Original Article

Individualized Outcome Prognostication for Patients With Laryngeal Cancer

Connor W. Hoban, BS [1]; Lauren J. Beesley, MS[2]; Emily L. Bellile, MS[2]; Yilun Sun, MS[2]; Matthew E. Spector, MD [1]; Gregory T. Wolf, MD[1]; Jeremy M. G. Taylor, PhD[2]; and Andrew G. Shuman, MD [1]

BACKGROUND: Accurate prognostication is essential to the optimal management of laryngeal cancer. Predictive models have been developed to calculate the risk of oncologic outcomes, but extensive external validation of accuracy and reliability is necessary before implementing them into clinical practice. **METHOD:** Four published prognostic calculators that predict 5-year overall survival for patients with laryngeal cancer were evaluated using patient information from a prospective epidemiology study cohort (n = 246; median follow-up, 60 months) with previously untreated, stage I through IVb laryngeal squamous cell carcinoma. **RESULTS:** Different calculators yielded substantially different predictions for individual patients. The observed 5-year overall survival was significantly higher than the averaged predicted 5-year overall survival of the 4 calculators (71.9%; 95% confidence interval [CI], 65%-78%] vs 47.7%). Statistical analyses demonstrated the calculators' limited capacity to discriminate outcomes for risk-stratified patients. The area under the receiver operating characteristic curve ranged from 0.68 to 0.72. C-index values were similar for each of the 4 models (range, 0.66-0.68). There was a lower than expected hazard of death for patients who received induction (bioselective) chemotherapy (hazard ratio, 0.46; 95% CI, 0.24-0.88; $P = .024$) or primary surgical intervention (hazard ratio, 0.43; 95 % CI, 0.21-0.90; $P = .024$) compared with those who received concurrent chemoradiation. **CONCLUSIONS:** Suboptimal reliability and accuracy limit the integration of existing individualized prediction tools into routine clinical decision making. The calculators predicted significantly worse than observed survival among patients who received induction chemotherapy and primary surgery, suggesting a need for updated consideration of modern treatment modalities. Further development of individualized prognostic calculators may improve risk prediction, treatment planning, and counseling for patients with laryngeal cancer. ***Cancer*** **2018;124:706-16.** *© 2017 American Cancer Society.*

KEYWORDS: calculator, larynx cancer, nomogram, prognostication, risk prediction.

INTRODUCTION

The multidisciplinary management of head and neck cancer (HNC) depends critically on accurate risk stratification and prediction of clinical outcomes.[1] Managing laryngeal cancer introduces additional challenges secondary to debilitating functional impairments that often accompany the primary disease and/or related therapeutic interventions.[2-5] Accepted standards for tumor staging and oncologic prognostication continue to be helpful in confronting these challenges and guiding decision making.[6,7] However, in order to further individualize treatment selections that enhance survival and minimize morbidity, more sophisticated methods that successfully capitalize on emerging discoveries related to tumor biology/genomics and patient factors are mandatory.[8,9] Treatment decision making is particularly complex for laryngeal cancer because of the various treatment options available and the differing short-term and long-term functional consequences that affect quality of life and survival. The management of laryngeal cancer has experienced substantial evolution, primarily driven by the implementation of chemotherapeutic modalities and novel organ-preservation strategies.[3,10-12] This further emphasizes the need for modernized tools when calculating prognostic estimates.

The heterogeneity of laryngeal cancer is 1 of the many factors that impose formidable challenges to the accurate prediction of individual survival.[13] Survival is influenced by numerous variables, including multiple and diverse tumor-

Corresponding author: Andrew G. Shuman, MD, Department of Otolaryngology-Head and Neck Surgery, Michigan Medicine, 1500 East Medical Center Drive, Ann Arbor, MI 48109; shumana@med.umich.edu

[1]Department of Otolaryngology-Head and Neck Surgery, University of Michigan Medical School, Ann Arbor, Michigan; [2]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan

We thank the many investigators in the University of Michigan Head and Neck Specialized Program of Research Excellence for their contributions to patient recruitment, assistance in data collection, and encouragement, without compensation, including: Carol R. Bradford, MD; Thomas E. Carey, PhD; Douglas B. Chepeha, MD; Sonia Duffy, PhD; Avraham Eisbruch, MD; Joseph Helman, DDS; Kelly M. Malloy, MD; Jonathan McHugh, MD; Scott A. McLean, MD; Tamara H. Miller, RN; Jeff Moyer, MD; Jacques E. Nor, PhD; Lisa Peterson, MPH; Mark E. Prince, MD; Nancy Rogers, RN; Laura Rozek, PhD; Nancy E. Wallace, RN, Heather Walline, PhD; Brent Ward, DDS; and Francis Worden, MD. We also thank our patients and their families, who tirelessly participated in our survey and specimen collections.

The funding organizations had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; the preparation, review, or approval of the article; or the decision to submit the article for publication.

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.31087, **Received:** July 10, 2017; **Revised:** September 13, 2017; **Accepted:** September 27, 2017, **Published online** November 7, 2017 in Wiley Online Library (wileyonlinelibrary.com)

specific (size, grade, genomics, biologic features, and stage) and patient-related (age, race, sex, immune status, smoking status, and medical comorbidities) factors.[14] The TNM (tumor-lymph node-metastasis) staging system defined by the American Joint Committee on Cancer (AJCC) is the current prognostic standard for HNC and predicts survival with reasonable accuracy.[15] Despite its trusted reputation and ubiquitous assimilation in clinical practice, TNM staging fails to incorporate many criteria that demonstrate prognostic value, thereby limiting its ability to tailor risk predictions to an individual patient.[15] Studies suggest that implementing additional tumor, patient, and treatment characteristics into risk calculations can promote superior prognostic accuracy across a diverse range of oncologic subspecialties.[16-19]

The current trend in attitudes toward electronic health information suggests that online versions of these calculators would be readily implemented into medical decision making.[20-24] As a means of regulating newly published risk predictors, the AJCC recently published 16 inclusion and exclusion criteria that are required for endorsement of any probability or risk model.[25] These benchmarks should help to ensure that performance metrics, compatibility, and clinical relevance are robust amid the expansion of new prediction tools.[25]

Investigators have worked to address the dearth of individualized clinical decision tools currently available to interdisciplinary teams that manage head and neck squamous cell carcinoma (HNSCC) by developing prognostic calculators specific to HNCs.[26-28] Risk calculators wield potential clinical value but have not yet been subjected to sufficient evaluation and validation to warrant their assimilation into routine practice. Moreover, many of these were generated with older data, which may not be directly relevant to current patients. To assess clinical prognostic tools, analyses that compare the calculators' predictions with each other and with modern observed outcomes are imperative.[29] These validation studies are best performed on independent patient cohorts that encompass diverse geographic regions and patient demographics.[30,31] The objective of the current study was to use an independent patient cohort to externally evaluate and validate published prognostic calculators designed for patients with laryngeal cancer.

MATERIALS AND METHODS

The University of Michigan Institutional Review Board evaluated and approved this study. All participants provided written informed consent at enrollment in this prospective epidemiology study (typically at the time of diagnosis).

Prognostic Calculators

Prognostic clinical decision tools were identified using online search engines and expert input. PubMed and Google Scholar were investigated for peer-reviewed publications using a combination of search terms representing disease (larynx, cancer), prognosis (ie, survival, risk, prediction, and outcome) and methodology (calculator, tool, model, and nomogram). Multidisciplinary HNC specialists were also surveyed to probe for existing or emerging prognostic tools not identified in the online search.[16]

Potential calculator candidates were evaluated for eligibility. Inclusion criteria mandated that the calculator used clinical data to predict 5-year overall survival for squamous cell carcinoma of the larynx. Four prognostic calculators were identified (MAASTRO, LifeMath, Leiden, and MyCancerJourney), and each model was reviewed for content and format.[26-28] The calculators' mathematical formulas were acquired from the original publication, supplementary online materials, or computational derivation. We note that the MyCancerJourney calculator does not have an associated peer-reviewed publication.

Table 1[26-28] summarizes each calculator and includes the period, sample size, and other characteristics that describe the original study cohorts. Each calculator functions according to an equation constructed to represent the relation between tumor characteristics, patient demographics, treatment modalities received, and observed survival outcomes. The calculators considered a distinct set of variables in their prognostic equation and assigned differing quantities of statistical weight to these variables (Table 2 and Supporting Tables 1-7; see online supporting information). The calculators were modeled from patient data contained in the Surveillance, Epidemiology, and End Results (SEER) registry; regional study cohorts; or a combination of 2 patient populations. The 4 study cohorts included patients who received treatment with curative intent between 1973 and 2009.[26-28]

Patients

The analysis data set was derived from a single-institution, prospectively maintained HNC epidemiologic study.[32-35] In total, 246 patients with biopsy-proven, previously untreated, AJCC stage I through IVb squamous cell carcinoma of the larynx who were diagnosed and treated with curative intent at the University of Michigan Health System between 2003 and 2014 were included. Table 2 provides additional summary demographics for the cohort. Patients were evaluated by our multidisciplinary team and discussed at our Tumor Board, where treatment

TABLE 1. Summary of Calculators

Calculator	Cancers in Training Data Set	Training Data Set	Validation Data Set	Model Type	Model Details
MAASTRO (Egelmeer 2011[27])	Larynx	994 patients with laryngeal carcinoma who received RT between 1977 and 2008 (89.9% N0)	Leuven, 109 patients who received RT between 2000 and 2006 (75.2% N0); VU Amsterdam, 178 patients who received RT between 2001 and 2007 (92.7% N0); NKI/AML Amsterdam, 205 patients who received RT between 2000 and 2008 (89.8% N0); Manchester, 403 patients who received RT between 1998 and 2005 (98.8% N0)	Cox regression	Main effects only
LifeMath (Emerick 2013[26])	HN	50,145 patients with HN cancer in SEER between 1980 and 2009	1362 patients at Massachusetts General Hospital between 1980 and 2009	Statistical-mechanistic model of cancer metastasis involving separate tumor and node contributions	Complicated formulas with many parameters and interactions
Leiden (Datema 2013[28])	HN	1371 patients (638 with laryngeal cancer) at Leiden University Medical Center between 1981 and 1999	598 patients at Barnes-Jewish Hospital between 1995 and 2000	Cox regression	Main effects only
MyCancerJourney[a]	Many cancers	Patients in SEER between 1973 and 1996 and 11,791 patients at Barnes-Jewish Hospital between 1995 and 2001	No validation data	Cox regression	Main effects and many interactions

Abbreviations: HN, head and neck; NKI/AML, Netherlands Cancer Institute/acute myeloid leukemia; RT, radiation therapy; SEER, Surveillance, Epidemiology, and End Results program of the US National Cancer Institute.

[a] This calculator is available online at: https://staging.mycancerjourney.com/myinsights/survival-curves. Accessed September 12, 2017.

recommendations were formulated. Patients with stage I or II disease generally underwent single-modality surgery (33 patients; 35.1%), or received radiotherapy alone (50 patients; 53.2%), or received concurrent chemoradiotherapy for deeply invasive T2 lesions (11 patients; 11.7%). Patients with stage III or IV disease either underwent primary surgery (32 patients; 20.1%), received a single cycle of induction chemotherapy (bioselective) followed by either combined chemoradiation for a tumor response >50% or total laryngectomy for a tumor response <50% (70 patients; 46.1%), or received definitive chemoradiation (50 patients; 32.9%). The median follow-up was 60 months. Tumor-specific, patient-specific, and treatment-specific variables were exported from the database and confirmed by chart abstraction.

The calculators were designed for utility before oncologic treatment. Consequently, pretreatment clinical information was used to populate the relevant variables. Pathologic information was only used as a substitute when clinical information was not available. Missing variables were populated using established algorithms, as described in the online supporting information. Exclusion criteria included carcinoma in situ, distant metastasis at the time of diagnosis, and synchronous primary tumors, not including basal or squamous cell carcinoma of the skin.

Statistical Analysis

Each calculator was used to individually predict 5-year overall survival for 246 patients in our independent external validation cohort. The arithmetic average of the predictions from the 4 calculators was tested as a distinct (fifth) calculator, which is referred to as the "mean" in subsequent analyses. The agreement between these

TABLE 2. Patient Characteristics, N = 246

Characteristic	No. (%) or Mean ± SD[a]	No. Missing (%)	Calculators Using Characteristic
Demographics			
Age at diagnosis, y	60.0 ± 10.2	0 (0)	All 4
Sex		0 (0)	All 4
Women	56 (22.7)		
Men	190 (77.2)		
Race		0 (0)	LifeMath; MyCJ
Black	9 (3.6)		
Other	9 (3.6)		
White	228 (92.6)		
Smoking status		1 (0.4)	None
Current, in past 12 mo	166 (67.4)		
Former, > 12 mo	61 (24.7)		
Never	18 (7.3)		
ACE comorbidities		0 (0)	MyCJ
None	49 (19.9)		
Mild	112 (45.5)		
Moderate	60 (24.3)		
Severe	25 (10.1)		
ACE comorbidities without prior tumors		0 (0)	Leiden
None	53 (21.5)		
Mild	118 (47.6)		
Moderate	58 (23.5)		
Severe	17 (6.9)		
Tumor information			
Primary site		0 (0)	Leiden, MAASTRO
Glottic	115 (46.7)		
Supraglottic	131 (53.2)		
Subglottic	0 (0)		
AJCC overall stage		0 (0)	None
I	60 (24.3)		
II	34 (13.8)		
III	53 (21.5)		
IV	99 (40.2)		
SEER stage		0 (0)	MyCJ
Localized	103 (41.7)		
Regional	92 (37.2)		
Distant	51 (20.6)		
Tumor classification		0 (0)	Leiden; MAASTRO after transformation
T1	64 (25.9)		
T2	52 (21.1)		
T3	71 (28.8)		
T4	59 (23.9)		
Lymph node classification		0 (0)	LifeMath; transformation used for Leiden and MAASTRO
N0	156 (63.4)		
N1	26 (10.5)		
N1b	1 (0.4)		
N2	1 (0.4)		
N2a	2 (0.8)		
N2b	24 (9.7)		
N2c	33 (13.4)		
N3	3 (1.2)		
No. of positive lymph nodes		0 (0)	LifeMath
0	157 (63.8)		
1	33 (13.4)		
2	24 (9.7)		
3	19 (7.7)		
4	10 (4.0)		
≥5	3 (1.2)		
Greatest tumor dimension, cm		103 (41.8)	LifeMath
Mean ± SD, cm	2.6 ± 1.49		
<1.5	31 (12.6)		
1.5-2.5	34 (13.8)		
2.5-3.5	43 (17.4)		
≥3.5	35 (14.2)		

TABLE 2. *Continued*

Characteristic	No. (%) or Mean ± SD[a]	No. Missing (%)	Calculators Using Characteristic
Grade		0 (0)	MyCJ
1, Well differentiated	38 (15.4)	(Unknown category)	
2, Moderately differentiated	117 (47.5)		
3, Poorly differentiated	40 (16.2)		
4, Undifferentiated	1 (0.4)		
Unknown	50 (20.3)		
Extracapsular spread		0 (0)	LifeMath
Irrelevant, no lymph nodes	157 (63.5)	(Unknown category)	
No	18 (7.3)		
Yes	43 (17.4)		
Unknown	28 (11.3)		
Margin status		163 (66.2)	None
Negative	78 (31.7)		
Positive	5 (2.0)		
Hemoglobin, g/dL	13.8 ± 1.69	45 (18.2)	MAASTRO
Treatment information			
Initial treatment plan		0 (0)	None
Induction chemotherapy	70 (28.4)		
No induction chemotherapy			
Surgery	65 (26.4)		
Chemoradiation	56 (22.7)		
RT only	55 (22.3)		
Surgery within 4 mo[b]		0 (0)	MyCJ
No	166 (67.4)		
Yes	80 (32.5)		
Chemotherapy within 4 mo[b]		0 (0)	MyCJ
No	107 (43.4)		
Yes	139 (56.5)		
Radiation within 4 mo[b]		0 (0)	MyCJ
No	42 (17.0)		
Yes	204 (82.9)		
Total radiation dose, Gy	68.1 ± 4.07	105 (42.6)	None
Radiation EQD_{2T}	58.6 ± 1.74	222 (90.2)	MAASTRO

Abbreviations: ACE, Adult Comorbidity Evaluation index; AJCC, American Joint Committee on Cancer; EQD_{2T}, equivalent dose; Gy, grays; MyCJ, MyCancerJourney; SD, standard deviation; SEER, Surveillance, Epidemiology, and End Results of the US National Cancer Institute.
[a] Percentages include missing values.
[b] This was treatment delivered within 4 months of original treatment initiation.

predictions was compared using scatterplots, Spearman correlation coefficients, and the proportion of 5-year overall survival predictions that differed by <0.10 between separate calculators. The calibration of each calculator was assessed using Kaplan-Meier plots, which stratified patients into equal-sized quintiles according to calculator-predicted risk. The average predicted risk for each quintile was compared with the estimated 5-year survival for that quintile in a calibration plot. The discriminatory ability of each calculator was assessed using both the area under the receiver operating characteristic (ROC) curve for the binary outcome of survival at 5 years[36] and the C-index. Both the C-index and the 5-year area under the ROC curve (AUC) measure the concordance between the predicted risk and the survival outcome and are frequently reported in the literature.

To assess which factors may be responsible for discrepancies between the predicted outcomes and observed survival, a separate Cox model was fit for each possible factor and adjusted for the predicted risk as measured by the mean prediction. All tests for statistical significance used 95% confidence intervals (CIs) and were 2-sided. For treatment factors, a calibration plot was used to elucidate which treatment modalities were not well calibrated with the predicted risk. The method used to construct the calibration plot is described in the online supporting information.

RESULTS

The patient cohort represented the typical distribution and epidemiology of patients with laryngeal cancer. Most patients were Caucasian, current/former smokers, and men with no or mild medical comorbidities. Slightly less than one-half had a glottic subsite, whereas the majority of tumors originated in the supraglottic larynx.

The observed 5-year overall survival was 71.9% (95% CI, 65%-78%), whereas each calculator predicted significantly worse outcomes and contributed to a mean predicted 5-year overall survival of 47.7%. Figure 1 describes each calculator's predicted 5-year survival for our patient population. Visual assessment of the estimates suggests that MAASTRO and MyCancerJourney have a tendency to predict worse outcomes; both have more predictions clustered at lower values than the other calculators. MAASTRO exhibits the greatest discrepancy in its prognostication for patients with high-risk disease. MyCancerJourney has more variation in its predictions. Comparisons of MyCancerJourney and MAASTRO with the remaining 2 calculators demonstrated less agreement and were characterized by lower correlation coefficients.

The Leiden and LifeMath calculators demonstrated the closest correlation coefficient ($\rho = 0.816$), and the weakest association occurred between LifeMath and MyCancerJourney ($\rho = 0.644$). Supporting Table 8 (see online supporting information) reports the percentage of patients for which a selected pair of calculators predicted 5-year overall survival within 0.10. For example, if 1 calculator predicted a 50% 5-year overall survival for a given patient, then the paired calculator was considered to be in consensus if it predicted between 40% and 60% 5-year overall survival for the same patient. Prognostic consensus

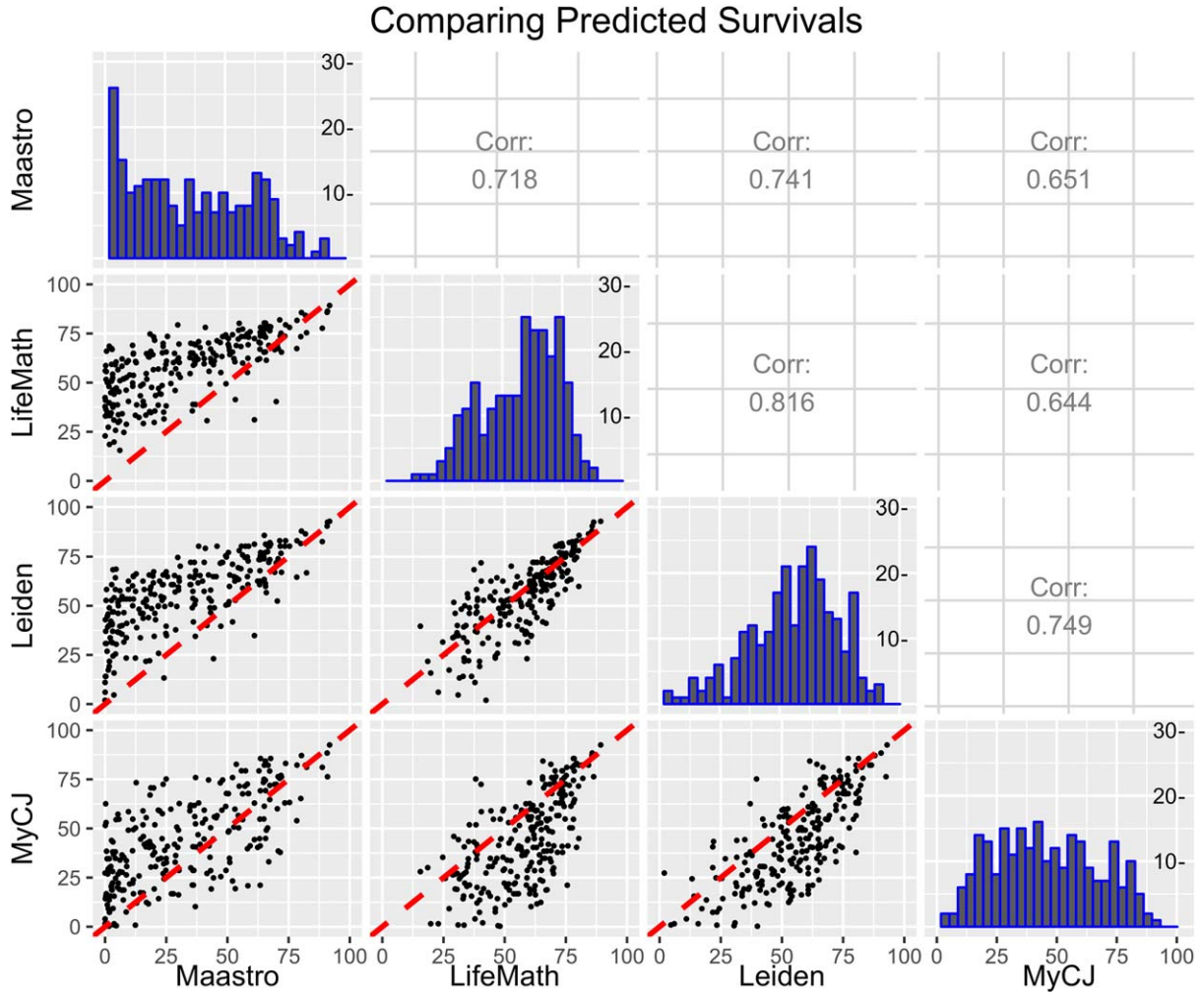

Figure 1. Scatterplots compare calculator predictions. The histograms on the diagonal illustrate the distribution of predicted 5-year survival from each calculator for the 246 patients from the University of Michigan. The scatterplots below the diagonal illustrate individual predictions of 5-year overall survival from pairs of calculators when applied to the University of Michigan patients. Points close to the 45-degree line are from patients with similar predictions from the 2 calculators. Correlation coefficients (Corr.) from the scatterplots are shown above the diagonal. MyCJ indicates MyCancerJourney.

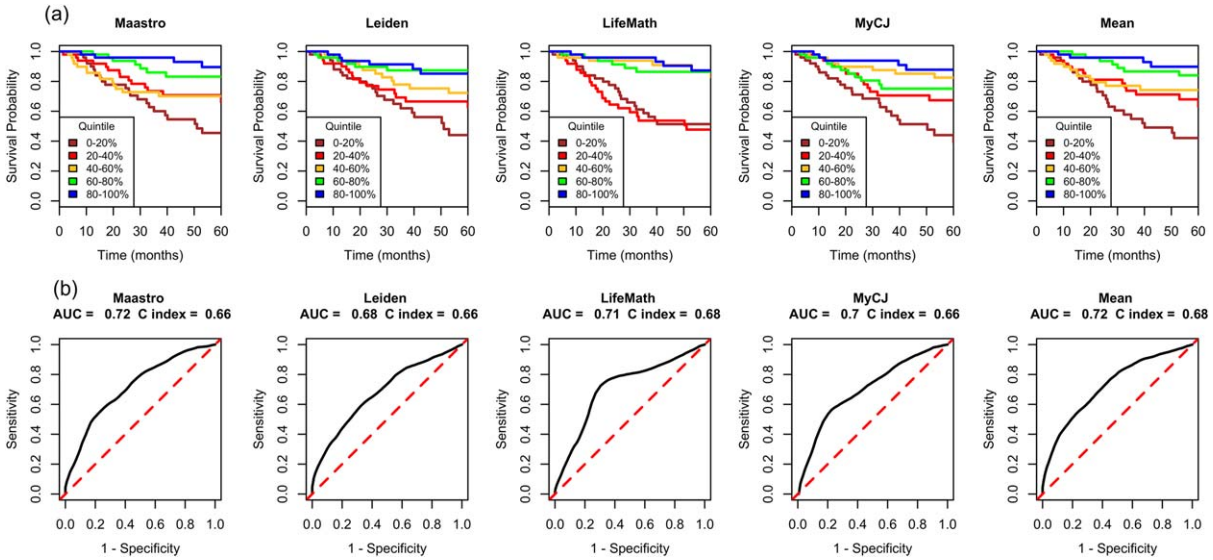

Figure 2. Calculator predictions are compared with observed outcomes. (a) Kaplan-Meier overall survival estimates, stratified by quintile of predicted 5-year survival, are shown for each calculator. (b) Receiver operating characteristic curves are illustrated for the sensitivity and specificity of each calculator-predicted 5-year survival compared with the observed 5-year survival. AUC indicates area under the receiver operating characteristic curve; MyCJ, MyCancerJourney.

between calculators was variable, and survival estimates typically agreed within 0.10 for < 50% of the cohort.

Figure 2a displays Kaplan-Meier survival plots that used calculator predictions to risk-stratify the patients into equal-portioned quintiles. The 4 calculators were reasonably effective in discriminating between these risk quintiles, but LifeMath and Leiden were less adroit in stratifying low-risk and high risk-patients, whereas MyCancerJourney and MAASTRO were less able to discriminate middle-tiered risk. The mean of the calculators more accurately stratified risk for each of the quintiles.

Figure 2b displays ROC curves and their accompanying AUC and C-index scores for each of the 4 calculators. These values provide informative measures of prognostic discrimination. The AUC values were similar for each of the 4 models and fell into a range from 0.68 (Leiden) to 0.72 (MAASTRO and Mean). The C-index values ranged from 0.66 (MAASTRO, Leiden, and MyCancerJourney) to 0.68 (LifeMath and Mean).

Calibration studies are summarized in Figure 3a and demonstrate lower survival estimates in relation to observed outcomes for each risk-stratified quintile. MAASTRO was especially pessimistic for high-risk patients, and LifeMath was well calibrated to observed outcomes for patients with middle-tiered risk.

Cox modeling identified male sex (hazard ratio [HR], 2.03; 95% CI, 1.00-4.11; $P = .031$) and initial planned treatment as additional factors that added predictive value, even after adjusting for the predicted risk from the calculators. Laryngeal subsite (glottis vs supraglottic) was not significantly related to survival after adjusting for calculator predictions (Supporting Table 9; see online supporting information). Figure 3b demonstrates that there was a lower than expected hazard of death for those patients in the cohort who received induction (bioselective) chemotherapy (HR, 0.46; 95% CI, 0.24-0.88; $P = .024$) or underwent primary surgical intervention (HR, 0.43; 95% CI, 0.21-0.90; $P = .024$) compared with those who received primary chemoradiation. Adjusted HRs for all other factors are listed in Supporting Table 9 (see online supporting information).

DISCUSSION

Available prognostic calculators generated variable predictions with inconsistent accuracy compared with observed outcomes in an external, prospectively maintained cohort of patients with laryngeal cancer. The 4 calculators were designed with various degrees of similarity, but substantial disparity in performance was evident. We have reported both the discriminatory ability and the calibration properties of the calculators, and these can be thought of as measures of the accuracy of the relative and absolute predictions, respectively. Relative predictions indicate whether the patients can be correctly ranked according to risk, and absolute predictions indicate whether the predicted probabilities of survival are correct. Both types of

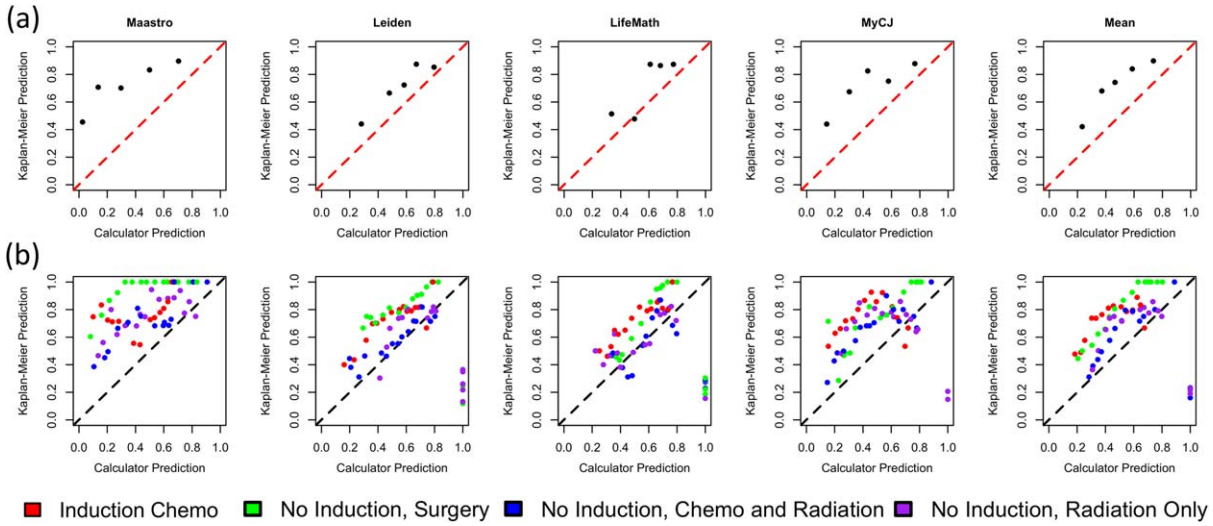

Figure 3. Calculator predictions were calibrated to observed outcomes. (a) The calibration of each calculator is illustrated. Each point represents a set of patients with similar predicted probability of 5-year survival (grouped by quintiles). The horizontal axes represent the average predicted probability of the group, and the vertical axes indicate the observed 5-year survival for the group obtained from Kaplan-Meier plots. A well calibrated calculator would have points near the diagonal line. (b) Calibration curves stratified by initial treatment plan are illustrated. Each point represents a small group of patients who had similar predicted 5-year survival and received the same initial treatment. MyCJ indicates MyCancerJourney.

predictions are important, and both were deficient for each of the 4 calculators.

Differences in patient cohorts, prognostic variables, statistical modeling, and inherent calculator limitations contributed to this deficiency and highlight many of the challenges associated with oncologic prognostication. LifeMath and MyCancerJourney both used the SEER population as their study cohort but exhibited the weakest association in risk prediction. This observation emphasizes the importance of incorporating and weighing prognostic variables accurately.[37] Comorbidity status has proven to be useful in predicting outcomes but was only integrated into the Leiden and MyCancerJourney calculators, providing further explanation for observed differences in calculator performance.[38] The AJCC established guidelines in 2016 that work to synchronize the development of future prognostic calculators, but meaningful discrepancies in the perceived importance of prognostic variables will likely persist.[25]

The MAASTRO calculator was developed in the Netherlands using a study cohort of 994 patients with laryngeal cancer between 1977 and 2008.[27] This calculator was designed for patients who only received radiation therapy as their primary treatment modality, and exclusion criteria included carcinoma in situ, distant metastases, and chemotherapy. MAASTRO's internal validation demonstrated an AUC of 0.73, whereas traditional TNM staging demonstrated an AUC of 0.62 for the study cohort.[27] Previous external validations yielded AUC values of 0.68, 0.74, 0.76, and 0.71. Compared with survival predictions for our entire external cohort, predictions for the 55 patients who exclusively received radiation demonstrated similar calibration, better correlation to predictions from the other calculators, and substantially improved AUC and C-index values (Supporting Table 10 and Supporting Figs. 1, 2, and 3; see online supporting information). Our study produced the first non-European external validation with an AUC of 0.72 when applied to all 246 patients and an AUC of 0.81 when applied to just those 55 patients who received radiation only. This calculator's pessimistic tendency could be explained by the evolving role of multimodality therapy, although the absolute incremental overall survival benefit of chemotherapy over radiation alone in laryngeal cancer is fairly modest (5%) outside of bioselection.[39] The calculator also has a noticeable, unfavorable HR for male sex and low hemoglobin count; however, many patients in our cohort were anemic but had better-than-predicted outcomes. In summary, the calculator's performance was fair considering the fundamental differences in the study cohort from which it was developed.

The LifeMath calculator was developed using 50,145 American patients with all sites of HNC in the SEER database between 1980 and 2009.[26] Internal

validation and external validation on 1362 patients from Massachusetts General Hospital were performed using an incomparable correlation metric. The current study provides an external validation with a C-index of 0.68 and an AUC of 0.71. The accuracy of this calculator may have been limited by its neglect of comorbidity status, a variable that is not available in the SEER database.

The Leiden calculator was developed using a study cohort of 1371 Dutch patients with several different sites of HNC between 1981 and 1998.[28] Internal validation demonstrated a C-index of 0.73, whereas an external validation on 598 patients from the Siteman Cancer Center yielded a C-index of 0.69. The current study provides an additional external validation with a C-index of 0.66. The model likely produced pessimistic survival estimates secondary to continued evolution and refinement of treatment over the past 4 decades.

MyCancerJourney used SEER data between 1973 and 1996 and Barnes-Jewish Hospital data between 1995 and 2001 to construct its model, but it did not have a publication to accompany its online calculator. The calculator performed with a C-index of 0.66 and an AUC of 0.70 in the current study. MyCancerJourney used a novel comorbidity metric to characterize patients. The majority of variations in outcome predictions could be explained by the grouping of patients according to treatment modality and other parameters, which resulted in substantial prognostic fluctuation.

Each of the calculators was designed using a training data set that included patients from over 35 years ago. Consequently, many of the patients in the study cohort did not receive treatment according to modern strategies. Reliance on older patient data likely contributed to the accumulative tendency to estimate worse than observed survival in the study cohort and underscores the importance of ensuring reliability and accuracy before adopting these into clinical practice.

In contrast to a previous study of currently available oral cavity cancer calculators, the patients with laryngeal cancer in our cohort demonstrated consistently better survival compared with the calculators' predicted outcomes.[30] This observation reinforces the hypothesis that individualized treatment paradigms for laryngeal cancer need to be considered when predicting survival. Whether this represents differences among the patients themselves or the individualized treatment approaches used remains speculative. However, both induction chemotherapy (bioselective) for subsequent treatment selection and/or primary surgical intervention were associated with survival benefits that were greater than expected after adjustments for calculator-estimated survival. Individualized treatment paradigms that integrate neoadjuvant bioselection are associated with a significant survival benefit and may account for the superior outcomes observed.[35] These findings reinforce the need for updated survival calculators and provide further evidence that oncologic interventions and institution-specific care are independent variables that affect the prognosis of patients with laryngeal cancer.

The absence of laryngeal cancer clinical practice guidelines that clearly specify preferred treatment modality may be contributing to disparities in calculator performance. Established practice guidelines help to optimize patient outcomes and help to standardize the value of care.[40-42] However, emerging evidence suggests that adherence to current guidelines established by the American Head and Neck Society and the National Comprehensive Cancer Network does not significantly improve outcomes.[43] Therapeutic regimens are complicated by the range of available treatment modalities and the need to individualize these based on patient, tumor, and institutional factors, making population-level recommendations challenging.[44]

There is a growing impetus for evaluating the value of cancer care. Judging value involves the balanced consideration of quality and outcome delivered and remains especially difficult to measure in oncology because of its multidisciplinary nature, need for prolonged follow-up, and consideration of post-treatment function as well as survival.[45] To address these challenges, prognostic calculators could be modeled from patient cohorts that received treatment according to optimized quality metrics. Once appropriately accurate, precise, and calibrated, such calculators could help to establish standardized expected outcomes for individually risk-stratified patients. Comparing an institution's outcomes with calculator predictions on an individualized and risk-stratified basis may serve as an effective method for evaluating and comparing relative quality and value.[45,46]

There are inherent limitations to this study, chiefly involving the single-institutional data, which may not reflect practice patterns or outcomes in other populations. The patients in our cohort also were treated by an experienced multidisciplinary team and had comprehensive follow-up, allowing appropriate salvage treatment for recurrence when necessary. Additional limitations may be linked to the improved outcomes generally associated with treatment in academic centers.[35,40,42,47-49] The accuracy of the SEER data, as well as the calculators relying on it, may be confounded if incurable patients treated for palliation were included or other inaccuracies in treatment

details were present. Missing variables were an additional source of error in this evaluation. However, derived values (for missing variable methods, see online supporting information) for hemoglobin, radiation dosage, and greatest tumor dimension rarely led to substantial differences in survival predictions, helping to mitigate this concern. The calculators do not consider the role of human papillomavirus, and the incidence and prognostic impact of human papillomavirus in laryngeal cancer is considerably lower than that in oropharyngeal cancer.[50]

There is a need for more accurate prognostic calculators that can predict individualized outcomes for patients with laryngeal cancer. Currently available prognostic calculators vary in their ability to consistently and accurately predict survival in an external cohort of patients with laryngeal cancer. Suboptimal reliability and accuracy limit the potential integration of existing individualized prediction tools into routine clinical practice. The calculators estimated significantly worse than observed survival among patients who received induction bioselective chemotherapy and underwent primary surgery, suggesting that modern treatment modalities must be better integrated into revised prediction tools. Deficiencies in calculator performance may be further explained by institutional variation in oncologic outcomes. Potential avenues to improve the performance of calculators include using contemporary patient cohorts, integrating biomarkers, and harnessing the promise of the genomic frontier as these data emerge. The use of statistical and machine-learning approaches when data sets are large is another intriguing possibility to create mechanisms that can more nimbly respond to exponentially complex and evolving data.[51]

These data raise questions about the inherent value of oncologic nomograms. We contend that they are useful for patients to estimate individualized prognosis and perhaps for comparing results across different cohorts. Predictive models that guide treatment selection might be of higher value and could witness increasing demand as the arsenal of available therapies continues to proliferate and individualize. Improved individualized calculators may help to assign value of oncologic care and will be critical in refining the ability of multidisciplinary teams to predict risk, plan shared treatment-related decision making, and counsel patients effectively.

FUNDING SUPPORT

This work was supported by the University of Michigan Head and Neck Specialized Program of Research Excellence program sponsored by the National Cancer Institute at the National Institutes of Health (P50 CA097248); the University of Michigan Cancer Center (P30 CA046592); the Cancer Biostatistics Training Program (T32 CA083654); and an internal University of Michigan Research Office MCubed grant.

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

AUTHOR CONTRIBUTIONS

Connor W. Hoban: Conceptualization, methodology, formal analysis, investigation, resources, data curation, writing–original draft, writing–review and editing, and visualization. **Lauren J. Beesley**: Methodology, software, formal analysis, investigation, data curation, writing–review and editing, and visualization. **Emily L. Bellile**: Methodology, software, formal analysis, investigation, data curation, writing–review and editing, and visualization. **Yilun Sun**: Methodology, software, formal analysis, investigation, data curation, writing–review and editing, and visualization. **Matthew E. Spector**: Conceptualization, methodology, investigation, and writing–review and editing. **Gregory T. Wolf**: Conceptualization, methodology, investigation, resources, data curation, writing–original draft, writing–review and editing, supervision, project administration, and funding acquisition. **Jeremy M. G. Taylor**: Conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing–original draft, writing–review and editing, supervision, project administration, and funding acquisition. **Andrew G. Shuman**: Conceptualization, methodology, investigation, resources, data curation, writing–original draft, writing–review and editing, supervision, project administration, and funding acquisition.

REFERENCES

1. Siegel R, Miller K, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin*. 2015;65:5-29.
2. Hutcheson KA, Lewin JS. Functional outcomes after chemoradiotherapy of laryngeal and pharyngeal cancers. *Curr Oncol Rep*. 2012; 14:158-165.
3. Forastiere AA, Zhang Q, Weber RS, et al. Long-term results of RTOG 91-11: a comparison of 3 nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol*. 2013;31:845-852.
4. Worden FP, Kumar B, Lee JS, et al. Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: response and survival positively associated with HPV16 copy number. *J Clin Oncol*. 2008;26:3138-3146.
5. Terrell JE, Fisher SG, Wolf GT. Long-term quality of life after treatment of laryngeal cancer. *Arch Otolaryngol*. 1998;124:964-971.
6. Amin MB, Greene FL, Edge SB, et al. The 8th ed. AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. 2017;67:93-99.
7. Bilimoria KY, Stewart AK, Winchester DP, Ko CY. The National Cancer Data Base: a powerful initiative to improve cancer care in the United States. *Ann Surg Oncol*. 2008;15:683-690.
8. Bradford CR, Kumar B, Bellile E, et al. Biomarkers in advanced larynx cancer. *Laryngoscope*. 2014;124:179-187.
9. Starska K, Forma E, Jozwiak P, et al. Gene and protein expression of glucose transporter 1 and glucose transporter 3 in human laryngeal cancer—the relationship with regulatory hypoxia-inducible factor-1α expression, tumor invasiveness, and patient prognosis. *Tumor Biol*. 2015;36:2309-2321.

10. Wolf GT, Hong WK, Fisher SG, et al. induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *N Engl J Med*. 1991;324:1685-1690.
11. Lefebvre JL, Andry G, Chevalier D, et al. Laryngeal preservation with induction chemotherapy for hypopharyngeal squamous cell carcinoma: 10-year results of EORTC trial 24891. *Ann Oncol*. 2012; 23:2708-2714.
12. Calais G, Pointreau Y, Alfonsi M, et al. Randomized phase III trial comparing induction chemotherapy using cisplatin (P) fluorouracil (F) with or without docetaxel (T) for organ preservation in hypopharynx and larynx cancer. Preliminary results of GORTEC 2000-01 [abstract]. *J Clin Oncol*. 2006;24(18 suppl). Abstract 5506.
13. Lonning PE. Breast cancer prognostication and prediction: are we making progress? *Ann Oncol*. 2007;18(suppl 8):viii3-viii7.
14. Vuralkan E, Akin I, Kuran G, Saka C, Akcayoz N, Alicura S. Prognostic factors in larynx cancers. *Kulak Burun Bogaz Ihtis Derg*. 2008; 18:221-226.
15. Sobin LH. TNM. evolution and relation to other prognostic factors. *Semin Surg Oncol*. 2003;21:3-7.
16. Rabin B, Gaglio B, Sanders T, et al. Predicting cancer prognosis using interactive online tools: a systematic review and implications for cancer care providers. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:1645-1656.
17. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol*. 2002;20:82-88.
18. Mackillop WJ, Quirt CF. Measuring the accuracy of prognostic judgments in oncology. *J Clin Epidemiol*. 1997;50:21-29.
19. Gorlia T, Wu W, Wang M, et al. New validated prognostic models and prognostic calculators in patients with low-grade gliomas diagnosed by central pathology review: a pooled analysis of EORTC/RTOG/NCCTG phase III clinical trials. *Neuro Oncol*. 2013;15: 1568-1579.
20. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. *N Engl J Med*. 2010;362:859-860.
21. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. *Patient Educ Couns*. 2010;78:329-336.
22. Higgins O, Sixsmith J, Barry MM, Domegan C. A Literature Review on Health Information-Seeking Behaviour on the Web: A Health Consumer and Health Professional Perspective. Insights into Health Communication. Stockholm, Sweden: European Centre for Disease Prevention and Control (ECDC); 2011. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/Literature%20review%20on%20health%20information-seeking%20-behaviour%20on%20the%20web.pdf. Accessed February 14, 2017.
23. McClymont K, Lee S, Schonberg M, Widera E, Miao Y, Smith AK. Usefulness and impact of online prognostic calculators. *J Am Geriatr Soc*. 2014;62:2444-2445.
24. Nolan T, Dack C, Pal K, et al. Patient reactions to a web-based cardiovascular risk calculator in type 2 diabetes: a qualitative study in primary care. *Br J Gen Pract*. 2015;65:e152-e160.
25. Kattan MW, Hess KR, Amin MB, et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin*. 2016;66:370-374.
26. Emerick KS, Leavitt ER, Michaelson JS, Diephuis B, Clark JR, Deschler DG. Initial clinical findings of a mathematical model to predict survival of head and neck cancer. *Otolaryngol Head Neck Surg*. 2013;149:572-578.
27. Egelmeer AG, Velazquez ER, de Jong JM, et al. Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiother Oncol*. 2011;100:108-115.
28. Datema FR, Ferrier MB, Vergouwe Y, et al. Update and external validation of a head and neck cancer prognostic model. *Head Neck*. 2013;35:1232-1237.
29. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting [serial online]. *BMC Med Res Methodol*. 2014;14:40.
30. Prince V, Bellile EL, Sun Y, et al. Individualized risk prediction of outcomes for oral cavity cancer patients. *Oral Oncol*. 2016;63:66-73.
31. Parks C, Heald J, Hall G, Kamaly-Asl I. Can the prognosis of individual patients with glioblastoma be predicted using an online calculator? *Neuro Oncol*. 2013;15:1074-1078.
32. Duffy SA, Taylor JM, Terrell JE, et al. Interleukin-6 predicts recurrence among head and neck cancer patients. *Cancer*. 2008;113:750-757.
33. Virani S, Bellile E, Bradford CR, et al. NDN and CD1A are novel prognostic methylation markers and differ by HPV status in patients with head and neck squamous carcinoma [serial online]. *BMC Cancer*. 2015;15:825.
34. Peterson LA, Bellile EL, Wolf GT, et al. University of Michigan Head and Neck Specialized Program of Research Excellence Program: cigarette use, comorbidities and prognosis in a prospective head and neck squamous cell carcinoma population. *Head Neck Surg*. 2016;381810-1820.
35. Wolf GT, Bellile E, Eisbruch A, et al. Survival rates using individualized bioselection treatment methods in patients with advanced laryngeal cancer. *JAMA Otolaryngol Head Neck Surg*. 2017;143:355-366.
36. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56: 337-344.
37. Wang SJ, Wissel AR, Ord CB, et al. Individualized estimation of conditional survival for patients with head and neck cancer. *Otolaryngol Head Neck Surg*. 2011;145:71-73.
38. Piccirillo JF, Lacy PD, Basu A, Spitznagel EL. Development of a new head and neck cancer-specific comorbidity index. *Arch Otolaryngol Head Neck Surg*. 2002;128:1172-1179.
39. Blanchard P, Baujat B, Holostenco V, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. *Radiother Oncol*. 2011;100:33-40.
40. Lewis CM, Weber RS. Standardizing treatment: a crisis in cancer care. *Otolaryng Clin North Am*. 2013;46:671-679.
41. Blayney DW, McNiff K, Eisenberg PD, et al. Development and future of the American Society of Clinical Oncology's Quality Oncology Practice Initiative. *J Clin Oncol*. 2014;32:3907-3913.
42. Lewis CM, Weber RS, Hanna EY. Quality of care in head and neck cancer. *Curr Oncol Rep*. 2011;13:120-125.
43. Graboyes EM, Townsend ME, Kallogjeri D, Piccirillo JF, Nussenbaum B. Evaluation of quality metrics for surgically treated laryngeal squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2016;142:1154-1163.
44. Forastiere AA, Weber RS, Trotti A. Organ preservation for advanced larynx cancer: issues and outcomes. *J Clin Oncol*. 2015;33:3262-3268.
45. Roman BR, Awad MI, Patel SG. Defining value-driven care in head and neck oncology [serial online]. *Curr Oncol Rep*. 2015;17:424.
46. Shellenberger TD, Madero-Visbal R, Weber RS. Quality indicators in head and neck operations: a comparison with published benchmarks. *Arch Otolaryngol Head Neck Surg*. 2011;137:1086-1093.
47. Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA*. 1998;280:1747-1751.
48. Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. *JAMA*. 2000;283:1159-1166.
49. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002; 346:1128-1137.
50. Fakhry C, Westra WH, Wang SJ, et al. The prognostic role of sex, race, and human papillomavirus in oropharyngeal and nonoropharyngeal head and neck squamous cell cancer. *Cancer*. 2017;123:1566-1575.
51. Meldolesi E, van Soest J, Damiani A, et al. Standardized data collection to build prediction models in oncology: a prototype for rectal cancer. *Future Oncol*. 2016;12:119-136.