Predicting oropharyngeal tumor volume throughout the course of radiation therapy from pretreatment computed tomography data using general linear models

Adam D. Yock^{a)}

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Arvind Rao

Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and the Graduate School of Biomedical Sciences, the University of Texas Health Science Center at Houston, Houston, Texas 77030

Lei Dong

Scripps Proton Therapy Center, San Diego, California 92121 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Beth M. Beadle and Adam S. Garden

Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030

Rajat J. Kudchadker

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

Laurence E. Court

Department of Radiation Physics and Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 and The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030

(Received 5 December 2013; revised 17 March 2014; accepted for publication 18 March 2014; published 15 April 2014)

Purpose: The purpose of this work was to develop and evaluate the accuracy of several predictive models of variation in tumor volume throughout the course of radiation therapy.

Methods: Nineteen patients with oropharyngeal cancers were imaged daily with CT-on-rails for image-guided alignment per an institutional protocol. The daily volumes of 35 tumors in these 19 patients were determined and used to generate (1) a linear model in which tumor volume changed at a constant rate, (2) a general linear model that utilized the power fit relationship between the daily and initial tumor volumes, and (3) a functional general linear model that identified and exploited the primary modes of variation between time series describing the changing tumor volumes. Primary and nodal tumor volumes were examined separately. The accuracy of these models in predicting daily tumor volumes were compared with those of static and linear reference models using leave-one-out cross-validation.

Results: In predicting the daily volume of primary tumors, the general linear model and the functional general linear model were more accurate than the static reference model by 9.9% (range: −11.6%–23.8%) and 14.6% (range: −7.3%–27.5%), respectively, and were more accurate than the linear reference model by 14.2% (range: −6.8%–40.3%) and 13.1% (range: −1.5%–52.5%), respectively. In predicting the daily volume of nodal tumors, only the 14.4% (range: −11.1%–20.5%) improvement in accuracy of the functional general linear model compared to the static reference model was statistically significant.

Conclusions: A general linear model and a functional general linear model trained on data from a small population of patients can predict the primary tumor volume throughout the course of radiation therapy with greater accuracy than standard reference models. These more accurate models may increase the prognostic value of information about the tumor garnered from pretreatment computed tomography images and facilitate improved treatment management. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4870437>]

Key words: tumor volume, general linear model, anatomical models, image analysis, modeling

1. INTRODUCTION

Modern intensity-modulated radiation therapy achieves highly conformal dose distributions that deliver a therapeutic dose to the target with minimal dose to neighboring healthy tissue. This conformality finds particular utility in head and neck radiation therapy, where the anatomy is complex and many organs-at-risk are in close proximity to the disease. Unfortunately, highly conformal dose distributions are also sensitive to variations in patient setup and anatomy, which may compromise the integrity of the treatment.

A significant departure of the delivered dose distribution from the planned dose distribution may result from anatomic changes that progress throughout the course of treatment. This discrepancy between the planned and delivered doses is particularly significant when the gross tumor volume (GTV) changes. Two examples of anatomic changes of the GTV that trend throughout the course of treatment are the decrease in its volume and the medial shift of its centroid.¹ Although changes in the GTV can, along with changes in surrounding normal tissue, compromise the precision of the delivered radiation dose, they are often considered encouraging signs of the patient's response to treatment and a proxy for clinical outcomes. The initial GTV volume (that is, the initial numeric volume encompassed by the GTV contour) and changes in the GTV volume have prognostic value regarding clinical outcomes for patients with cancers at a number of head and neck sites.²⁻⁹

As the patient's anatomy departs from that depicted in the pretreatment simulation computed tomography (CT) images used for treatment planning, the delivered dose distribution resembles the planned dose distribution less and less. Adaptive radiation therapy (ART) has emerged as a treatment strategy in which patients are reimaged and the treatments are replanned during the course of therapy to compensate for anatomic changes.¹⁰ In principle, the delivered dose distribution would remain very conformal if the patient's anatomy was imaged at each treatment fraction and a new fraction-specific plan was optimized to that anatomy. However, this workflow is prohibitively labor intensive, so current implementations of ART are generally limited to one or two replans. Each plan is delivered without adjustment until the next replan or the end of treatment. When limited to only a few replans, the number and timing of these adjustments are also important considerations.^{11,12}

Although the process of frequent reimaging and replanning is not clinically feasible, it may be possible to maintain dose conformality by anticipating trending anatomic changes and performing replanning accordingly. While it is impossible to predict with certainty any particular patient's response to treatment, statistical models have the potential to anticipate the changing volume of the GTV throughout the entire course of treatment. Implementation of these models could assist in identifying candidates who would benefit most from ART, suggest the appropriate number and timing of replans, or be used to estimate anatomic changes between adaptive adjustments.

Presented here are three experimental models designed to predict the changing volume of the GTV throughout the

course of radiation therapy. The first describes a linear change at a constant rate. The second uses the relationship between the volume of the GTV at a particular treatment fraction and its initial volume. Finally, the third uses the primary modes of variation between time series describing the changing GTV volumes. The predictive accuracy of each model was evaluated and compared with static and linear reference models.

2. METHODS

Three models were generated to predict the changing volume of the GTV throughout the entire course of radiation therapy. Parameters of each model were determined from a training set of patient data and applied to a test set of patient data by means of leave-one-out cross-validation. The predictive accuracy of each model was then compared with that of two reference models.

2.A. Patient population and target volumes

Nineteen patients were retrospectively selected randomly from an institutional ART protocol described by Schwartz *et al.*¹² Patient information is provided in Table I. Selection criteria for this protocol included stage III, IVa, or IVb squamous cell carcinoma of the oropharynx as defined by the American Joint Committee on Cancer and an Eastern Cooperative Oncology Group performance status of 0–2. All patients were treated with similar prescription doses and fractionation schemes—either 69.96 Gy in 33 fractions of 2.12 Gy, or 70.00 Gy in 35 fractions of 2.00 Gy. Each patient exhibited anatomic changes that warranted at least one reoptimized treatment plan over the course of treatment. In accordance with the protocol, patients were imaged prior to the delivery of each treatment fraction using a CT-on-rails system. This system shares a single patient couch between a 2100 EX linear accelerator (Varian Medical Systems, Palo Alto, CA) and a SmartGantry CT scanner (GE Healthcare, Waukesha, WI).¹³ The “daily” CT images were acquired for image-guided patient alignment using the C2 vertebral body as the alignment target.

For these 19 patients, an aggregate of 35 GTVs that had been delineated on the planning CT images by physicians were identified as either a primary GTV ($n = 17$) or a nodal GTV ($n = 18$). Primary and nodal GTVs were all intended to receive the prescription dose. Each GTV was considered an independent region of interest, ignoring the fact that some originated from the same individuals.

2.B. Deformable image registration

Deformable image registration was used to facilitate the generation of daily GTV contours. Each patient's daily CT images were registered to the appropriate planning CT image using inhouse software, which was based on the demons algorithm by Thirion¹⁴ and improved by Wang *et al.*¹⁵ When applied to head and neck anatomy deformed according to a known mathematical transformation, this algorithm resulted

TABLE I. Patient information.

Patient	Age (years)	Disease site	Disease stage ^a	Concurrent systemic therapy
1	62	Base of tongue	T1 N2A	Cisplatin
2	41	Tonsil	T3 N2B	Cisplatin & Carboplatin
3	53	Tonsil	T2 N2C	Cetuximab
4	56	Base of tongue	T3 N2C	Cisplatin, Carboplatin, & Paclitaxel
5	57	Base of tongue	T3 N2A	Cisplatin
6	58	Base of tongue	T4 N2B	Cisplatin & Cetuximab
7	50	Base of tongue	T2 N2B	Cetuximab
8	49	Tonsil	T1 N2B	Cisplatin
9	42	Base of tongue	T2 N2A	Cisplatin
10	54	Base of tongue	T4 N0	Cisplatin
11	62	Tonsil	T4 N0	Cetuximab
12	56	Base of tongue	T2 N2B	Cisplatin
13	50	Base of tongue	T2 N2A	Cisplatin
14	51	Tonsil	T3 N2B	Cisplatin
15	50	Base of tongue	T4 N2B	Cisplatin
16	38	Tonsil	T1 N2B	Cisplatin
17	68	Tonsil	T2 N2B	Cisplatin
18	69	Tonsil	T3 N1	Cisplatin
19	44	Base of tongue	T2 N2B	Cetuximab

^aNo patient had metastatic disease.

in 99% of errors being of magnitude less than 2 mm, with a mean error of 0.2 mm.¹⁵ The deformation vector fields resulting from the registration were used to propagate the GTV contours from the planning CT images onto the daily images. After propagation, each of the deformed contours was individually reviewed by one of the authors (A.D.Y.). Any residual contour inaccuracies were considered acceptable for our analysis. For data analysis, only the first 32 treatment fractions were considered due to the unavailability of the final daily images for several of the patients.

2.C. GTV volume models

Three experimental models were generated from training sets of daily GTV volumes in order to predict the volume of a test GTV at each treatment fraction during radiation therapy.

2.C.1. Linear_{Median} model

The first and most simple of the experimental models was a linear model that assumed a constant rate of change in the GTV volume over time. This model predicted the volume of the test GTV at a given treatment fraction t ($V_{t,\text{test}}$) based on its initial volume ($V_{0,\text{test}}$) and a single parameter (β) describing the rate of change [Eq. (1)]. This predicted rate of change was that which resulted in a total change in volume equal to the median relative change observed at the end of treatment in a set of training GTVs [Eq. (2)]. This model is referred to as the linear_{Median} model,

$$V_{t,\text{test}} = V_{0,\text{test}}(1 + \beta t), \quad (1)$$

$$\beta = \text{median} \left(\frac{V_{32,\text{train}} - V_{0,\text{train}}}{V_{0,\text{train}}} \right) \left(\frac{1}{32} \right). \quad (2)$$

2.C.2. General linear model (GLM)

The second model is a GLM. We observed that the GTV volume at a given treatment fraction t was related to the initial GTV volume through a power fit (data not shown). Furthermore, the exponent of the fit was effectively described as a quadratic function of t . These relationships differed visibly between primary and nodal GTVs. As a result, two separate GLMs were considered, one for primary GTVs and one for nodal GTVs, both of the form of Eq. (3),

$$V_{t,\text{test}} = \beta_0 V_{0,\text{test}}^{(\beta_1 + \beta_2 + \beta_3 t^2)}, \quad (3)$$

where, again, t and $V_{0,\text{test}}$ denote the treatment fraction and the initial volume of the test GTV, respectively. The four model parameters are denoted by β_n . Equations of the form of Eq. (3) can be linearized by taking the logarithm of both sides

$$\begin{aligned} \log_{10}(V_{t,\text{test}}) &= \log_{10}(\beta_0) + \beta_1 \log_{10}(V_{0,\text{test}}) \\ &+ \beta_2 t \log_{10}(V_{0,\text{test}}) + \beta_3 t^2 \log_{10}(V_{0,\text{test}}). \end{aligned} \quad (4)$$

The parameters of Eq. (4) were fit to a set of training GTVs according to the leave-one-out cross-validation procedure (described below in Sec. 2.D) using ordinary least squares optimization. These were then used to find the parameters of Eq. (3). The result was an empirical relationship for the volume of the GTV at any treatment fraction given the target's initial volume and identification as a primary or nodal GTV.

2.C.3. Functional general linear model (functional GLM)

The third experimental model focused on the changing GTV volume as a continuous function. Model parameters were determined by operating on the GTV volume function

as a whole rather than considering daily data points individually. This model was thus considered a functional GLM.

To determine the functional GLM, we first smoothed the raw volume data for each GTV using cubic splines. The smoothing parameter to be included in the spline fitting objective function was qualitatively selected to emphasize a smooth fit and remained constant across all GTVs. Next, a feature vector describing the volume function was created by sampling the function at the treatment fraction time points. This feature vector was defined as the coordinates of a single point in 33-dimensional space. The dimensionality of this space corresponded to the number of sample volumes per GTV (that is, volumes at each of the 32 treatment fractions plus one from the pretreatment CT image).

Principal component analysis (PCA) was performed on a training set of these 33-dimensional data points to determine the primary modes of variation between the GTV volume functions. PCA involves the Eigen-decomposition of the covariance matrix of a set of data vectors, here being the GTV volume function feature vectors. The result of PCA is a set of Eigenvectors paired with corresponding Eigenvalues. The Eigenvectors are the principal components (PCs) describing the different modes of variation within the data. The set of PCs represents an alternative orthogonal basis describing the high-dimensional data. The Eigenvalues are the variances of the dataset along each PC. It is common practice to use the Eigenvalues to identify the subset of PCs that describes a certain proportion of the total variance of the data. For the purposes of data smoothing and/or dimensionality reduction, the user may set a desired proportion of the total variance as a threshold, retain the minimum number of PCs required to exceed that threshold, and discard the rest. Projecting the original data onto a basis consisting of this subset of PCs allows the user to simplify the problem at hand while retaining the desired amount of information.

The number of PCs necessary to account for at least 90% of the total variance in the training GTV volume function data was determined. These PC coordinates were plotted against the initial volume of the GTV and fit with either a linear, polynomial, or logarithmic regression function as appropriate. The regression functions were then used to find the coordinate of each PC that corresponded to the initial volume of a test GTV. If there was no relationship between a particular PC and the initial volumes of the training GTVs, then the mean training PC coordinate, necessarily equal to zero as part of the analysis, was considered the PC coordinate of the test GTV. The PC coordinates were subsequently transformed back to the feature vector space, resulting in a function that predicted the test GTV volume at each treatment fraction.

2.C.4. Static and linear_{-1.8%} reference models

Two reference models were generated and compared with the linear_{Median} model, the GLM, and the functional GLM. The first reference model was a static model in which the predicted GTV volume at each treatment fraction was simply the initial volume of the test GTV. The second model assumed a constant decrease of 1.8% of the initial GTV volume per

day, as described by Barker *et al.*¹ This model is referred to as the linear_{-1.8%} model to distinguish it from the linear_{Median} model.

2.D. Evaluation of prediction accuracy with leave-one-out cross-validation

The prediction accuracy of each of the five models was evaluated using leave-one-out cross-validation, in which one of the GTVs was left out as a test dataset, while the remaining GTVs (16 primary GTVs or 17 nodal GTVs) were used as training datasets to derive values for the model parameters. The models generated predictions of the test GTV volume at each treatment fraction. Prediction error was quantified as the root mean squared error (RMSE) between the predicted volume and the true volume of the test GTV at each fraction. This leave-one-out process was conducted a total of 17 times and 18 times for primary GTVs and nodal GTVs, respectively, so that each GTV served as the test dataset. In the end, the accuracy of each model was reflected by a set of RMSE values, one for each instance of the leave-one-out protocol.

The linear_{Median} model, GLM, and functional GLM were compared to the static and linear_{-1.8%} reference models using a two-sided sign test to determine whether or not there was a statistically significant difference ($p < 0.050$) in the prediction accuracy of each model. The sign test was selected because it is a nonparametric, paired test that does not assume normality or that the distribution of differences is symmetric around the median.

3. RESULTS

Initial volumes of the primary and nodal GTVs are depicted in Fig. 1. Model parameters for the linear_{Median} model, GLM, and functional GLM are listed in Table II. The parameters for the linear_{Median} model and the GLM predicted the volume of the test GTV via Eqs. (1) and (3).

In generating the functional GLM, two and three PCs were necessary to exceed a 90% variance threshold for the primary and nodal GTVs, respectively. However, only the PC accounting for the largest variance of the primary GTV data was observed to relate to the initial GTV volume. A logarithmic regression function was used for this PC, while test coordinates of the remaining PCs were set to zero.

FIG. 1. Initial volumes of the primary and nodal GTVs. Boxes range from the 25th to 75th percentiles. Whiskers denote the 10th and 90th percentiles.

TABLE II. Experimental model parameters.

Experimental model	Parameter	Primary GTVs (1 Standard Deviation)	Nodal GTVs (1 Standard Deviation)	Units
Linear _{Median}	β	0.0050 (0.0004)	0.0126 (0.0009)	day ⁻¹
GLM	β_0	1.4065 (0.0181)	0.7691 (0.0182)	...
	β_1	0.9146 (0.0037)	1.0882 (0.0083)	...
	β_2	-0.0071 (0.0003)	-0.0002 (0.0006)	day ⁻¹
	β_3	0.0001 (0.0000)	-0.0002 (0.0000)	day ⁻²
Functional GLM ^a	β_0	290.3788 (7.9604)	...	cm ³
	β_1	-200.0491 (5.1636)	...	cm ³

^aModel parameters refer to the first PC describing changes in primary GTVs as it was the only one to vary as a function of initial GTV volume.

To calculate the predicted primary GTV volume at each treatment fraction according to the functional GLM, the coordinate of the first PC ($PC1_{\text{pri}}$) corresponding to the initial volume of the test GTV ($V_{0,\text{test}}$) was calculated using Eq. (5). The resulting coordinate was multiplied by fraction-specific loading values (l_t , Fig. 2) that represent the direction of the first PC in the original feature vector space. The product was added to the fraction-specific average volume of the training GTVs ($\bar{V}_{t,\text{train}}$) depicted in Fig. 3(a) [Eq. (6)]. The predicted volumes of the nodal GTVs were equal to $\bar{V}_{t,\text{train}}$ since the PC scores were equal to zero,

$$PC1_{\text{pri}} = \beta_0 \log_{10}(V_{0,\text{test}}) + \beta_1, \quad (5)$$

$$V_{t,\text{test}} = \bar{V}_{t,\text{train}} + l_t PC1_{\text{pri}}. \quad (6)$$

The relative prediction error (normalized to the initial GTV volume) for the three experimental and two reference models are depicted in Figs. 4(a) and 4(b) for the primary and nodal GTVs, respectively. The median changes in prediction error provided by each model are presented in Table III. Differences in the prediction error of the two reference models were not statistically significant for either the primary or nodal GTVs.

The three primary GTV experimental models resulted in significant decreases in prediction error compared to both reference models in all cases except the comparison between the linear_{Median} model and the linear_{-1.8%} model. The largest difference among these comparisons (14.6% of the initial pri-

mary GTV volume) was observed between the functional GLM and the static models. For the nodal GTV experimental models, only the decrease in prediction error provided by the functional GLM compared to that of the static model (14.4%) was statistically significant.

Among the primary GTV experimental models, the GLM and functional GLM both resulted in significantly lower prediction errors than the linear_{Median} model (5.8% and 6.2%, respectively), but they were not statistically different from each other. Conversely, for the nodal GTV experimental models, neither the GLM nor the functional GLM provided a significant decrease in prediction error compared to the linear_{Median} model; however, the functional GLM decreased the prediction error compared to the GLM by a small but statistically significant 1.4%.

FIG. 2. Loading values for the first PC of the primary GTVs. Error bars represent one standard deviation of the loading values observed during each iteration of the leave-one-out cross-validation.

FIG. 3. Mean normalized volumes of the primary (a) and nodal (b) training GTVs. Error bars represent one standard deviation of the average values observed during each iteration of the leave-one-out cross-validation.

FIG. 4. Primary (a) and nodal (b) GTV prediction error relative to the initial GTV volume according to model. Boxes range from the 25th to 75th percentiles. Whiskers denote the 10th and 90th percentiles.

4. DISCUSSION

Presented here are three experimental models that predict the volume of the GTV throughout the course of radiation therapy with accuracy greater than that provided by two reference models. All five models provide volume estimates

TABLE III. Median difference in prediction error as a percentage of initial test GTV volume.

Model A	Model B	Median Δ RMSE ^a (<i>p</i> -value)	
		Primary GTVs	Nodal GTVs
Linear _{-1.8%}	Static	7.8 (1.000)	− 8.2 (0.238)
Linear _{Median}	Static	− 7.2 (0.049)	− 13.0 (0.238)
GLM	Static	− 9.9 (0.049)	− 10.1 (0.096)
Functional GLM	Static	− 14.6 (0.013)	− 14.4 (0.031)
Linear _{Median}	Linear _{-1.8%}	− 14.5 (0.332)	− 5.6 (0.096)
GLM	Linear _{-1.8%}	− 14.2 (0.002)	− 3.7 (0.481)
Functional GLM	Linear _{-1.8%}	− 13.1 (0.002)	− 7.8 (0.096)
GLM	Linear _{Median}	− 5.8 (0.013)	1.6 (0.815)
Functional GLM	Linear _{Median}	− 6.2 (0.013)	− 2.1 (0.096)
Functional GLM	GLM	− 1.2 (0.332)	− 1.4 (0.000)

^a Δ RMSE values were determined by subtracting the Model B values from the Model A values.

for each treatment fraction based only on the initial GTV volume and identification of the volume as a primary or a nodal GTV. This information is readily available from the pretreatment CT image, possibly improving its prognostic value.

The potential of information garnered from observing tumor dimensions has led many to investigate the correlation between parameters such as the initial tumor volume and the clinical outcome. These correlations have been observed at multiple head and neck disease sites but have been most extensively demonstrated for tumors of the larynx.^{2,6-9,16} The literature regarding the prognostic value of the initial volume of oropharyngeal tumors is less consistent.

Several authors have found little or no correlation between the initial oropharyngeal tumor volume and clinical outcomes. Mendenhall *et al.*² conducted multivariate analysis of a number of predictors applied to glottic, supraglottic, oropharyngeal, and hypopharyngeal disease. While tumor volume was observed to influence local control rates for glottic and supraglottic disease, for tumors of the oropharynx and hypopharynx, this effect was not observed. The results of Nathu *et al.*¹⁷ corroborated those of Mendenhall *et al.*, suggesting initial tumor volume was not predictive of local control in the oropharynx. Similarly, in their institutional review, Been *et al.*¹⁸ found no correlation between the rate of locoregional failure and the initial volume of tumors in the oropharynx. According to Hermans *et al.*,¹⁹ total tumor volume was not significantly correlated with the rate of locoregional control, and primary tumor volume carried marginal value in predicting local control when stratified by volume quartile but not when compared within each T stage. While Hermans *et al.* observed a correlation between nodal tumor volumes and the rate of regional control, Mishra *et al.*²⁰ observed none.

On the contrary, other authors have reported significant correlations between the initial oropharyngeal tumor volume and clinical outcomes. Chao *et al.*³ found that the initial volumes of the primary and nodal GTVs were independent risk factors for disease-free survival and the achievement of locoregional control, while the initial volume of the primary GTV (but not the nodal GTV) was predictive of distant metastasis-free survival. Lok *et al.*⁴ found an association between the initial primary tumor volume with the rate of overall survival, local failure, and distant metastatic failure. However, like Chao *et al.*, Lok *et al.* found the prognostic value of the initial nodal GTV volume to be limited. The most accurate predictor of local control, according to Studer *et al.*,⁵ was the initial volume of the primary GTV, although the total GTV volume was a more powerful predictor of nodal and distant control, as well as disease-free survival.

That initial tumor volume appears less predictive of clinical outcomes for patients with cancers of the oropharynx than for those with cancers at other head and neck sites has been ascribed to the greater variability of cancers of the oropharynx. This wide variability may in part be due to inherent variations in the radiosensitivity of oropharyngeal tumors. Mendenhall *et al.*² noted that within a particular stage, malignancies located at the base of tongue have a greater probability of local control than those of the oral tongue. Similarly, local control of cancers of the tonsillar fossa/posterior pillar is more likely

than that of tumors in the anterior tonsillar pillar.²¹ In addition, it has become clear that human papillomavirus is of great significance in head and neck cancer, influencing the patient's risk profile and clinical outcome.²² Human papillomavirus infection status and smoking history were both more critical in a recursive partitioning analysis-derived risk model than T and N staging,²² which are more size-oriented for oropharyngeal tumors than for laryngeal tumors where staging focuses on disease invasion. While it may provide additional value in predicting patient-specific treatment outcomes, human papillomavirus infection status was not considered in the reports cited above. Treatment modality and treatment technique also appear pertinent. Been *et al.*¹⁸ and Studer *et al.*⁵ observed that in the works cited above, those that did not observe a correlation between the initial oropharyngeal tumor volume and clinical outcomes had used conventional external beam therapy, while those that did observe a correlation had used intensity-modulated radiation therapy.

The work presented here did not consider clinical outcomes but focused on the models' ability to predict the changing tumor volume. Anticipating the dynamic response of the tumor from the initial tumor volume may strengthen the prognostic value of the pretreatment CT images because, according to Lee *et al.*,²³ the rate of volume reduction is of greater predictive power than the initial volume alone. Additional prognostic value may be derived by incorporating additional clinical or dosimetric parameters into a similar model. We limited our analysis to tumor volume in order to avoid additional complexity warranting a patient dataset larger than that available. An increased patient dataset would also improve the statistical power of our analysis. As a guide, the statistical power of the tests between the static and functional GLM models for the primary and nodal GTVs were calculated to be 83% and 72%, respectively. To achieve power for the nodal GTV comparison of greater than 80% (a common benchmark) would have required a sample size of at least 21 nodal GTVs.

Direct patient measurements acquired midtreatment are more accurate than predictive models in describing a patient-specific response to treatment; however, the models provided here have the potential to spare patients the burden of additional imaging exams. Midthreatment images could also be acquired in conjunction with the application of the described models in order to evaluate the applicability of the models to an individual patient. Alternatively, predictive models may assist clinicians in determining the appropriate timing of subsequent imaging exams to optimize adaptive adjustments to therapy.

It is also important to note that the experimental models described here are purely statistical and are not based on any underlying radiobiological mechanism. Incorporating radiobiological considerations, particularly if patient-specific values can be discerned, might lead to even more accurate predictions of GTV volume change and clinical outcome. Chvetsov *et al.* have used two-compartment²⁴ and four-compartment²⁵ models to describe the observed change in GTV volume by modeling tumors as populations of cells that are either oxygenated or hypoxic and either viable or lethally damaged.

With these models, those investigators were able to derive radiobiological parameter estimates similar to those of other clinical studies.

The potential to predict tumor volume using models similar to those presented here is not limited to tumors of the oropharynx; the described workflow may be applied more generally. However, the models for other tumor sites need to be generated based on the appropriate training patient data, and the predictive accuracy and clinical utility of the resulting models may vary.

5. CONCLUSION

Statistical models have demonstrated the ability to anticipate changes in tumor volumes throughout the course of radiation therapy. These models extend the use and applicability of pre- and midtreatment imaging examinations. Understanding the dynamic manner in which head and neck tumors and anatomy respond to radiation provides information beyond that available in the patient-specific CT images. This new information has the potential to improve treatment management by identifying ART candidates, suggesting the appropriate number and timing of adaptive adjustments, and estimating the extent of anatomic changes throughout treatment.

ACKNOWLEDGMENTS

The authors thank Jinzhong Yang, Ph.D., and Joy Zhang, Ph.D., for computational support of this work and Kathryn Carnes for editing this paper. The authors also acknowledge financial support from the American Legion Auxiliary and the University of Texas Graduate School of Biomedical Sciences at Houston.

^aE-mail: ADYock@mdanderson.org

¹J. L. Barker, A. S. Garden, K. K. Ang, J. C. O'Daniel, H. Wang, L. E. Court, W. H. Morrison, D. I. Rosenthal, K. S. C. Chao, S. L. Tucker, R. Mohan, and L. Dong, "Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system," *Int. J. Radiat. Oncol., Biol., Phys.* **59**, 960–970 (2004).

²W. M. Mendenhall, C. G. Morris, R. J. Amdur, R. W. Hinerman, and A. A. Mancuso, "Parameters that predict local control after definitive radiotherapy for squamous cell carcinoma of the head and neck," *Head Neck-J. Sci. Spec.* **25**, 535–542 (2003).

³K. S. C. Chao, G. Ozyigit, A. I. Blanco, W. L. Thorstad, J. O. Deasy, B. H. Haughey, G. J. Spector, and D. G. Sessions, "Intensity-modulated radiation therapy for oropharyngeal carcinoma: Impact of tumor volume," *Int. J. Radiat. Oncol., Biol., Phys.* **59**, 43–50 (2004).

⁴B. H. Lok, J. Setton, N. Caria, J. Romanyshyn, S. L. Wolden, M. J. Zelefsky, J. Park, N. Rowan, E. J. Sherman, M. G. Fury, A. Ho, D. G. Pfister, R. J. Wong, J. P. Shah, D. H. Kraus, Z. G. Zhang, K. D. Schupak, D. Y. Gelblum, S. D. Rao, and N. Y. Lee, "Intensity-modulated radiation therapy in oropharyngeal carcinoma: Effect of tumor volume on clinical outcomes," *Int. J. Radiat. Oncol., Biol., Phys.* **82**, 1851–1857 (2012).

⁵G. Studer, U. M. Lutolf, M. El-Bassiouni, V. Rousson, and C. Glanzmann, "Volumetric staging (VS) is superior to TNM and AJCC staging in predicting outcome of head and neck cancer treated with IMRT," *Acta Oncol.* **46**, 386–394 (2007).

⁶I. Doweck, D. Denys, and T. Robbins, "Tumor volume predicts outcome for advanced head and neck cancer treated with targeted chemoradiotherapy," *Laryngoscope* **112**, 1742–1749 (2002).

⁷D. E. Freeman, A. A. Mancuso, J. T. Parsons, W. M. Mendenhall, and R. R. Million, "Irradiation alone for supraglottic larynx-carcinoma—Can

- CT findings predict treatment results?," *Int. J. Radiat. Oncol., Biol., Phys.* **19**, 485–490 (1990).
- ⁸A. A. Mancuso, S. K. Mukherji, I. Schmalfuss, W. Mendenhall, J. Parsons, F. Pameijer, R. Hermans, and P. Kubilis, "Preradiotherapy computed tomography as a predictor of local control in supraglottic carcinoma," *J. Clin. Oncol.* **17**, 631–637 (1999).
- ⁹F. A. Pameijer, A. A. Mancuso, W. M. Mendenhall, J. T. Parsons, S. K. Mukherji, R. Hermans, and P. S. Kubilis, "Evaluation of pretreatment computed tomography as predictor of local control in T1/T2 pyriform sinus carcinoma treated with definitive radiotherapy," *Head Neck-J. Sci. Spec.* **20**, 159–168 (1998).
- ¹⁰D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Phys. Med. Biol.* **42**, 123–132 (1997).
- ¹¹D. L. Schwartz and L. Dong, "Adaptive radiation therapy for head and neck cancer—Can an old goal evolve into a new standard?" *J. Oncol.* **2011**, 690595.
- ¹²D. L. Schwartz, A. S. Garden, J. Thomas, Y. P. Chen, Y. B. Zhang, J. Lewin, M. S. Chambers, and L. Dong, "Adaptive radiotherapy for head-and-neck cancer: Initial clinical outcomes from a prospective trial," *Int. J. Radiat. Oncol., Biol., Phys.* **83**, 986–993 (2012).
- ¹³L. Court, I. Rosen, R. Mohan, and L. Dong, "Evaluation of mechanical precision and alignment uncertainties for an integrated CT/LINAC system," *Med. Phys.* **30**, 1198–1210 (2003).
- ¹⁴J. P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Imag. Anal.* **2**, 243–260 (1998).
- ¹⁵H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.* **50**, 2887–2905 (2005).
- ¹⁶R. W. Gilbert, D. Birt, H. Shulman, J. Freeman, D. Jenkin, R. Mackenzie, and C. Smith, "Correlation of tumor volume with local-control in laryngeal carcinoma treated by radiotherapy," *Ann. Oto. Rhinol. Laryn.* **96**, 514–518 (1987).
- ¹⁷R. M. Nathu, A. A. Mancuso, T. C. Zhu, and W. M. Mendenhall, "The impact of primary tumor volume on local control for oropharyngeal squamous cell carcinoma treated with radiotherapy," *Head Neck-J. Sci. Spec.* **22**, 1–5 (2000).
- ¹⁸M. J. Been, J. Watkins, R. M. Manz, L. R. Gentry, G. E. Levenson, P. M. Harari, and G. K. Hartig, "Tumor volume as a prognostic factor in oropharyngeal squamous cell carcinoma treated with primary radiotherapy," *Laryngoscope* **118**, 1377–1382 (2008).
- ¹⁹R. Hermans, K. Op de beeck, W. Van den Bogaert, A. Rijnders, L. Staelens, M. Feron, and E. Bellon, "The relation of CT-determined tumor parameters and local and regional outcome of tonsillar cancer after definitive radiation treatment," *Int. J. Radiat. Oncol., Biol., Phys.* **50**, 37–45 (2001).
- ²⁰S. Mishra, A. Hammond, N. Read, V. Venkatesan, E. Winkquist, A. Nichols, A. Warner, and D. Palma, "Can radiological changes in lymph node volume during treatment predict success of radiation therapy in patients with locally advanced head and neck squamous cell carcinoma?" *J. Med. Imag. Radiat. Oncol.* **57**, 603–609 (2013).
- ²¹W. M. Mendenhall, R. J. Amdur, S. P. Stringer, D. B. Villaret, and N. J. Cassisi, "Radiation therapy for squamous cell carcinoma of the tonsillar region: A preferred alternative to surgery?," *J. Clin. Oncol.* **18**, 2219–2225 (2000).
- ²²K. K. Ang, J. Harris, R. Wheeler, R. Weber, D. I. Rosenthal, P. F. Nguyen-Tan, W. H. Westra, C. H. Chung, R. C. Jordan, C. Lu, H. Kim, R. Axelrod, C. C. Silverman, K. P. Redmond, and M. L. Gillison, "Human papillomavirus and survival of patients with oropharyngeal cancer," *N. Engl. J. Med.* **363**, 24–35 (2010).
- ²³H. Lee, Y. C. Ahn, D. Oh, H. Nam, Y. I. Kim, and S. Y. Park, "Tumor volume reduction rate measured during adaptive definitive radiation therapy as a potential prognosticator of locoregional control in patients with oropharyngeal cancer," *Head & Neck* **36**, 499–504 (2014).
- ²⁴A. V. Chvetsov, "Tumor response parameters for head and neck cancer derived from tumor-volume variation during radiation therapy," *Med. Phys.* **40**, 034101 (8pp.) (2013).
- ²⁵A. V. Chvetsov, L. Dong, J. R. Palta, and R. J. Amdur, "Tumor-volume simulation during radiotherapy for head-and-neck cancer using a four-level cell population model," *Int. J. Radiat. Oncol., Biol., Phys.* **75**, 595–602 (2009).