

A Model and Nomogram to Predict Tumor Site Origin for Squamous Cell Cancer Confined to Cervical Lymph Nodes

Arif N. Ali, MD¹; Jeffrey M. Switchenko, PhD²; Sungjin Kim, MS²; Jeanne Kowalski, PhD²; Mark W. El-Deiry, MD³; and Jonathan J. Beitler, MD^{1,3,4}

BACKGROUND: The current study was conducted to develop a multifactorial statistical model to predict the specific head and neck (H&N) tumor site origin in cases of squamous cell carcinoma confined to the cervical lymph nodes (“unknown primaries”). **METHODS:** The Surveillance, Epidemiology, and End Results (SEER) database was analyzed for patients with an H&N tumor site who were diagnosed between 2004 and 2011. The SEER patients were identified according to their H&N primary tumor site and clinically positive cervical lymph node levels at the time of presentation. The SEER patient data set was randomly divided into 2 data sets for the purposes of internal split-sample validation. The effects of cervical lymph node levels, age, race, and sex on H&N primary tumor site were examined using univariate and multivariate analyses. Multivariate logistic regression models and an associated set of nomograms were developed based on relevant factors to provide probabilities of tumor site origin. **RESULTS:** Analysis of the SEER database identified 20,011 patients with H&N disease with both site-level and lymph node-level data. Sex, race, age, and lymph node levels were associated with primary H&N tumor site (nasopharynx, hypopharynx, oropharynx, and larynx) in the multivariate models. Internal validation techniques affirmed the accuracy of these models on separate data. **CONCLUSIONS:** The incorporation of epidemiologic and lymph node data into a predictive model has the potential to provide valuable guidance to clinicians in the treatment of patients with squamous cell carcinoma confined to the cervical lymph nodes. *Cancer* 2014;120:3469-76. © 2014 The Authors. *Cancer* published by Wiley Periodicals, Inc. on behalf of *American Cancer Society*. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

KEYWORDS: unknown primary, Surveillance, Epidemiology, and End Results (SEER), cervical lymph nodes, predictive model, radiation.

INTRODUCTION

Squamous cell carcinoma (SCC) confined to cervical lymph nodes, also known as head and neck (H&N) cancer of an unknown primary (CUP), refers to a clinical scenario in which a patient presents with biopsy-proven cervical adenopathy yet appropriate workup reveals no evidence of the primary H&N malignancy. SCC from an unknown primary tumor represents approximately 2% to 4% of all H&N malignancies.^{1,2} The standard workup in a patient presenting with a histologically confirmed H&N SCC cervical mass without an obvious primary tumor is careful history and physical examination, 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography (PET)/computed tomography (CT), and then evaluation under anesthesia with appropriate directed (“blind”) biopsies and elective tonsillectomies.³ The use of combination 18F-FDG-PET/CT has become the standard of care for the noninvasive component of the diagnostic evaluation of a CUP.⁴ A recent meta-analysis found that the diagnostic detection rate of 18F-FDG-PET/CT was 37%, with a sensitivity of 84% and a specificity of 84%.⁵ Cianchetti et al found that panendoscopy with directed biopsies successfully identified the primary H&N tumor site in 29.2% of patients with H&N cancer with a CUP.⁶ In addition, the use of an ipsilateral tonsillectomy has been shown to have a higher diagnostic yield than a simple deep tonsil biopsy.⁷⁻¹⁰ There has also been recent enthusiasm for the incorporation of transoral robotic biopsy for evaluation of CUP.¹¹⁻¹³ Although the identification rate for the unknown primary site may increase to 77.3%,¹² which is certainly promising, the question of morbidity after elective lingual tonsillectomy, either acute or in terms of swallowing dysfunction after radiotherapy (or chemoradiation), has, to our knowledge, not yet been well addressed to date. As with any many new techniques in oncology, “the good news always comes first.”¹⁴

Corresponding author: Arif N. Ali, MD, Department of Radiation Oncology, Winship Cancer Institute, Emory University, 1365 Clifton Rd NE, Suite A1300, Atlanta, GA, 30322; Fax: (404) 778-4139; aali24@emory.edu

¹Department of Radiation Oncology, Winship Cancer Institute, Emory University, Atlanta, Georgia; ²Department of Biostatistics and Bioinformatics, Winship Cancer Institute, Emory University, Atlanta, Georgia; ³Department of Otolaryngology, Winship Cancer Institute, Emory University, Atlanta, Georgia; ⁴Department of Hematology and Oncology, Winship Cancer Institute, Emory University, Atlanta, Georgia

Additional Supporting Information may be found in the online version of this article.

DOI: 10.1002/cncr.28901, **Received:** March 17, 2014; **Revised:** May 19, 2014; **Accepted:** May 22, 2014, **Published online** July 24, 2014 in Wiley Online Library (wileyonlinelibrary.com)

Among patients with an unknown primary site who were treated with surgery alone, the primary lesion appeared after treatment in only 20% of patients (with the hypopharynx and oropharynx being the most common sites). It is important to know that this low rate of delayed appearance of the primary tumor was demonstrated in an article written in 1973, well before both contemporary imaging and flexible endoscopy were available. The combination of surgery and radiotherapy improved the ipsilateral neck control from 76% to 86% and the addition of radiation improved the contralateral neck control from 84% to 100%.¹⁵

For CUP, physicians in clinical practice correlate the involved cervical lymph node level(s) at the time of presentation with the most likely corresponding H&N site of primary malignancy and treat the patient accordingly.¹⁶ This intuitive correlation is based on work originally published by Lindberg in 1972.¹⁷ Although much has changed since that time, including quantum leaps in imaging technology (PET, magnetic resonance imaging, CT, etc) and state-of-the-art diagnostic and surgical techniques, the fundamentals of that work remain to this day.¹⁷

If a CUP workup fails to identify a primary H&N site for SCC confined to the cervical lymph nodes, patients are often treated with neck dissection followed by definitive external beam radiotherapy to a wide field. The traditional “usual suspects” have led radiation oncologists to electively treat the nasopharynx, oropharynx, and hypopharynx/larynx.^{3,18,19} The target selection is based partly on the accessibility of the oral cavity and the low probability of an unknown primary tumor being located there as well as the relative difficulty in evaluating the nasopharynx and the hypopharynx, especially during the pre-fiber optic period. Given the significant acute and late toxicities associated with comprehensive mucosal irradiation and the low incidence of larynx/hypopharynx primary malignancies, some institutions are routinely only irradiating the oropharynx ± the nasopharynx for unknown H&N primary malignancies.²⁰⁻²³

Therefore, the purpose of the current study was to present a multifactorial logistical regression model incorporating the involved cervical lymph node level at the time of presentation and epidemiologic data to calculate the probability of an H&N tumor site of origin for patients with SCC confined to the cervical lymph nodes.

MATERIALS AND METHODS

Data Source

Involved cervical lymph node levels, age, race, and sex were included in the model used in the current study.

These factors were all obtained from the Surveillance, Epidemiology, and End Results (SEER) 17 registry, which was released in April 2013 and includes tumors diagnosed between 2004 and 2011.²⁴ This database was accessed using SEER*Stat (version 7.1.0). The inclusion criteria for the case listing session required that all cases have a known age, race, sex, cervical lymph node level, and primary H&N tumor site. Cases included only a patient's first diagnosis of cancer and those registered at the time of autopsy or death certificate only were excluded from the analysis. Lymph node-level data were parsed from Collaborative Stage site-specific factors 3 and 4 of the SEER data for H&N tumor site, which included information regarding lymph node involvement for levels 1, 2, and 3 and levels 4, 5, and retropharyngeal lymph nodes, respectively. Race was categorized into white (non-Hispanic), Hispanic, black, and Asian (Asian or Pacific Islander) using SEER race recode variables. Age at the time of diagnosis was used for patient age.

Statistical Analysis

Descriptive statistics were generated for the SEER data, along with an analysis of race, sex, age, and lymph node level by primary tumor site using either chi-square tests or analysis of variance where appropriate.²⁵ A split-sample internal validation procedure was performed to establish that the models worked sufficiently among patients other than those whose data generated the model. Observations were assigned a random number between 0 and 1, and those observations whose random number fell between 0 and 0.5 were placed in the test set. Observations whose random number fell between 0.5 and 1 were placed in the validation (or training) set. Univariate and multivariate logistic regression models were fit for each primary tumor site separately for the test set. For example, for the larynx model, the outcome was either larynx or not larynx. As a result, we determined the effect of race, sex, age, and each cervical lymph node on the presence of each primary site while adjusting for the other covariates. Each primary site multivariate model contained the same covariates for consistency. Graphical nomograms, derived from the multivariate logistic regression model, were created for each site using the R statistical package rms (R Foundation for Statistical Computing, Vienna, Austria). Plots of observed versus predicted estimates at each decile for the validation set were produced, and the root mean square error (RMSE) was calculated for both the test set and validation set using the model generated from the test set. RMSE is the square root of the difference between the observed and predicted values squared divided by the number of

observations, and is a measure of accuracy. Similar RMSE values between the test and validation sets indicate adequate model validation. Logistic regression model fit

TABLE 1. SEER Data Descriptive Statistics

Variable	Level	N = 20,011	%
Sex	Female	3799	19.0
	Male	16,212	81.0
Primary tumor site	Oropharynx	12,829	64.1
	Nasopharynx	1650	8.2
	Hypopharynx	1854	9.3
	Larynx	3678	18.4
Race	Asian	1307	6.5
	Black	2458	12.3
	Hispanic	1289	6.4
	White	14,957	74.7
Level 1 lymph nodes	Not involved	15,294	76.4
	Involved	4717	23.6
Level 2 lymph nodes	Not involved	5612	28.0
	Involved	14,399	72.0
Level 3 lymph nodes	Not involved	12,723	63.6
	Involved	7288	36.4
Level 4 lymph nodes	Not involved	16,587	82.9
	Involved	3424	17.1
Level 5 lymph nodes	Not involved	17,210	86.0
	Involved	2801	14.0
Retropharyngeal lymph nodes	Not involved	19,456	97.2
	Involved	555	2.8
Age, y	Mean	59.37	—
	Median	59	—
	Minimum	1	—
	Maximum	100	—
	SD	11.34	—

Abbreviation: SD, standard deviation; SEER, Surveillance, Epidemiology, and End Results.

TABLE 2. Relationship of Head and Neck Primary Tumor Site With SEER Variables

Covariate	Statistics	Level	Oropharynx N=12,829	Nasopharynx N=1650	Hypopharynx N=1854	Larynx N=3678	Parametric <i>P</i> ^a
Sex	No. (row %)	Female	2108 (55.49)	460 (12.11)	332 (8.74)	899 (23.66)	<.001
	No. (row %)	Male	10,721 (66.13)	1190 (7.34)	1522 (9.39)	2779 (17.14)	
Race	No. (row %)	Asian	382 (29.23)	682 (52.18)	116 (8.88)	127 (9.72)	<.001
	No. (row %)	Black	1240 (50.45)	210 (8.54)	319 (12.98)	689 (28.03)	
	No. (row %)	Hispanic	758 (58.81)	141 (10.94)	129 (10.01)	261 (20.25)	
	No. (row %)	White	10,449 (69.86)	617 (4.13)	1290 (8.62)	2601 (17.39)	
Level 1 lymph nodes	No. (row %)	Not involved	9680 (63.29)	1240 (8.11)	1471 (9.62)	2903 (18.98)	<.001
	No. (row %)	Involved	3149 (66.76)	410 (8.69)	383 (8.12)	775 (16.43)	
Level 2 lymph nodes	No. (row %)	Not involved	3230 (57.56)	510 (9.09)	650 (11.58)	1222 (21.77)	<.001
	No. (row %)	Involved	9599 (66.66)	1140 (7.92)	1204 (8.36)	2456 (17.06)	
Level 3 lymph nodes	No. (row %)	Not involved	8534 (67.08)	1107 (8.7)	1021 (8.02)	2061 (16.2)	<.001
	No. (row %)	Involved	4295 (58.93)	543 (7.45)	833 (11.43)	1617 (22.19)	
Level 4 lymph nodes	No. (row %)	Not involved	10,934 (65.92)	1309 (7.89)	1436 (8.66)	2908 (17.53)	<.001
	No. (row %)	Involved	1895 (55.34)	341 (9.96)	418 (12.21)	770 (22.49)	
Level 5 lymph nodes	No. (row %)	Not involved	11,396 (66.22)	1112 (6.46)	1539 (8.94)	3163 (18.38)	<.001
	No. (row %)	Involved	1433 (51.16)	538 (19.21)	315 (11.25)	515 (18.39)	
Retropharyngeal lymph nodes	No. (row %)	Not involved	12,586 (64.69)	1475 (7.58)	1784 (9.17)	3611 (18.56)	<.001
	No. (row %)	Involved	243 (43.78)	175 (31.53)	70 (12.61)	67 (12.07)	
Age, y	No.		12,829	1650	1854	3678	<.001
	Mean		59.1	52.23	63.29	61.52	
	Median		58	53	62	61	

Abbreviation: SEER, Surveillance, Epidemiology, and End Results.

^aThe parametric *P* value was calculated using the analysis of variance for numerical covariates and the chi-square test for categorical covariates.

Bold values indicate statistical significance less than 0.05.

also was assessed using the *c*-statistic. Assumptions for chi-square tests, analysis of variance, and logistic regression were checked and verified. The descriptive statistics, comparisons, and logistic regression models were produced using SAS statistical software (version 9.3; SAS Institute Inc, Cary, NC), and significance was set at $\alpha = .05$.²⁶⁻²⁸

RESULTS

In the analysis of the SEER data, a total of 20,011 cases were identified with complete data regarding the primary tumor site, cervical lymph node level, age, sex, and race (Table 1). The data set consisted of 16,212 males (81.0%) and the mean age was 59 years. The majority of cases (64.1%) were located in the oropharynx (Table 1). As seen in Table 2, it was found that race, sex, and age were all significantly associated with H&N primary sites ($P < .001$), thus indicating that these variables would likely be useful in differentiating between H&N primary tumor sites in an unknown primary predictive model. In addition, in Table 2, it is shown that all cervical lymph node levels were found to be significantly associated with H&N primary sites ($P < .001$). This finding indicates that the level of cervical lymph node involvement at the time of presentation of an H&N unknown primary malignancy would also be useful in differentiating between H&N primary sites in an unknown primary predictive model.

TABLE 3. Multivariate Analysis of SEER Variables

Covariate	Level	Oropharynx		Nasopharynx		Hypopharynx		Larynx	
		OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>
Sex	Male	1.54 (1.39-1.71)	<.001	0.51 (0.42-0.61)	<.001	1.15 (0.96-1.38)	.126	0.72 (0.63-0.81)	<.001
	Female	—	—	—	—	—	—	—	—
Race	Asian	0.16 (0.13-0.19)	<.001	27.6 (22.4-33.9)	<.001	1.00 (0.74-1.34)	.996	0.48 (0.37-0.64)	<.001
	Black	0.44 (0.38-0.49)	<.001	2.09 (1.66-2.64)	<.001	1.60 (1.32-1.93)	<.001	1.86 (1.61-2.15)	<.001
	Hispanic	0.61 (0.52-0.72)	<.001	2.74 (2.08-3.61)	<.001	1.29 (0.99-1.68)	.056	1.15 (0.94-1.41)	.175
	White	—	—	—	—	—	—	—	—
Age		0.99 (0.99-0.99)	<.001	0.95 (0.94-0.96)	<.001	1.04 (1.03-1.04)	<.001	1.02 (1.02-1.03)	<.001
Level 1 lymph nodes	Involved	1.34 (1.20-1.49)	<.001	0.96 (0.79-1.17)	.687	0.79 (0.66-0.94)	.008	0.77 (0.68-0.88)	<.001
Level 2 lymph nodes	Involved	1.46 (1.32-1.61)	<.001	0.97 (0.80-1.18)	.771	0.71 (0.61-0.83)	<.001	0.74 (0.66-0.83)	<.001
Level 3 lymph nodes	Involved	0.77 (0.70-0.84)	<.001	0.71 (0.59-0.86)	<.001	1.37 (1.19-1.59)	<.001	1.38 (1.24-1.54)	<.001
Level 4 lymph nodes	Involved	0.79 (0.70-0.89)	<.001	1.07 (0.86-1.33)	.567	1.22 (1.03-1.45)	.024	1.24 (1.08-1.42)	.002
Level 5 lymph nodes	Involved	0.71 (0.63-0.80)	<.001	2.42 (1.98-2.96)	<.001	1.22 (1.01-1.47)	.040	0.92 (0.79-1.07)	.281
Retropharyngeal lymph nodes	Involved	0.59 (0.46-0.76)	<.001	3.46 (2.50-4.81)	<.001	1.20 (0.81-1.78)	.352	0.60 (0.41-0.87)	.008

Abbreviations: 95% CI, 95% confidence interval; OR, odds ratio; SEER, Surveillance, Epidemiology, and End Results. Bold values indicate statistical significance less than 0.05.

A univariate logistic regression was performed (see online supporting information). All variables were statistically significant for the oropharynx. However, for the nasopharynx, involvement of level 1 lymph nodes did not statistically predict for involvement of the nasopharynx, whereas all other variables were found to be statistically significant. Likewise, for the larynx, involvement of level 5 lymph nodes was not statistically as useful as the other variables. In the hypopharynx, the epidemiological factors of sex and race were not found to be statistically significant, but all the cervical lymph node levels did reach statistical significance.

Multivariate logistic regression models were derived using the SEER patient data set with the inclusion of sex, race, age, and cervical lymph node involvement as independent variables (Table 3). As seen in Table 3, for the oropharynx, all variables considered were statistically significant. However, only the epidemiologic variables (sex, age, and race) and level 3, 5, and retropharyngeal lymph nodes were significant for the nasopharynx site. In the larynx model, all variables were significant except for the level 5 lymph nodes, whereas in the hypopharynx model, all variables were significant except for sex and the retropharyngeal lymph nodes.

The resulting coefficients from the multiple logistic regression models (see online supporting information) were then used to create nomograms for each of the H&N tumor sites, which yield the probability of involvement of that site. For example, in comparing Table 3 with Figure 1, the higher magnitude odds ratio variables correlate with the largest point allocation on each of the H&N site nomograms. An

interesting observation is the robust strength of Asian race as a predictor for nasopharyngeal involvement. Likewise, young age, male sex, and level 1 or 2 lymph node involvement would appear to indicate a strong probability of the oropharynx as the site of origin.

After the creation of the multivariate logistic regression models from the SEER patient data set, model validation was performed using split-sample internal validation with separate training and validation data sets. As seen in Figure 2, there is excellent agreement between predicted and observed probabilities for all deciles of each of the 4 H&N tumor sites. RMSE values for both the training and validation sets were calculated and are reported in Table 4. In all 4 models, the RMSE value calculated for the validation set was close to the value generated from the training set using the training set model estimates (differences ranging from 0.4%-1.3%). This indicates a high level of model validation and accuracy. C-statistic values for the oropharynx, nasopharynx, hypopharynx, and larynx models were 0.67, 0.81, 0.65, and 0.64, respectively, indicating adequate model fit, particularly for the nasopharynx model.

DISCUSSION

The data used to construct the model in the current study were derived from the SEER database and compare favorably with independent data sources and prospective trials. For example, as seen in Table 2, there was a significant difference in age between the nasopharyngeal group (median, 53 years) and the rest of the sites (median age range, 58 years-62 years). This compares well with both the Intergroup 0099 study (nasopharyngeal median age

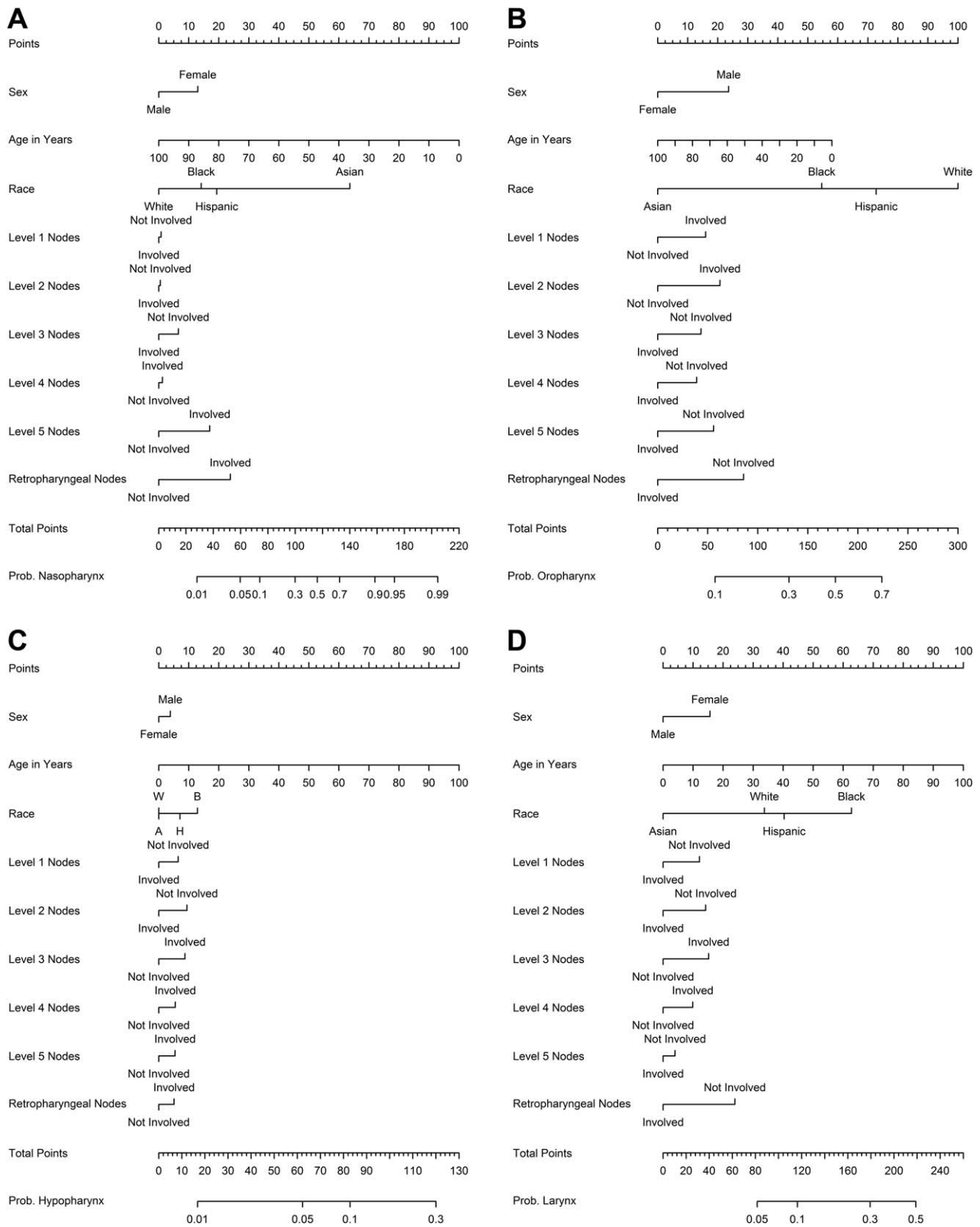


Figure 1. Predictive nomograms are shown for (A) nasopharynx primary site, (B) oropharynx primary site, (C) hypopharynx primary site, and (D) larynx primary site. Prob indicates probability.

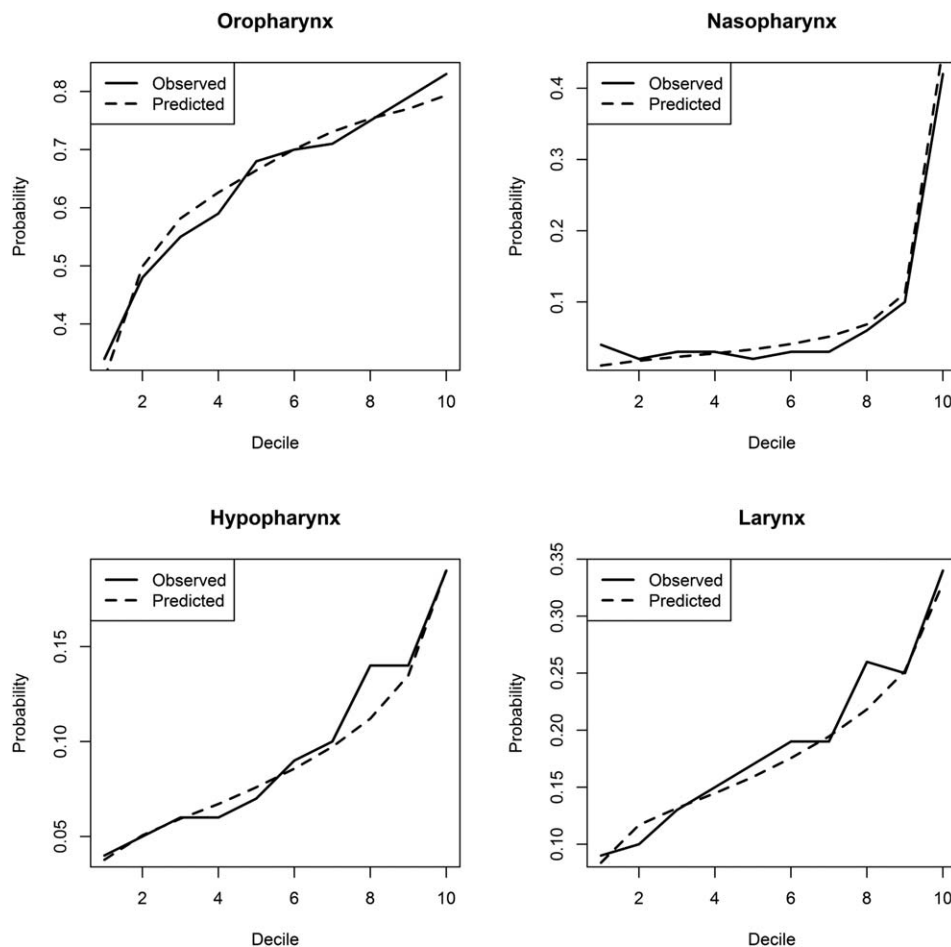


Figure 2. Validation plots for assessing the predictive ability of each primary site model through split-sample internal validation are shown.

range, 50 years-52 years)²⁹ and the more recent phase 3 cetuximab H&N trial by Bonner et al (oropharynx, larynx, and hypopharynx median age, 56 years-58 years).³⁰ Likewise, the cervical lymph node level correlation with primary tumor site compares well with prior published data. For example, in Online Supporting Information Table 5, the hypopharynx had the highest percentage (23%) of involved level 4 cervical lymph nodes out of all the considered sites. Chao et al similarly listed the hypopharynx as the site with the highest incidence of level 4 cervical lymph nodes.³¹ In addition, the oropharynx appeared most likely to metastasize to the level 2 or 3 cervical lymph nodes (75% and 33%, respectively) (see online supporting information). Again, this compares favorably with the study by Chao et al, which also listed the cervical level 2 and 3 lymph nodes as the most likely levels of metastasis from a clinically lymph node-positive oropharynx primary tumor (tonsil, 74% and 31%,

respectively).³¹ Likewise, Lindberg listed the upper, middle, and lower jugular lymph nodes as the most involved cervical metastatic sites from primary tumors of the soft palate, tonsil, base of tongue, and oropharyngeal wall.¹⁷

Although the SEER database has the advantage of large patient numbers and a favorable comparison with prior published data, a potential limitation of a model derived from the SEER database would be the finding that the model would primarily be applicable only to patients in the United States because other parts of the world are likely to have different epidemiologic percentages. In addition, the model does not include information regarding viral biomarkers or social risk factors such as smoking, alcohol and marijuana use, and number of oral sexual partners. SEER is currently tracking human papillomavirus (HPV) status, but to our knowledge has not yet publicly released those data. When these data are

TABLE 4. RMSE Estimates for Validation and Training Sets Using Validation Model Estimates

Site	Validation RMSE	Training RMSE	Absolute Difference	% Difference
Nasopharynx	0.235	0.233	0.002	0.9%
Oropharynx	0.457	0.459	0.002	0.4%
Hypopharynx	0.285	0.287	0.002	0.7%
Larynx	0.378	0.383	0.005	1.3%

Abbreviation: RMSE, root mean square error.

available, it is anticipated that the current model will be easily updated with that additional variable. The inclusion of HPV will likely be most relevant to the oropharynx site. Prior studies have shown a strong correlation between HPV status and oropharyngeal primary tumors.³²⁻³⁵ Thus, the inclusion of the HPV variable and a positive HPV result would likely increase the calculated probability of the oropharyngeal site to the exclusion of the other H&N sites compared with calculated probabilities in the current model. Likewise, Epstein-Barr virus has been shown to be correlated primarily with the nasopharyngeal site and inclusion of that variable would similarly affect the calculated probabilities of the nasopharynx site.³⁶

A minor limitation to the current study that must also be noted is the finding that the other 3 primary tumor sites provide the reference level for a given primary site's logistic regression model. For example, the combined totals of oropharynx, nasopharynx, and hypopharynx serve as the reference level for the larynx model. As a result, the sum of the predicted probabilities for the 4 models will be approximately 1 for a given set of patient characteristics, but not exactly 1.

The model presented in the current study delivers a value for each of the 4 H&N sites (nasopharynx, oropharynx, larynx, and hypopharynx) representing the probability that a patient's unknown H&N primary malignancy arose from that site. This probability value may be graphically derived from the respective nomogram (Fig. 1) or calculated directly from the multiple logistic regression equations (see online supporting information). Such a probability value has the potential for many uses, including more personalized radiation treatment planning and/or further guidance for higher spatial resolution biopsies. In reviewing the nomograms, it can be observed that the resulting probabilities compare favorably with clinical intuition based on prior studies. For example, if a white man aged 62 years presented with a mass primarily centered at approximately the level 3 cervical lymph nodes, there would be a 63.7% probability that the primary H&N site was the oropharynx, a 15.6% probability for the hypopharynx, a 22.4% probability for the larynx, and a 2.0% probability for the nasopharynx. If a radiation

oncologist's personal cutoff is 15% for covering a given H&N primary tumor site, they might choose to omit the nasopharynx in that case. Consider another example in which a 42-year-old Asian woman presents with a level V cervical mass, but no known H&N primary tumor: there would be a 91.2% probability that the primary site was the nasopharynx, a 17.1% probability for the oropharynx, a 6.0% probability for the hypopharynx, and a 8.0% probability for the larynx. This might prompt a more thorough investigation of the nasopharynx with another nasopharyngoscopy, additional biopsies of multiple locations within the nasopharynx by the otolaryngologist, and/or a magnetic resonance image of that region. If still no H&N primary tumors were identified, a radiation oncologist might choose to omit the larynx and hypopharynx from the mucosal field.

The current study reports on the use of the SEER database to create a multivariate logistic regression model that provides predicted probabilities of involvement of H&N tumor sites based on age, race, sex, and cervical lymph node metastasis level. The model has demonstrated a high level of validation and accuracy when subjected to internal split-sample validation with separate training and validation data sets. Although good clinical judgment must remain the foundation of any treatment decision, it is hoped that the probabilities derived from this model may provide clinicians with additional insight into the site of origin of an unknown H&N primary tumor.

FUNDING SUPPORT

Research reported in this publication was supported in part by the Biostatistics and Bioinformatics Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292 (W. Curran). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

REFERENCES

1. Waltonen JD, Ozer E, Hall NC, Schuller DE, Agrawal A. Metastatic carcinoma of the neck of unknown primary origin: evolution and

- efficacy of the modern workup. *Arch Otolaryngol Head Neck Surg.* 2009;135:1024-1029.
2. Grau C, Johansen LV, Jakobsen J, Geertsen P, Andersen E, Jensen BB. Cervical lymph node metastases from unknown primary tumours. Results from a national survey by the Danish Society for Head and Neck Oncology. *Radiother Oncol.* 2000;55:121-129.
 3. Strojjan P, Ferlito A, Langendijk JA, et al. Contemporary management of lymph node metastases from an unknown primary to the neck: II. a review of therapeutic options. *Head Neck.* 2013;35:286-293.
 4. Roh JL, Kim JS, Lee JH, et al. Utility of combined (18)F-fluorodeoxyglucose-positron emission tomography and computed tomography in patients with cervical metastases from unknown primary tumors. *Oral Oncol.* 2009;45:218-224.
 5. Kwee T, Kwee R. Combined FDG-PET/CT for the detection of unknown primary tumors: systematic review and meta-analysis. *Eur Radiol.* 2009;19:731-744.
 6. Cianchetti M, Mancuso AA, Amdur RJ, et al. Diagnostic evaluation of squamous cell carcinoma metastatic to cervical lymph nodes from an unknown head and neck primary site. *Laryngoscope.* 2009;119:2348-2354.
 7. Waltonen JD, Ozer E, Schuller DE, Agrawal A. Tonsillectomy vs. deep tonsil biopsies in detecting occult tonsil tumors. *Laryngoscope.* 2009;119:102-106.
 8. Righi PD, Sofferan RA. Screening unilateral tonsillectomy in the unknown primary. *Laryngoscope.* 1995;105:548-550.
 9. Lapeyre M, Malissard L, Peiffert D, et al. Cervical lymph node metastasis from an unknown primary: is a tonsillectomy necessary? *Int J Radiat Oncol Biol Phys.* 1997;39:291-296.
 10. Randall DA, Johnstone PA, Foss RD, Martin PJ. Tonsillectomy in diagnosis of the unknown primary tumor of the head and neck. *Otolaryngol Head Neck Surg.* 2000;122:52-55.
 11. Abuzeid WM, Bradford CR, Divi V. Transoral robotic biopsy of the tongue base: a novel paradigm in the evaluation of unknown primary tumors of the head and neck. *Head Neck.* 2013;35:E126-E130.
 12. Durmus K, Rangarajan SV, Old MO, Agrawal A, Teknos TN, Ozer E. Transoral robotic approach to carcinoma of unknown primary. *Head Neck.* 2014;36:848-852.
 13. Mehta V, Johnson P, Tassler A, et al. A new paradigm for the diagnosis and management of unknown primary tumors of the head and neck: a role for transoral robotic surgery. *Laryngoscope.* 2013;123:146-151.
 14. Hellman S. It's too soon to know. *J Natl Cancer Inst.* 1990;82:250-251.
 15. Jesse RH, Perez CA, Fletcher GH. Cervical lymph node metastasis: unknown primary cancer. *Cancer.* 1973;31:854-859.
 16. Werner JA, Dunne AA, Myers JN. Functional anatomy of the lymphatic drainage system of the upper aerodigestive tract and its role in metastasis of squamous cell carcinoma. *Head Neck.* 2003;25:322-332.
 17. Lindberg R. Distribution of cervical lymph node metastases from squamous cell carcinoma of the upper respiratory and digestive tracts. *Cancer.* 1972;29:1446-1449.
 18. Harper CS, Mendenhall WM, Parsons JT, Stringer SP, Cassisi NJ, Million RR. Cancer in neck nodes with unknown primary site: role of mucosal radiotherapy. *Head Neck.* 1990;12:463-469.
 19. Mendenhall WM, Mancuso AA, Amdur RJ, Stringer SP, Villaret DB, Cassisi NJ. Squamous cell carcinoma metastatic to the neck from an unknown head and neck primary site. *Am J Otolaryngol.* 2001;22:261-267.
 20. Barker CA, Morris CG, Mendenhall WM. Larynx-sparing radiotherapy for squamous cell carcinoma from an unknown head and neck primary site. *Am J Clin Oncol.* 2005;28:445-448.
 21. Lu H, Yao M, Tan H. Unknown primary head and neck cancer treated with intensity-modulated radiation therapy: to what extent the volume should be irradiated. *Oral Oncol.* 2009;45:474-479.
 22. Wallace A, Richards GM, Harari PM, et al. Head and neck squamous cell carcinoma from an unknown primary site. *Am J Otolaryngol.* 2011;32:286-290.
 23. Hsing CY, Liu SA, Wang CC. Management of unknown primary head and neck squamous cell carcinoma. *Am J Otolaryngol.* 2012;33:637-638; author reply 638-639.
 24. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program. Bethesda, MD: National Cancer Institute; 2011.
 25. Neter J, Kutner M, Wasserman W, Nachtsheim C. Applied Linear Statistical Models. 4th ed. Chicago: McGraw-Hill; 1996.
 26. R Project for Statistical Computing. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
 27. SAS Institute Inc. Base SAS 9.3 Procedures Guide. Cary, NC: SAS Institute Inc; 2011.
 28. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons; 2000.
 29. Al-Sarraf M, LeBlanc M, Giri PG, et al. Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized Intergroup study 0099. *J Clin Oncol.* 1998;16:1310-1317.
 30. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med.* 2006;354:567-578.
 31. Chao KS, Wippold FJ, Ozyigit G, Tran BN, Dempsey JF. Determination and delineation of nodal target volumes for head-and-neck cancer based on patterns of failure in patients receiving definitive and postoperative IMRT. *Int J Radiat Oncol Biol Phys.* 2002;53:1174-1184.
 32. Begum S, Gillison ML, Ansari-Lari MA, Shah K, Westra WH. Detection of human papillomavirus in cervical lymph nodes. *Clin Cancer Res.* 2003;9:6469-6475.
 33. Begum S, Gillison ML, Nicol TL, Westra WH. Detection of human papillomavirus-16 in fine-needle aspirates to determine tumor origin in patients with metastatic squamous cell carcinoma of the head and neck. *Clin Cancer Res.* 2007;13:1186-1191.
 34. El-Mofty S, Zhang M, Davila R. Histologic identification of human papillomavirus (HPV)-related squamous cell carcinoma in cervical lymph nodes: a reliable predictor of the site of an occult head and neck primary carcinoma. *Head Neck Pathol.* 2008;2:163-168.
 35. Zhang MQ, El-Mofty SK, Davila RM. Detection of human papillomavirus-related squamous cell carcinoma cytologically and by in situ hybridization in fine-needle aspiration biopsies of cervical metastasis. *Cancer (Cancer Cytopathol).* 2008;114:118-123.
 36. Singhi AD, Califano J, Westra WH. High-risk human papillomavirus in nasopharyngeal carcinoma. *Head Neck.* 2012;34:213-218.