Advances in Radiation Oncology (2023) 8, 101295

advances in radiation oncology

www.advancesradonc.org

Research Letter

Scorecards: Quantifying Dosimetric Plan Quality in Pancreatic Ductal Adenocarcinoma Stereotactic Body Radiation Therapy

Kareem Rayn, MD,[a,b,*] Ryan Clark, MS, CMD,[b] Anthony Magliari, MS, CMD,[b] Brian Jeffers, BS,[a] Elizaveta Lavrova, BS,[a] Ingrid Valencia Lozano, PhD,[a] Michael J. Price, PhD,[a] Lesley Rosa, MS, CMD,[b] and David P. Horowitz, MD[a]

[a]*Department of Radiation Oncology, Columbia University Irving Medical Center, New York, New York; and* [b]*Varian Medical Systems Inc, Palo Alto, California*

Received 23 January 2023; accepted 9 June 2023

Purpose: A scoring mechanism called the scorecard that objectively quantifies the dosimetric plan quality of pancreas stereotactic body radiation therapy treatment plans is introduced.

Methods and Materials: A retrospective analysis of patients with pancreatic ductal adenocarcinoma receiving stereotactic body radiation therapy at our institution between November 2019 and November 2020 was performed. Ten patients were identified. All patients were treated to 36 Gy in 5 fractions, and organs at risk (OARs) were constrained based on Alliance A021501. The scorecard awarded points for OAR doses lower than those cited in Alliance A021501. A team of 3 treatment planners and 2 radiation oncologists, including a physician resident without plan optimization experience, discussed the relative importance of the goals of the treatment plan and added additional metrics for OARs and plan quality indexes to create a more rigorous scoring mechanism. The scorecard for this study consisted of 42 metrics, each with a unique piecewise linear scoring function which is summed to calculate the total score (maximum possible score of 365). The scorecard-guided plan, the planning and optimization for which were done exclusively by the physician resident with no prior plan optimization experience, was compared with the clinical plan, the planning and optimization for which were done by expert dosimetrists, using the Sign test.

Results: Scorecard-guided plans had, on average, higher total scores than those clinically delivered for each patient, averaging 280.1 for plans clinically delivered and 311.7 for plans made using the scorecard ($P = .003$). Additionally, for most metrics, the average score of each metric across all 10 patients was higher for scorecard-guided plans than for clinically delivered plans. The scorecard guided the planner toward higher coverage, conformality, and OAR sparing.

Conclusions: A scorecard tool can help clarify the goals of a treatment plan and provide an objective method for comparing the results of different plans. Our study suggests that a completely novice treatment planner can use a scorecard to create treatment plans with enhanced coverage, conformality, and improved OAR sparing, which may have significant effects on both tumor control and toxicity. These tools, including the scorecard used in this study, have been made freely available.

Introduction

Significant variations in treatment planning quality exist between treatment centers and treatment planners.[1-4] Comprehensive plan quality metrics that incorporate multiple parameters (eg, plan scores) may facilitate objective plan

Sources of support: Kareem Rayn received an American Society of Radiation Oncologists/Varian–funded fellowship.

Corresponding author: Kareem Rayn, MD; E-mail: Knr9008@nyp.org

https://doi.org/10.1016/j.adro.2023.101295

comparison and reduce subjectivity.[1,5] Such plan quality metrics have been implemented in several studies and have detected interinstitutional differences in plan quality.[2,3]

In addition to scorecards, plan quality may be assessed in comparison to automatically generated knowledge-based plans[6] or to other automatically generated metrics such as feasibility dose-volume histograms.[4] Additionally, automated treatment planning has been proposed to help standardize plan quality.[7] Although there is no consensus yet on the best method to evaluate plan quality, all plan quality assessment should include dose metrics (such as target coverage, organ-at-risk [OAR] constraints, conformity, homogeneity, among others) as well as plan robustness and complexity.[5]

In this study, we introduce a scoring mechanism called the dosimetric scorecard that objectively quantifies improvements in plan quality. It is important to note that the scorecard does not claim to identify the best clinical plan; such an assessment would require comparing plan quality metrics to outcome data. However, scorecards allow physicians to express their clinical intent exhaustively and precisely. These scorecards were first introduced to judge treatment plan quality competitions in 2011.[2,7] In previous studies, judgments of plan quality were subjective, and scorecards were well received by the treatment planner community.[2] The scorecard provides a transparent and fair platform for comparing plan results and removes any ambiguity about physician preference.

Methods and Materials

A single-institution retrospective analysis of patients with pancreatic ductal adenocarcinoma receiving stereotactic body radiation therapy between November 2019 and November 2020 was performed. Ten patients were identified. Patients were simulated supine with a 4-dimensional computed tomography scan with Vacloc immobilization and abdominal compression. Multiphase intravenous contrast agent and 4-dimensional reconstruction were used for the planning computed tomography. A 2-arc volumetric arc therapy technique with full coplanar arcs or a 10-field static intensity-modulated radiation therapy technique was used. All patients were treated to 36 Gy in 5 fractions, and OARs were constrained based on Alliance A021501[8] with institutional review board approval.

The PlanScoreCard Eclipse Scripting Application Programming Interface tool, available free on the Varian Medical Affairs Applied Solutions GitHub, created scoring metrics and automatically generated additional optimization and evaluation structures, scoring candidate plans throughout the process.[7] These dosimetric scorecards use established scoring methodology of multiple piecewise linear score functions, which measure specific plan quality metrics. The scorecard included target coverage and was based on dose constraints from Alliance A021501.[8] The scorecard awarded points for OAR doses lower than those cited in Alliance A021501. Aspirational ranges and total points awarded varied based on as low as reasonably achievable principles and physician preference. To this end, a team of 3 treatment planners and 2 radiation oncologists, including a physician resident without plan optimization experience, discussed the relative importance of the goals of the treatment plan and added additional metrics for OARs and plan quality indexes, including ring and evaluation structures, not listed in the protocol to create a more rigorous scoring mechanism that accounts for target coverage as well as conformality and heterogeneity. The scorecard for this study consisted of 42 metrics, each with a unique piecewise linear scoring function which is summed to calculate the total score (maximum possible score of 365). An example scorecard is shown in Fig. 1, and the details of each scorecard metric are summarized in Table 1. The Digital Imaging for Communications in Medicine for an example scorecard-guided case along with this scorecard is freely available to view and download.[9]

This study's data meets the requirements of using the Sign test (independent matched-pairs of scores, where the scores and differences between the scores are measured on a continuous level) for statistical analysis.[10,11] The scorecard-guided plan, the planning and optimization for which were done retrospectively, exclusively by the physician resident with no prior plan optimization experience, was compared with the clinical plan, the planning and optimization for which were done by expert dosimetrists.

Results

Scorecard-guided plans had, on average, higher total scores (out of 365) than those clinically delivered for each patient, averaging 280.1 for plans clinically delivered and 311.7 for plans made using the scorecard (P = .003; Table 2). The most pronounced increase in score was for patient 1, whose scorecard-guided plan had a score of 331.6 out of 365 (90.9%) compared with the clinically delivered plan which had a score of 253.2 out of 365 (69.4%), a 78.4 point (21.5%) improvement (Fig. 1). Overall, the scorecard-guided plan enforced higher coverage, conformality, and OAR sparing (Figs. 1-4).

Increases in target coverage contributed to the most significant increase in scores between the plans (Table 1). There are marginal but quantifiable improvements with the lower doses to the OARs that also contributed to the overall rise in score. For example, for patient 2, the scorecard-guided plan reduced dose at 0.03 cc to the stomach to 26.99 Gy from 31.62 Gy in the clinical plan, reduced stomach volume at 12 Gy to 16.83% from 34.3%, and reduced stomach volume at 20 Gy to 1.49 cc from 8.96 cc.

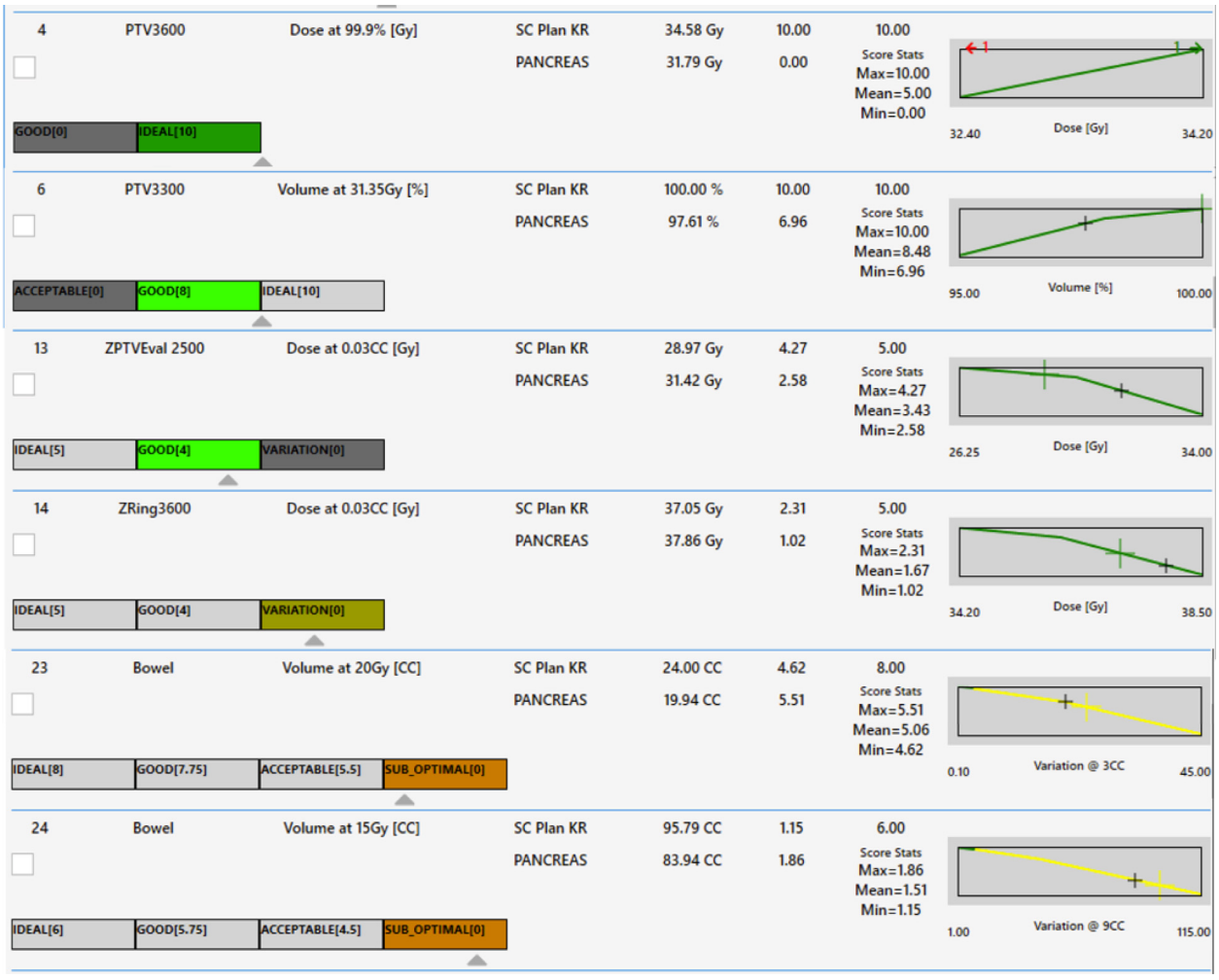


Figure 1 Selected scorecard metric examples for patient 1, where SC plan KR is the scorecard plan and PANCREAS is the clinically delivered plan. These metrics show higher scores for the scorecard plan's PTV coverage, homogeneity, and conformality at the expense of a slightly lower score due to increased bowel dose when compared to the clinical plan. The total score (out of 365) for the scorecard and clinically delivered plans for this patient are 331.6 (90.9%) and 253.2 (69.4%), respectively. *Abbreviations:* KR = kareem rayn (physician resident); max = maximum; min = minimum; PTV = planned target volume; SC = scorecard.

Additionally, Table 1 shows that, for most metrics, the average score of each metric across all 10 patients was higher for scorecard-guided plans than for clinically delivered plans. Scorecard-guided plan scores differed significantly ($P < .05$) from their paired clinical plan scores for 6 metrics. Of these 6 metrics, average score was increased in scorecard-guided plans compared with clinical plans for volume constraint metrics for PTV_3600 and PTV_3300, and average score was decreased for the max dose (0.03 cc) constraint metrics for PTV_3600 and skin (Table 1).

Discussion

The purpose of this study is to show that clearly expressed clinical intent guides the treatment planner and removes subjectivity in treatment planning. This approach can even help when clinicians with less treatment planning experience are tasked with creating the plan, as in this study, where the novice treatment planner with no optimization experience surpassed experienced clinical dosimetrists by generating plans with significantly improved scores. Dosimetric scorecards provide objective measures to continue improving the plan quality (score) per the physician's preference.

In current practice, treatment planners are expected to know physician preferences from prior experience without all quality metrics explicitly stated. It is common for a treatment planner's clinical directives (physician's prescription) to only include a minimal set of single-point dose-volume histogram metric goals without clear or precise intra and intergoal prioritization. Because of this, several expected treatment

Table 1 Average values and scores of all 10 patients by score metric for clinical plan and scorecard plan

Structure	Score metric	Clinical plan		Scorecard plan		*P* value
		Avg value	Avg score	Avg value	Avg score	
ITV	Volume at 25 Gy (%)	99.95	9.54	99.97	9.73	.564
PTV_3600	Volume at 34.2 Gy (%)	99.47	9.23	99.93	9.93	.020*
	Volume at 36 Gy (%)	93.98	7.63	98.48	18.80	.034*
	Dose at 99.9% (Gy)	33.76	6.88	34.92	9.03	.180
	Dose at 0.03 cc (Gy)	38.53	4.57	39.56	3.42	.008*
PTV_3300	Volume at 31.35 Gy (%)	99.49	9.43	99.96	9.96	.020*
	Volume at 33 Gy (%)	95.69	11.95	98.73	19.28	.034*
	Dose at 99.9% (Gy)	30.73	6.23	32.09	9.86	.059
PTVEval_3300	Dose at 0.03 cc (Gy)	36.93	3.23	37.21	2.88	.739
PTV_2500	Volume at 25 Gy (%)	99.61	19.99	99.72	20.00	.317
	Volume at 23.75 Gy (%)	99.91	9.91	99.97	9.97	.655
	Dose at 99.9% (Gy)	26.11	9.00	25.98	9.46	.317
PTVEval_ 2500	Dose at 0.03 cc (Gy)	31.55	2.27	30.69	2.96	.317
Ring_3600	Dose at 0.03 cc (Gy)	37.33	1.87	37.68	1.58	.317
Ring_3300	Dose at 0.03 cc (Gy)	34.66	1.97	34.82	1.99	.739
Ring_2500	Dose at 0.03 cc (Gy)	32.53	1.37	32.20	2.05	.257
Duodenum	Volume at 20 Gy (cc)	15.21	3.66	14.75	3.87	.480
	Volume at 15 Gy (cc)	34.71	2.60	34.17	2.63	.480
	Dose at 0.03 cc (Gy)	30.79	9.46	30.04	10.95	.096
	Mean dose (Gy)	11.64	1.58	11.62	1.57	.739
Bowel	Volume at 33 Gy (cc)	0.02	14.99	0.00	15.00	.317
	Volume at 30 Gy (cc)	0.72	11.50	0.28	11.77	.083
	Volume at 20 Gy (cc)	31.92	3.13	26.16	3.88	.096
	Volume at 15 Gy (cc)	99.33	1.61	89.72	1.87	.480
	Dose at 0.03 cc (Gy)	30.68	8.22	29.86	10.28	.096
	Mean dose (Gy)	5.00	2.75	4.96	2.75	.739
Stomach	Volume at 33 Gy (cc)	0.00	15.00	0.00	15.00	.096
	Volume at 30 Gy (cc)	0.10	11.90	0.07	11.93	.739
	Volume at 20 Gy (cc)	4.42	3.27	3.59	3.47	.739
	Volume at 15 Gy (cc)	16.13	3.31	14.97	3.37	.739
	Volume at 12 Gy (%)	14.88	5.13	14.40	5.28	.739
	Dose at 0.03 cc (Gy)	25.06	10.80	23.71	11.33	.480
	Mean Dose (Gy)	4.73	2.76	4.62	2.77	.739
Liver	Volume at 12 Gy (cc)	52.26	5.31	57.78	5.05	.257
	Dose at 0.03 cc (Gy)	22.87	2.34	22.89	2.40	.317
	Mean dose (Gy)	2.46	2.80	2.51	2.79	.739
	Dose to 700 cc volume of liver (Gy)	1.44	14.26	1.47	14.20	.317
Total kidneys	Volume at 12 Gy (%)	7.29	9.58	6.36	9.64	.096
	Mean dose (Gy)	4.76	2.34	4.59	2.37	.739
Spinal canal	Dose at 0.03 Gy (Gy)	12.50	8.30	12.17	8.35	.739
Skin	Dose at 0.03 ccc (Gy)	10.97	4.19	12.11	4.04	.020*
	Dose at 10 cc (Gy)	7.59	4.26	7.65	4.25	.739

Abbreviations: ITV = internal target volume; PTV = planned target volume.

* *P* values <.05.

P values found comparing clinical and scorecard score distributions using Sign test.

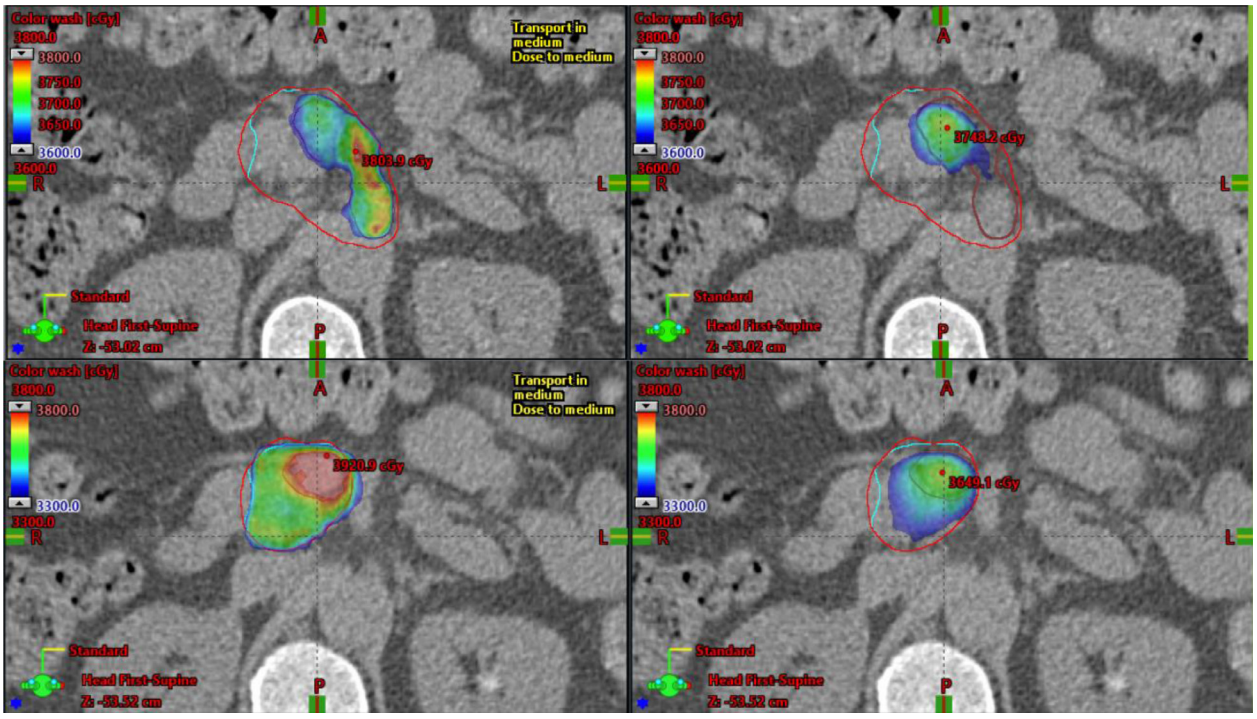


Figure 2 Representative isodose colorwash showing PTV_3600 (brown) for scorecard-guided plan (top left) versus clinically delivered plan (top right). Bottom panel: PTV_3300 (cyan) for scorecard-recommended plan (bottom left) versus clinically delivered plan (bottom right). *Abbreviations:* PTV = planning target volume.

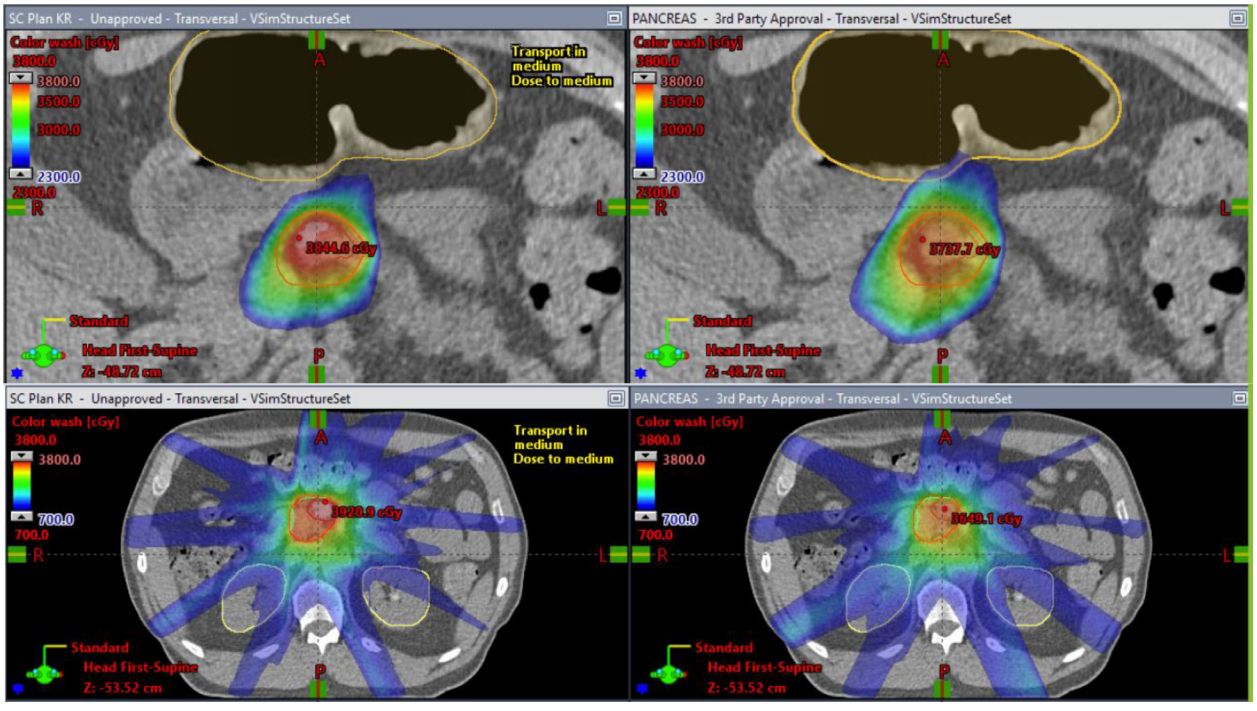
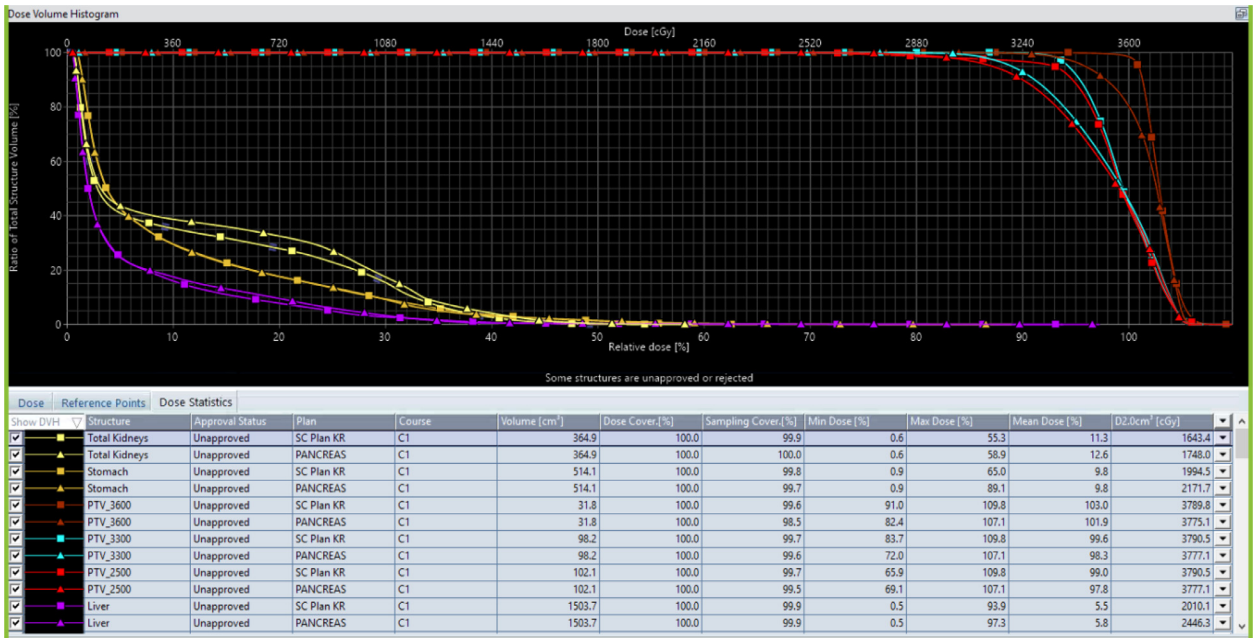


Figure 3 Representative isodose colorwash showing improved moderate level (23 Gy) dose conformality around PTV_2500 (red) resulting in decreased dose to the stomach for scorecard plan (top left) versus clinically delivered plan (top right). Improved low-dose (7 Gy) conformality around PTV_2500 (red), resulting in decreased dose to the kidney for scorecard plan (bottom left) versus clinically delivered plan (bottom right). *Abbreviations:* KR = kareem rayn (physician resident); PTV = planned target volume; SC = scorecard.



Show DVH	Structure	Approval Status	Plan	Course	Volume [cm³]	Dose Cover.[%]	Sampling Cover.[%]	Min Dose [%]	Max Dose [%]	Mean Dose [%]	D2.0cm³ [cGy]
	Total Kidneys	Unapproved	SC Plan KR	C1	364.9	100.0	99.9	0.6	55.3	11.3	1643.4
	Total Kidneys	Unapproved	PANCREAS	C1	364.9	100.0	100.0	0.6	58.9	12.6	1748.0
	Stomach	Unapproved	SC Plan KR	C1	514.1	100.0	99.8	0.9	65.0	9.8	1994.5
	Stomach	Unapproved	PANCREAS	C1	514.1	100.0	99.7	0.9	89.1	9.8	2171.7
	PTV_3600	Unapproved	SC Plan KR	C1	31.8	100.0	99.6	91.0	109.8	103.0	3789.8
	PTV_3600	Unapproved	PANCREAS	C1	31.8	100.0	98.5	82.4	107.1	101.9	3775.1
	PTV_3300	Unapproved	SC Plan KR	C1	98.2	100.0	99.7	83.7	109.8	99.6	3790.5
	PTV_3300	Unapproved	PANCREAS	C1	98.2	100.0	99.6	72.0	107.1	98.3	3777.1
	PTV_2500	Unapproved	SC Plan KR	C1	102.1	100.0	99.7	65.9	109.8	99.0	3790.5
	PTV_2500	Unapproved	PANCREAS	C1	102.1	100.0	99.5	69.1	107.1	97.8	3777.1
	Liver	Unapproved	SC Plan KR	C1	1503.7	100.0	99.9	0.5	93.9	5.5	2010.1
	Liver	Unapproved	PANCREAS	C1	1503.7	100.0	99.9	0.5	97.3	5.8	2446.3

Figure 4 Representative DVH for scorecard-guided plan (SC Plan KR; squares) versus clinically delivered plan (PANCREAS; triangles). *Abbreviations:* DVH = dose-volume histogram; KR = kareem rayn (physician resident); max = maximum; min = minimum; PTV = planned target volume; SC = scorecard.

planning quality goals, such as conformality of certain isodose levels, homogeneity, etc, are left unexpressed during plan evaluation. Clearly defined scorecards, created at the protocol level, are needed to improve plan quality and standardization. As concluded by Olch et al,[12] future protocol authors would be well advised to include a dosimetric scorecard as a superset of the standard "variation acceptable" and "variation unacceptable" dose constraints to reduce plan quality variability, which may improve outcomes.

Table 2 Total scores (percent out of maximum score of 365, %) for scorecard and clinical plans by patient

Patient	Clinically delivered (%)	Scorecard guided (%)
1	330.5 (90.6)	346.5 (94.9)
2	293.4 (80.4)	334.3 (91.6)
3	298.2 (81.7)	298.2 (81.7)
4	281.4 (77.1)	293.2(80.3)
5	284.7 (78.0)	326.3 (89.4)
6	297.7 (81.6)	307.7 (84.3)
7	231.2 (63.3)	282.7 (77.5)
8	253.2 (69.4)	331.6 (90.9)
9	277.4 (76.0)	311.8 (85.4)
10	253.7 (69.5)	284.5 (77.9)
Average	280.1 (76.7)	311.7 (85.4)

Due to the retrospective nature of this study, the clinical plans could not be as competitive because the prior treatment planners did not have precise articulation of clinical intent. In this study, the physician could exhaustively quantify each relative dose coverage or dose-limiting metric after learning how to use the scorecard editing tool. It should be noted that the clinical plans optimized by the experienced dosimetrists were under time and clinic pressures of generating a treatment plan with a specific start date in mind. Although the novice physician resident had a learning curve for planning the first cases, time was used to iteratively improve plans without clinical constraints. However, after getting past the learning curve, time spent optimizing was an average timeframe of 1 to 2 hours. The small number of plans reviewed[10] limited our statistical analysis.

Scorecards enable plan quality quantification, which can be helpful when conducting future studies with retrospective analysis or evaluating dosimetric improvements from new technologies. Furthermore, when treatment plan quality improvement is defined simply (higher score equals a better plan), and when/if these dosimetric scorecard methods are popularized, medical devices can be designed to implement plan optimization on physicians' precise clinical intent directly.

Dosimetric plan scorecards are not new to the field of radiation oncology and have been used for over 10 years but have never been as accessible as they are now.[2,13] There are free scorecard tools available that can directly connect to a treatment planning system through an

application programming interface[14] or cloud-based solutions that can be used anywhere.[15] The scorecard used in this study as well as several other example scorecards can be found online.[9]

Conclusion

Using a scorecard tool can reduce uncertainty about the goals of a treatment plan and provide a consistent method for comparing the dosimetric quality of different plans. Our study suggests that a completely novice treatment planner can use a scorecard to create treatment plans with enhanced coverage, conformality, and improved OAR sparing. Dosimetric scorecards can help enhance dosimetric quality and/or reduce variability if included prospectively with future clinical trials. These tools, including the scorecard used in this study, have been made freely available.

Disclosures

Ryan Clark, Anthony Magliari, and Lesley Rosa are employed by Varian.

References

1. Giglioli FR, Garibaldi C, Blanck O, et al. Dosimetric multicenter planning comparison studies for stereotactic body radiation therapy: Methodology and future perspectives. *Int J Radiat Oncol Biol Phys*. 2020;106:403-412.
2. 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.
3. Blanck O, Wang L, Baus W, et al. Inverse treatment planning for spinal robotic radiosurgery: An international multi-institutional benchmark trial. *J Appl Clin Med Phys*. 2016;17:313-330.
4. Ahmed S, Liu C-W, LaHurd D, et al. Using feasibility dose-volume histograms to reduce intercampus plan quality variability for head-and-neck cancer. *J Appl Clin Med Phys*. 2023;24:e13749.
5. Hernandez V, Hansen CR, Widesott L, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. *Radiother Oncol*. 2020;153:26-33.
6. Zhang J, Wu QJ, Ge Y, et al. Knowledge-based statistical inference method for plan quality quantification. *Technol Cancer Res Treat*. 2019;18:1-10.
7. Liu H, Clark R, Magliari A, et al. RapidPlan hippocampal sparing whole brain model version 2—How far can we reduce the dose? *Med Dosim*. 2022;47:258-263.
8. Katz MHG, Shi Q, Meyers J, et al. Efficacy of preoperative mFOLFIRINOX versus mFOLFIRINOX plus hypofractionated radiotherapy for borderline resectable adenocarcinoma of the pancreas: The A021501 phase 2 randomized clinical trial. *JAMA Oncol*. 2022;8:1263-1270.
9. Varian. Pancreas SBRT 36Gy SIB (Alliance A021501). Available at: https://medicalaffairs.varian.com/pancreas-sbrt-36gy-sib-vmat2. Accessed December 20, 2022.
10. Laerd Stastics. Sign test in SPSS Statistics. Available at: https://statistics.laerd.com/spss-tutorials/sign-test-using-spss-statistics.php. Accessed April 14, 2023.
11. Social Science Stastistics. Sign test calculator. Available at: https://www.socscistatistics.com/tests/signtest/default.aspx. Accessed April 14, 2023.
12. Olch AJ, Gopalakrishnan M, Murphy ES, MacDonald SM, Hua, C-H. Toward systematic assessment and improvement of radiation therapy plan quality of cooperative group trial submissions: A report from the Children's Oncology Group [e-pub ahead of print]. *Pract Radiat Oncol*. https://doi.org/10.1016/j.prro.2023.02.008, Accessed March 5, 2023.
13. Sun Nuclear. PlanIQ. Available at: https://www.sunnuclear.com/products/planiq. Accessed December 20, 2022.
14. GitHub. Varian medical affairs applied solutions plan score card. Available at: https://github.com/Varian-MedicalAffairsAppliedSolutions/MAAS-PlanScoreCard. Accessed December 20, 2022.
15. Elekta ProKnow. Available at: https://proknow.com/. Accessed December 20, 2022.