Advances in Radiation Oncology (2024) 9, 101533

advances
in radiation oncology

www.advancesradonc.org

Scientific Article

A Visualization and Radiation Treatment Plan Quality Scoring Method for Triage in a Population-Based Context

Alexandra O. Leone, MBS,[a,*] Abdallah S.R. Mohamed, MD, PhD,[b] Clifton D. Fuller, MD, PhD,[b] Christine B. Peterson, PhD,[c] Adam S. Garden, MD,[b] Anna Lee, MD, MPH,[b] Lauren L. Mayo, MD,[b] Amy C. Moreno, MD,[b] Jay P. Reddy, MD, PhD,[b] Karen Hoffman, MD,[b] Joshua S. Niedzielski, PhD,[a] Laurence E. Court, PhD,[a,c] and Thomas J. Whitaker, PhD[a]

[a]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas; [b]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; and [c]UTHealth Houston Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas

Received 2 January 2024; accepted 16 March 2024

Purpose: Our purpose was to develop a clinically intuitive and easily understandable scoring method using statistical metrics to visually determine the quality of a radiation treatment plan.

Methods and Materials: Data from 111 patients with head and neck cancer were used to establish a percentile-based scoring system for treatment plan quality evaluation on both a plan-by-plan and objective-by-objective basis. The percentile scores for each clinical objective and the overall treatment plan score were then visualized using a daisy plot. To validate our scoring method, 6 physicians were recruited to assess 60 plans, each using a scoring table consisting of a 5-point Likert scale (with scores ≥3 considered passing). Spearman correlation analysis was conducted to assess the association between increasing treatment plan percentile rank and physician rating, with Likert scores of 1 and 2 representing clinically unacceptable plans, scores of 3 and 4 representing plans needing minor edits, and a score of 5 representing clinically acceptable plans. Receiver operating characteristic curve analysis was used to assess the scoring system's ability to quantify plan quality.

Results: Of the 60 plans scored by the physicians, 8 were deemed as clinically acceptable; these plans had an 89.0th ± 14.5 percentile value using our scoring system. The plans needing minor edits or deemed unacceptable had more variation, with scores falling in the 62.6nd ± 25.1 percentile and 35.6th ± 25.7 percentile, respectively. The estimated Spearman correlation coefficient between the physician score and treatment plan percentile was 0.53 ($P < .001$), indicating a moderate but statistically significant correlation. Receiver operating characteristic curve analysis demonstrated discernment between acceptable and unacceptable plan quality, with an area under the curve of 0.76.

Conclusions: Our scoring system correlates with physician ratings while providing intuitive visual feedback for identifying good treatment plan quality, thereby indicating its utility in the quality assurance process.

© 2024 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Sources of support: This work was funded by Varian Medical Systems.

Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

*Corresponding author: Alexandra O. Leone, MBS; Email: aoleone@mdanderson.org

https://doi.org/10.1016/j.adro.2024.101533

2452-1094/© 2024 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Current radiation oncology clinical workflows often rely on peer review to prevent poor-quality plans from being used for treatment. This process involves formal intra- or interinstitutional evaluation of a treatment plan for a given patient, including but not limited to the organ-at-risk (OAR) contours, target volumes, beams, dose and fractionation, and dosimetry. Unfortunately, a standard for conducting peer review of treatment plans is lacking, resulting in subjectivity in the quality assurance of treatment plans.[1] Therefore, many in the radiation oncology community have pushed to create an objective metric for standardized grading of plan quality.[2-7]

Using dose-volume histogram (DVH) metrics for treatment plan evaluation is standard practice. Achieving normal tissue constraints indicates that a radiation treatment plan is safe; however, it does not indicate that it is the best normal tissue–sparing plan. Researchers have defined many metrics not based on DVHs to assist in the peer review process for treatment plans. For example, conformality indices have been suggested for assessing congruence between the planning target volume and prescription isodose,[7] and other researchers have implemented a ratio-based approach comparing the planned target and OAR doses with the associated tolerance values.[8] These methods provide insight into the geometric distribution of dose that is missing when only using DVH metrics for plan evaluation. Combining geometric properties to a DVH metric, known as a "vectorized DVH," has been explored by one research group.[9] These evaluation methods are helpful but, on their own, do not provide the clinical context needed to assess each value. Alongside the push for novel evaluation metrics, there has been an increase in the use of graphical methods to display the associated results. For example, radar plots have demonstrated utility in showing data across different dimensions, such as plan quality criteria.[8,10] The radar or spider plot has 4 or more axes integrated into a single radial figure on which data for 1 or more cases can be presented simultaneously. One study used this type of plot to display key differences between their scoring system's chosen plan and the plan chosen by the physician.[10] Evaluation of this method was consistent with radiation oncologist evaluation, demonstrating the value of using plan quality metrics in conjunction with a graphical display.[11]

Although these methods described are promising, they only indicate whether a clinical objective was met; they do not convey if the clinical goals are likely improvable. Integrating knowledge-based planning or population-based methods would help provide the needed clinical meaning to the evaluation results. For example, if a plan does not meet a specific clinical objective, it only demonstrates that the plan may not be safe. However, by using a population of similarly treated patients as a reference, we can determine whether the clinical objective is commonly violated.

Even with these metrics in place, chart rounds in the peer review process are not always successful. One previous study demonstrated that the longer physicians spend conducting peer reviews, the less likely they are to identify errors in a plan.[12] With 1 in 2 physicians reporting symptoms of burnout,[13-16] an efficient, well-understood plan quality assessment tool can be a beneficial supplement to the treatment plan review process by reducing physician workload. We sought to develop clinically intuitive and easily understandable statistical metrics to determine the quality of a radiation treatment plan and display it visually.

Methods and Materials

Materials

A data set was used that consisted of 111 patients with a variety of head and neck cancers to the tongue, tonsils, and larynx treated with 70.0 Gy in 35 fx using a variety of intensity modulated radiation therapy (including static gantry techniques and volumetric modulated arc therapy). This cohort represents 85 males and 36 females ranging in age from 37 to 85 years, with a median age of 59 years, who underwent treatment from 2006 to 2017. All patients in this cohort were auto-contoured using an in-house model[17] retrospective of treatment for OAR consistency across plans. This ensured all patients had each of the 12 OARs evaluated and that they were drawn without human bias. Evaluation of each plan was based on the institutional head and neck cancer treatment planning directive shown in Table 1.

Computation of the population statistics and the individual plan report required no specialized hardware and was completed on a local desktop (RAM 32.0 GB, Intel Core i7-8665U CPU @ 1.90GHz, 2.11 GHz, 64-bit operating system) in approximately 2 minutes.

Statistical methods

For each planning objective, a density plot was created to visualize the distribution of dose values for each specified metric (Fig. 1). Once the distributions were reviewed, an empirical cumulative density function (eCDF) per objective was created. The eCDFs normalize for the inherent differences in objectives (eg, mean parotid dose vs maximum dose to the brain stem), thus allowing all objectives to be visualized on the same scale. The empirical cumulative distribution function, also known as the "empirical distribution function," is a step function that jumps by 1/n at each data point, representing the fraction of observations less than or equal to a specified value. In this context, the inverse of the eCDF function is taken to

Table 1 Head and neck cancer treatment planning directive for patients treated with 70.0 Gy in 35 fractions

ROI	Clinical goal
Brain stem	Maximum dose <54 Gy
Spinal cord	Maximum dose <45 Gy
Brain	Maximum dose <50 Gy
Left parotid	Mean dose <26 Gy
Right parotid	Mean dose <26 Gy
Mandible	Maximum dose <70 Gy
Left cochlea	Maximum dose <35 Gy
Right cochlea	Maximum dose <35 Gy
Left submandibular gland	Mean dose <30 Gy
Right submandibular gland	Mean dose <30 Gy
Larynx	Mean dose <30 Gy
Esophagus	Maximum dose <40 Gy
Esophagus	Mean dose <30 Gy

Abbreviation: ROI = region of interest.

translate each dose value in the population to a percentile value ranging from 0 for the maximum dose to 100 for the minimum dose in the population. For example, as visualized in Fig. 1, a dose value of 32.6 Gy was translated into a percentile value of 0.61 (or 61.0%) by the eCDF. As a result, each DVH constraint value in each plan had a corresponding percentile rank.

The plan score was determined by first calculating per patient the geometric mean of the percentile value for each of the 13 clinical DVH objectives in the planning directive (Table 1). A density plot of the geometric means for all patients in the population was then created to visualize the distribution of plan scores within the population. An eCDF was then created, mapping each geometric mean directly to a percentile value ranging from 0 for the minimum plan score to 100 for the maximum plan score (Fig. 2).

Plan quality visualization

For visualization of the results for our treatment plan scoring method, an interpretation of the radar plot, called the "daisy plot," was used. Similar to a radar plot, our daisy plot was designed to integrate multiple axes into a single radial figure. Each bar or petal on the daisy plot displays a nominal categorical variable. To demonstrate the daisy plot, a patient was selected from our population and evaluated based on the physician-requested DVH metric limits for each clinical objective. The daisy plot was created by looking up the percentile values for the physician-requested DVH values and achieved patient DVH values from each objective eCDF. The overall treatment plan score was then found by looking up the percentile rank of the geometric mean of the patient's achieved objective-eCDFs (Fig. 3). Each planning objective score for the patient is displayed as an individual bar on the plot. The height of each bar represents the achieved percentile value for the listed objective, and all bars are arranged from lowest (worst-performing structure) to highest (best-performing structure). A black dot alongside each bar also

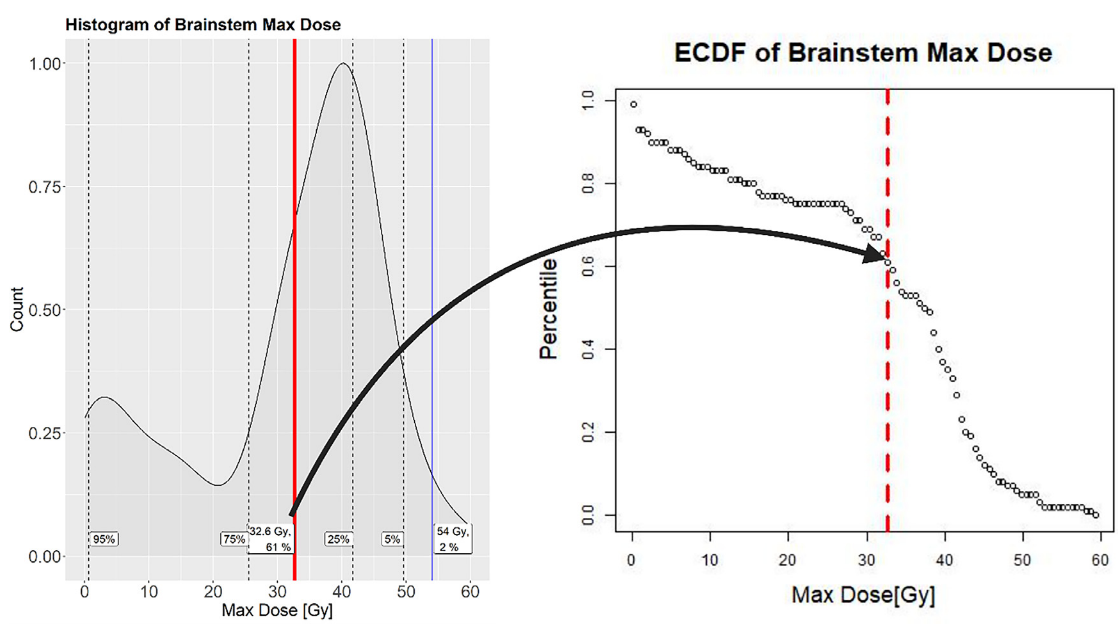

Figure 1 Sample distribution of the clinical objective maximum brain stem dose in a histogram and its translation to a percentile using an inverse empirical cumulative density function. The red lines in both plots represent a patient's brain stem objective value.

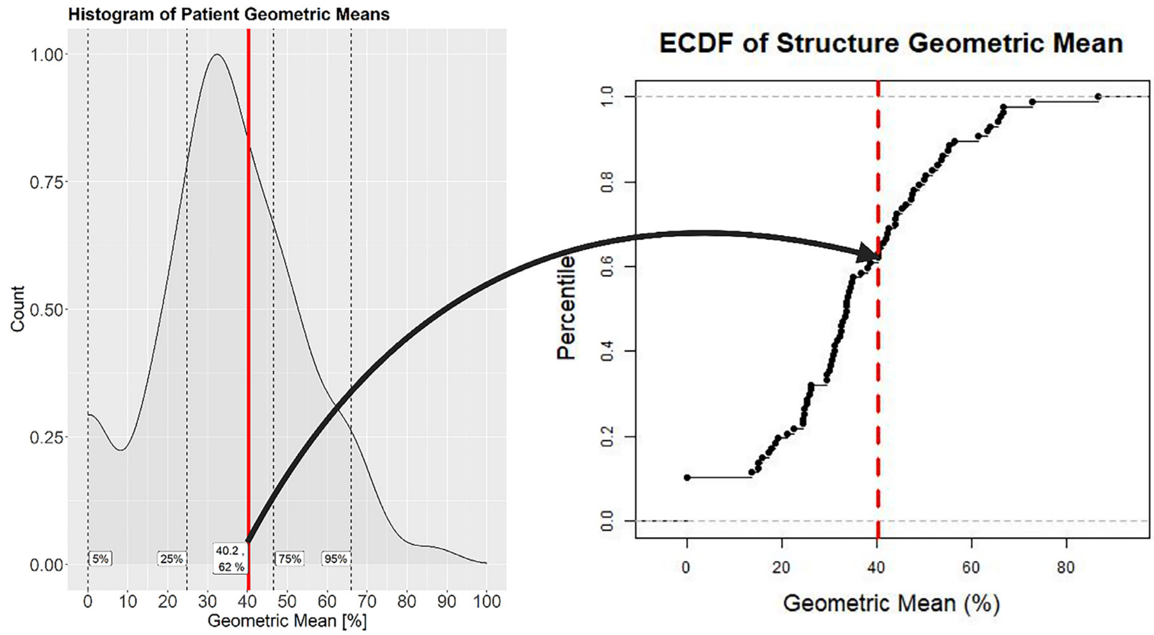

Figure 2 Translation of the geometric mean histogram distribution to a percentile in an empirical cumulative density function. The red line in both plots represents the treatment plan value for a sample patient, with a geometric mean of 40.2 representing an overall plan percentile of 62%.

denotes each achieved objective value to display it more definitively than the height of the bar alone. A gray dot in each bar indicates the percentile rank of the physician-requested value.

The patient's overall plan percentile, or score, is located in the middle of the circle in the daisy plot. The difference between the achieved and requested percentile values is indicated by the coloring of the bars. The values used to select the color of each bar range from -100% to +100% in increments of 25%. A -100% difference signifies that the achieved dose is significantly higher than the requested value, whereas a +100% difference signifies that the achieved value is significantly lower than the requested value (Fig. 4). The daisy plot graphics were created in the R computing language using the ggplot2 package.[18,19]

Physician treatment plan review

To validate our treatment plan scoring system, 6 head and neck radiation oncologists were recruited to evaluate 60 plans randomly chosen from the 111 identified in the data set that met our criteria for analysis. A further randomization was performed to assign 10 plans to each of

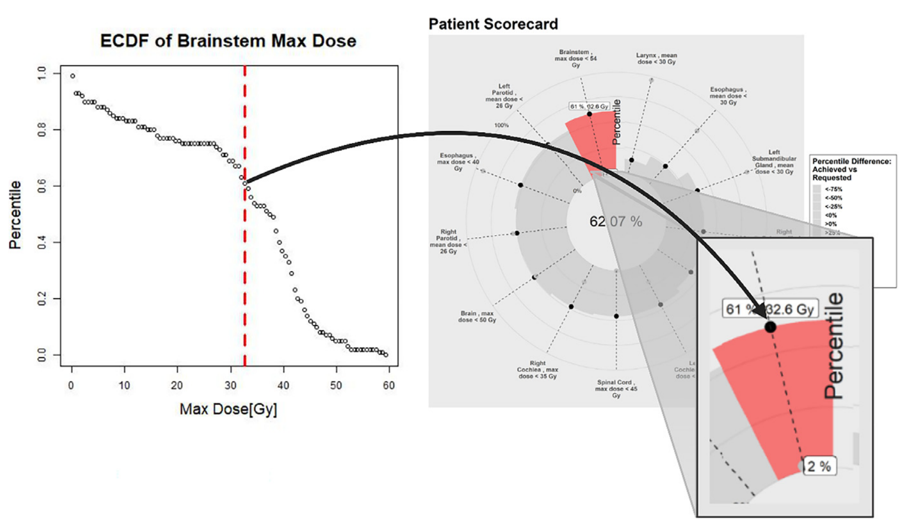

Figure 3 Display of a patient's maximum brain stem dose percentile in a daisy plot on the patient's treatment plan scorecard.

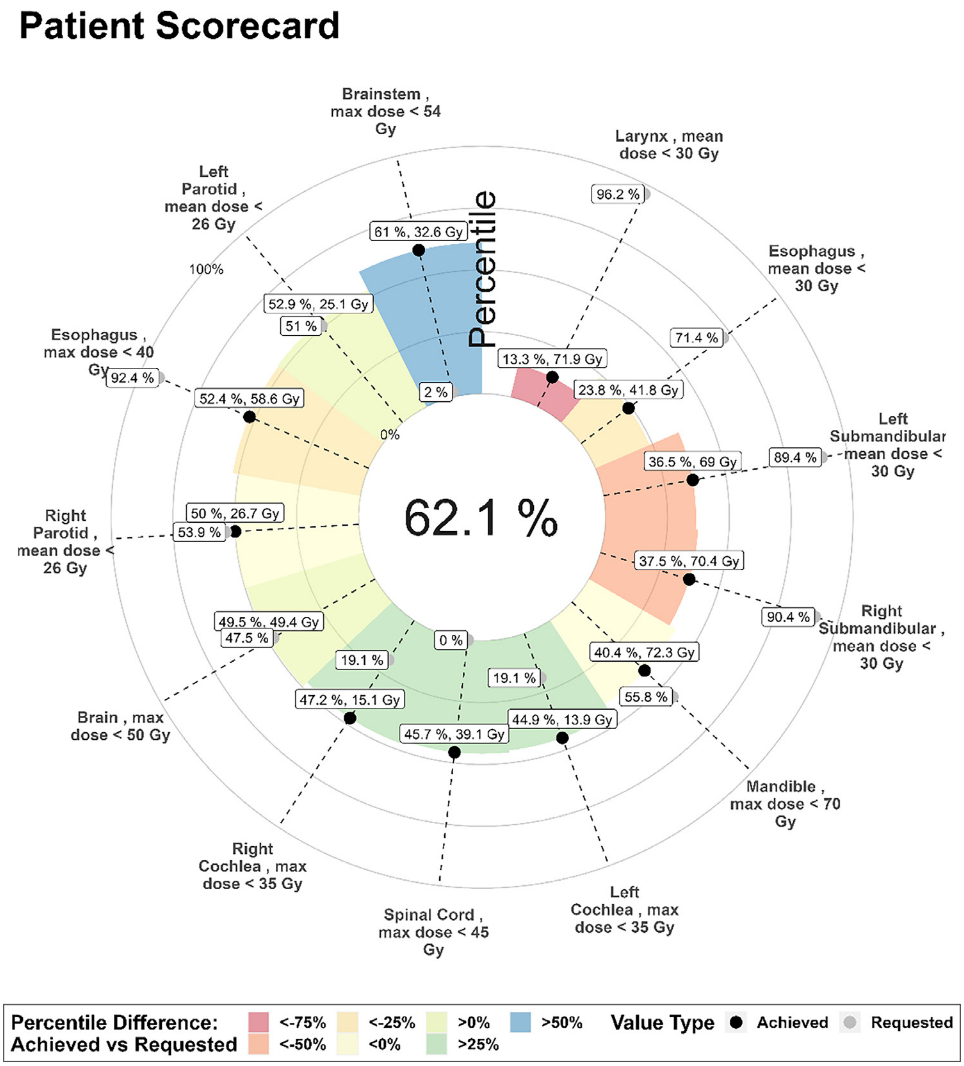

Figure 4 A sample treatment plan scorecard for a patient with a malignant neoplasm in the supraglottis.

the 6 physicians. The physicians were provided the 5-point Likert-style scale in Table 2 to evaluate plan quality, in which a score of 5 represents a plan deemed clinically acceptable and a score of 1 represents a plan that fails to meet the outlined clinical criteria. Spearman correlation analysis was conducted to examine the association, if any, between increasing treatment plan percentile rank and physician rating, with Likert scores of 1 and 2 representing clinically unacceptable plans, scores of 3 and 4 representing plans that may require or have optional preferential minor edits, and a score of 5 representing clinically acceptable plans. Also, a receiver operating characteristic curve was used to assess the scorecard's ability to identify a treatment plan needing minimal edits. In this assessment, scores of 1 and 2 were deemed as true failures, indicating that the plans would require major editing or fail to meet the clinical objectives. Scores of 3, 4, and 5 were the true positive values, indicating that plans can be used for treatment, with only some requiring minor or stylistic edits.

Table 2 Likert scoring system provided to physicians upon review of treatment plans

Score	Description
5	Acceptable – Would not make any changes and happy to sign off on plan
4	Minor improvements possible – Would request possible improvements if time permits
3	Minor improvements needed – Would request improvements before allowing treatment
2	Major improvements needed – Major improvements needed before allowing treatment
1	Unacceptable – Fails to meet clinical criteria

Results

Of the 60 plans scored using our system, 8 plans were clinically acceptable (Likert score 5) and had a median plan score of 89th ± 14.5 percentile. There were 35 plans needing minor edits and 17 that were deemed unacceptable. These groups had more variation, with the former having a median score falling in the 62.6nd ± 25.1 percentile and the latter having a median score falling in the 35.6th ± 25.7 percentile (Fig. 5). A Spearman correlation coefficient of 0.53 ($P < .001$) further quantified the correlation between our plan scores and the physician grades. Plans that received high scores were discernable as being acceptable by physicians. This is supported by the minor overlap of the IQR of the plans marked acceptable and those of the plans needing minor edits in Fig. 5. The scoring system's ability to separate plans requiring minor edits from those that are unacceptable was less strong as demonstrated by the IQR overlap. Discernment between acceptable (Likert 4 and 5) and unacceptable (Likert 1, 2, and 3) plan quality was further displayed by our receiver operating characteristic curve analysis, with an area under the curve of 0.76 (Fig. 6).

Discussion

Radiation treatment plan quality is often determined by answering 2 questions. The first question—"Is the plan safe?" —is frequently answered using a checklist of DVH-based planning objectives. A plan is considered safe if the resulting DVH values are below the maximum allowed objective value. If the plan is deemed safe, the second question, "Is the plan optimal?," is often less formalized. Rather, the answer is determined by a subjective review of the plan performed by a physician in which a mental checklist of plan attributes is evaluated. These attributes may include dose-target conformity or an ideal value for a DVH objective that is not documented but rather the physician's preference over the typical clinical value. Our treatment plan scoring method is an objective way to determine radiation treatment plan quality using a library

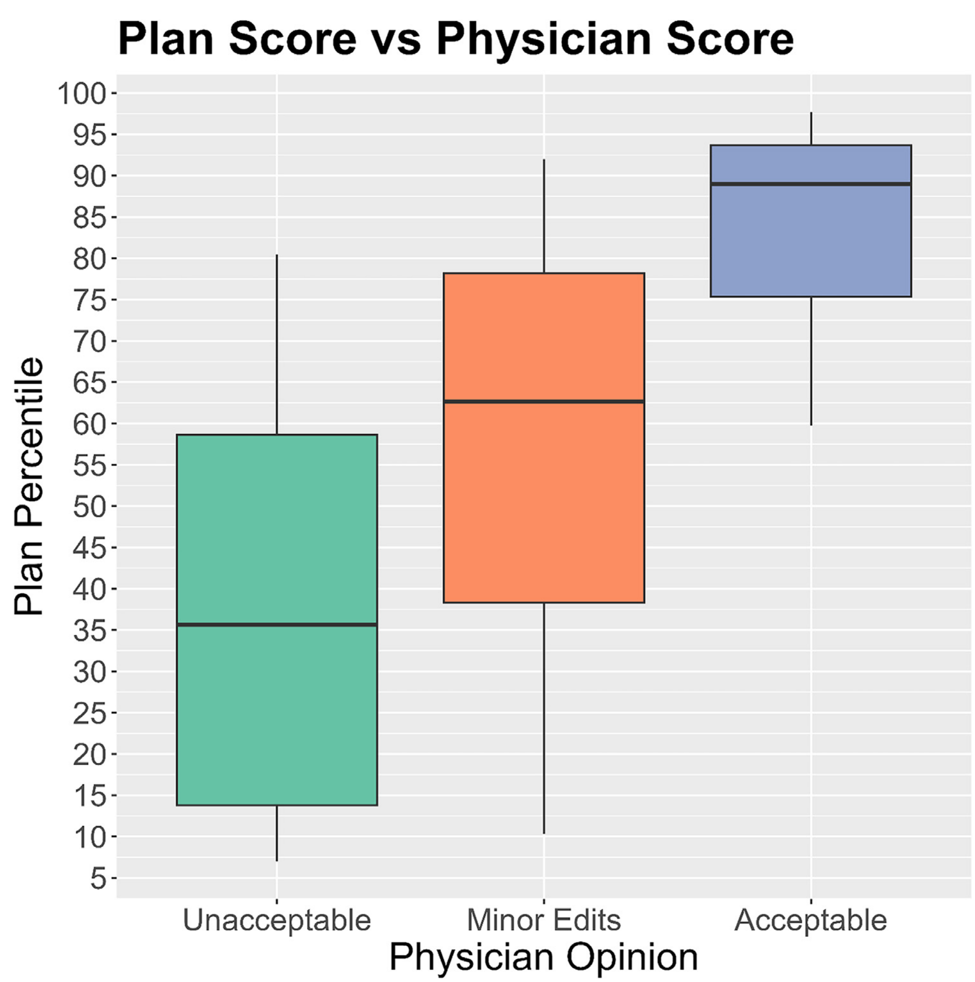

Figure 5 Box and whisker plot of our treatment plan scorecard percentiles versus physician opinion category based on Likert scores.

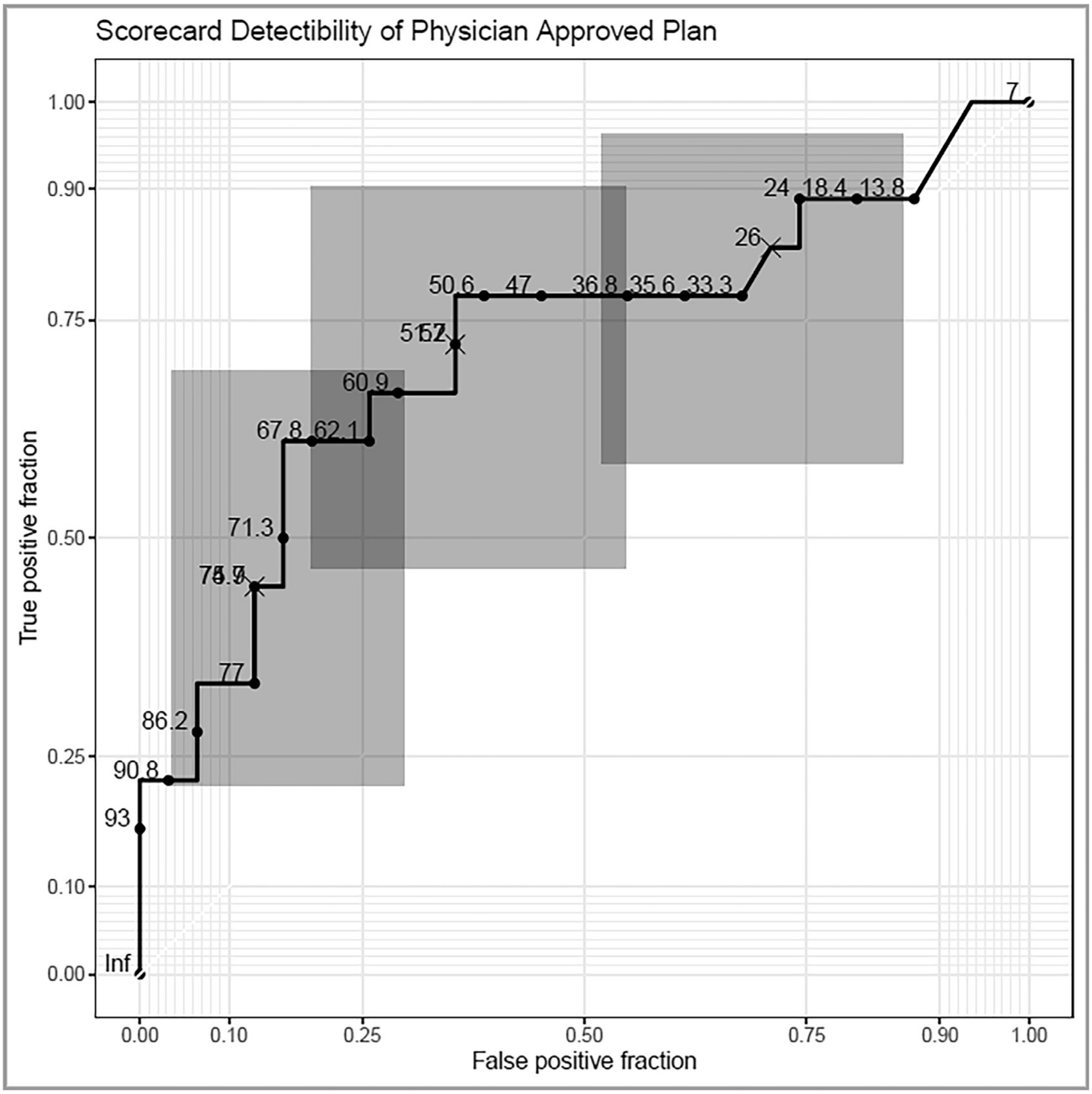

Figure 6 Receiver operating characteristic curve of true positive (Likert score 3-5) versus false positive (Likert score 1 and 2) for overall treatment plan percentile.

of similar treatment plans. The results obtained using this method correlate with those of physician review regarding these plans and provide insight into the safety metrics used to evaluate plans via intuitive visual feedback.

The first goal of the present study was to provide a single visual representation of multiple radiation treatment plan evaluation criteria. In our example, we display and evaluate each treatment plan based on the outlined 13 DVH objectives from the head and neck cancer treatment planning directive, each of which has its own endpoint that can vary significantly from those of the other objectives. For simultaneous visualization of all the objectives, we used percentile ranking to normalize the scale of all objectives. Because of its ubiquity, a percentile scale has the benefit of being an easily understandable statistical measure. A percentile scale also provides information not found in a simple objective checklist by conveying an understanding of the likelihood of a requested DVH objective value and the resulting DVH objective value. For example, the planning directive demonstrates a physician requested that the maximum brain stem dose be less than 54 Gy. Based on the location of the gray dot in the brain stem region in Fig. 4, we can see that this constraint is only violated, that is, maximum dose to the brain stem is greater than 54 Gy in only 2.00% of the population; because violation of this constraint is rare, we have little insight regarding the optimal maximum dose to the brain stem for any given patient. The right submandibular gland region in Fig. 4 illustrates a similar but opposing problem. The requested mean dose of less than 30 Gy for the right submandibular gland was violated 90.0% of the time in the tested population. Again, we are left without intuition regarding the likely mean dose to the right submandibular gland. Using our treatment plan scoring method, a reviewer can see that the specific patient achieved brain stem dose of 32.6 Gy, as represented by the black dot within the brain stem petal of Fig. 4, has a percentile rank of 61.0%, and thus is a bit better than average as well as much better than the clinician's requested value. Likewise, the patient's achieved mean dose to the right submandibular gland was 69.0 Gy, with a percentile rank of 39.0%. Therefore, the mean dose to the right submandibular gland was higher than typically achieved and should be reviewed.

For calculating the overall radiation treatment plan percentile, we preferred a geometric mean over the arithmetic mean as the former reduces the weight of extreme observations and is more robust in comparing multiplicative changes. For example, the patient represented in Fig. 4 had a larynx dose of 71.9 Gy, which fell in the 13th percentile. This patient also had a variety of structures in the 40th to 50th percentile range. As demonstrated by the patient's overall plan score of 62.1%, the poor larynx result did not heavily impact the plan score as would have an arithmetic mean (46.0%).

The second goal of the present study was to validate our treatment plan score. To do this, we asked 6 head and neck radiation oncologists to grade a unique set of treatment plans. Figure 5 and our Spearman correlation analysis demonstrated moderate correlation between our scores and the physicians' scores regarding plans of acceptable quality. The differences in the rank-based plan scores for plans requiring minor edits compared with those that were unacceptable were less strong as demonstrated by overlap of their IQR. We believe this to be a result of 2 potential differences. The first is a large variation in our plan scores for the unacceptable plans, which can likely be attributed to the physicians having nonnegotiable dose constraints. The second is uncertainty regarding when a plan must be edited and when it is unacceptable, as replanning is performed under both scenarios. Despite this, the median plan scores for each category (acceptable, minor edits, and unacceptable) are well separated, indicating that our treatment plan score provides some differentiation between unacceptable plans and plans needing minor edits.

Visual demonstration of our results was an important factor in designing our scorecard tool, as we wanted the results to be easily understood. In a recent study, Ventura et al[8] used radar plots to evaluate radiation treatment plan quality in a fashion similar to that in the present study through the graphical display of an objective metric. Despite the similarity, our scorecard tool has 2 unique advantages. The first is the scoring metric used. Whereas Ventura and colleagues compared the planned clinical value with the physician-requested value, we compared the planned dose with the doses in a library of similarly treated plans to gain insight into the quality of the plan. The second difference is our use of a daisy plot instead of a radar plot. Although both radar and daisy plots can display a wealth of information in a single plot, we found the daisy plot easier to read. By having the height of a bar represent the achieved value alongside the black dot in a daisy plot, we can better visualize weaknesses and strengths in plan quality, with the highest bar in a plot representing the best-performing structure and the shortest bar representing the worst-performing structure. Simply reviewing these 2 structures, next to each other on a plot, may identify structures that can be more or less spared in the plan. We also felt that including the clinician's requested value in the daisy plot is a unique advantage, as it provides a safety barrier for each OAR and some insight into that barrier as described previously.

The main limitation of this study, as with most studies of knowledge-based planning tools, was the use of a library of plans. The 111 patients in our population received treatment over a wide time span (2005-2022) using intensity modulated radiation therapy including static gantry techniques and volumetric modulated arc therapy. However, because the objective of our study was to define a methodology and compare our treatment plan scoring method with physician review of treatment plans, this cohort was sufficient. A second limitation of this study was that each plan was scored by only 1 physician. However, in this study, having a larger quantity of graded plans was emphasized over quantity of grades per plan.

Our treatment plan scorecard tool is designed to serve as a feedback mechanism for treatment planning. It can be used by a physician reviewer, for whom the daisy plot would highlight the primary areas to review or by a dosimetrist as an objective means of determining what is typically achievable for a patient of similar diagnosis at the time of planning, reducing disagreement between the physician and the planning team. Future work on this methodology can include modifications of the library for clinical use, such as expanding the patient population for more specific divisions (eg, specific diagnoses and/or demographics). This would allow for comparison of future plans with those in a highly specific population, potentially using plans created using machine learning tools (such as automated planning and dose prediction algorithms). This can be done using machine learning based tools to expand the population within a specific diagnosis and/or demographic using pre-existing patient information, allowing for comparison of future plans with those in a highly specific population. Also, integration of non-DVH measures, such as spatial characteristics of the dose distribution and applying Tumor Control Probability/Normal Tissue Complication Probability modeling, are potential next steps that would allow incorporated location and toxicity risk information into the daisy plot.

Conclusion

The results of our radiation treatment plan scoring method demonstrated moderate correlation with those of physician reviews of treatment plans, demonstrating the utility of our method in the plan quality assurance process as a normalized, physician-reviewed scoring system. We also demonstrated that our scoring system is an effective visual means of identifying good plan quality. The combination of easily comprehensible metrics and visual aids provides a framework for a treatment plan quality assessment tool that has the potential to determine whether a plan is of good quality in an easily understood, intuitive manner.

Disclosures

This project was performed by the radiation planning team at MD Anderson, which receives funding from the NCI, the Cancer Prevention & Research Institute of Texas, the Wellcome Trust, and Varian Medical Systems. A premarket notification 510(k) for the Radiation Planning Assistant has been cleared for commercial distribution by the U.S. Food and Drug Administration; however, it is yet to be marketed. Authors of this work also receive funds in whole or in part from the NIH, Elekta AB, Oncospace Inc, and Phillips Medical Systems.

Acknowledgments

The authors acknowledge the support of the High-Performance Computing for Research facility at the University of Texas MD Anderson Cancer Center for providing computational resources that have contributed to the research results reported in this article. This study was supported by Varian Medical Systems. The manuscript was edited by Donald Norwood of the Research Medical Library at MD Anderson.

References

1. Lewis PJ, Court LE, Lievens Y, Aggarwal A. Structure and processes of existing practice in radiotherapy peer review: A systematic review of the literature. *Clin Oncol (R Coll Radiol)*. 2021;33:248-260.
2. Krishnan J, Shetty J, Rao S, Hegde S, Shambhavi C. Comparison of rapid arc and intensity-modulated radiotherapy plans using unified dosimetry index and the impact of conformity index on unified dosimetry index evaluation. *J Med Phys*. 2017;42:14-17.
3. Thieke C, Küfer K-H, Monz M, et al. A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. *Radiother Oncol*. 2007;85:292-298.
4. Sung K, Choi YE. Dose gradient curve: A new tool for evaluating dose gradient. *PloS One*. 2018;13: e0196664.
5. Yoon M, Shin D, Park J, et al. A new evaluation method of target volume coverage and homogeneity for IMRT treatment planning. *Phys Med*. 2006;22:43-51.
6. Qingya P, Yizhong F, Fuke Z, Shengjie L, Linzhao T, Yuling L. Standard-deviation based conformity index for evaluating treatment plan of intensity modulated radiotherapy in lung cancer. *Radiat Prot Dosimetry*. 2023;199:87-94.
7. Patel G, Mandal A, Choudhary S, Mishra R, Shende R. Plan evaluation indices: A journey of evolution. *Rep Pract Oncol Ratiother*. 2020;25:336-344.
8. Ventura T, Lopes MD, Ferreira BC, Khouri L. Spiderplan: A tool to support decision-making in radiation therapy treatment plan assessment. *Rep Pract Oncol Ratiother*. 2016;21:508-516.
9. Mayo CS, Zankowski C, Herman M, et al. A method to vectorize the dose distribution, the dose volume histogram and create a dose vector histogram. *Med Phys*. 2013;40: 011717.
10. Craft DL, Hong TS, Shih HA, Bortfeld TR. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2012;82:e83-e90.
11. Ventura T, Dias J, Khouri L, et al. Clinical validation of a graphical method for radiation therapy plan quality assessment. *Radiat Oncol*. 2020;15:64.
12. Talcott WJ, Lincoln H, Kelly JR, et al. A blinded, prospective study of error detection during physician chart rounds in radiation oncology. *Pract Radiat Oncol*. 2020;10:312-320.
13. Shanafelt TD, Hasan O, Dyrbye LN, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. *Mayo Clin Proc*. 2015;90:1600-1613.
14. Royce TJ, Davenport KT, Dahle JM. A burnout reduction and wellness strategy: Personal financial health for the medical trainee and early career radiation oncologist. *Pract Radiat Oncol*. 2019;9:231-238.
15. Bui S, Pelosi A, Mazzaschi G, et al. Burnout and oncology: An irreparable paradigm or a manageable condition? Prevention strategies to reduce burnout in oncology health care professionals. *Acta Biomed*. 2021;92: e2021091.
16. Mampuya WA, Matsuo Y, Nakamura A, Hiraoka M. Evaluation of the prevalence of burnout and psychological morbidity among radiation oncologist members of the Kyoto Radiation Oncology Study Group (KROSG). *J Radiat Res*. 2017;58:217-224.
17. Rhee DJ, Cardenas CE, Elhalawani H, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys*. 2019;46:5086-5097.
18. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed June 28, 2023. https://www.R-project.org/.
19. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag; 2016.