A knowledge-based quantitative approach to characterize treatment plan quality: Application to prostate VMAT planning

Buthayna Alnaalwa  
Department of Radiation Oncology, Universitätsmmediizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, Mannheim D-68167, Germany

Obioma Nwankwo  
Strahlentherapie RheinMainNahe, Standort Rüsselsheim, August-Bebel-Str. 59d, Rüsselsheim 65428, Germany

Yasser Abo-Madyan, Frank A. Giordano and Frederik Wenz  
Department of Radiation Oncology, Universitätsmmediizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, Mannheim D-68167, Germany

Gerhard Glatting  
Department of Nuclear Medicine, Universität Ulm, Albert-Einstein-Allee 23, Ruprecht-Karls-Universität Heidelberg, Ulm 89081, Germany

[Correction added on January 17, 2021, after first online publication: Ruprecht-Karls-Universität Heidelberg, was added for Gerhard Glatting.]

(Received 26 June 2020; revised 21 September 2020; accepted for publication 15 October 2020; published 21 November 2020)

Purpose: To characterize treatment plan (TP) quality, a quantitative quality control (QC) tool is proposed. The tool is validated using volumetric modulated arc therapy (VMAT) plans for treatment of prostate cancer by estimating the achievable organ at risk (OAR) sparing, based on the knowledge learned from prior plans.

Methods: Prostate TP quality was investigated by evaluating the achieved OAR sparing in the rectum and bladder, based on their proximity to target surface. The knowledge base used in this work comprises 450 plans, consisting of 181 homogenous prostate plans and 269 simultaneous integrated boost (SIB) prostate plans. A knowledge-based algorithm was used to relate the absorbed doses of the OARs (rectum and bladder) and their proximity to the planning target volume (PTV). A metric ($M_{q,r}$ value) was calculated to characterize the OAR sparing based on the weighted differences of the mean doses at binned distances to the PTV surface. The 90% probability ellipse of the normally distributed OARs $M_{q,r}$ values was considered to define a threshold above which the treatment plan was re-optimized.

Results: Following re-optimization, 8/11 of the homogenous plans and 6/13 of the SIB plans outside the 90% probability ellipse could be re-optimized to gain better OAR sparing while achieving the same or better target coverage. However, 3/4 of the homogenous TPs and 1/9 of the SIB TPs between 80% and 90% were improved. $M_{q,r}$ values of bladder and rectum after re-optimizing the plans in both groups of homogenous and SIB showed lower values compared to the corresponding values before re-optimization, which implies that better OARs sparing was achieved.

Conclusions: This work demonstrates an effective anatomy-specific QC tool for identifying suboptimal plans and determining the achievable OAR sparing for each individual patient anatomy. © 2020 The Authors. Medical Physics published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine [https://doi.org/10.1002/mp.14564]

Key words: knowledge-based radiation therapy treatment planning, OARs dose sparing, quantitative quality control algorithm, VMAT treatment plans re-optimization

1. INTRODUCTION

Inverse treatment planning is a semi-automated process that is subjective and time consuming. The planner sets and adapts the dose objectives for the targets and organ at risks (OARs) to fit dose–volume constraints using an iterative approach. Running more iterations to improve the treatment plan (TP) can be counterproductive if unachievable constraints are used. The most common dose–volume constraints have typically been determined using the Radiation Therapy Oncology Group protocol recommendations. These population-based recommendations can be meaningful in many situations. However, based on these constraints, while the generated plan can be clinically acceptable, it can be far from being the best optimal plan, if the dose to the normal tissue is not minimized to the best extent possible, particularly when inexperienced planners are involved. Thus, plans at radiation therapy centers with limited experience may not be clinically optimal. This unsatisfactory situation highlights the need for a quality control (QC) method to estimate the achievable OARs dose sparing. Such a QC algorithm should not entirely rely on personal judgment but should consider the geometric variations among patients.
Knowledge-based radiation therapy (KBRT) is an approach to improve efficiency and to reduce variability in treatment planning.\(^5\)\(^6\)\(^7\)\(^8\) A variety of KBRT approaches have been reported. Some KBRT approaches are based on search and retrieval of prior optimized patient plans to guide the generation of new plans based on the similarities of their geometries.\(^5\)\(^6\)\(^9\)\(^10\) For example, Wu et al.\(^6\) retrieved a prior patient plan based on similarities of overlapping volume histograms (OVHs), to guide the optimization of the new plan. Chanyavanich et al.\(^9\) used the beam’s eye view of the treatment region of a new case to find the patient with the most similar anatomy from a database of previously treated patients. The planning constraints of the matching plan were used to define the planning goals of the new plan. Another approach predicts the dose–volume histograms (DVHs) of the OARs based on their distance to target histograms (DTHs).\(^12\)\(^13\) More specifically, in this approach, a quantitative evaluation tool is proposed based on machine learning to characterize the relation between the DVH and anatomical features for each patient plan. This information could be used to estimate the DVH for a new TP or to optimize plan quality.\(^14\)\(^15\)

Various efforts using these KB approaches were presented to improve treatment planning consistency.\(^15\)\(^16\)\(^17\)\(^18\) including Rapid plan\(^16\) in Eclipse, multi-criteria optimization (MCO) in RaySearch.\(^17\) RapidPlan is the commercially available DVH guidance approach. In this approach, a mathematical model that correlates the geometrical features of patients to previously achieved dosimetry is used to derive the achievable DVHs for prospective cases.\(^19\) The key limitation of DVH-based approaches is the lack of spatial information. Planners may need extra work to deal with a case with uncommon OAR/target geometry. Furthermore, the DVH only predicts the delineated regions. Optimizing the dose to improve the conformality of tissues outside the delineated region might not be considered. Another approach called MCO tries to find the optimal trade-offs between target coverage and normal tissues sparing. Thus, it helps less-experienced planners to produce high-quality IMRT plans. In MCO, a Pareto surface containing a spectrum of fitting plans is automatically generated by the optimizer. Then, planners navigate through the combinations based on specified trade-offs objectives to choose a Pareto-optimal plan.\(^20\)\(^21\) However, Pareto-optimal plans can be clinically highly undesirable, and do not indicate clinically optimal plans. Pareto optimal is still the best clinically acceptable plan in MCO. The limitation of this approach is the large number of automatically generated Pareto-optimal plans investigating many parameters to optimize the different clinical objectives. Generating a large number of optimal plans needs intensive computation resources, in addition to the accompanied difficulties in the selection of an optimal plan. However, using prior knowledge to automatically predict the DVH of the new plan cannot provide any reasonable estimation of the achievable dose distributions among different patients.\(^22\)

Another KBRT approach described by Nwankwo et al.\(^23\) predicts the dose at voxel level of the OARs. The algorithm was derived from the analysis of a database of high-quality TPs. It can also predict the three-dimensional (3D) dose distribution in the organ of interest. In this work, based on Nwankwo et al.\(^23\) we proposed a QC quantitative algorithm for VMAT prostate planning that estimates a threshold to spot the suboptimal plans to achieve further dose sparing in rectum and bladder using knowledge learned from prior plans.\(^23\) More specifically, the dose of the binned voxels according to their distant location from the PTV surface between the query plan and the reference plans were compared. As a result, this method objectively considers patient-specific anatomical variations in the treatment planning process, and thus enables the comparison of plans irrespective of patient anatomical variations.

2. MATERIALS AND METHODS

2.A. Ethics approval and data base of plans

This study was approved by the Medical Ethics Commission of the Medical Faculty Mannheim, Heidelberg University (2015-816R-MA).

In total, 450 VMAT plans that were approved for prostate cancer treatment at the Department of Radiation Oncology, University Medical Centre, Mannheim, Germany, were used to perform this study. These VMAT prostate plans were randomly selected from the database of patient plans that were treated between 2007 and 2017. The plans were anonymized for use in this study. The plans were split into two groups: 181 homogenous prostate VMAT plans and 269 SIB prostate VMAT plans. The target volumes in the homogenous plans were the prostate and seminal vesicles while the target volumes in the SIB plans were the prostate and seminal vesicles and a boost volume. The prescribed doses to the PTV were variable. These plans were calculated with a fully inverse treatment planning system (commercially available as “Monaco,” CMS, Elekta-Group, UK). The template used for prostate planning\(^24\) specifies the desired PTV coverage and the acceptable doses to the normal tissues. These constraints are iteratively adjusted in a trial-and-error fashion until the dose distribution is deemed to be optimal for the given patient anatomy.

2.B. Extraction of information from the treatment plans

The plans were imported into a MATLAB computing environment using the Computational Environment for Radiotherapy Research (CERR) software.\(^25\) In each patient plan, the following variables were extracted for each voxel in the OARs (rectum and bladder)\(^23\)\(^26\)\(^27\).

1. The coordinates \((x, y, z)\) of each voxel of an OAR. Both the rectum and bladder were the considered OARs in this study.
2. The dose of each voxel which was normalized to the prescribed planning target volume (PTV) dose. Thus,
the dose does not depend on the absolute value of the prescribed dose to the PTV.

3. The distance-to-PTV of each voxel. It is the shortest Euclidean distance between a voxel of an OAR and the surface of the PTV. This value is set negative if the voxel is shared by both the OAR and PTV. The dose at distance to PTV was calculated only for the voxel within the beam path.

Two matrices, one for each OAR (rectum and bladder), were computed from each plan. A row of the matrix represents a voxel while the column specifies its parameters (dose, slice level, distance-to-PTV, and the Cartesian coordinates).

2.C. Data analysis

2.C.1. The mean dose-at-distance function

The mean dose-at-distance function of an organ was computed from the in-field voxels of the OARs. The voxels were first binned (0.5-mm bin size according to their distance-to-PTV). The mean dose-at-distance value was computed for each distance-to-PTV bin, more specifically the mean dose-at-distance values of the in-field voxels was calculated against the center of the corresponding distance-to-PTV bin.

This function relates the mean dose of all the in-field voxels within a distance-to-PTV bin to the median distance of the voxels to the surface of the PTV. The equation for calculating this variable is given in Ref. [23].

2.C.2. The reference set versus the query set

For both groups, homogenous VMAT plans and SIB VMAT plans, the reference sets and the query sets were generated from the same plan matrices of the group. The 181) homogenous plan matrices were used to generate the query and reference of homogenous plans, and, the 269 SIB plan matrices, were used to generate the query and reference of SIB plans. For each OAR, a single merged reference matrix out of (181) homogenous plan matrices, and a single merged reference matrix out of (269) SIB matrices, were derived. The merged reference matrix was generated by grouping all voxels located at approximately same distance to PTV bin (bin size = 0.5 mm) from all reference plans into the same distance to PTV bin. The query set was made up of the same plans (181) homogenous plan matrices and (269) SIB plan matrices, but this time each plan matrix was used separately without being merged (voxels at approximately same distance to PTV bin (bin size = 0.5 mm) in each query plan were grouped into same distance to PTV bin). The mean dose-at-distance function was calculated for the homogenous reference matrix and the SIB reference matrix, as well as, for each single query plan matrix. A schematic representation which summarizes the workflow is reported in Fig. 1.

2.C.3. Plan evaluation metrics: \( M_{q,r} \) value

A comparison of the query and reference dose distributions of both, the homogenous and the SIB VMAT plans, was performed by calculating the \( M_{q,r} \) value, which characterizes the quality of OAR sparing of the query plan (index q) relative to the merged reference plans (index r) according to

\[
M_{q,r} = \frac{1}{N_q} \sum_{i} n_{q,i} \sqrt{\frac{s_{q,r}^2 + s_{r,q}^2}{n_{q,i} + n_{r,i}}} \tag{1}
\]

where \( d_{r,i} \) is the mean absorbed dose at distance-to-PTV bin \( i \) of the reference matrix \( r \) of each OAR (bladder, rectum) and \( d_{q,i} \) is the mean dose at distance-to-PTV bin \( i \) of the query plan \( q \) of each OAR. This mean dose difference considers the shift between each query plan with respect to the reference matrix in each group for each OAR. Negative differences indicate less mean dose in the query plan than the reference matrix, that is, a better query plan quality. \( s_{q,r}^2 \) and \( s_{r,q}^2 \) are the variances of the dose distributions at a given distance-to-PTV bin \( i \) for the query plan and the reference matrix. \( n_{q,i} \) and \( n_{r,i} \) are the numbers of voxels at distance-to-PTV bins \( i \) for the query plan and the reference matrix, respectively. This metric is derived in analogy to a t-test as it compares the sample dose distribution over each single bin \( i \) in a query plan and the sample dose distribution in the corresponding bin in the reference matrix. This sample dose distribution difference over each single bin is weighted by the number of voxels of the query plan at the same bin and summed over all the bins. Finally, the summed value is divided by the total number of voxels \( N_q \) of the query plan which is calculated as

\[
N_q = \sum_{i} n_{q,i} \tag{2}
\]

\( M_{q,r} \) of the rectum and \( M_{q,r} \) of the bladder were calculated for every query plan of each treatment group (homogenous and SIB plans) to characterize the quality of the plans by evaluating the achievable OAR dose sparing for each patient plan. \( M_{q,r} \) values for both OARs represent the data points of this work.

2.C.4. Statistical analysis

Mean values and standard deviations were calculated for \( M_{q,r} \) of both OARs for each group (Table III). Thereafter, the (50%, 80%, 85%, 90%, and 95%) confidence ellipses were drawn for the 2D normally distributed data points, \( M_{q,bladder} \) vs \( M_{q,rectum} \).

2.C.5. Model validation

Patient plans between 80%–85%, 85%–90%, 90%–95%, and 95%–100% probability ellipses to the upper right corner of longitudinal major axis (relatively high \( M_{q,r} \) of both rectum and bladder) and to the lower right corner of longitudinal major axis (relatively high \( M_{q,r} \) of bladder) were exported to TPS to be checked and re-optimized to possibly obtain better OARs sparing while achieving the same or a better target
coverage [Fig. 1]. The replanning process was subjectively performed depending on both the experience of the oncologist and physicist in a trial and error fashion using MONACO® TPS. All the re-optimized plans were accepted for clinical treatment by an experienced physician. Thereafter, \( M_{q,r} \) versus \( M_{q,bladder} \) versus \( M_{q,rectum} \) of the re-optimized plans were calculated and 2D normally distributed data points of re-optimized plans were added to the original probability ellipse bivariate normal distribution for each group. The numbers of improved plans after re-optimization to the total number of plans outside 80%, 85%, 90%, and 95% probability ellipses were calculated. Such quantification of successfully re-optimized plans is a step toward defining a certain threshold above which re-examination and re-optimization of the plans is highly recommended, as they should have a relatively high chance to be improved.

The DVHs were computed for the query plans outside 80%, 85%, 90%, and 95% probability ellipses for the comparison of the plans before and after re-optimization.

2.C.6. Refinement of database

Suboptimal plans possibly get accepted for patient treatment in the absence of a metric that incorporates patient anatomy to aid the treatment planning and acceptance process. To discover such suboptimal plans, the \( M_{q,r} \) value was calculated for all the plans in the original data base. Plans outside 80% probability ellipses to the right of longitudinal major axis were considered to be suboptimal and were replaced by the re-optimized plans to yield an improved database of optimal plans [Fig. 1].

3. RESULTS

The results of the calculated \( M_{q,r} \) values before and after re-optimization are shown in Figs. 2 and 3.

A summary of the findings are as follows:

1. Homogenous plans: Of 11 TPs, 8 located outside the 90% (including 4 of 5 TPs) located outside the 95% ellipse) could be improved after replanning to gain better OARs sparing while achieving the same or better target coverage. Thus, 73% of the plans outside the 90% probability ellipse were improved [Fig. 2(b)]. Of four plans, three located outside the 80% [including one out of two located outside the 85% ellipse] could be improved after re-planning [Fig. 2(c)].
2. SIB plans: Of 13 TPs, 6 located outside the 90% probability ellipse (including 5 of 8 TPs outside the 95% ellipse) were improved after replanning which means
45% of the plans out of the 90% were improved [Fig. 3(b)]. One plan of nine located outside the 80% (including one of seven located outside the 85% ellipse) could be improved after replanning [Fig. 3(c)].

Tables I and II present the $M_{q,r}$ values for the homogenous and SIB plans before and after re-optimization. Most of the plans gain better quality for both rectum and bladder with respect to the $M_{q,r}$ values. However, a relatively small increase of $M_{q,rectum}$ was noticed in the homogenous group of prostate plans [Table I], for example, for the hexagram and right-sided triangle data points outside of the 90% ellipse and the pentagram data point out of the 95% ellipse [Fig. 2(b)]. However, this was within the clinically acceptable ranges and did not degrade the gained quality of the re-calculated plans.

Figures 4 and 5 show seven homogenous plans and four SIB plans, respectively, using their DVHs. Judging from the rectum $M_{q,r}$ difference of homogenous plans (last column of Table I), inverted triangle and triangle plans showed the best and worst $M_{q,r}$ difference to the rectum outside 95%, respectively, while square and right-sided triangle plans showed the best and worst $M_{q,r}$ difference to the rectum outside 90%, respectively. However, regarding homogenous plans outside 80% probability ellipse, circle and hexagram plans showed the best and worst $M_{q,r}$ difference to the bladder suggests that inverted triangle plan has the best $M_q$ difference while circle plan has the worst between plans before and after re-optimization outside 95%. Only the DVH of the 13 plans of the both sets are presented for the sake of space.

Finally, the reference set of both groups for both OARs were refined by replacing all the plans outside 80% probability ellipse with the re-optimized plans. Table III presents the $M_{q,r}$ values for the homogenous and the SIB plans before and after the database was refined.
4. DISCUSSION

We propose a quantitative KBRT algorithm to investigate a threshold for the achievable dose sparing in OARs to help radiotherapy planners to identify suboptimal TPs. The results of this work showed that the majority of the re-optimized plans were of greater quality when compared with the original plans in the database for both SIB and homogenous VMAT prostate TPs especially outside 90%. Although much of plan quality assessment is subjective, and is best made by planners, 8 of the 11 homogenous prostate plans as well as 6 of the 13 SIB prostate plans outside the 90% probability ellipse had superior dose sparing of OARs with more uniform PTV coverage after re-optimization. In addition, three out of the four homogenous TPs and one of the nine SIB TPs located outside 80% could be improved after replanning. Thus, all the re-optimized plans would have potentially resulted in clinical improvements or clinically negligible differences when compared to the original plans. Re-optimizing the plans outside the 90% probability ellipse in both data groups led to more satisfying plans that better matched the planning dose–volume objectives. The quantitative and qualitative ($M_q,r$ and DVHs) comparison of plans before and after re-optimized showed the efficiency and accuracy of this metric in evaluating the agreement between the compared plans.

The replanning, although not the main focus of the study, was performed to confirm that plans falling below certain thresholds defined by this method could be significantly improved. It should be noted that, the observed dose reduction of OARs after the re-optimization could also have been achieved by modifying the planning constraints and objectives during optimization without guidance of a KBRT QC algorithm. However, the fact that the planners of the original plans did not achieve lower OARs doses and the original plans were approved, indicates the difficulty of finding the right set of constraints and objectives for planning and judging the quality of TP based on experience and intuition alone. This emphasizes the usefulness of a QC model for treatment.
Data point shape | Before re-optimization $M_{q,\text{bladder}}$ | After re-optimization $M_{q,\text{bladder}}$ | Difference 
--- | --- | --- | --- 
95%–100% | | | 
Pentagram | 97.0 | 37.7 | 59.3 
Circle | 88.5 | −73.2 | 161.7 
Inverted triangle | 33.9 | −39.9 | 73.8 
Triangle | 65.8 | −66.6 | 132.4 
90%–95% | | | 
Square | 28.3 | −4.3 | 32.6 
Hexagram | 40.7 | −39.4 | 80.1 
Right-sided triangle | 52.2 | −40.9 | 93.1 
Diamond | 79.0 | −0.8 | 79.8 
85%–90% | | | 
Square | 66.4 | 29.9 | 36.5 
80%–85% | | | 
Circle | 22.5 | 11.3 | 11.2 
Hexagram | 32.2 | −49.9 | 82.1 
| $M_{q,\text{rectum}}$ | $M_{q,\text{rectum}}$ | Differences 
--- | --- | --- | --- 
95%–100% | | | 
Pentagram | −1.9 | 0.1 | −2.0 
Circle | 12.9 | −1.0 | 13.9 
Inverted triangle | 66.5 | 19.6 | 46.9 
Triangle | −20.2 | −16.9 | −3.3 
90%–95% | | | 
Square | −30.5 | −56.8 | 26.3 
Hexagram | −28.4 | −15.8 | −12.6 
Right-sided triangle | −23.4 | −2.7 | −20.7 
Diamond | 9.7 | −3.9 | 13.6 
85%–90% | | | 
Square | 10.2 | 2.3 | 7.9 
80%–85% | | | 
Circle | 25.4 | −4.5 | 29.9 
Hexagram | −24.1 | 2.4 | −26.5 

common to increase the sparing of one OAR at the expense of the other OAR. In our approach this is considered: For example, $M_{q,\text{bladder}}$ of a homogenous plan (Table I, pentagram) before re-optimization is 97, while for $M_{q,\text{rectum}}$, it is −1.9; this suggests that a prioritized sparing of rectum compared to bladder may be possible. Even if having negative scores of $M_{q,r}$ to both OARs indicating that both OARs are better spared compared to the reference plans, still one OAR should be lower than the other to indicate the priority of sparing in the competing OARs. An example for this is the homogenous plan (Table I, circle) after re-optimization with $M_{q,\text{bladder}} = −73.2$, $M_{q,\text{rectum}} = −1$ which implies a prioritized sparing of bladder compared to rectum in the re-optimized plan. This adds a considerable worth to the calculated metric.

Furthermore, significant deviations of the OAR dose of a query plan from the reference values implies either a better sparing (lower dose) of query plan compared to the reference or worse sparing (higher dose). However, it should also be emphasized that we do not necessarily conclude that a worse sparing for an OAR in a plan implies an overall poor plan quality. The identified deficiencies in the suboptimal plans fell in one of two categories; insufficient modulation level

**TABLE I.** $M_{q,r}$ values of bladder and rectum, before and after re-optimization for homogenous VMAT prostate plans. Lower $M_{q,r}$ values indicate better quality, which implies better OAR sparing was achieved. Plans between 80%–85%, 85%–90%, 90%–95%, and 95%–100% probability ellipses were considered.

**TABLE II.** $M_{q,r}$ values of bladder and rectum, before and after re-optimization for SIB VMAT prostate plans. Lower $M_{q,r}$ values indicate better quality, which implies better OAR sparing was achieved. Plans between 80%–85%, 85%–90%, 90%–95%, and 95%–100% probability ellipses were considered. Both rectum and bladder in all the SIB plans gained better sparing (lower $M_{q,r}$ values).
which leads to wider dose gradients than optimally possible, and insufficient use of cost-functions especially regarding the balance between bladder and rectum that sometimes led to unnecessary dose spillage in one of the two organs. More specifically, this is due to inexperience and the absence of a tool that incorporates the anatomies of different patients to help the planner in the assessment of the plan quality. In addition, it can be due to the compromised sparing of an OAR for the benefit of another competing OAR which is deemed unsatisfactory focused.

In addition, this method showed that the chance to improve plans diminishes as we move inward toward the central part of ellipses. In homogenous plans (4/5 TPs) located outside 95% and (5/7 TPs) between 90% and 95% were able to be re-optimized. In SIB (5/8 TPs) outside 95%, as well as, (1/5 TPs) between 90% and 95% achieved better OARs sparing when re-optimized. However, it is still suggestive to check and improve the quality of the plans within 80%, 50%, etc. In fact, it could be argued that with sufficient time and experience, most TPs could be improved. Although this is true, a cost-benefit consideration must be made: Is it worth investing hours to improve a plan to a level (relative to a reference) that is clinically irrelevant? As the gains are the smaller the more we move towards the center of the ellipses, a threshold ellipse must be defined outside of which the plans are to be improved. It is also important to note that the use of clinically accepted plans which have relatively high quality is quite important because the algorithm is based on the mean values of dosimetric features of the used reference plans. We believe that this method can be used to eliminate suboptimal plans iteratively from the reference set which will strict the characterizing capability of the used method and increase its efficiency. Therefore, the reference set of both groups for both OARs were refined by replacing all the plans outside 80% probability ellipse with the matching re-optimized plans with improved quality.
Several studies have demonstrated the importance of adoption of QC methods to predict and quantify the achievable OAR sparing to provide planning consistency.\(^6,7\) Wu et al.\(^6\) proposed a QC method to optimize the DVHs of the OARs in new TPs based on an OVH descriptor to identify related patients. They reported a clinically significant excess radiation dose that was delivered to patients as a result of insufficient plan QC. Other recent advances in knowledge-based methods used machine learning.\(^12,28\) Zhu et al.\(^12,13\) proposed a quantitative evaluation tool based on machine learning to characterize the relation between DVH and anatomical features, which could be used later to optimize plan quality.\(^14,15\) Although these approaches help planners to achieve better OAR sparing based on the DVH-guidance, the interpatient anatomical variations and their impact on the OARs sparing need more than DVH objectives to achieve lowest possible dose to OARs.

An interesting approach that has been investigated to overcome these issues is voxel-based dose prediction. Nwankwo et al.\(^23\) proposed a KB QC approach depending on the mean

---

**Table III.** Summary of \(M_q\) values — Mean ± Standard Error (SE), STD ± SE, the correlation parameter between bladder and rectum \(M_q\) values (rho factor) — for both treatment groups before and after the database was refined. Lower \(M_q\) values indicate better quality, which implies that better OAR sparing was achieved.

| Data-base refined | Homogenous TPs | SIB TPs |
|-------------------|---------------|---------|
|                    | Before        | After   | Before        | After   |
| \(M_q,\text{rectum}\) |              |         |              |         |
| Mean ± SE         | −2.2 ± 1.1    | −2.1 ± 1.0 | −1.5 ± 1.1    | −1.5 ± 1.1 |
| STD ± SE          | 14.6 ± 5.7    | 13.5 ± 0.7 | 17.8 ± 0.8    | 18.2 ± 0.8 |
| \(M_q,\text{bladder}\) |              |         |              |         |
| Mean ± SE         | −0.4 ± 2.6    | −0.4 ± 2.4 | 11.4 ± 2.1    | 10.7 ± 2.0 |
| STD ± SE          | 35.5 ± 1.9    | 33.0 ± 1.7 | 34.5 ± 1.5    | 34.0 ± 1.4 |
| Rho factor        | 0.17          | 0.17     | 0.20          | 0.23     |

---

**Figure 5.** Validation of the model of the bladder and rectum of the SIB plans. The DVHs were computed for both OARs of the query plans before and after re-optimization. Within each percentile range (column), the first two rows showed the best and worst plans based on the \(M_q\) difference of plans (titles of subplots) before and after re-optimization.
dose difference at voxel level between the compared plans. Although, in the proposed approaches of this work and that of Nwankwo et al., the same KB algorithm was used, our approach is unique as it introduces a threshold to spot the sub-optimal plans that are highly recommended to be re-optimized using the $M_{q,r}$ metric. In addition to the mean dose difference at binned distance to PTV, other factors that are expected to have an impact on the $M_{q,r}$ metric were considered, to calculate more realistic outcomes. For example, the number of estimated standard deviations of the sample mean dose from its expected value were considered. This estimation includes different variables that directly affect the $M_{q,r}$ metric values. First, the variances of the dose distributions at a given distance-to-PTV bin for the query plan and the reference matrix that give a feedback on the dose scattering around the mean over the compared bins. Second, the numbers of voxels at distance-to-PTV bins for both query and reference, which corresponds to the number of observations, since this metric was derived in analogy to a t-test. If we have a large number of observations and all of these observations are close to the sample mean (large $n$, small $s$), we can be confident that our estimate of the sample dose distribution is fairly accurate, which results in a small $M_{q,r}$ value. The large number of plans used in this approach database enables the approach making more accurate calculations, since it allows more observed variations in organ geometries. However, a limitation of this approach is that the quality of the new plan strongly depends on the quality of the reference plans.

Finally, this KBRT QC approach is expected to be efficient even with more complicated treatment sites and their corresponding treatment techniques, for example, head and neck tumors and Hyper-Arch technique. Even though these techniques improve the sparing of the adjacent critical tissue while delivering a more conformal dose to the PTV, they still need to fulfill the DVH-guidance. However, judging the real benefits of integrating this algorithm for more complex treatment sites/techniques is a subject that requires further investigations. We believe that the adoption of this QC method in the planning process will provide better planning efficiency in radiation therapy centers, especially ones with a lower experience level.

5. CONCLUSION

In this work, a knowledge-based QC algorithm using clinically approved prostate VMAT TPs successfully characterized the achievable OARs dose reductions based on individual patient anatomy. The algorithm correctly identified suboptimal plans, a high percentage of which showed further OAR sparing after re-optimization. The adoption of such a method will advance the quality of current planning by providing better treatment planning consistency.

ACKNOWLEDGMENTS

BA is grateful for full funding support for this research study by the Islamic Development Bank-Merit Scholarship Program (funding code 600031512). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

REFERENCES

1. Thieke C, Bortfeld T, Niemierko A, Nill S. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. Med Phys. 2003;30:2332–2339.
2. 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.
3. Das IJ, Cheng CW, Chopra KL, Mitra RK, Srivastava SP, Glatstein E. Intensity-modulated radiation therapy dose prescription, recording, and delivery: patterns of variability among institutions and treatment planning systems. J Natl Cancer Inst. 2008;100:300–307.
4. Moore KL, Brame RS, Low DA, Mutic S. Quantitative metrics for assessing plan quality. Int J Radiat Oncol Biol Phys. 2012;22:62–69.
5. Good D, Lo J, Lee WR, Wu QJ, Yin FF, 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.
6. 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.
7. 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.
8. 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.
9. Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys. 2011;38:2515–2522.
10. 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.
11. Petit SF, Wu B, Kazhdan M, et al. Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma. Radiother Oncol. 2012;102:38–44.
12. 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.
13. Zhu X, Ge Y, Li T, Thongphiew D, Yin FF, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys. 2011;38:719–726.
14. Appendolzer 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.
15. Schubert C, Waletzko O, Weiss C, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. PLoS One. 2017;12:e0178034.
16. Tol JP, Delaney AR, Dahele M, Slotman BJ, Verbakel WF. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol Biol Phys. 2015;91:612–620.
17. Zieminski S, Khandekar M, Wang Y. Assessment of multi-criteria optimization (MCO) for volumetric modulated arc therapy (VMAT) in...
hippocampal avoidance whole brain radiation therapy (HA-WBRT). J Appl Clin Med Phys. 2018;19:184–190.
18. Craft DL, Halabi TF, Shih HA, Bortfeld TR. Approximating convex pareto surfaces in multiobjective radiotherapy planning. Med Phys. 2006;33:3399–3407.
19. Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol. 2016;120:473–479.
20. Monz M, Kufer KH, Bortfeld TR, Thieke C. Pareto navigation: algorithmic foundation of interactive multi-criteria IMRT planning. Phys Med Biol. 2008;53:985–998.
21. Teichert K, Suss P, Serna JI, Monz M, Kufer KH, Thieke C. Comparative analysis of Pareto surfaces in multi-criteria IMRT planning. Phys Med Biol. 2011;56:3669–3684.
22. Shiraishi S, Tan J, Olsen L, Moore K. Knowledge-based prediction of plan quality metrics in intracranial stereotactic radiosurgery. Med Phys. 2015;42:908.
23. Nwankwo O, Sihono DS, Schneider F, Wenz F. A global quality assurance system for personalized radiation therapy treatment planning for the prostate (or other sites). Phys Med Biol. 2014;59:5575–5591.
24. Nwankwo O, Mekdah H, Sihono DS, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. Radiat Oncol. 2015;10:111.
25. Deasy JO, Blanco AL, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys. 2003;30:979–985.
26. Brown WE, Sung K, Aleman DM, Moreno-Centeno E, Purdie TG, McIntosh CJ. Guided undersampling classification for automated radiation therapy quality assurance of prostate cancer treatment. Med Phys. 2018;45:1306–1316.
27. Stanhope C, Wu QJ, Yuan L, et al. Utilizing knowledge from prior plans in the evaluation of quality assurance. Phys Med Biol. 2015;60:4873–4891.
28. Babier A, Mahmood R, McNiven AL, Diamant A, Chan TCY. Knowledge-based automated planning with three-dimensional generative adversarial networks. Med Phys. 2020;47:297–306.
29. Kim TK. $T$ test as a parametric statistic. Korean J Anesthesiol. 2015;68:540–546.
30. Boutilier JJ, Craig T, Sharpe MB, Chan TC. Sample size requirements for knowledge-based treatment planning. Med Phys. 2016;43:1212–1221.