Statistical Analysis of Treatment Planning Parameters for Prediction of Delivery Quality Assurance Failure for Helical Tomotherapy

Kyung Hwan Chang, PhD[1], Young Hyun Lee, MS[2],
Byung Hun Park, MS[2], Min Cheol Han, PhD[1], Jihun Kim, PhD[1],
Hojin Kim, PhD[1], Min-Seok Cho, MS[3], Hyokyeong Kang, MS[1],
Ho Lee, PhD[4], Dong Wook Kim, PhD[1], Kwangwoo Park, PhD[1],
Jaeho Cho, MD, PhD[1], Yong Bae Kim, MD, PhD[1],
Jin Sung Kim, PhD[1], and Chae-Seon Hong, PhD[3]

Abstract

Purpose: This study aimed to investigate the parameters with a significant impact on delivery quality assurance (DQA) failure and analyze the planning parameters as possible predictors of DQA failure for helical tomotherapy. **Methods:** In total, 212 patients who passed or failed DQA measurements were retrospectively included in this study. Brain (n = 43), head and neck (n = 37), spinal (n = 12), prostate (n = 36), rectal (n = 36), pelvis (n = 13), cranial spinal irradiation and a treatment field including lymph nodes (n = 24), and other types of cancer (n = 11) were selected. The correlation between DQA results and treatment planning parameters were analyzed using logistic regression analysis. Receiver operating characteristic (ROC) curves, areas under the curves (AUCs), and the Classification and Regression Tree (CART) algorithm were used to analyze treatment planning parameters as possible predictors for DQA failure. **Results:** The AUC for leaf open time (LOT) was 0.70, and its cut-off point was approximately 30%. The ROC curve for the predicted probability calculated when the multivariate variable model was applied showed an AUC of 0.815. We confirmed that total monitor units, total dose, and LOT were significant predictors for DQA failure using the CART. **Conclusions:** The probability of DQA failure was higher when the percentage of LOT below 100 ms was higher than 30%. The percentage of LOT below 100 ms should be considered in the treatment planning process. The findings from this study may assist in the prediction of DQA failure in the future.

Keywords

leaf open time, DQA failure, prediction, planning parameters, helical tomotherapy

Introduction

Pretreatment patient-specific delivery quality assurance (DQA) is essential for the verification of accurate dose delivery in advanced radiation therapy techniques, such as intensity-modulated radiation therapy (IMRT), volumetric modulated arc therapy, stereotactic body radiation therapy, and helical tomotherapy (HT).[1-4]

An HT system is an IMRT delivery system that delivers the radiation beam with a high degree of conformity; it is an image-guided system using megavoltage computed tomography (CT).[5] The radiation beam is delivered using a 6-MV linear accelerator continually while the couch is moving into the gantry, and intensity modulation is achieved with a binary multileaf collimator (MLC).[6] DQA and planning quality for

[1] Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, South Korea
[2] Eretec Inc. 401, Simin-daero, Dongan-gu, Anyang-si, Gyeonggi-do, South Korea
[3] Department of Radiation Oncology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, South Korea
[4] Department of Radiation Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea

Corresponding Authors:
Chae-Seon Hong, Department of Radiation Oncology, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin-si, Gyeonggi-do 16995, South Korea.
Email: cs.hong@yuhs.ac

Jin Sung Kim, Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemoon-gu, Seoul, South Korea.
Email: jinsung@yuhs.ac

HT are influenced by various planning parameters such as pitch, field width (FW), the modulation factor (MF), and the leaf open time (LOT).[7] Pitch is defined as the couch travel distance for a complete gantry rotation (GR) relative to the axial beam width at the axis of rotation in helical CT.[8] FW is defined as the longitudinal width of the beam at the central axis. Finally, MF is defined as the ratio of the maximal LOT to the mean (non-zero) LOT.[9]

Several studies have presented the optimal planning parameters between pitch, FW, and MF to maintain the optimal plan quality and treatment time for clinical cases using HT, and they have studied the correlation between planning parameters and DQA results.[9-18] However, to the best of our knowledge, there is no published report using comprehensive statistical analyses such as logistic regression, receiver operating characteristic (ROC) curves, and the Classification and Regression Tree (CART) algorithm for the impact of the various planning parameters on DQA failure, to date.

This study aimed to investigate the parameters with a significant impact on DQA failure and to analyze the planning parameters as possible predictors of DQA failure for HT.

Methods

Patient Selection and Characteristics

In total, 212 patients who had passing or failing DQA measurements were randomly selected for inclusion in this retrospective study (Table 1). All selected patients were treated with Tomotherapy (TomoHDA, Accuray; Sunnyvale, CA, USA) in our institution.

Patient-Specific DQA in HT

Treatment planning for all patients was performed using the HT planning station (Accuray). The convolution/superposition algorithm was used for dose calculation, and a fine dose-calculation grid size (0.195×0.195 cm^2) was employed in all the DQA plans. To generate the DQA plan, a cheese phantom was selected for all treatment plans. The center of the Exradin A1SL ionization chamber (Standard Imaging, Middleton, WI) positioned in the cylindrical Solid Water^{TM} phantom ("Cheese phantom," Accuray) was moved to the low-dose gradient region or target point. A red laser was placed at the center of the ionization chamber, and the doses were calculated. We used a cheese phantom with an ionization chamber and Gafchromic EBT3 film (International Specialty Products, Wayne, NJ) to measure the absolute dose and gamma values for all HT plans.[19] The differences between the calculated and measured point doses and dose distributions were computed using tomotherapy DQA software (Accuracy).[19] The point dose difference (DD) and global gamma passing rate (GPR) for all patients were analyzed. The threshold for analysis was set at 10% of the global maximum. The tolerance range of DD was within $\pm 5\%$ and that of GPR was $> 90\%$ based on the criteria

of 3%/3 mm. If one criterion failed, then DQA was considered to have failed.[19]

Analysis of Planning Parameters

In this study, various treatment planning parameters were analyzed in 8 clinical cases. Additionally, the mean and standard deviation (SD) of individual parameters were evaluated based on DQA failure. Finally, the proportion of LOT below 100 ms was assessed.

Logistic Regression Analysis

The correlation between DQA results and the treatment planning parameters were analyzed using an Akaike information criterion stepwise logistic regression. Logistic regression analysis was used to conduct univariate and multivariate analyses to determine the correlation between DQA failure and treatment planning parameters. ROC analysis and areas under the curves (AUCs) were used to identify the sensitivity and specificity of treatment planning parameters cut-off points for the prediction of DQA failure. The optimal cut-off points were defined as the value that maximizes the sum of sensitivity and specificity using the Youden index ($J =$ sensitivity$+$specificity-1).[20]

ROC curves were used to evaluate treatment planning parameters as possible predictors for DQA failure. The higher the AUC, the better the model was considered to be in distinguishing between DQA failure and treatment planning parameters. An AUC of 0.5 meant no discrimination above chance and an AUC of 1.0 indicated perfect discrimination. Generally, AUCs of 0.9-1.0, 0.8-0.9, 0.7-0.8, and 0.6-0.7 indicated excellent, very good, good, and poor discrimination, respectively.[21] All statistical analyses were performed using the R software package (v.3.4.3, R Foundation; Vienna, Austria). A *p*-value of <0.05 was considered statistically significant.

Classification and Regression Tree

CART was used to derive decision rules to predict DQA failure from various treatment planning parameters. Fifteen treatment planning parameters were considered as predictor variables and DQA failure was considered as the predictive variable. Since the number of cases of DQA failure was small (21 cases), the number of failures was increased by 3 times using the bootstrapping method. The importance of a variable was determined using CART for a comprehensive understanding of the influence of different variables (treatment planning parameters) on DQA failure. The variable importance scores were calculated within the CART method using the improvement measure attributable to each variable in its role as a surrogate to the primary split. The values of these improvements are summed over each node and the nodes were then summed together. Then, they were scaled relative to the best performing variable. The variable with the highest sum of improvement

Table 1. Average Values for the Treatment Planning Parameters Used in This Study.

Treatment sites	Brain (43)		Head and Neck (37)		Spine (12)		Prostate (36)		Rectum (36)		Pelvis (13)		Large field RT (24)		Others (11)	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
Target volume [cc]	830.94	795.28	270.58	204.14	330.38	234.50	492.44	611.13	832.55	347.95	927.71	792.31	1647.48	1221.91	785.15	1763.85
Total fractions	19.72	9.13	24.57	7.27	13.25	7.12	22.53	5.51	22.53	6.64	20.77	6.72	20.04	6.03	18.64	8.77
Total dose [cGy]	4523.26	1157.34	5151.57	1473.99	3936.67	871.60	5710.00	1055.53	4753.06	879.11	4346.15	877.13	3887.08	1235.02	4543.64	1595.29
Fractional dose [cGy]	258.60	70.76	214.16	34.23	344.17	98.09	261.81	41.72	237.78	95.86	232.31	87.00	196.04	45.08	307.27	174.82
Pitch	0.33	0.07	0.39	0.08	0.25	0.04	0.34	0.06	0.31	0.05	0.36	0.09	0.38	0.07	0.29	0.10
Field width [cm]	2.27	0.55	2.51	0.53	3.77	1.32	3.00	1.01	2.72	0.71	4.46	1.12	4.94	0.52	2.97	1.03
Modulation factor [Planning]	2.49	0.37	2.30	0.17	2.48	0.34	2.43	0.19	2.22	0.26	2.48	0.32	2.46	0.32	2.25	0.25
Modulation factor [Actual]	2.07	0.39	1.79	0.26	1.71	0.19	1.67	0.19	1.87	0.24	1.58	0.16	1.78	0.30	1.87	0.41
Gantry rotations	19.73	6.45	17.82	7.95	17.76	7.86	17.49	7.56	20.71	6.43	15.81	8.87	28.55	18.93	24.08	15.08
Gantry period [s]	16.62	3.60	16.16	4.03	19.37	5.27	17.05	4.09	25.58	33.66	17.54	3.93	17.46	4.31	18.08	7.81
Time [s]	320.10	104.91	273.54	96.46	320.66	124.18	287.02	125.38	416.50	156.84	279.72	183.03	435.20	213.10	369.94	156.54
Couch travel [cm]	13.99	5.44	16.58	5.70	17.10	10.79	17.35	9.53	16.92	2.71	22.44	7.90	50.42	29.41	20.13	14.47
Couch speed [cm/s]	0.04	0.01	0.06	0.03	0.05	0.03	0.08	0.11	0.05	0.02	0.10	0.04	0.11	0.02	0.05	0.02
Total Monitor Units [MUs]	4325.42	1450.02	3628.24	1362.21	4361.50	1780.42	3747.61	1814.06	5796.06	2181.68	3779.38	2579.60	5942.04	2975.78	5205.27	2188.84
LOT [%]	29.74	17.04	30.92	14.36	23.16	21.15	26.10	6.84	19.82	9.91	25.74	6.81	28.52	17.99	25.25	21.16

Abbreviations: MF, modulation factor; SD, standard deviation; LOT, the percentage of leaf open time below 100 ms. Large field, cranial spinal irradiation field or treatment fields with lymph nodes.

Table 2. Summary of the Treatment Planning Parameters in the Passing and Failing Delivery Quality Assurance Groups.

		Target volume [cc]	Total fractions	Total dose [cGy]	Fractional dose [cGy]	Pitch	FW [cm]	MF [Planning]	MF [Actual]	GR	GP [s]	Treatment time [s]	Couch Travel [cm]	Couch speed [cm/s]	Total MUs	LOT [%]	DD [%]	GPR [%]
Total (212)	Average	743.60	21.20	4758.43	246.53	0.34	3.07	2.38	1.82	20.06	17.59	335.90	20.65	0.07	4552.58	26.65	1.26	93.95
	SD	849.43	7.61	1284.94	83.16	0.08	1.16	0.29	0.32	10.27	4.84	151.41	16.11	0.05	2149.88	14.63	1.65	4.92
Pass (191)	Average	686.28	21.78	4879.29	245.10	0.34	3.00	2.37	1.82	19.35	17.64	324.21	19.49	0.06	4380.28	25.85	1.23	95.05
	SD	810.50	7.34	1210.64	81.14	0.08	1.13	0.29	0.32	9.64	4.85	139.12	15.30	0.06	1983.90	14.19	1.61	3.20
Fail (21)	Average	1264.88	15.90	3659.24	259.48	0.33	3.71	2.48	1.82	26.44	17.17	442.21	31.15	0.08	6119.71	33.94	1.56	83.96
	SD	1027.02	8.18	1444.65	101.03	0.08	1.30	0.28	0.25	13.49	4.90	211.71	19.65	0.03	2916.99	16.77	1.99	6.42

Abbreviation: FW, field width; MF, modulation factor; GR, gantry rotations; GP, gantry periods; MUs, monitor units; LOT, the percentage of leaf open time below 100 ms; DD, dose difference; GPR, gamma passing rate; SD, standard deviation.

was scored 100, and all other variables had decreasingly lower scores. All statistical analyses were performed using Minitab (v18, Minitab; State College, PA).

Results

Analysis of Treatment Planning Parameters and DQA Results

Treatment planning parameters for the 212 patients included are summarized in Table 1. These included patients with brain (n = 43), head and neck (H&N) (n = 37), spinal (n = 12), prostate (n = 36), rectal (n = 36), pelvis (n = 13), cranial spinal irradiation and treatment field including lymph nodes (n = 24), and other (n = 11) cancers. Table 1 shows the patient characteristics for various cases and characteristics of planning parameters such as prescription dose, FW, pitch, MF, LOT, and treatment time. There were 191 (90.1%) and 21 (9.9%) patients who had passing and failing DQA measurements, respectively (Table 2). The average DD and GPR in the passing and failing groups were 1.24% and 94.9%, and 1.4% and 84.5%, respectively (Table 2).

Correlation Between DQA Failure and Treatment Planning Parameters

In the univariate analysis, we found that the total number of fractions, total dose, FW, GR, treatment time, couch travel, total monitor units (MUs), and LOT were significantly related to DQA failure (Table 3). Significant correlations were seen for the total dose ($p < 0.003$), total MUs ($p < 0.040$), and LOT ($p < 0.001$) when multivariate analysis was performed including significant parameters from the univariate analysis (Table 3). As shown in Table 4, the AUCs ranged from 0.52-0.74 for all parameters. The cut-off point was determined as the corresponding value (>0.7) of the AUC value of each parameter. An AUC greater than 0.7 means acceptable discrimination. The AUCs for couch travel, total dose, and LOT were 0.74, 0.71, and 0.70, respectively. The cut-off points for couch travel, total dose, and LOT were 20.15 cm, 4170 cGy, and 29.67%, respectively. When these cut-off values were exceeded, the probability of DQA failure increased. Figure 1 shows the ROC curve for the predicted probability calculated when the multivariate variable model was applied, and the AUC value was 0.82. We confirmed that total dose, couch travel, and LOT were significant predictors of DQA failure. The significant scores were plotted against the variables for treatment planning in Figure 2; the 3 most significant variables were total MUs, total dose, and LOT.

Summary of the Significant Variables for DQA Failure in Clinical Cases

The total dose, total MUs, and the percentage of LOT less than 100 ms were significantly associated with DQA failure in both

Table 3. Univariate and Multivariate Logistic Regression Analysis of the Factors Associated With Delivery Quality Assurance Results.
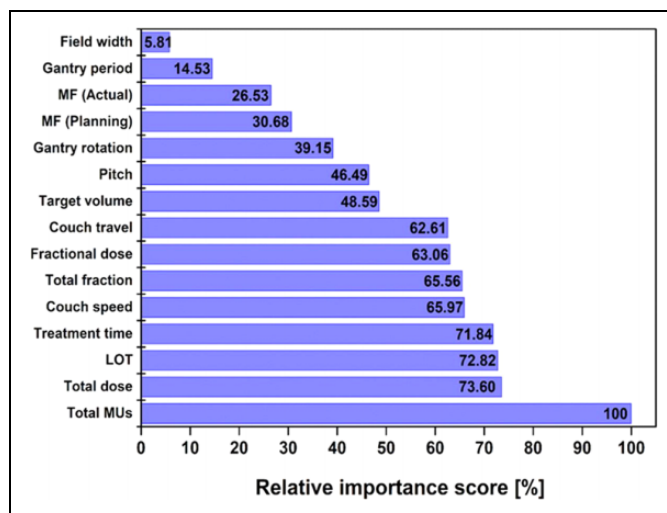
Treatment planning parameters	Univariate analysis				Multivariate analysis			
	OR	LB	UB	*p*-value	OR	LB	UB	*p*-value
Target volume [cc]	1.000	0.999	1.001	0.059	-	-	-	-
Total fractions	0.928	0.881	0.976	0.004*	-	-	-	-
Total dose [cGy]	0.999	0.999	1.001	0.000*	0.999	0.999	1.000	0.003*
Fractional dose [cGy]	1.001	0.997	1.006	0.599	-	-	-	-
Pitch	0.039	0.000	7.194	0.223	-	-	-	-
FW [cm]	1.634	1.188	2.252	0.003*	1.393	0.972	2.000	0.071
MF [Planning]	2.347	0.614	9.009	0.212	-	-	-	-
MF [Actual]	0.627	0.164	2.40	0.495	-	-	-	-
GR	1.044	1.010	1.079	0.010*	0.942	0.867	1.023	0.152
GP [s]	0.951	0.864	1.046	0.298	-	-	-	-
Treatment time [s]	1.003	1.001	1.005	0.015*	0.974	0.945	1.002	0.071
Couch travel [cm]	1.026	1.006	1.046	0.086*	-	-	-	-
Couch speed [cm/s]	17.85	0.082	100	0.290	-	-	-	-
Total MUs	1.001	1.000	1.001	0.011*	1.002	0.996	1.004	0.043*
LOT [%]	1.036	1.011	1.062	0.005*	1.065	1.025	1.106	0.001*

Abbreviation: SD, standard deviation; OR, odds ratio; LB, lower bound; UB, upper bound; FW, field width; MF, modulation factor; GR, gantry rotations; GP, gantry periods; MUs, monitor units; LOT, the percentage of leaf open time below 100 ms. *Indicates a statistically significant difference.
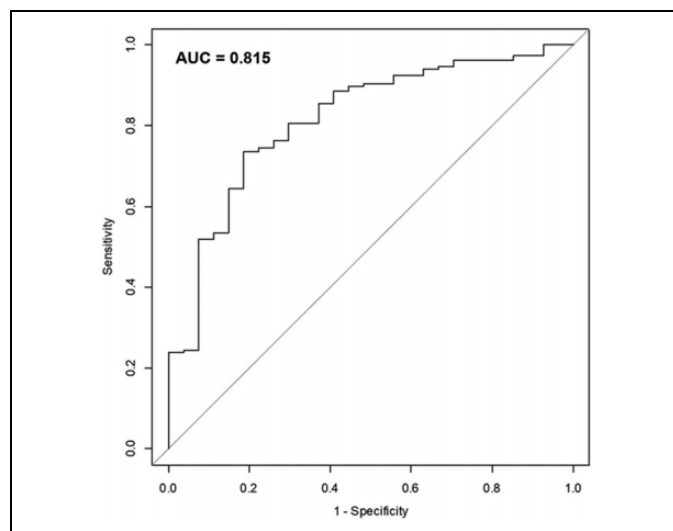
Figure 1. Receiver operating characteristic (ROC) curve and area under the curve (AUC) for multivariate analysis.

Figure 2. Variable importance of parameters influencing delivery quality assurance (DQA) failure. CART (Classification and Regression Tree) was used to identify variable importance. Fifteen parameters were considered as predictor variables and DQA failure as the predictive variable. The total monitor units was determined to be the most important variable (100%), followed by total dose (73.6%) and leaf open time (72.8%).

multivariate (Table 3) and CART analysis (Figure 2). The total dose, couch travel, and LOT were significant parameters of DQA failure as shown in Table 4. Therefore, the results of each clinical case are described with a focus on the parameters mentioned above in Table 5.

Table 5 shows the results of the 4 significant variables (total dose, total MUs, couch travel, and LOT) for DQA failure in various clinical cases. Total dose was lower in cases involving DQA failure than in cases that passed DQA. However, the total MUs and dose were higher in cases involving DQA failure than in cases that passed DQA. DQA failure occurred in cases with a relatively large couch travel value (i.e., in those with a large or long target size). Among all cases, the proportion of cases with LOT below 100 ms in the failure group was higher than that in the passing group, with the exception of H&N and rectal cancer cases. In these groups, the proportion was greater than 30%.

Discussion

In this study, we analyzed the correlation between treatment planning parameters and the DQA measures based on HT DQA

results. We determined that the probability of DQA failure was higher when the proportion of LOT less than 100 ms was higher than approximately 30%. Additionally, in the ROC analysis, we presented the sensitivity values of the optimized plan parameters at each treatment site. Therefore, we presented the results for each planning parameter to reduce the rate of DQA failure. With reference to these results, the values are given so that the planning parameters can be considered prior to treatment to reduce DQA failure. Since there are various treatment planning parameters in Tomotherapy, it is time-consuming and labor-intensive to use routine clinical practice to determine new treatment plans and DQA measurements to change these

Table 4. Areas Under the Curve (AUCs) and Cut-Off Point for Each Treatment Parameter.

Treatment planning parameters	AUC	Cut-off point
Target volume [cc]	0.62	308.73
Total fractions	0.67	17.50
Total dose [cGy]	**0.71***	4170.00
Fractional dose [cGy]	0.51	205.00
Pitch	0.60	0.29
FW [cm]	0.55	3.76
MF [Planning]	0.57	2.43
MF [Actual]	0.52	2.04
GR	0.65	17.11
GP [s]	0.57	14.86
Treatment time [s]	0.59	501.15
Couch travel [cm]	**0.74***	20.15
Couch speed [cm/s]	0.65	0.07
Total MUs	0.61	6842.50
LOT [%]	**0.70***	29.67

Abbreviation: AUC, area under the curve; MF, modulation factor; GR, gantry rotations; GP, gantry periods; MUs, monitor units; LOT, the percentage of leaf open time below 100 ms. *AUC greater than 0.7 means acceptable discrimination.

planning parameters. Although it is difficult to use statistical analysis in routine clinical practice, the planners or dosimetrists can predict the DQA results according to the treatment planning parameters in advance.

Binny et al. investigated the DD and GPR value in cases of brain, H&N, and pelvic cancer.[18] They found that the DD and GPR values in the brain, H&N, and pelvic cases were $0.01 \pm 1.04\%$ and $95.94 \pm 2.70\%$, $1.21 \pm 1.43\%$ and $96.30 \pm 3.06\%$, and $0.44 \pm 0.88\%$ and $95.35 \pm 2.99\%$, respectively. As shown in Table 5, the DD results for the passing group ranged from 0.3% (H&N) to 2.0% (pelvis) from a previously reported study, whereas the GPRs in passing groups were consistent with the results. However, in the previous paper, there was no data for failure cases. Therefore, we did not compare and evaluate the DQA failure cases with this study.[18]

In this study, the proportion of cases with LOT below 100 ms was assessed in the passing and failing groups. The average LOT in the passing and failing groups were $25.85 \pm 14.19\%$ and $33.94 \pm 16.77\%$, respectively (Table 2). As mentioned above, Accuray recommends that the percentage of LOT below 100 ms is maintained at less than 30% due to the risk of increased MLC errors.[16] We showed that the AUC of LOT was approximately 0.7; this represented acceptable discrimination between the AUC and DQA failure, and its cut-off point was approximately 30% as shown in Table 4. When the proportion was higher than this value, it was confirmed that DQA failure increased. This study is the first to investigate the correlation between DQA failure and treatment planning parameters using ROC curves and AUC values.

Westerly et al. showed that plans with small mean open times increased the rates of DQA failure among 6 patients. When the mean LOT was increased from 29.8% to 83.1% relative to the initial plans, DDs were reduced by approximately 2-4.4% for re-planned plans. They explained that the impact of small LOTs should be considered since LOTs less

Table 5. Summary of the Important Variables for Delivery Quality Assurance Failure in Clinical Cases.

Cases		Total dose [cGy]	Total MUs	Couch travel [cm]	LOT [%]	DD [%]	GPR [%]
Brain	Pass (40)	4537.50 ± 1148.84	4312.18 ± 1472.76	13.84 ± 5.57	28.41 ± 16.61	1.05 ± 1.44	95.76 ± 3.38
	Fail (3)	4333.33 ± 1527.53	4502.00 ± 1344.32	16.03 ± 3.24	47.47 ± 14.58	1.04 ± 1.61	86.98 ± 3.63
H&N	Pass (34)	5363.94 ± 1291.52	3566.09 ± 1401.64	16.10 ± 5.71	31.09 ± 14.98	1.50 ± 1.43	96.09 ± 3.29
	Fail (3)	2744.67 ± 1459.06	4332.67 ± 418.48	21.93 ± 1.21	28.91 ± 2.14	0.47 ± 1.74	82.63 ± 11.42
Pelvis	Pass (10)	4570.00 ± 802.84	3698.00 ± 2638.31	20.81 ± 6.65	24.70 ± 7.46	2.58 ± 2.07	94.10 ± 2.86
	Fail (3)	3600.00 ± 793.73	4050.67 ± 2908.52	27.87 ± 10.86	29.21 ± 2.03	3.06 ± 1.10	82.37 ± 6.05
Prostate	Pass (32)	5868.44 ± 889.75	3496.62 ± 1459.24	15.83 ± 7.99	25.60 ± 7.03	1.42 ± 1.23	94.25 ± 3.03
	Fail (4)	4442.50 ± 1551.82	5755.50 ± 3215.88	29.58 ± 13.23	30.08 ± 3.35	2.29 ± 1.92	80.66 ± 7.53
Rectum	Pass (33)	4882.12 ± 713.40	5429.79 ± 1777.10	16.59 ± 2.54	20.08 ± 10.20	0.50 ± 1.28	94.08 ± 3.08
	Fail (3)	3333.33 ± 1443.38	9825.00 ± 2490.69	20.53 ± 2.02	16.93 ± 6.16	1.13 ± 1.71	85.38 ± 2.51
Spine	Pass (12)	3936.67 ± 871.60	4361.50 ± 1780.42	17.10 ± 10.79	23.16 ± 21.15	2.73 ± 2.21	95.98 ± 3.71
	Fail (0)	-	-	-	-	-	-
Large field	Pass (20)	4127.50 ± 1156.37	5523.50 ± 2940.27	47.76 ± 30.58	24.48 ± 12.40	0.64 ± 1.83	94.18 ± 3.02
	Fail (4)	2685.00 ± 955.46	8043.75 ± 2476.01	63.73 ± 20.66	48.68 ± 29.28	0.55 ± 2.81	88.61 ± 1.26
Others	Pass (10)	4368.00 ± 1565.47	5203.60 ± 2307.23	19.95 ± 15.24	24.76 ± 22.24	1.35 ± 4.24	95.91 ± 1.19
	Fail (1)	6300	5222	21.90	30.18	4.24	73.98

Abbreviations: MUs, monitor units; LOT, the percentage of leaf open time below ± 100 ms; DD, dose difference; GPR, gamma passing rate; H & N, head and neck; Large field, cranial spinal irradiation field or treatment fields with lymph nodes.

than 100 ms are not accounted for in radiation treatment planning.[13] Conversely, Binny et al. reported that the proportion of LOT below 100 ms was more than 30% for all cases. For brain cases, the LOT value (53.89 ± 9.25%) was the highest; it was approximately 20% higher than that among the H & N and pelvis cases. Binny et al. also found that there was no correlation between LOT and DD or gamma value. They described that the LOT was determined at the maximum MF, and the brain cases involved a complicated treatment site; a pitch of 0.287 was thus used to reduce the thread effect. However, they also described that it is necessary to reduce the percentage of LOT to <100 ms to improve the efficiency of treatment delivery.[18] We showed that the proportion of LOT below 100 ms in the passing group was higher than that in the failing group for H&N and rectal cases, as shown in Table 5. This result opposed the findings acquired using statistical analysis in this study. It has been explained that DQA failure could be determined using other treatment planning parameters such as target volume, couch travel, MU, or treatment time.[14] In the future, it is necessary to investigate the contributing weights of each treatment planning parameter in DQA failure. Additionally, an analysis including more DQA failure cases should be conducted to improve the accuracy of prediction for DQA failure.

Westerly et al. recommended the proper pitch value be selected so that GP is at least 15 s to reduce the impact of inaccuracies of MLC. This could improve the efficiency of the treatment delivery as well as reduce the delivery time.[13] It was confirmed that the average GP values in various clinical cases were greater than 15 s.[18] In the current study, we also found that the GP values were above 15 s regardless of the passing or failing DQA cases. We found that GP values were not a significant factor that affected DQA failure, since the average GP values in the passing and failing groups were >15 s.

The FW, GR, and treatment time were correlated with DQA failure in the univariate analysis, but they were not correlated in the multivariate analysis. In this study, we found that DQA failure was correlated with LOT, couch travel, total dose, and total MUs in the univariate and multivariate analyses as shown in Tables 3 and 4. Couch travel is related to irradiation length. The MLC inaccuracies, such as tongue and groove/penumbra effect and MLC leaf latencies,[13] can be greater with increasing number of individual leaf behavior to open and close the leaf to modulate the beam in the treatment field.[22] In addition, the thread effect due to the helical field junction is increased. An increased number of MUs results in increased leakage radiation.[23] These various uncertainties can lower the GPR within the measurement area.[9] We confirmed that the total dose was relatively smaller in the failure cases than in the passing cases. And it was confirmed that DQA failure decreased as the total dose increased as shown in the odds ratio value (0.999) of univariate and multivariate analysis (Table 3). Also, we confirmed that the total dose was small, the fractional dose was high, and the couch travel was long for failure cases. As a result, as treatment time increases, the MU also increases, and it is assumed that DQA failure eventually occurs. Therefore, it cannot be determined whether a single parameter has an effect

on DQA failure, and we assumed that various factors work in combination and affect DQA failure. We are currently collecting patient data to analyze this issue. In addition, Binny et al. showed that factors contributing to gantry speed are MF, pitch, and FW. MF, pitch, and FW can be set by the planner; however, the actual MF, projections, and LOT are largely determined by the optimization technique. Therefore, it is important to collect and analyze treatment parameter outcomes based on DQA for each treatment site to make informed decisions for user-controlled parameters specific to each treatment region.[18] Figure 1 shows the ROC curve for the predicted probability calculated when the multivariate variable model was applied, and the AUC value was 0.82. We confirmed that total dose, couch travel, and LOT were significant predictors of DQA failure. The significant scores were plotted against the variables for treatment planning as shown in Figure 2. The relative importance score suggests that total MUs has the most significant impact on DQA failure. In addition, total dose, LOT, and treatment time among 15 treatment planning parameters were important factors. Thus, when we comprehensively analyze the results of Figures 1 and 2, we believe that LOT and total dose were the most important factors for DQA failure. The LOT and total dose factors were consistent with the results in work by previous reports.[14,18] Since total dose and couch travel are parameters that were determined by the physicians before the treatment planning process, both parameters are factors that the planner cannot change in the treatment planning process. Conversely, among the 3 parameters influencing DQA failure, LOT is the only parameter that could be changed in the treatment planning process, although planners require re-optimization to change the LOT value. Therefore, to reduce the probability of DQA failure, it is necessary to maintain the percentage of LOT below 100 ms to be less than 30%. No other studies have used logistic regression, ROC, AUC, or CART analyses for the investigation of DQA failure for HT. Therefore, although it is difficult to directly compare these results with previous studies, it was found that, as was similar to the findings from previous studies, the parameter affecting DQA failure is LOT.[13,14,18]

HT plans are quite complicated because there are various planning parameters and each parameter affects other parameters. Therefore, planners or dosimetrists need to understand the impact and relevance of various parameters when creating a plan. Although it is difficult to routinely apply these methods to all cases in clinical practice due to the small number of DQA failures, we suggest that our planning guideline acquired through statistical analysis could assist in devising treatment plans by referring to Tables 4 and 5.

This study has some limitations inherent in retrospective studies. First, in this study, all DQA measurements were taken using EBT3 film and the effect of the scanner light may have introduced sources of various uncertainties, including warming of the scanner lamp, film homogeneity, scan-to-scan stability, long-term stability of the scanner, light scattering, film calibration, phantom setup, measurement position, and human errors. These uncertainties may contribute to decreases in the accuracy of DQA.[24,25] Second, the number of failing cases (n = 21) was

only approximately 1/10 of the number of passing cases. Since there were only a few DQA failure cases, determining the parameters affecting DQA failure was difficult, and the statistical accuracy of the predictors may be decreased. To overcome this problem, the numbers of DQA failure cases was increased by using randomly selected bootstrapped samples for CART analysis. In this analysis, it was confirmed that the variables affecting DQA failure were similar to those variables (total dose, total MUs, and LOT) acquired using the logistic regression (multivariate analysis), AUC, ROC, and CART analysis methods (Figure 2, Tables 3 and 4). However, to predict the risk of DQA failure in treatment planning procedures or without re-measurement of DQA, the prediction accuracy should be improved by increasing the number of included cases. Finally, the correlation between treatment planning parameters and the DQA failure pattern was not verified with the acquired statistical model. This requires further validation in a study with a larger number of enrolled patients. Moreover, the use of artificial intelligence or a machine-learning model might improve the accuracy of DQA by accurately and quickly analyzing large numbers of HT plans and a more considerable amount of DQA data. This requires further investigation in a future study.

Conclusion

We confirmed that the probability of DQA failure was higher when the proportion of LOT below 100 ms was greater than 30%. Based on the findings from this study, it is evident that planners should consider the proportion of LOT below 100 ms when they create treatment plans. Although it is difficult to use statistical analysis in routine clinical practice, our findings may contribute to the prediction of DQA failure in advance.

Ethical Statement

Our study did not require an ethical board approval because it did not contain human or animal trials.

ORCID iD

Chae-Seon Hong 🔟 https://orcid.org/0000-0001-9120-6132

References

1. Chang KH, Ji Y, Kwak J, et al. Clinical implications of high definition multileaf collimator (HDMLC) dosimetric leaf gap (DLG) variations. *Prog Med Phys*. 2016;27(3):111-116.

2. Cho B. Intensity-modulated radiation therapy: a review with a physics perspective. *Radiat Oncol J*. 2018;36(1):1-10.

3. Chang KH, Lee S, Jung H, et al. Development of a 3D optical scannerfor evaluating patient-specific dose distributions. *Physica Medica*. 2015;31(5):553-559.

4. Thiyagarajan R, Nambiraj A, Sinha SN, et al. Analyzing the performance of ArcCHECK diode array detector for VMAT plan. *Rep Pract Oncol Radiother*. 2016;21(1):50-56.

5. Mackie TR, Balog J, Ruchala K, et al. Tomotherapy. *Semin Radiat Oncol*. 1999;9(1):108-117.

6. Mackie TR, Kapatoes J, Ruchala K, et al. Image guidance for precise conformal radiotherapy. *Int J Radiat Oncol Biol Phys*. 2003;56(1):89-105.

7. Cao YJ, Lee S, Chang KH, et al. Patient performance–based plan parameter optimization for prostate cancer in tomotherapy. *Med Dosim*. 2015;40(4):285-289.

8. Kissick MW, Fenwick J, James JA, et al. The helical tomotherapy thread effect. *Medical Physics*. 2005;32(5):1414-1423.

9. Chen M, Chen Y, Chen Q, Lu W. Theoretical analysis of the thread effect in helical TomoTherapy. *Med Phys*. 2011;38(11):5945-5960.

10. Shimizu H, Sasaki K, Tachibana H, et al. Analysis of modulation factor to shorten the delivery time in helical tomotherapy. *J Appl Clin Med Phys*. 2017;18(3):83-87.

11. Skorska M, Piotrowski T. Optimization of treatment planning parameters used in tomotherapy for prostate cancer patients. *Phys Med*. 2013;29(3):273-285.

12. De Kerf G, Van Gestel D, Mommaerts L, Van den Weyngaert D, Verellen D. Evaluation of the optimal combinations of modulation factor and pitch for Helical TomoTherapy plans made with TomoEdge using Pareto optimal fronts. *Radiat Oncol*. 2015;10(1):191.

13. Westerly DC, Soisson E, Chen Q, et al. Treatment planning to improve delivery accuracy and patient throughput in helical tomotherapy. *Int J Radiat Oncol Biol Phys*. 2009;74(4):1290-1297.

14. Binny D, Lancaster CM, Harris S, Sylvander SR. Effects of changing modulation and pitch parameters on tomotherapy delivery quality assurance plans. *J Appl Clin Med Phys*. 2015;16(5):87-105.

15. Thomas SJ, Geater AR. Implications of leaf fluence opening factors on transfer of plans between matched helical tomotherapy machines. *Biomed Phys Eng Express*. 2017;4(1):017001.

16. Vaden RSN. Efficient planning for efficient treatments on tomotherapy. 2016.

17. Boyd R, Jeong K, Tome WA. Determining efficient helical IMRT modulation factor from the MLC leaf-open time distribution on precision treatment planning system. *J Appl Clin Med Phys*. 2019;20(5):64-74.

18. Binny D, Lancaster CM, Byrne M, Kairn T, Trapp JV, Crowe SB. Tomotherapy treatment site specific planning using statistical process control. *Phys Med*. 2018;53:32-39.

19. Chang KH, Kim DW, Choi JH, et al. Dosimetric comparison of four commercial patient-specific quality assurance devices for helical tomotherapy. *J Korean Phys Soc*. 2020;76(3):257-263.

20. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005;16(1):73-81.

21. Lemeshow S, Hosmer DW. *Applied Logistic Regression*. John Wiley & Sons, Inc; 2000:156-164.

22. Zhao YL, Mackenzie M, Kirkby C, Fallone BG. Monte Carlo calculation of helical tomotherapy dose delivery. *Med Phys*. 2008;35(8):3491-3500.

23. Lárraga-Gutiérrez JM, Galván de la Cruz OO, García-Garduño OA, Ballesteros-Zebadúa P. Comparative analysis of several detectors for the measurement of radiation transmission and leakage from a multileaf collimator. *Phys Med*. 2014;30(3): 391-395.

24. Martisikova M, Ackermann B, Jakel O. Analysis of uncertainties in Gafchromic EBT film dosimetry of photon beams. *Phys Med Biol*. 2008;53(24):7013-7027.

25. Tomori S, Kadoya N, Takayama Y, et al. A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance. *Med Phys*. 2018;45(9):4055-4065.