVH in the external contour. This result is an obvious one when helical-delivered tomotherapy plans are included with 3D techniques including as few as 4 fixed beams.

Our approach did not explicitly account for variations in delivery techniques. It may be more valuable to use class-based prior dosimetry, with each class consisting only of plans of similar delivery techniques. Knowledge-based planning may only consider one delivery technique and clinical site.[2] However, such a technique is fundamentally flawed if the goal is to show achievable dosimetry. For example, by only including helical or arc delivered treatments, the ability to significantly reduce the low dose will not be shown. This study's results are based on single-institution data and a modest sample of 130 patients with a similar diagnosis treated with photon therapy. These factors will introduce cognitive bias in our interpretation of quality treatment plans, which may vary between clinicians and institutions. However, using our method of data query and DVH data reduction, interphysician and interinstitution plan quality may be readily shown according to data quartiles.

As far as limiting prior data to subsets of similar patients, we found that varying the number of patients did not affect our ability to predict DVH and determine whether a plan was acceptable or an outlier. Whether we included the 5 or the 50 most similar patients, the study findings showed consistent information. Other researchers have suggested that robust and larger databases are needed during knowledge-based planning; Ge and Wu[16] evaluated more than 70 articles on this topic and concluded that "larger data sets collected through multi-institutional collaboration will enable the development of more advanced models," but this study's data suggest smaller, more focused models that capture variations in planning strategies may be more valuable in showing achievable dose than larger sets with similar planning approaches. Specific to focused models, all models should show that OARs that are adjacent to but not overlapping the PTV can be completely avoided via beam placement or preference weighting to within the geometric penumbra of the beam and energy.

Conclusions

Use of prior patient data for prospective patient treatment is an area of increased interest, but there are several hurdles and limitations that could affect the predictive power of model and data-driven approaches. Assessment of overall plan quality is reasonable using population averages of prior RT data. Prospective planning (or treatment plan optimization) based on prior data, however, has the potential to adversely affect personalized patient care and shift plans toward population averages. In plan optimization, achievable dose is more important than population averages. This study's data showed that DVH prediction is reasonable in organs that contain the tumor and have a known overlap.

References

1. Nelms BE, Robinson G, Markham J, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. *Pract Radiat Oncol*. 2012;2:296–305.
2. Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—A review of recent innovations. *BJR*. 2018;91: 20180270.
3. Chung HT, Lee B, Park E, Lu JJ, Xia P. Can all centers plan intensity-modulated radiotherapy (IMRT) effectively? An external audit of dosimetric comparisons between three-dimensional conformal radiotherapy and IMRT for adjuvant chemoradiation for gastric cancer. *Int J Radiat Oncol Biol Phys*. 2008;71:1167–1174.
4. Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys*. 2012;39:7446–7461.
5. Ahmed S, Nelms B, Gintz D, et al. A method for a priori estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning. *Med Phys*. 2017;44:5486–5497.
6. Mayo CS, Yao J, Eisbruch A, et al. Incorporating big data into treatment plan evaluation: Development of statistical DVH metrics and visualization dashboards. *Adv Radiat Oncol*. 2017;2: 503–514.
7. Skarpman Munter J, Sjölund J. Dose-volume histogram prediction using density estimation. *Phys Med Biol*. 2015;60:6923–6936.
8. Sheng Y, Ge Y, Yuan L, et al. Outlier identification in radiation therapy knowledge-based planning: A study of pelvic cases. *Med Phys*. 2017;44:5617–5626.
9. Zhang HH, Meyer RR, Shi L, D'Souza WD. The minimum knowledge base for predicting organ-at-risk dose-volume levels and plan-related complications in IMRT planning. *Phys Med Biol*. 2010;55:1935–1947.
10. Good D, Lo J, Lee WR, Wu QJ, Yin F-F, Das SK. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: An example application to prostate cancer planning. *Int J Radiat Oncol Biol Phys*. 2013;87:176–181.
11. Yuan L, Ge Y, Lee WR, Yin FF, Kirkpatrick JP, Wu QJ. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. *Med Phys*. 2012;39:6868–6878.
12. Wu B, Ricchetti F, Sanguineti G, et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. *Med Phys*. 2009;36:5497–5505.
13. Moore KL, Brame RS, Low DA, Mutic S. Experience-based quality control of clinical intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys*. 2011;81:545–551.
14. Wu B, Ricchetti F, Sanguineti G, et al. Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys*. 2011;79:1241–1247.

15. Zhu X, Ge Y, Li T, Thongphiew D, Wu QJ, Yin F-F, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys*. 2011;38:719–726.
16. Ge Y, Wu QJ. Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches. *Med Phys*. 2019;46:2760–2775.
17. Wang Y, Zolnay A, Incrocci L, et al. A quality control model that uses PTV-rectal distances to predict the lowest achievable rectum dose, improves IMRT planning for patients with prostate cancer. *Radiother Oncol*. 2013;107:352–357.
18. Zhang J, Wu QJ, Ge Y, et al. Knowledge-based statistical inference method for plan quality quantification. *Technol Cancer Res Treat*. 2019;18: 1533033819857758.
19. Oncospace. Available at: https://oncospace.radonc.jhmi.edu/. Accessed September 3, 2021.
20. McNutt T, Evans J, Moore W, et al. WE-G-108-02: Oncospace: A database designed for personalized medicine in radiation oncology. *Med Phys*. 2013;40:501.
21. Schreibmann E, Fox T. Prior-knowledge treatment planning for volumetric arc therapy using feature-based database mining. *J Appl Clin Med Phys*. 2014;15:4596.
22. Craft D, Halabi T, Bortfeld T. Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys Med Biol*. 2005;50:5857–5868.
23. Craft D, McQuaid D, Wala J, Chen W, Salari E, Bortfeld T. Multicriteria VMAT optimization. *Med Phys*. 2012;39:686–696.
24. Serna JI, Monz M, Küfer KH, Thieke C. Trade-off bounds for the Pareto surface approximation in multi-criteria IMRT planning. *Phys Med Biol*. 2009;54:6299–6311.
25. Craft D, Halabi T, Shih HA, Bortfeld T. An approach for practical multiobjective IMRT treatment planning. *Int J Radiat Oncol Biol Phys*. 2007;69:1600–1607.
26. Kamran SC, Mueller BS, Paetzold P, et al. Multi-criteria optimization achieves superior normal tissue sparing in a planning study of intensity-modulated radiation therapy for RTOG 1308-eligible non-small cell lung cancer patients. *Radiother Oncol*. 2016;118:515–520.
27. Miguel-Chumacero E, Currie G, Johnston A, Currie S. Effectiveness of multi-criteria optimization-based trade-off exploration in combination with RapidPlan for head & neck radiotherapy planning. *Radiat Oncol*. 2018;13:229.
28. Xiao J, Li Y, Shi H, et al. Multi-criteria optimization achieves superior normal tissue sparing in intensity-modulated radiation therapy for oropharyngeal cancer patients. *Oral Oncol*. 2018;80:74–81.
29. Watkins WT, Nourzadeh H, Siebers JV. Multiobjective, multidelivery optimization for radiation therapy treatment planning. *Adv Radiat Oncol*. 2020;5:279–288.
30. Reese AS, Das SK, Curle C, Marks LB. Integral dose conservation in radiotherapy. *Medical Physics*. 2009;36:734–740.
31. Nguyen D, Dong P, Long T, et al. Integral dose investigation of non-coplanar treatment beam geometries in radiotherapy. *Med Phys*. 2014;41: 011905.
32. Yang R, Xu S, Jiang W, Xie C, Wang J. Integral dose in three-dimensional conformal radiotherapy, intensity-modulated radiotherapy and helical tomotherapy. *Clin Oncol (R Coll Radiol)*. 2009;21:706–712.
33. Ślosarek K, Osewski W, Grządziel A, et al. Integral dose: Comparison between four techniques for prostate radiotherapy. *Rep Pract Oncol Radiother*. 2015;20:99–103.