Sensitivity of IROC phantom performance to radiotherapy treatment planning system beam modeling parameters based on community-driven data

Mallory C. Glenn
Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
The University of Texas MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA

Christine B. Peterson
The University of Texas MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA
Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Rebecca M. Howell, David S. Followill, Julianne M. Pollard-Larkin, and Stephen F. Kry[a)]
Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
The University of Texas MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA

(Received 13 April 2020; revised 29 June 2020; accepted for publication 6 July 2020;
published 16 August 2020)

Purpose: Treatment planning system (TPS) dose calculations have previously been shown to be sensitive to modeling errors, especially when treating with complex strategies like intensity-modulated radiation therapy (IMRT). This work investigates the dosimetric impact of several dosimetric and nondosimetric beam modeling parameters, based on their distribution in the radiotherapy community, in two commercial TPSs in order to understand the realistic potential for dose deviations and their clinical effects.

Methods and materials: Beam models representing standard 120-leaf Varian Clinac-type machines were developed in Eclipse 13.5 (AAA algorithm) and RayStation 9A (v8.99, collapsed-cone algorithm) based upon median values of dosimetric measurements from Imaging and Radiation Oncology Core (IROC) Houston site visit data and community beam modeling parameter survey data in order to represent a baseline linear accelerator. Five clinically acceptable treatment plans (three IMRT, two VMAT) were developed for the IROC head and neck phantom. Dose distributions for each plan were recalculated after individually modifying parameters of interest (e.g., MLC transmission, percent depth doses [PDDs], and output factors) according to the 2.5th to 97.5th percentiles of community survey and machine performance data to encompass the realistic extent of variance in the radiotherapy community. The resultant dose distributions were evaluated by examining relative changes in average dose for thermoluminescent dosimeter (TLD) locations across the two target volumes and organ at risk (OAR). Interplay was also examined for parameters generating changes in target dose greater than 1%.

Results: For Eclipse, dose calculations were sensitive to changes in the dosimetric leaf gap (DLG), which resulted in differences from −5% to +3% to the targets relative to the baseline beam model. Modifying the MLC transmission factor introduced differences up to ± 1%. For RayStation, parameters determining MLC behaviors likewise contributed substantially; the MLC offset introduced changes in dose from −4% to +7%, and the MLC transmission caused changes of −4% to +2%. Among the dosimetric qualities examined, changes in PDD implementation resulted in the most substantial changes, but these were only up to ±1%. Other dosimetric factors had <1% impact on dose accuracy. Interplay between impactful parameters was found to be minimal.

Conclusion: Factors related to the modeling of the MLC, particularly relating to the leaf offset, can cause clinically significant changes in the calculated dose for IMRT and VMAT plans. This should be of concern to the radiotherapy community because the clinical effects of poor TPS commissioning were based on reported data from clinically implemented beam models. These results further reinforce that dose errors caused by poor TPS calculations are often involved in IROC phantom failures.
 [https://doi.org/10.1002/mp.14396]

Key words: beam modeling, MLC, quality assurance, treatment planning

1. INTRODUCTION

Radiation dose calculation accuracy is contingent upon how well the treatment planning system (TPS) mathematically represents the physical photon beam under the conditions used for radiation therapy. Good commissioning and validation of the beam model is fundamental, for once established, this model is used to calculate the dose for all treatments with the radiation beam. However, modern technologies such as intensity modulated radiation therapy (IMRT) and volumetric

modulated arc therapy (VMAT) pose a particular challenge for TPS dose calculations due to the modulation and constraints needed to provide highly conformal dose distributions. This challenge results in increased uncertainty in the dose calculation.

In order to properly assess the accuracy of IMRT, external validation tests are suggested for individual institutions, as well as for clinical trials.[1–3] The Imaging and Radiation Oncology Core Houston Quality Assurance Center (IROC-H) provides anthropomorphic phantom credentialing for National Cancer Institute-sponsored multi-institutional clinical trials for IMRT to ensure treatments are delivered as intended while minimizing uncertainty. Over the years, IROC-H has observed a broad range of IMRT performance.[4] In particular, recent works from IROC-H indicate a substantial number of phantom results showing systematic dose errors,[5] and poor TPS dose calculations in failing phantom cases.[6] Additionally, there exists substantial evidence that standard quality assurance (QA) methods, including IMRT QA, fail to detect unacceptable plans and errors related to the TPS.[7–10]

Due to these challenges, interest has developed in understanding how beam modeling, and which specific factors within the model, can contribute to poor plan performance. Previous studies have investigated the relative errors that several modeling factors related to the multileaf collimator (MLC) can contribute to the overall accuracy, as well as the detectability of these errors.[11–13] While generally informative, such works have been relative to single clinical systems, and thus cannot provide wide-ranging context into other clinical scenarios. More problematically, the magnitude of change in each parameter (i.e., how much error is introduced into the MLC offset) and associated effect size, have not been based on clinically realistic values. That is, the ranges of values for modeling parameters used in these works are, in general, arbitrary and may not necessarily be relevant to current practice.

Instead, this study evaluated the impact of beam modeling errors (both dosimetric and nondosimetric) that are consistent with the errors seen clinically, or are consistent with the range of values used in clinical practice. TPS errors in basic dosimetric data, such as percent depth dose (PDD) measurement, have been previously reported by IROC-H based on measurements of over 1000 linear accelerators (linacs).[14] Nondosimetric data, such as MLC leaf offset and MLC transmission factor, have been compiled in a recent IROC-H survey that included information from over 2800 beam models from 642 institutions.[15] These data, collected from January 2018 to January 2019, describe the most up-to-date modeling descriptions for the radiotherapy community, including those of both large academic centers and smaller community clinics. In this study, we used these values from the community to determine the degree of change introduced into each parameter. In this way, this study investigated the potential dosimetric impact of using beam modeling parameter values that are either erroneous or at least deviate from typical as established by the radiation oncology community. These data can inform the ways that errors can and do manifest for IMRT and VMAT treatments across the community at large. Understanding the expected error contributions of erroneous or atypical parameter values can also help explain and rectify the ongoing suboptimal IROC-H phantom performance rates by providing more in-depth guidance to the dose calculation variations that may exist.

2. METHODS AND MATERIALS

2.A. TPS beam model creation and validation

Beam models representing a 6 MV beam on a Varian Clinac-type machine with Millennium 120 multileaf collimator (MLC) were developed in Eclipse v13.5 with AAA algorithm (Varian Medical Systems, Palo Alto, CA) and RayStation 9A v8.99 with collapsed cone algorithm (RaySearch Laboratories, Stockholm, Sweden). Dosimetric characteristics were tailored to the linac-specific reference data from IROC-H's site visit program, which encompasses 23 output measurements (including several percent depth dose curves, output factors, and off-axis factors).[16] Nondosimetric modeling parameters (e.g., source size, MLC leaf-tip offset, etc.) were defined to match median beam modeling parameters as reported by the radiation oncology community in an IROC-H survey (Glenn, et al.[15]). This was specific for a Varian Base class linac, described in detail in Kerns, et al.,[16] with standard 120-leaf MLC. In this way, the most representative linac (of a widely used model) was created.

Baseline beam models for Eclipse and RayStation were then validated via two IROC-H head and neck (H&N) phantom irradiations on a clinical Varian Trilogy linac. The two plans, a standard nine-field IMRT and two-arc VMAT, were assessed for agreement between thermoluminescent dosimeter (TLD) dose and TPS-reported dose as calculated by both Eclipse and RayStation to ensure clinical applicability and reasonability of the baseline models prior to further manipulation and study.

2.B. Phantom plan development

The IROC-H H&N phantom was scanned on a CT simulator following standard clinical workflow, and five IMRT plans (three IMRT, two VMAT) were developed in Eclipse 13.5 and imported to RayStation for consistency across platforms. A grid size of 0.25 cm was used for all plans in both TPS. Following IROC protocol, the plans were designed to deliver 660 cGy to 95% of the primary target and 540 cGy to 95% of the secondary target while maintaining organ-at-risk (OAR) dose below 450 cGy. Within the targets and OAR are eight volumes defining TLD locations: four locations distributed anterior/posterior and superior/inferior in the primary target, two locations distributed superior/inferior in the secondary target, and two locations distributed superior/inferior in the OAR. These plans were designed for dynamic IMRT delivery in a single fraction and follow dose prescription guidelines provided by IROC-H (Fig. 1). Monitor units (MU) and general plan setup for each of the five plans are detailed

in Table I. All IMRT fields were planned about the center of the H&N phantom and spaced equidistantly, as is common in clinical practice. Likewise, the VMAT plans were developed for isocentric delivery with full 360° arcs. These plans were developed with a variety of complexities and beam angles to encompass a range of treatment strategies and plans as previously observed by IROC-H.[17]

2.C. Parameter manipulation (simulated beam model deviations) and evaluation

The baseline models, in Eclipse and RayStation, had all parameters at the 50th percentile community value. In order to simulate the impact on dose agreement from variations that have been shown to exist in the radiotherapy community, alternative versions of the baseline beam models were created by individually manipulating parameters of interest within each TPS environment. The parameters of interest considered in this study, including both basic dosimetric characteristics and TPS-specific modeling parameters, are outlined in Table II. When manipulations of a parameter were introduced, all other parameters were maintained at the baseline (50th percentile) value in order that effects may be isolated. Variations in beam modeling parameters were introduced as the 2.5th, 25th, 50th, 75th, and 97.5th percentiles of beam modeling parameter survey responses.[15] In order to characterize variance in dosimetric characteristics, the baseline beam model was modified in RayStation by changing the photon spectrum, square-field output factor correction factors (jaw-defined fields), and off-axis factors. Changes in dosimetric parameters were made for the same percentiles (2.5th to 97.5th), based on reported measurement accuracy versus TPS calculation from IROC-H's site visit program.[14]

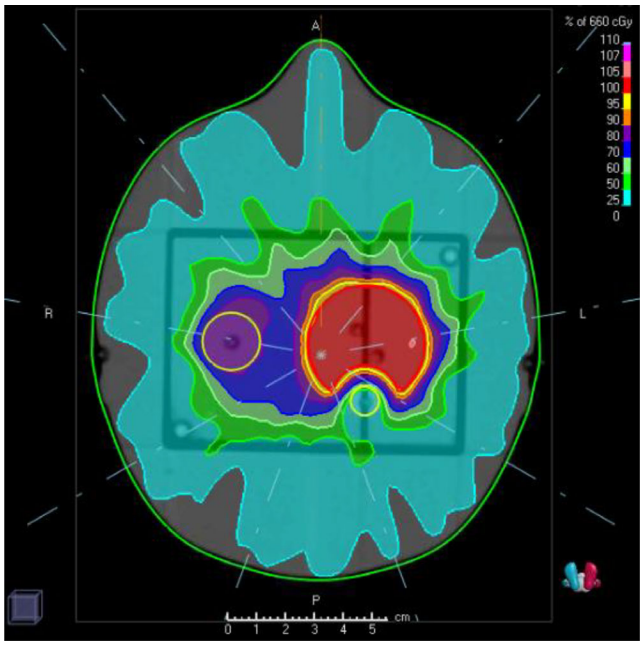

FIG. 1. Axial slice of the nine-field intensity modulated radiation therapy plan developed for the IROC-H H&N phantom. The phantom contains primary and secondary targets planned to be treated at 660 and 540 cGy, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Summary of H&N phantom plans developed for testing.

Plan name	Delivery type	Number of beams	Total MU
IMRT5	IMRT	5	3741
IMRT7	IMRT	7	2470
IMRT9	IMRT	9	2729
VMAT1	VMAT	1 arc (360°)	1990
VMAT2	VMAT	2 arcs (each 360°)	2130

Note: Phantoms were prescribed 660 cGy to the primary target, delivered in a single fraction. All plans were developed for dynamic delivery.

Following beam model modifications, each of the five H&N phantom treatment plans was recalculated (plans were not re-optimized in this process) and compared to the calculated dose from the baseline model. These recalculations were evaluated for average change in dose across the six TLD locations distributed throughout the primary (four TLD locations) and secondary targets (two TLD locations) in the phantom and average change in dose of the two TLD locations in the organ-at-risk (OAR).

2.D. Parameter interplay

To understand potential interdependencies in beam modeling parameters, we investigated the effect of changing two parameters simultaneously. We could then evaluate if, for example, two different parameters, both set simultaneously to the 97.5th percentile, had a different impact on the dose distribution than simply the sum of the two effects when each parameter was set sequentially to the 97.5th percentile. Parameters that were investigated were only those that, by themselves, had a sizeable impact on the dose distribution; in particular, those parameters that introduced changes in average dose for the TLD locations greater than 1% across the five treatment plans were selected. These impactful parameters were then varied pairwise against other impactful parameters in the same TPS. For this analysis, the most extreme values (2.5th and 97.5th percentiles) were adopted for one

TABLE II. Beam modeling parameters investigated for sensitivity.

Eclipse	RayStation	Dosimetric characteristics
Effective target spot size	Primary source size	Percent depth dose (PDD)
MLC transmission factor	MLC transmission	Small-field output factors[a]
Dosimetric leaf gap (DLG)	Leaf tip width	Off-axis factors
	Tongue and groove	
	MLC leaf tip offset	
	MLC gain	
	MLC curvature	

[a]Jaw-defined small field output factors.

parameter while the other was varied across the distribution to assess the greatest extent to which interplay can occur. Interplay between parameters was assessed through linear regression analysis including terms for potential interaction effects. Additionally, the dosimetric effect of the combined scenario was visually compared to simply summing the average effects together (thus representing complete independence).

3. RESULTS

The baseline TPS models, used to irradiate the phantom for validation, showed agreement between measurement and calculation of within 5% across all locations for both planning systems and both plans. This level of agreement exceeds IROC-H's acceptability criterion (7%/4 mm) and indicated that the baseline beam models are a good representation of a standard linac.

3.A. Eclipse

Table III describes the numeric parameter modifications implemented in Eclipse, based on the values from Glenn, et al.[15] Figure 2 shows the average percent difference in calculated dose for the TLD locations (over the six TLD locations within the two targets, averaged across all five plans) between the baseline and modified beam models as a function of the percentile score for the parameters of interest as reported by the radiation oncology community. Based on the range of values from the community, the variations in DLG have the greatest impact on dosimetric accuracy. This was true for all endpoints examined: changes up to 6% were observed for target TLD locations, and changes up to 10% were observed in the OAR TLD locations.

The dosimetric impact is plotted against actual DLG value in Fig. 3. From this figure it can be seen that the average dose to the targets changes approximately linearly with DLG value for all plans examined (Fig. 3), suggesting systematic effects on dose. Moreover, for all plans except IMRT5, the relative change in dose caused by DLG variations was consistent despite differences in treatment complexity. IMRT5, the plan that was very highly modulated, had a notably different slope than the others, and was more sensitive to changes in DLG. The other beam modeling parameters produced much smaller dose deviations; changes in dose caused by manipulation of the effective target spot size and MLC transmission factor were generally less than 1% in the target volume doses and <5% change in OAR dose.

To further investigate these results, we assessed the interdependence of the DLG and MLC transmission factor, given that both produced differences in average dose for the target TLD locations >1%. Figure 4 depicts a comparison between the cases for which only the DLG was varied and the MLC transmission factor was left at the 50th percentile value (i.e., MLCT 50th) and when the DLG and MLC were varied together (with the MLC transmission factor defined at the 2.5th and 97.5th percentiles). Regression modeling of the calculated dose values determined that the two parameters have statistically significant linear effects on dose ($P < 0.001$), while their interaction was not significant ($P = 0.233$). That is, there is no evidence to suggest that the effect of DLG on dose is different for different values of MLC transmission factor. Additionally, these were compared with the cases where average change in dose caused by variation in MLC transmission factor (here, −1.14% for the 2.5th percentile and +0.92% for the 97.5th percentile) was simply summed with the DLG-only case, thereby assuming the DLG and MLC transmission factor behaved independently (the "expected" cases). The calculated and expected dose difference curves are consistent, further emphasizing the linearity and predictability of dose response with respect to these parameters.

TABLE III. Parameter changes implemented for the 6 MV Varian Clinac-type machine simulated in Eclipse.

Percentile	Effective target spot size X (mm)	Effective target spot size Y (mm)	Dosimetric leaf gap (cm)	MLC transmission factor
2.5th	0.000	0.000	0.1000	0.0118
25th	0.000	0.000	0.1550	0.0145
50th	0.000	0.000	0.1700	0.0158
75th	0.000	0.000	0.1900	0.0165
97.5th	1.250	1.000	0.2300	0.0200

3.B. RayStation

Table IV describes the numeric parameter modifications implemented in RayStation, based on the values from Glenn, et al.[15] Figure 5 shows the changes in average dose for the TLD locations in the target relative to the baseline (50th percentile values) as a function of percentile score. Analogous to the DLG in Eclipse, dose calculation accuracy had the strongest dependence on the MLC position offset: changes in MLC position offset produced as much as a 13% change in the target TLD locations and 25% in the OAR TLD locations.

The dosimetric impact across the target TLD locations is plotted against MLC offset value in Fig. 6. From this figure, it can be seen that the average dose to the target changes linearly with MLC offset for all plans examined (Fig. 6), suggesting systematic effects on dose. As with DLG, the MLC offset was most sensitive to the IMRT5 plan, which was very highly modulated, and was uniformly sensitive to all of the other plans.

Other parameters related to the MLC, including the MLC transmission factor, tongue and groove, and leaf tip width, also greatly impacted the resultant dose recalculations. For MLC transmission, average dose changes in the target TLD locations ranged from −4% to +2% with singular points up to 10% off in the OAR. Changing the leaf tip width generated more moderate changes in the target TLD locations (up to 3%) but produced changes as high as 15% in the OAR. Interestingly, the tongue and groove produced negligible changes in any of the IMRT plans, but introduced uniform and

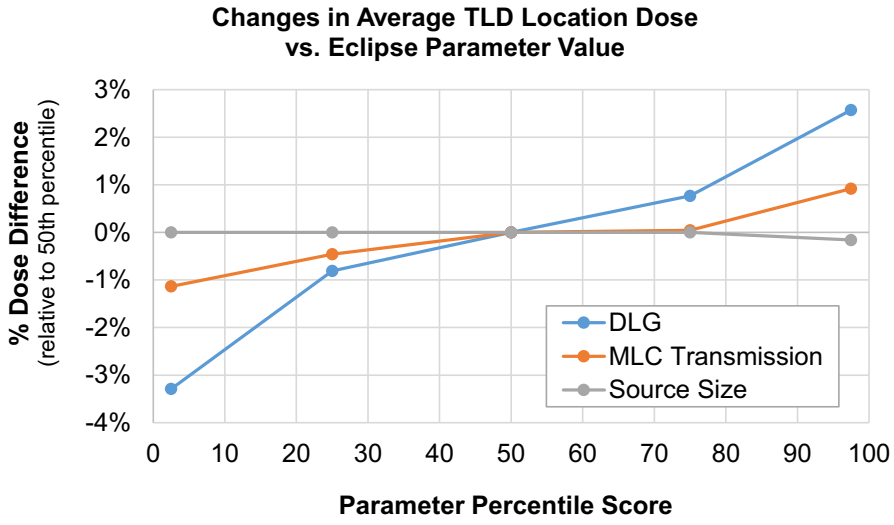

FIG. 2. Average changes in dose calculated to target thermoluminescent dosimeter locations for IROC-H head and neck phantom plans when parameters of interest are manipulated in Eclipse. Percentile score corresponds to community reported values.[15] DLG = dosimetric leaf gap. [Color figure can be viewed at wileyonlinelibrary.com]

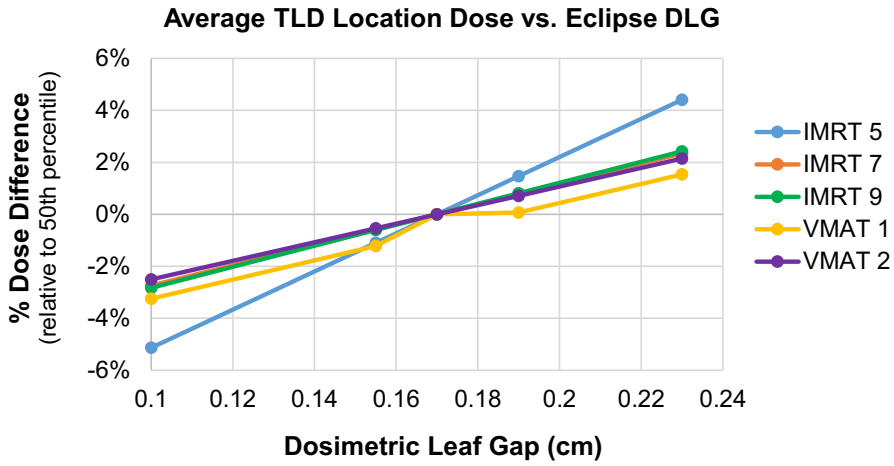

FIG. 3. Changes in average dose for the target thermoluminescent dosimeter locations calculated for each of the five IROC-H head and neck phantom plans following manipulation of the dosimetric leaf gap in Eclipse. Points on the curve correspond to percentile scores from the radiation oncology community for a Varian Base Class machine with a Millennium 120 leaf multileaf collimator. [Color figure can be viewed at wileyonlinelibrary.com]

consistent changes in the target TLD locations up to 3% in both VMAT plans.

In general, the differences in dose as a result of model manipulation were relatively consistent across the five plans examined, with exception to the tongue and groove, which had the greatest effect on VMAT plans. The other parameters examined, based on the distributions of community data, produced minimal changes in plan dose.

Interplay was examined for the MLC position offset, MLC transmission factor, leaf tip width, and tongue and groove using the IMRT9 plan. Figure 7 shows several tested relationships where the actual effect of combining parameter changes were compared against cases for which the average change in dose was simply summed based on the impact of each individual parameter (assuming they were independent). The results, shown in Figure 7, illustrate consistently that the two approaches (calculating interplay and simply adding effects assuming they are independent) yielded similar results. This is corroborated through linear regression and interaction modeling. Among the interaction terms considered, only the interaction between MLC transmission factor and leaf tip width [Fig. 7(c)] was found to be significant ($P < 0.001$); however, the adjusted coefficients of determination for the models with and without the interaction term included were very similar (R^2 equal to 0.998 vs 0.994, respectively), indicating that the magnitude of the contribution of this interaction effect to the overall change in dose is not large. All other examined parameter combinations demonstrate linear effects on dose ($P < 0.001$) with no significant interaction among the parameter pairs considered. That is, there was little if any additional compounding or suppressing of changes in average dose relative to the expected cases, demonstrating that these parameters had minimal interdependence over the range used clinically.

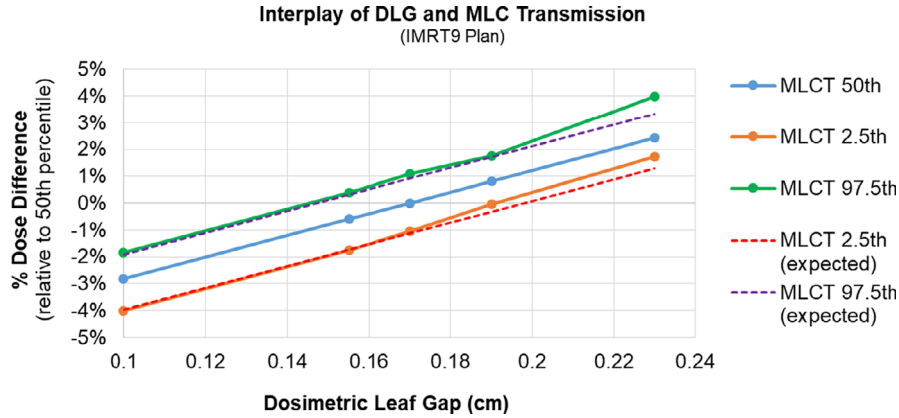

FIG. 4. Interplay between dosimetric leaf gap (DLG) and multileaf collimator transmission factor (MLCT). "DLG only" varies DLG while the MLCT remains at the 50th percentile value. MLCT 2.5th and MLCT97.5th illustrate varying DLG while the MLC transmission is set to those percentile values. "Expected" cases describe the change in dose if individual, average parameter effects were simply summed together. Note that this comparison is for the IMRT9 treatment plan [Color figure can be viewed at wileyonlinelibrary.com]

TABLE IV. Parameter changes implemented for the 6 MV Varian Clinac-type machine simulated in RayStation.

Percentile	Primary source size X [mm]	Primary source size Y [mm]	Tongue and groove [cm]	Leaf Tip width [cm]	MLC transmission	MLC position offset [cm]	MLC position gain	MLC position curvature [1/cm]
2.5th	0.04000	0.05000	0.01	0.177	0.007	0	0	0
25th	0.05350	0.05200	0.03	0.200	0.016	0.022	0	0
50th	0.05700	0.07000	0.04	0.320	0.018	0.040	0.0015	0
75th	0.10300	0.07250	0.05	0.360	0.022	0.040	0.0015	0.0008
97.5th	0.12345	0.10075	0.05	0.500	0.025	0.116	0.0150	0.0010

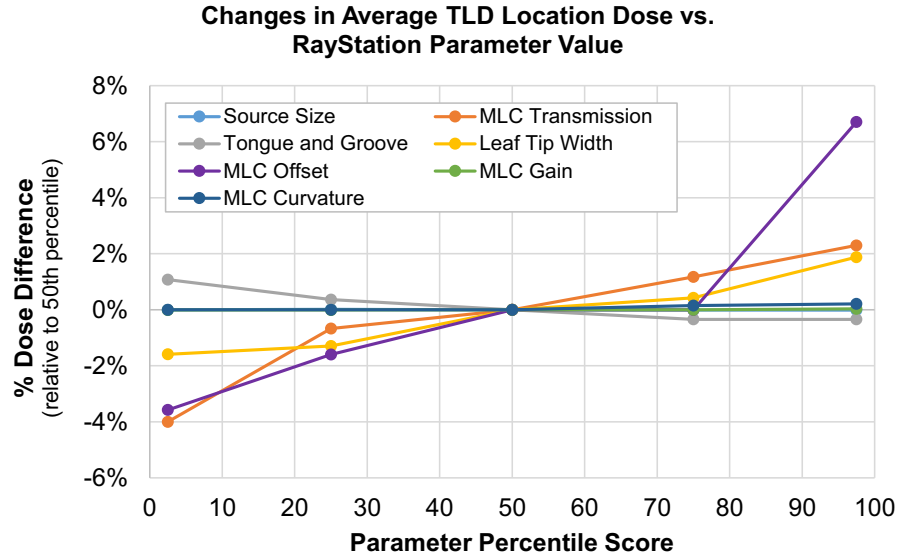

FIG. 5. Average changes in dose calculated to target thermoluminescent dosimeter locations for IROC-H head and neck phantom plans when parameters of interest are manipulated in RayStation. Percentile score corresponds to community reported values.[15] [Color figure can be viewed at wileyonlinelibrary.com]

3.C. Dosimetric parameters

Table V describes the numeric dosimetric modifications implemented in the beam modeling to achieve the variance observed in the community between modeled and true values.[16] Figure 8 shows the relative changes in average dose to TLD locations introduced by modifying the underlying beam model according to the percentile value for each parameter.

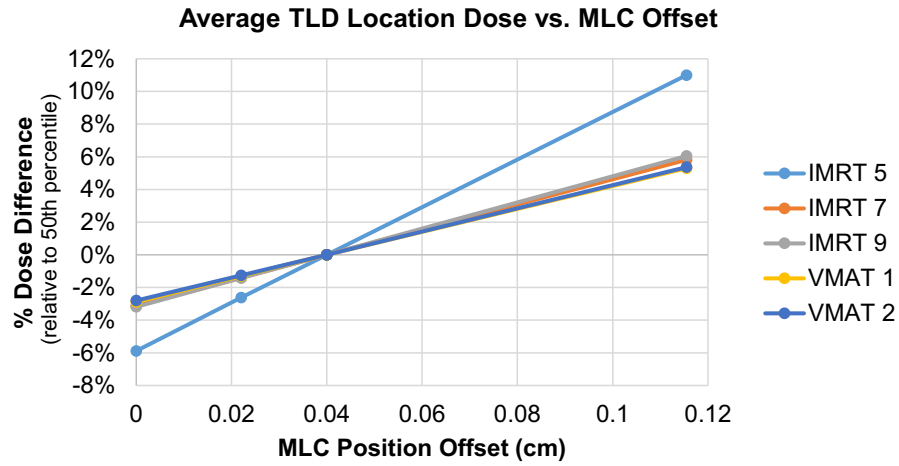

FIG. 6. Changes in average dose for the target thermoluminescent dosimeter locations calculated for each of the five IROC-H head and neck phantom plans following manipulation of the multileaf collimator (MLC) offset in RayStation. Points on the curve correspond to percentile scores from the radiation oncology community for a Varian Base Class machine with a Millennium 120 leaf MLC. [Color figure can be viewed at wileyonlinelibrary.com]

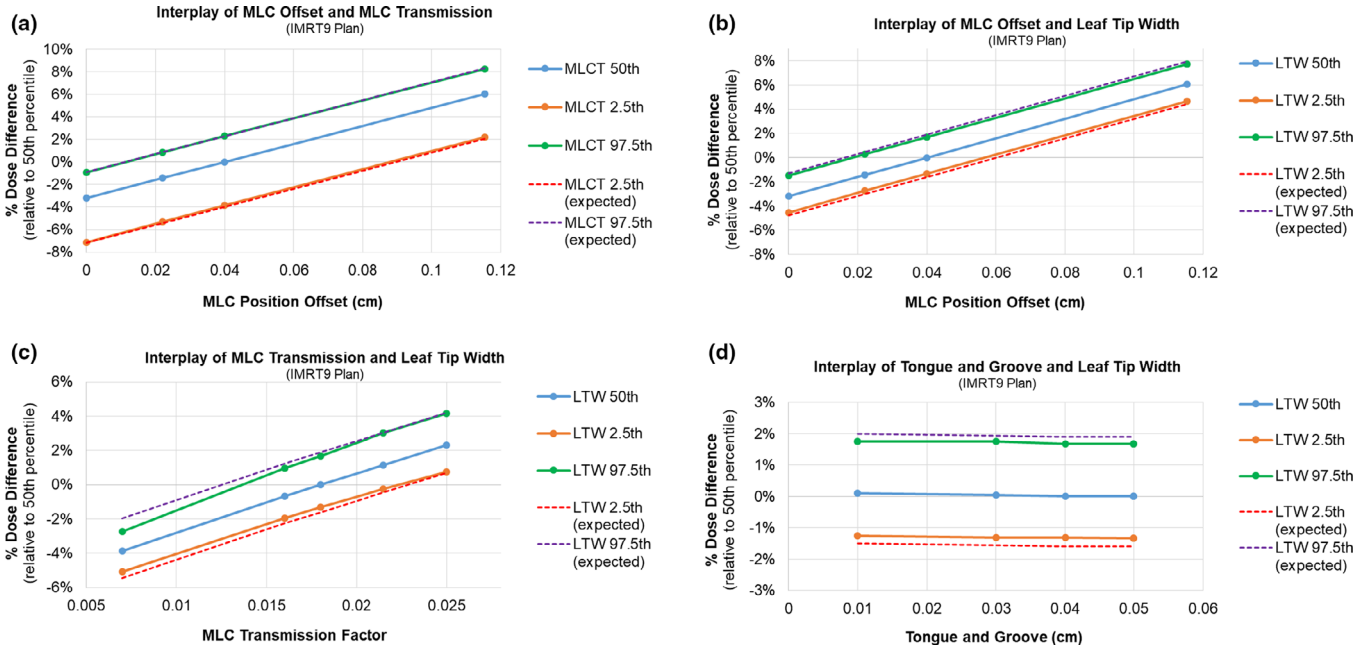

FIG. 7. Select cases demonstrating the extent of interplay between (a) multileaf collimator (MLC) offset (MLCO) and MLC transmission factor (MLCT), (b) MLC offset and leaf tip width (LTW), (c) MLC transmission factor and leaf tip width, and (d) tongue and groove (T&G) and leaf tip width in RayStation using the IMRT9 plan. Each value was evaluated at the 2.5th and 97.5th percentile and compared to the "expected" cases, which described the change in dose if individual parameter effects were simply added together. [Color figure can be viewed at wileyonlinelibrary.com]

The only parameter to introduce dose changes in the target TLD locations in excess of 1% was the PDD, which on average ranged from −1.1% to +1.2% with a maximum change of 2.5% in the OAR TLD locations. Unsurprisingly, changes in target TLD locations were less than the PDD errors introduced because the depths of the targets were <10 cm. The small field output factors had a wide range of errors as measured by IROC;[14] however, these errors did not manifest as a variation in TLD location dose in this study. This is because the output factors defined in Table V are defined by the jaw. The span across the primary and secondary targets is approximately 6–8 cm across, so no small jaw-defined fields are typically used in treating the phantom. Similarly, the off-axis factors have known errors as shown in Table V, but these did not materialize in the phantom cases. This is because the targets are relatively small and close to central axis.

Because PDD was the only factor to contribute plan changes greater than 1%, interplay effects among dosimetric characteristics were not assessed.

4. DISCUSSION

This study explored the dosimetric effects of changing common beam modeling parameters on clinically acceptable

H&N phantom treatment plans to understand how these changes may contribute to dose calculation accuracy or inaccuracies in the context of IMRT and VMAT planning. Small variations in MLC offset, as described by either the DLG in Eclipse or MLC leaf-tip offset in RayStation, can have substantial impacts on the resultant plan performance, affecting both the targets and OAR doses substantially. Other parameters modeling the MLC and radiation source characteristics, based on community-reported data, were impactful to a lesser degree.

The magnitudes of dose differences caused by variations in beam modeling are consistent with previous studies in Eclipse and RayStation. Like that reported here, McVicker, et al.[12] demonstrated changes in IROC-H H&N phantom target TLD dose on the order of 5% or more for changes in DLG greater than 1 mm while the MLC transmission factor more greatly affected OAR TLDs than those in the target structures. Additionally, Kielar, et al.[18] demonstrated the linear proportionality between error in defining the DLG and dose error, much like that shown in Fig. 2. For RayStation, Koger, et al.[11] also demonstrated the linear response of MLC position offset and observed, despite using a different machine setup, that additional offsets of 1 mm can produce changes in TLD dose in excess of 10%, much like Fig. 1. Likewise the leaf tip width was found to be a parameter of importance in accurate MLC modeling, generating TLD dose differences of up to 2% in the targets for the range of values examined.

At first glance, the results from this work may seem unsurprising; multiple previous studies have likewise noted a strong dependence of the dose calculation for modern treatments (IMRT and VMAT) with the MLC leaf position.[7,13,18-22] However, the current work is unique in that the dose differences calculated herein are based upon the actual variations in beam modeling adopted by the radiotherapy community, making the magnitude of the dose deviations particularly relevant to clinical practice. Moreover, these differences were evaluated in a reference geometry (the IROC-H H&N phantom) where the radiation oncology community is known to struggle in the clinical trial credentialing process. To maximize the breadth of impact, these evaluations were conducted using several common IMRT delivery methods. Additionally, this work examines interplay among parameters as a factor of IMRT performance, which has not been extensively investigated to date.

In a previous study, we reported that the beam modeling parameters that had the greatest spread among the community are those representing the MLC characteristics.[15] Interestingly, these very same parameters, namely the MLC transmission factor, DLG in Eclipse, and the MLC offset in RayStation, generated the most considerable dose changes among the treatment plans studied herein. But more importantly, these very same factors were also ones that could be theoretically measured, thus underscoring the need for physicists to be extremely cautious when assigning these values in their clinical TPS. In fact, current guidance for TPS commissioning is limited to several tests to validate the overall TPS performance; recommendations for individual parameter assignment are limited to the TPS vendor. It is solely up to the physicist to understand the intended effects of parameter assignment and consequently understand what values will generate the most robust model. For example, both Kim et al.[13] and Kielar et al.[18] determined that it was necessary to adjust the measured physical DLG values to reduce dose calculation errors for their system. To compound this, measured DLG values can be different based on the measurement settings (e.g., field size, depth, and ion chamber).[23] This work can help physicists to better understand how each of these parameters generally contributes to IMRT plan accuracy and the interplay among major parameters in order to make informed decisions in beam model commissioning.

What is important to keep in context is that deviations in dose calculation accuracy are generally very difficult to detect using conventional QA methods. Studies in detectability, such as that of Koger et al.,[11] McVicker et al.,[12] Nelms et al.,[7] and Kry et al.[10] point to the gross lack of sensitivity for traditional IMRT QA methods in identifying beam modeling inaccuracies. Only external validation through an independent phantom, such as that from the IROC-H phantom program, can sufficiently capture the extent of delivery errors caused by poor dose calculations. However, more work is needed to determine whether discrepancies in phantom dose distributions are connected with atypical beam modeling parameter selection.

This work is limited that it only focuses on one of the most common clinical systems: a Varian Clinic-type accelerator. It is possible that these results may differ from other clinical units with different physical configurations and geometries. Additionally, dose calculations were performed for the IROC-H H&N phantom, which is a simplified representation of human anatomy with little heterogeneity. Inherently, some dose effects may not have manifested through this choice of experimental setup. For instance, off-axis factor errors were likely not fully evident because the phantom is smaller than the dimension for which the greatest change in modeling was implemented (10 cm from central axis). Likewise, changes in dose due to PDD variations were less impactful, as the targets are located within 10 cm depth in the phantom. The use of a larger or more inhomogeneous phantom habitus (e.g., pelvis) may further reveal the greatest extent of changes caused by

TABLE V. Dosimetric parameter changes implemented for the 6 MV Varian Clinac-type machine simulated in RayStation.

Percentile	PDD (20 cm depth)	Small-field output factors (at 2x2 cm)	Off-axis factor (10 cm from central axis)
2.5th	−2.3%	−5.8%	−2.0%
25th	−0.7%	−0.5%	−0.4%
50th	0%	0%	0%
75th	+0.6%	+0.4%	+0.5%
97.5th	+2.5%	+2.3%	+1.6%

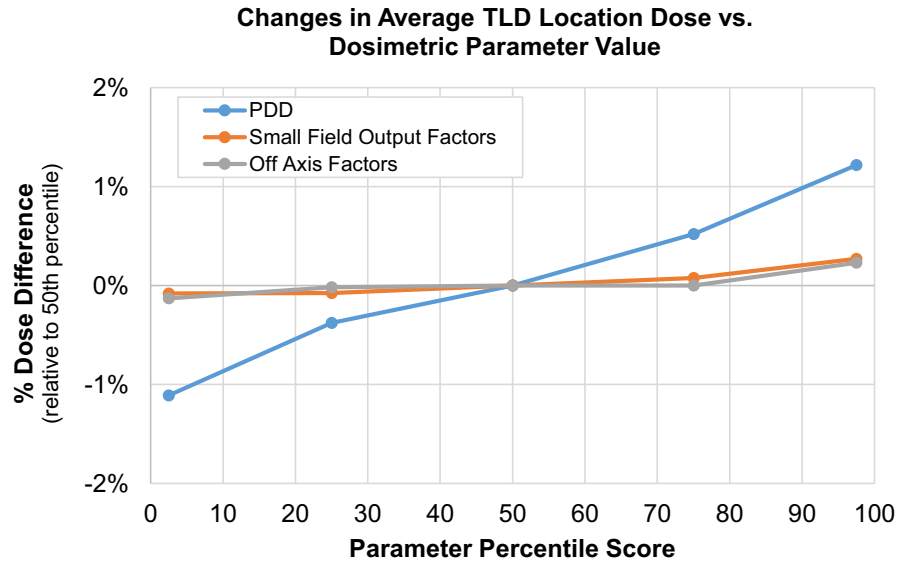

FIG. 8. Average changes in dose calculated to target thermoluminescent dosimeter locations for IROC-H head and neck phantom plans when dosimetric characteristics are manipulated in RayStation according to the documented variations in measurement accuracy from IROC-H site visit data.[14] [Color figure can be viewed at wileyonlinelibrary.com]

these parameter variations. Finally, this work's focus on average dose calculation outcome does not evaluate all aspects of dose calculation perturbation; changes in dose were not evaluated beyond the phantom target and OAR TLD locations. However, the linearity of such perturbations (Figs. 3 and 6) suggest that changes in dose occur throughout the phantom. It may be of value to investigate the potential effects of dose modeling variance on other phantom setups or geometries, where a variety treatment strategies may further highlight or unmask the true effects of modeling on clinical care.

It also is important to note that while this work intends to describe the range of dose calculation variations among the radiotherapy community, the results of this study do not imply that the use of atypical beam modeling parameters is totally inappropriate. This work is based upon average survey data and can only provide so much information regarding appropriateness for use. In fact, machine models that are intended for specific purposes (e.g., VMAT-dedicated units, radiosurgery units, etc.) may require use of parameter values that deviate from the norm.

5. CONCLUSIONS

In this study, several beam modeling parameters encompassing both basic dosimetry data (i.e., PDDs, output factors, and off-axis factors) and other nondosimetric modeling elements were assessed for their potential effects on IROC-H H&N phantom performance, based on the radiotherapy community's range of measured and reported values. Of interest, the parameters related to the modeling of the MLC, specifically the DLG for Eclipse and the MLC Offset for RayStation, demonstrated substantial impact with regards to dose to the target, corroborating well with previous works. By applying the most extreme parameter values used clinically by the radiotherapy community, differences from the baseline, average performance beam model produced clinically compromising dose calculation errors. This result implies that these parameters can have a substantial clinical impact on the overall development and accuracy of IMRT plans and are of the utmost importance to commission correctly.

The quality and accuracy of the TPS radiation beam model is essential to providing high quality treatments. It is clear that, despite fundamental differences in TPS modeling formalisms, that parameters defining the MLC provide the greatest challenge to commission correctly. Understanding the ways in which these parameters influence the resultant dose calculation, both individually and collectively, can assist both IROC-H and the radiotherapy community at large to adopt the most appropriate values for modeling and improve IMRT performance.

ACKNOWLEDGMENTS

This work was supported by Public Health Service Grants CA180803 and CA214526 awarded by the National Cancer Institute, United States Department of Health and Human Services. Mallory C. Glenn is supported by the Rosalie B. Hite Graduate Fellowship in Cancer Research and the American Legion Auxiliary Fellowship awarded by The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences. Christine B. Peterson is partially supported by NIH/NCI CCSG grant P30CA016672.

CONFLICT OF INTEREST

The authors have no conflict of interests to disclose.

[a]Author to whom correspondence should be addressed. Electronic mail: sfkry@mdanderson.org.

REFERENCES

1. Smilowitz JB, Das IJ, Feygelman V, et al. AAPM medical physics practice guideline 5.a.: commissioning and QA of treatment planning dose calculations - megavoltage photon and electron beams. *J Appl Clin Med Phys*. 2015;16:14–34.
2. Moran JM, Dempsey M, Eisbruch A, et al. Safety considerations for IMRT: executive summary. *Pract Radiat Oncol*. 2011;1:190–195.
3. Palta JR, Deye JA, Ibbott GS, et al. Credentialing of institutions for IMRT in clinical trials. *Int J Radiat Oncol Biol Phys*. 2004;59:1257–1259.
4. Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. *Med Phys*. 2013;40:022101.
5. Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. *Med Phys*. 2016;43:6491–6496.
6. Kerns JR, Stingo F, Followill DS, Howell RM, Melancon A, Kry SF. Treatment planning system calculation errors are present in most Imaging and Radiation Oncology Core-Houston phantom failures. *Int J Radiat Oncol Biol Phys*. 2017;98:1197–1203.
7. Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys*. 2011;38:1037–1044.
8. Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. *Med Phys*. 2010;37:2516–2524.
9. Nelms BE, Chan MF, Jarry G, et al. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Med Phys*. 2013;40:111722.
10. Kry SF, Molineu A, Kerns JR, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. *Int J Radiat Oncol Biol Phys*. 2014;90:1195–1201.
11. Koger B, Price R, Wang D, Toomeh D, Geneser S, Ford E. Impact of the MLC leaf-tip model in a commercial TPS: dose calculation limitations and IROC-H phantom failures. *J Appl Clin Med Phys*. 2020;21:82–88.
12. McVicker D, Yin F-F, Adamson JD. On the sensitivity of TG-119 and IROC credentialing to TPS commissioning errors. *J Appl Clin Med Phys*. 2016;17:34–48.
13. Kim J, Han JS, Hsia AT, Li S, Xu Z, Ryu S. Relationship between dosimetric leaf gap and dose calculation errors for high definition multi-leaf collimators in radiotherapy. *Phys Imaging Radiat Oncol*. 2018;5:31–36.
14. Kerns JR, Followill DS, Lowenstein J, et al. Agreement between institutional measurements and treatment planning system calculations for basic dosimetric parameters as measured by the Imaging and Radiation Oncology Core-Houston. *Int J Radiat Oncol Biol Phys*. 2016;95:1527–1534.
15. Glenn MC, Peterson CB, Followill DS, Howell RM, Pollard-Larkin JM, Kry SF. Reference dataset of users' photon beam modeling parameters for the Eclipse, Pinnacle, and RayStation treatment planning systems. *Med Phys*. 2020;47:282–288.
16. Kerns JR, Followill DS, Lowenstein J, et al. Technical report: reference photon dosimetry data for Varian accelerators based on IROC-Houston site visit data. *Med Phys*. 2016;43:2374–2386.
17. Glenn MC, Hernandez V, Saez J, et al. Treatment plan complexity does not predict IROC Houston anthropomorphic head and neck phantom performance. *Phys Med Biol*. 2018;63:205015.
18. Kielar KN, Mok E, Hsu A, Wang L, Luxton G. Verification of dosimetric accuracy on the TrueBeam STx: rounded leaf effect of the high definition MLC. *Med Phys*. 2012;39:6360–6371.
19. Rangel A, Dunscombe P. Tolerances on MLC leaf position accuracy for IMRT delivery with a dynamic MLC. *Med Phys*. 2009;36:3304–3309.
20. Luo W, Li J, Price RA, et al. Monte Carlo based IMRT dose verification using MLC log files and R/V outputs. *Med Phys*. 2006;33:2557–2564. https://doi.org/10.1118/1.2208916.
21. Lee J-W, Hong S, Kim Y-L, et al. Effects of static dosimetric leaf gap on MLC-based small beam dose distribution for intensity modulated radiosurgery. *J Appl Clin Med Phys*. 2007;8:2397.
22. Yao W, Farr JB. Determining the optimal dosimetric leaf gap setting for rounded leaf-end multileaf collimator systems by simple test fields. *J Appl Clin Med Phys*. 2015;16:65–77.
23. Wasbø E, Valen H. Dosimetric discrepancies caused by differing MLC parameters for dynamic IMRT. *Phys Med Biol*. 2008;53:405–415.