Original Article

Individualized Survival Prediction for Patients with Oropharyngeal Cancer in the Human Papillomavirus Era

Lauren J. Beesley, PhD[1]; Peter G. Hawkins, MD, PhD[2]; Lahin M. Amlani[3]; Emily L. Bellile, MS[1]; Keith A. Casper, MD[3]; Steven B. Chinn, MD[3]; Avraham Eisbruch, MD[2]; Michelle L. Mierzwa, MD[2]; Matthew E. Spector, MD[3]; Gregory T. Wolf, MD[3]; Andrew G. Shuman, MD[3]; and Jeremy M. G. Taylor, PhD[1]

BACKGROUND: Accurate, individualized prognostication in patients with oropharyngeal squamous cell carcinoma (OPSCC) is vital for patient counseling and treatment decision making. With the emergence of human papillomavirus (HPV) as an important biomarker in OPSCC, calculators incorporating this variable have been developed. However, it is critical to characterize their accuracy prior to implementation. **METHODS**: Four OPSCC calculators were identified that integrate HPV into their estimation of 5-year overall survival. Treatment outcomes for 856 patients with OPSCC who were evaluated at a single institution from 2003 through 2016 were analyzed. Predicted survival probabilities were generated for each patient using each calculator. Calculator performance was assessed and compared using Kaplan-Meier plots, receiver operating characteristic curves, concordance statistics, and calibration plots. **RESULTS**: Correlation between pairs of calculators varied, with coefficients ranging from 0.63 to 0.90. Only 3 of 6 pairs of calculators yielded predictions within 10% of each other for at least 50% of patients. Kaplan-Meier curves of calculator-defined risk groups demonstrated reasonable stratification. Areas under the receiver operating characteristic curve ranged from 0.74 to 0.80, and concordance statistics ranged from 0.71 to 0.78. Each calculator demonstrated superior discriminatory ability compared with clinical staging according to the seventh and eighth editions of the American Joint Committee on Cancer staging manual. Among models, the Denmark calculator was found to be best calibrated to observed outcomes. **CONCLUSIONS**: Existing calculators exhibited reasonable estimation of survival in patients with OPSCC, but there was considerable variability in predictions for individual patients, which limits the clinical usefulness of these calculators. Given the increasing role of personalized treatment in patients with OPSCC, further work is needed to improve accuracy and precision, possibly through the identification and incorporation of additional biomarkers. ***Cancer* 2019;125:68-78**. © *2018 American Cancer Society.*

KEYWORDS: calculator, human papillomavirus (HPV), oropharyngeal squamous cell carcinoma (OPSCC), prediction, prognosis.

INTRODUCTION

Head and neck cancer (HNC) comprises a diverse group of malignancies that arise from multiple subsites and vary with regard to presentation, treatment, and prognosis.[1] Accurate prognostication in patients with HNC is critical to provide effective counseling and individualize optimal treatment. Prognostication in HNC typically is based on tumor-node-metastasis (TNM) characteristics as captured in traditional staging systems.[2] As the prognostic importance of additional biomarkers and clinical features has become better appreciated, the need for improved decision-making tools has become apparent.[3-8]

This is particularly salient for patients with oropharyngeal squamous cell carcinoma (OPSCC). Our understanding of the clinical behavior and management of OPSCC has evolved due to the increasing prevalence of human papillomavirus (HPV)–associated disease.[9] Compared with smoking-related and alcohol-related OPSCC, HPV positivity

Corresponding author: Jeremy M.G. Taylor, PhD, Department of Biostatistics, University of Michigan School of Public Health, M4509 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109; jmgt@umich.edu

[1]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan; [2]Department of Radiation Oncology, University of Michigan Medical School, Ann Arbor, Michigan; [3]Department of Otolaryngology–Head and Neck Surgery, University of Michigan Medical School, Ann Arbor, Michigan.

The first 2 authors contributed equally to this article.

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; 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.

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

DOI: 10.1002/cncr.31739, **Received:** May 23, 2018; **Revised:** August 7, 2018; **Accepted:** August 8, 2018, **Published online** October 6, 2018 in Wiley Online Library (wileyonlinelibrary.com)

is associated with an improved prognosis.[10,11] Due to its distinct presentation and prognosis, a novel staging system for HPV-associated OPSCC has been adopted for the eighth edition of the American Joint Committee on Cancer (AJCC) staging manual.[2,12,13] Although the AJCC eighth edition staging system for OSPCC accounts for HPV status, its authors explicitly avoided the incorporation of additional personalized features, favoring instead the more generalizable TNM framework.[14] Despite the prognostic value added by HPV status, subsets of patients continue to demonstrate outcomes discordant with their disease stage. As such, there remains a need for accurate risk stratification that incorporates HPV status as well as other personalized features.

With the goal of providing methods for individualized outcome prediction, numerous multifactorial, patient-specific calculators have been developed for multiple cancer types.[15,16] To assist with calculator evaluation, the AJCC has published criteria for the endorsement of any probability or risk model.[17] Among the prognostic calculators published in recent years, several have been developed for OPSCC in the HPV era.[18-21] Although these calculators incorporate additional prognostic factors such as age, smoking history, and TNM classifications in a heterogeneous manner, they all appear to demonstrate impressive accuracy in their respective study cohorts. However, to our knowledge, the generalizability, consistency, and accuracy of these calculators in predicting outcomes for individual patients in diverse populations remain unclear. These uncertainties are critical to address so as to optimally implement these tools into clinical practice. Thus, we sought to characterize and compare the accuracy and precision of existing OPSCC individualized prognostic calculators.

MATERIALS AND METHODS

Patient Selection

Patient data collection, extraction, and analysis were approved by the institutional review board of the University of Michigan. Patient data were extracted from 2 overlapping data sets at the University of Michigan: 1) a prospectively collected epidemiologic database of patients with HNC[6,7]; and 2) a database of patients treated with radiotherapy (RT) or chemoradiotherapy (CRT) for OPSCC.[22] Patient data that existed in both data sets were checked individually for agreement, and any conflicting values were resolved by reference to the primary medical record. Patients were diagnosed and treated from 2003 through 2016 as per institutional practices, which were consistent with National Comprehensive Cancer Network guidelines. Because the selected calculators were designed to be used prior to any intervention, clinical information was analyzed, with pathological data substituted only in cases for which clinical information was unavailable.

Prognostic Calculator Selection

Candidate prognostic calculators were identified through a systematic literature search and assessed for eligibility. Inclusion criteria stipulated applicability to OPSCC, provision of 5-year overall survival (OS) prediction, and the inclusion of HPV status as determined by detection of HPV DNA and/or p16.

We identified 4 calculators: 1) one that was developed at the MAASTRO Clinic ("MAASTRO")[20];

TABLE 1. Summary of Data Sets and Models for Each Calculator

Calculator	Cancers In Training Data Set	Training Data Set	Validation Data Set	Reported C-Indices and/or AUCs
MAASTRO[20]	OPSCC	168 patients, MAASTRO Clinic: 2000-2011	189 patients, VUMC: 2000-2006	Training C-index: 0.82; external C-index: 0.73
RTOG[18]	OPSCC	493 patients, multiple North American centers on RTOG 0129 and 0522: 2002-2009	153 patients, multiple North American centers on RTOG 9003: 1991-1997	Training C-index: 0.76[a]; external C-index: 0.68
Erasmus[21]	Multiple HNC subsites	Cohort size unknown, EMC: 2006-2013	None	Not available
Denmark[19]	OPSCC	1542 patients, eastern Denmark: 2000-2014	None	Training AUC: 0.8[a]

Abbreviations: AUC, area under the receiver operating characteristic curve; C-index, concordance statistic; Denmark, eastern Denmark; EMC, Erasmus Medical Center; HNC, head and neck cancer; MAASTRO, MAASTRO Clinic; OPSCC, oropharyngeal squamous cell carcinoma; RTOG, Radiation Therapy Oncology Group; VUMC, Vrije University Medical Center.
[a]Uncorrected values; bias-corrected values are available in the corresponding published reports.[18,19]

2) one that was based on data from Radiation Therapy Oncology Group ("RTOG") trials[18]; 3) one that was based on patients treated in eastern Denmark ("Denmark")[19]; and 4) one that was developed at Erasmus Medical Center ("Erasmus").[21,23] Table 1 summarizes the data sources of each calculator, with additional details provided in Supporting Tables 1 to 4. The definition of HPV status varied among calculators. For model analysis, we applied whichever definition was used in the development and validation of each respective model. For example, for the MAASTRO model, HPV DNA was used to classify patients as HPV positive or HPV negative, regardless of p16 results. For the Erasmus calculator, because to our knowledge the method of HPV status definition has not been reported to date, we elected to use p16 as is recommended in the eighth edition of the AJCC staging manual.[24] In our data set, HPV DNA and p16 results were discordant for only 23 patients (2.7%). These patients were differentially classified as either HPV positive or HPV negative depending on the calculator being evaluated. Inputs for each calculator are summarized in Table 2 and Supporting Tables S5–7. Additional details are presented in the Supporting Information.

Statistical Analysis

The predicted 5-year OS was computed for each patient using each calculator. Agreement between predictions was assessed using scatter plots and Spearman correlation coefficients. The ability of each calculator to risk-stratify patients was assessed by dividing subjects into quintiles (5 equally sized groups) based on their predicted risk and using Kaplan-Meier methods to plot their corresponding observed OS. The discriminatory ability of each model was assessed by calculating areas under the receiver operating characteristic (ROC) curve (AUCs) at 5 years[25] and concordance statistics (C-indices).[26] Absolute prediction accuracy was assessed by calibration plots generated using a "moving window" method. For this analysis, patients were divided into multiple overlapping risk groups with calculator-predicted 5-year survival probabilities within a moving window of 0.25.[27] The average predicted 5-year OS for each group then was plotted against the corresponding Kaplan-Meier estimates of 5-year OS.

TABLE 2. Input Factors for Each Calculator

	MAASTRO	RTOG	Erasmus	Denmark
Age	Not included	Dichotomized (≤50 y vs >50 y)	Continuous	Continuous
Sex	Included	Not included	Included	Not included
Comorbidity	ACE-27, dichotomized (none-mild vs moderate-severe)	Not included	ACE-27, ordinal (grades 0-3)	Not included
Performance status	Not included	ECOG (0 vs 1)	Not included	ECOG, ordinal (0-4)
Smoking status	Pack-y, ordinal (none, moderate [1-30 pack-y], or heavy [>30 pack-y])	Pack-y, dichotomized (≤10 pack-y vs >10 pack-y)	Not included	Pack-y, continuous
Education	Not included	Dichotomized (≤high school vs >high school)	Not included	Not included
Determination of HPV status	HPV DNA	p16	Not specified[a]	HPV DNA and p16
Stage[b]	Ordinal T classification and dichotomized N classification (N0-N2a vs N2b-N3)	Dichotomized T classification (T2-3 vs T4) and N classification (N0-2b vs N2c-3)	Ordinal T, N, and M classifications	Ordinal T and N classifications
Hemoglobin	Continuous	Dichotomized (≤13.5 g/dL vs >13.5 g/dL for men, and ≤12.5 g/dL vs >12.5 g/dL for women)	Not included	Not included
Treatment	Not included	Not included	Receipt of chemotherapy, dichotomized (yes vs no)	Categorical (RT, CRT, palliative, or no treatment)

Abbreviations: ACE-27, Adult Comorbidity Evaluation-27; CRT, chemoradiotherapy; Denmark, eastern Denmark; ECOG, Eastern Cooperative Oncology Group; Erasmus, Erasmus Medical Center; HPV, human papillomavirus; MAASTRO, MAASTRO Clinic; RT, radiotherapy; RTOG, Radiation Therapy Oncology Group.
[a]To the authors' knowledge, the method used to define HPV status for the Erasmus calculator has not been reported to date. For this analysis of the Erasmus calculator, HPV status was defined using p16.
[b]Stage herein refers to American Joint Committee on Cancer seventh edition criteria, although the RTOG calculator allows for use of the seventh or eighth edition.

TABLE 3. Patient Characteristics (N = 856)

Characteristic	Value
Mean age at diagnosis, y (SD)	58.4 (9.65)
Sex, no. (%)	
Male	725 (84.7)
Female	131 (15.3)
Hemoglobin, mean (SD)	13.9 g/dL (1.53)
Unknown, no. (%)	119 (13.9)
ACE-27 comorbidity, no. (%)	
None	206 (24.1)
Mild	259 (30.3)
Moderate	112 (13.1)
Severe	44 (5.1)
Unknown	235 (27.4)
Smoking status, no. (%)	
Never	283 (33.1)
Former	297 (34.7)
Current	271 (31.7)
Unknown	5 (0.5)
Pack-y	
Mean (SD)	20.3 (25.2)
Median (range)	10 (0-150)
Unknown, no. (%)	42 (4.9)
ECOG performance status, no. (%)	
0	409 (47.8)
1	52 (6.1)
2	1 (0.1)
Unknown	394 (46.0)
Maximum level of education, no. (%)	
≤High school	379 (44.3)
>High school	162 (18.9)
Unknown	315 (36.8)
Race, no. (%)	
White	348 (40.7)
Black	10 (1.2)
Other	2 (0.2)
Unknown	496 (57.9)
T classification (AJCC 7th), no. (%)	
T0/Tis	0
T1	195 (22.8)
T2	284 (33.2)
T3	138 (16.1)
T4	236 (27.6)
Unknown	3 (0.3)
T classification (AJCC 8th), no. (%)	
T0/Tis	0
T1	195 (22.8)
T2	284 (33.2)
T3	138 (16.1)
T4	236 (27.6)
Unknown	3 (0.3)
N classification (AJCC 7th), no. (%)	
N0	110 (12.9)
N1	90 (10.5)
N2	45 (5.3)
N2a	65 (7.6)
N2b	327 (38.2)
N2c	139 (16.2)
N3	80 (9.3)
N classification (AJCC 8th), no. (%)	
N0	108 (12.6)
N1	384 (44.9)
N2	177 (20.7)
N3	79 (9.2)
Unknown[a]	108 (12.6)
Clinical stage (AJCC 7th), no. (%)	
0	0
I	18 (2.1)
II	38 (4.4)
III	98 (11.4)
IV	701 (81.9)
Unknown	1 (0.1)
Clinical stage (AJCC 8th), no. (%)	
I	275 (32.1)
II	113 (13.2)
III	158 (18.5)
IV	65 (7.6)
Unknown[b]	245 (28.6)
Viral markers, no. (%)	
HPV+/p16+	394 (46.0)
HPV+/p16-	8 (0.9)
HPV+/p16 missing	93 (10.9)
HPV-/p16+	15 (1.8)
HPV-/p16-	68 (7.9)
HPV-/p16 missing	13 (1.5)
HPV missing/p16+	21 (2.5)
HPV missing/p16-	10 (1.2)
HPV missing/p16 missing	234 (27.3)
Treatment modality, no. (%)	
CRT	651 (76.1)
RT alone	42 (4.9)
Surgery plus adjuvant CRT	33 (3.9)
Surgery plus adjuvant RT	35 (4.0)
Surgery alone	23 (2.7)
Chemotherapy alone	14 (1.6)
Palliative, unknown	58 (6.8)

Abbreviations: +, positive; -, negative; ACE-27, Adult Comorbidity Evaluation-27; AJCC 7th, American Joint Committee on Cancer seventh edition; AJCC 8th, American Joint Committee on Cancer eighth edition; CRT, chemoradiotherapy; ECOG, Eastern Cooperative Oncology Group; HPV, human papillomavirus; RT, radiotherapy; SD, standard deviation.
[a]The AJCC 8th N classification for these patients was either N1 or N2.
[b]The AJCC 8th group stage was unknown in these patients due to unknown N classification or HPV status.

To handle missing covariates, we used a multiple imputation approach based on a multistate cure model (Supporting Fig. 1).[28] We used the substantive model compatible–fully conditional specification approach, which involves imputing each covariate with missing values iteratively from a distribution proportional to the likelihood for the multistate cure model and a model for that covariate given the other covariates.[29] Additional details regarding this novel methodology are provided in the Supporting Information.

RESULTS

A total of 856 patients were identified for calculator assessment. Patient and disease characteristics are summarized in Table 3. The median follow-up was 61 months. Approximately 74% of patients and 61% of patients, respectively, had at least 3 years and 5 years of follow-up, or died before 3 years and 5 years. Figure 1 shows the distribution of predictions from the calculators (on the diagonal), scatter plots demonstrating the agreement between pairs of calculators (below the diagonal), and correlation coefficients between predictors (above the diagonal). The distributions showed similar ranges of predictions for each calculator, with the exception of MAASTRO, which tended to predict lower 5-year OS. The scatter plots and correlation coefficients demonstrated variable degrees of association between calculator pairs. The Denmark and Erasmus calculators demonstrated the strongest correlation with each other (rho, 0.907), whereas the RTOG and MAASTRO calculators demonstrated the weakest correlation with each other (rho, 0.634). Table 4 lists the percentages of patients for whom each pair of calculators yielded predictions within 10% of each other. For only 3 of the 6 pairings were predicted outcomes within 10% of each other for at least 50% of patients.

We next sought to characterize the relative accuracy of each calculator by assessing its ability to stratify patients into risk categories. To do this, we divided patients into equally sized quintile groups based on predicted risk and generated Kaplan-Meier plots of their observed outcomes. Although all calculators yielded the expected distribution of survival outcomes, the MAASTRO and RTOG calculators exhibited relatively poorer differentiation of the 2 lowest risk groups (Fig. 2).

We next generated ROC curves and calculated AUCs and C-indices (Fig. 3). There was a range of discriminatory ability among the 4 calculators, with the Denmark model demonstrating the best performance (AUC, 0.80; C-index, 0.78) and the MAASTRO calculator yielding the worst (AUC, 0.74; C-index, 0.71). By comparison, AUCs and C-indices based on clinical stage alone were 0.53 and 0.51, respectively, using the AJCC seventh edition criteria, and were 0.72 and 0.68, respectively, using the AJCC eighth edition criteria (see Supporting Fig. 2), indicating inferior discriminatory ability compared with each of the 4 calculators.

We next assessed the absolute predictive accuracy, or calibration, of these calculators by plotting calculator-predicted versus Kaplan-Meier–estimated rates of 5-year OS (Fig. 4). The Denmark calculator demonstrated the best calibration whereas the MAASTRO calculator underestimated survival and the RTOG and Erasmus calculators overestimated survival for patients at intermediate and low risk, respectively.

DISCUSSION

Using a large patient database, we assessed individualized calculators designed to predict 5-year OS in patients with OPSCC. Although these models demonstrated reasonable risk stratification and discriminatory abilities, we observed suboptimal consistency of predicted outcomes between calculators. In general, the AUCs and C-indices computed in the current analysis were lower than the training values and similar to or higher than the external cohort values previously reported for these calculators, although no external AUC was provided for the Denmark calculator and to our knowledge no training or validation values have been published for the Erasmus one. AUCs and C-indices from each calculator were higher than those obtained from the seventh and eighth editions of the AJCC staging manual. The assessment of absolute predictive accuracy indicated that the Denmark calculator had the best calibration.

Although the OPSCC calculators exhibited reasonable accuracy in the current study, the variability observed among them indicates that the predicted prognosis for a given patient could vary substantially depending on which calculator is used. Communication of accurate information is vital to patient counseling and shared decision making, which has been shown to improve perceived quality of care and patient satisfaction.[30-32] The results of the current study suggest that although currently available OPSCC calculators may outperform TNM staging, they nevertheless pose the risk of providing inaccurate predictions of prognosis. These results underscore another important issue regarding risk prognostication and patient education. In addition to OS, there are numerous other oncologic and functional outcomes that can be predicted by individualized risk calculators.[15] How to optimally integrate and present different predicted outcomes is unclear at this time, and future work should seek to better define physician and patient preferences and abilities to synthesize this complex information.

In addition to patient education, the accurate prediction of prognosis is becoming increasingly important for determining treatment recommendations and for identifying patients in need of treatment optimization. This is particularly relevant to recent efforts to deintensify treatment in patients with low-risk OPSCC.[33-35] Although these approaches consistently incorporate HPV

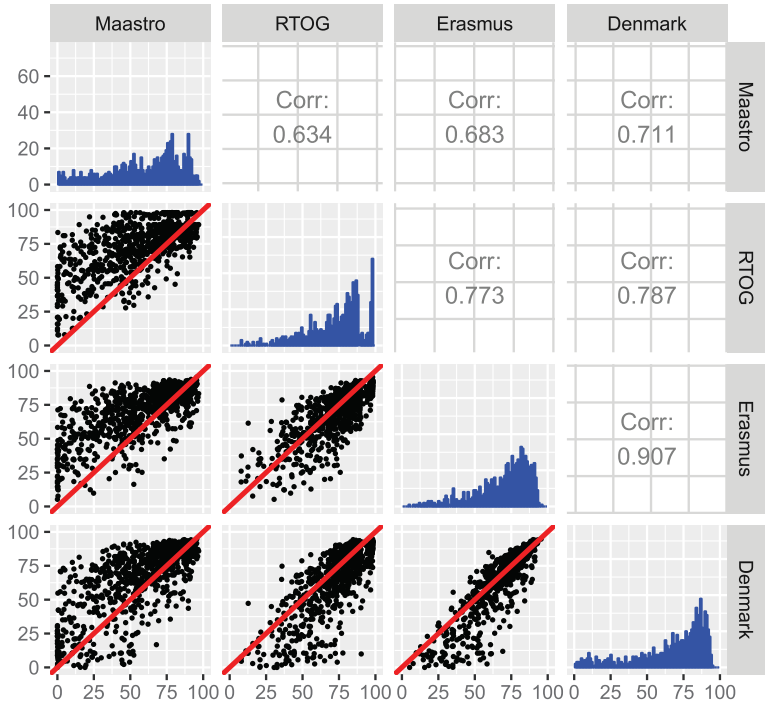

Figure 1. Distribution of predicted outcomes (diagonal), scatter plots (below diagonal), and correlation coefficients (above diagonal) for analyzed calculators. Corr indicates correlation; RTOG, Radiation Therapy Oncology Group.

status, they heterogeneously account for other prognostic factors.[36] The current study results indicate that individualized risk calculators improve prognostication beyond AJCC staging, even the eighth edition, which incorporates HPV status, and may represent a better method for selecting patients for deintensification. However, the observed variability among calculators suggests that the optimal combination of prognostic variables has not been identified to date. Therefore, further work is needed to improve prognostic calculators with which to guide individualized treatment.

The variability among calculators was likely a function of differences in the variables included, the manner in which the variables were modeled, and the data sources from which the variables were identified. For example, performance of the MAASTRO model may have been impaired by the inclusion of patients treated only with RT or CRT, and not surgery.[20] However, in a separate analysis of the 692 patients treated with RT or CRT in the current study, the MAASTRO calculator again demonstrated inferior performance in comparison with the other models (Supporting Figs. 3-5, Supporting Table S8). In addition, the MAASTRO calculator was developed using the smallest training cohort and was the only calculator not to include age, a variable that has been correlated with OS in multiple studies.[37]

The RTOG calculator was developed using data from patients treated on RTOG trials 0129 and 0522, and validated in patients treated on the RTOG 9003 trial.[18] Although these trials included patients with HNC of multiple subsites, only patients with OPSCC were used to generate this model. This data source is advantageous in that patient, treatment, and outcome data would be expected to be relatively homogenous, complete, and accurate. However, the inclusion of only those patients treated on clinical trials and the exclusion of patients with an Eastern Cooperative Oncology Group performance status >1 may have limited generalizability.[38,39] In

TABLE 4. Percentage of Patients With Predicted Survival Rates Within 10% of Each Other for Pairs of Calculators

	RTOG	Erasmus	Denmark
MAASTRO	35.6%	42.6%	41.5%
RTOG		62.9%	61.8%
Erasmus			78.9%

Abbreviations: Denmark, eastern Denmark; Erasmus, Erasmus Medical Center; MAASTRO, MAASTRO Clinic; RTOG, Radiation Therapy Oncology Group.

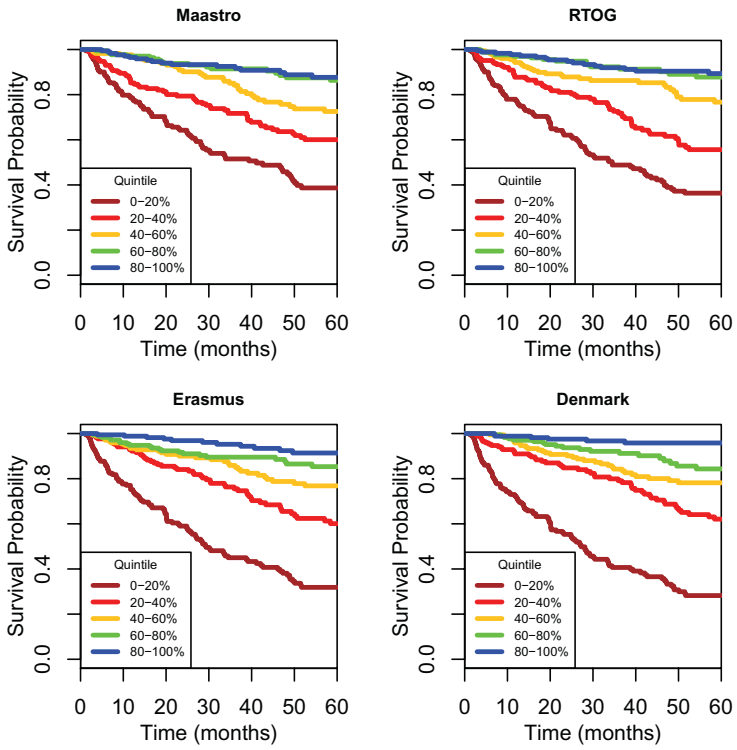

Figure 2. Kaplan-Meier plots of 5-year overall survival in risk-stratified quintile groups of equal patient numbers defined by each model. RTOG indicates Radiation Therapy Oncology Group.

addition, in contrast to the other calculators, which more commonly modeled variables as ordinal or continuous, the RTOG model used dichotomization of all variables, which may have impaired performance.

The Erasmus model is related to a previously published calculator based on patients treated at Leiden University Medical Center.[40] This "Leiden" calculator performed poorly in an initial analysis (data not shown), most likely because it does not incorporate HPV status. Therefore, we chose to evaluate the Web-based Erasmus calculator, which does include HPV status and is based on a larger, more modern cohort treated at Erasmus Medical Center.[21] This was the only calculator we evaluated for which a full description has not been published to date, although some cohort details have been reported in an analysis of patients with laryngeal cancer.[23] Although the online interface is convenient, the lack of a corresponding publication describing study patients and model performance limits its clinical usefulness. The Erasmus calculator was the only calculator derived from patients with HNC of multiple subsites, with different coefficients being assigned for each subsite. Because the prognostic impact of a given factor could vary among subsites, it is possible that this model could be less accurate than a model developed using only patients with OPSCC or a model that allowed coefficients of other factors to vary by subsite. In addition, this was the only calculator that did not include smoking status, which is an important prognostic factor in both patients with HPV-related and non–HPV-related OPSCC.[7,10,41]

Of the 4 calculators, the Denmark calculator yielded the highest AUC and C-index, and was the best calibrated. Although to the best of our knowledge there is no consensus, an AUC of ≥0.80 commonly is used to denote "good" discriminatory ability with a high potential for clinical usefulness.[42,43] With a value of 0.80, the Denmark calculator met this threshold. The Denmark calculator was 1 of 2 models to include treatment modality and did so in a more detailed manner than the Erasmus calculator.

Although these calculators all include HPV status, there are important differences with regard to how this is defined for each. The MAASTRO calculator defines HPV positivity based on the presence of HPV DNA, whereas the RTOG calculator considers only p16. The Denmark calculator considers p16 and HPV DNA separately. The Web site for the Erasmus calculator does not stipulate how HPV status should be determined and, as

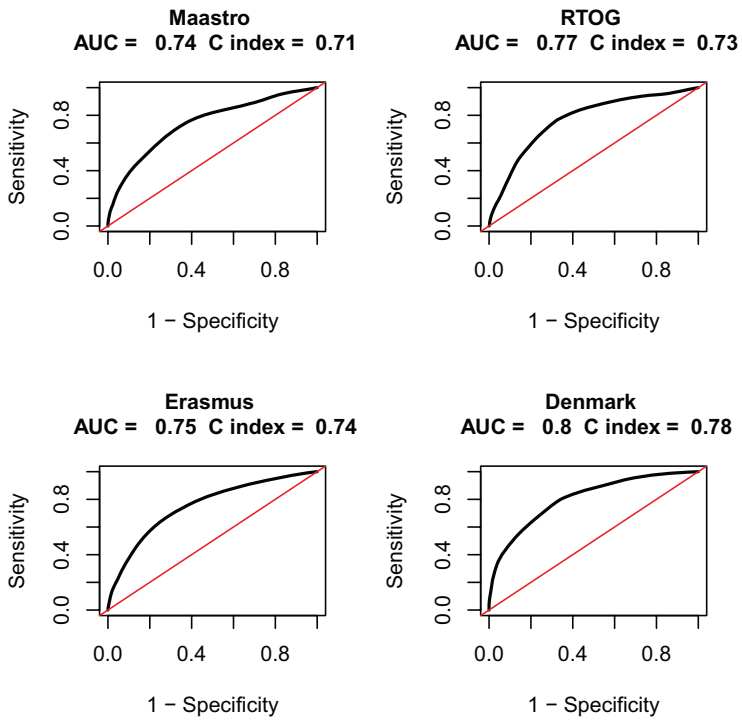

Figure 3. Receiver operating characteristic (ROC) curves with associated areas under the ROC and concordance statistics (C-indices) for each calculator. RTOG indicates Radiation Therapy Oncology Group.

discussed above, there is no corresponding publication describing how this was defined in model building. It is important to consider how this variability affects the implementation of these calculators. HPV DNA and p16 positivity are correlated but do not completely overlap,[44,45] and both tests are not performed consistently at all centers. The eighth edition of the AJCC staging manual stipulates that p16 staining be used preferentially to define HPV status.[24] Until it is clear that HPV DNA adds prognostic information beyond p16, it most likely is reasonable to use p16 to define HPV status when using these calculators, with HPV DNA being used as an alternative when p16 is unavailable. If neither HPV DNA nor p16 status are available, caution should be exercised if using these calculators. For staging purposes, in the absence of p16 and HPV status, the eighth edition of the AJCC staging manual recommends staging as if the patient were HPV negative.[24] A similar approach could be taken when using a calculator, although one also could consider calculating an average of predicted outcomes for HPV-positive and HPV-negative iterations or making an educated estimation of the patient's HPV status based on clinical and demographic factors. If any of these is done, the patient should be informed of the limitations of the resulting prediction. This challenge can arise when any prognostic variable is not available, which may limit the usefulness of these models.

The current study is similar to our previous efforts to analyze individual predictors of survival in patients with oral cavity and laryngeal cancer.[27,46] In the current study, we found OPSCC calculators to yield higher AUCs and C-indices than did oral cavity and larynx models. This is likely due to the availability and incorporation of HPV, a robust biomarker, in the OPSCC calculators. There also were important differences in the data sets used for the larynx and oral cavity compared with OPSCC calculators. Certain of the larynx and oral cavity calculators were developed using the Surveillance, Epidemiology, and End Results database, which, although large, to our knowledge lacks details regarding certain prognostic factors, thereby limiting its usefulness for calculator development.[47] In contrast, the data sets used for the OPSCC calculators were based on institutional and cooperative group databases that contained more detailed information regarding potential prognostic variables. However, all these studies demonstrated substantial variability among calculators that limits their current value.

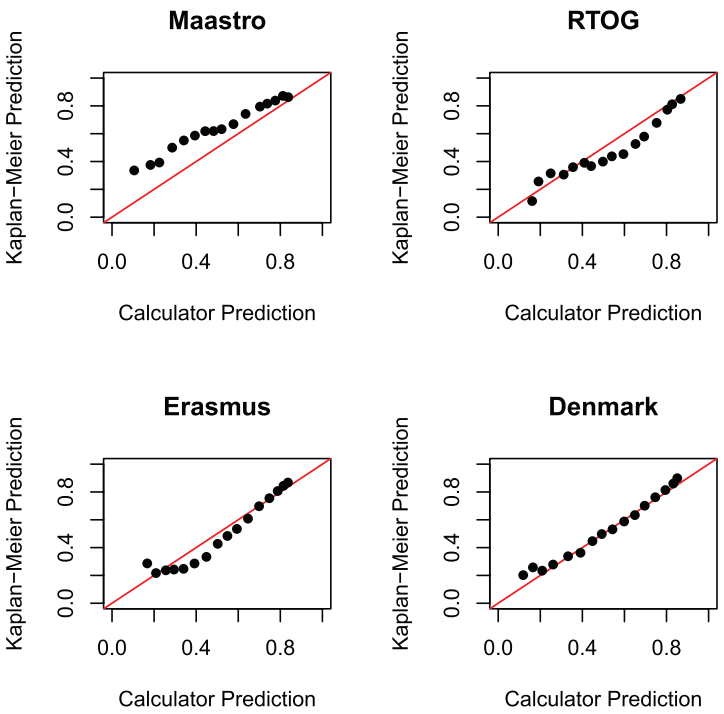

Figure 4. Calibration plots of predicted outcomes obtained using each model versus observed outcomes. Each dot represents a group of patients within a different risk group, or window, as defined using each calculator. RTOG indicates Radiation Therapy Oncology Group.

There are limitations to the current study that require consideration. Although the patient data set herein was reasonably comprehensive, there was some degree of missing data (Table 3). Because each calculator required different inputs, excluding patients with any missing variable would have reduced the number of evaluable patients substantially. As such, we elected to compensate for the missing covariates via a substantive model compatible–fully conditional specification multiple imputation technique.[29] Although this methodology is rigorous and theoretically justified, it is based on additional statistical models that, although fit to the data, do add additional modeling assumptions. A strength of this approach is that it uses known data to impute related missing covariates. Another limitation to the data set in the current study was that it was comprised of patients who were seen at a single academic institution in the midwestern United States. It is interesting to note that calculators derived from US populations were not found to be better calibrated to the current study patients compared with those developed in Europe. Regardless, it remains unclear how generalizable the results of the current study may be in relation to other institutions. It also is important to note that we evaluated predictions at only one time point: 5 years. Although each calculator also predicts survival at other intervals, 5 years was chosen for this analysis because it is the only time point shared by all 4 models. As a result, we were unable to draw conclusions regarding the performance of these calculators in predicting survival prior to or beyond 5 years.

Although existing OPSCC calculators have demonstrated reasonable predictive accuracy and superior discriminatory ability compared with AJCC TNM staging, a lack of precision limits their usefulness for predicting risk in individual patients. Additional work is needed to improve accuracy and consistency, possibly through the identification of additional biomarkers.[48,49] With further refinement, multifactorial, patient-specific risk calculators may prove beneficial for individualizing care and improving outcomes in individuals with OPSCC.

FUNDING SUPPORT

Supported by grants from the National Cancer Institute at the National Institutes of Health (P50 CA097248 [University of Michigan Head and Neck Specialized Program of Research Excellence], P30 CA46592 [University of Michigan Cancer Center], IU01 CA183848 [to Avraham Eisbruch], and T32 CA083654 [Cancer Biostatistics Training Program]) and by a Young Investigator Award from the American Head and Neck Society (to Andrew G. Shuman).

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

AUTHOR CONTRIBUTIONS

Lauren J. Beesley: Data curation, formal analysis, methodology, software, visualization, writing–original draft, and writing–review and editing. **Peter G. Hawkins**: Data curation, investigation, methodology, writing–original draft, and writing–review and editing. **Lahin Amlani**: Data curation and methodology. **Emily L. Bellile**: Data curation, formal analysis, methodology, software, and writing–review and editing. **Keith A. Casper**: Investigation, methodology, and writing–review and editing. **Steven B. Chinn**: Investigation, methodology, and writing–review and editing. **Avraham Eisbruch**: Investigation, methodology, resources, and writing–review and editing. **Michelle L. Mierzwa**: Investigation, methodology, and writing–review and editing. **Matthew E. Spector**: Investigation, methodology, and writing–review and editing. **Gregory T. Wolf**: Investigation, methodology, resources, and writing–review and editing. **Andrew G. Shuman**: Conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, supervision, writing–original draft, and writing–review and editing. **Jeremy M.G. Taylor**: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, resources, supervision, writing–original draft, and writing–review and editing.

REFERENCES

1. Harrison LB, Sessions RB, Hong WK. Head And Neck Cancer: A Multidisciplinary Approach. Philadelphia: Lippincott Williams & Wilkins; 2009.
2. O'Sullivan B, Lydiatt WM, Haughey BH, et al. Chapter 10. HPV-Mediated (p16+) Oropharyngeal Cancer. In: Amin MB, Edge SB, Greene FL, et al, eds. AJCC Cancer Staging Manual. 8th ed. New York: Springer International Publishing; 2017:113-121.
3. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer*. 2018;18:269-282.
4. Hayes DN, Van Waes C, Seiwert TY. Genetic landscape of human papillomavirus–associated head and neck cancer and comparison to tobacco-related tumors. *J Clin Oncol*. 2015;33:3227-3234.
5. Datema FR, Ferrier MB, van der Schroeff MP, Baatenburg de Jong RJ. Impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. *Head Neck*. 2010;32:728-736.
6. Duffy SA, Taylor JM, Terrell JE, et al. Interleukin-6 predicts recurrence and survival among head and neck cancer patients. *Cancer*. 2008;113:750-757.
7. 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*. 2016;38:1810-1820.
8. Sobin LH. TNM: evolution and relation to other prognostic factors. *Semin Surg Oncol*. 2003;21:3-7.
9. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol*. 2011;29:4294-4301.
10. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363:24-35.
11. Fakhry C, Westra WH, Li S, et al. Improved survival of patients with human papillomavirus–positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst*. 2008;100:261-269.
12. O'Sullivan B, Huang SH, Su J, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. *Lancet Oncol*. 2016;17:440-451.
13. Huang SH, Xu W, Waldron J, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus–related oropharyngeal carcinomas. *J Clin Oncol*. 2015;33:836-845.
14. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and neck cancers–major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2017;67:122-137.
15. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16:e173-e180.
16. Rabin BA, 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. Kattan MW, Hess KR, Amin MB, et al; members of the AJCC Precision Medicine Core. 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.
18. Fakhry C, Zhang Q, Nguyen-Tan PF, et al. Development and validation of nomograms predictive of overall and progression-free survival in patients with oropharyngeal cancer. *J Clin Oncol*. 2017;35:4057-4065.
19. Larsen CG, Jensen DH, Carlander AF, et al. Novel nomograms for survival and progression in HPV+ and HPV- oropharyngeal cancer: a population-based study of 1,542 consecutive patients. *Oncotarget*. 2016;7:71761-71772.
20. Rios Velazquez E, Hoebers F, Aerts HJ, et al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging. *Radiother Oncol*. 2014;113:324-330.
21. Erasmus Medical Center. Prediction of survival in patients with HNSCC. https://erasmusmc.thirdwave.nl/model/. Accessed February 15, 2018.
22. Vainshtein JM, Spector ME, McHugh JB, et al. Refining risk stratification for locoregional failure after chemoradiotherapy in human papillomavirus–associated oropharyngeal cancer. *Oral Oncol*. 2014;50:513-519.
23. Te Riele RJLM, Dronkers EAC, Wieringa MH, et al. Influence of anemia and BMI on prognosis of laryngeal squamous cell carcinoma: development of an updated prognostic model. *Oral Oncol*. 2018;78:25-30.
24. Huang SH, O'Sullivan B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr Treat Options Oncol*. 2018;18:40.
25. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61:92-105.
26. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543-2546.
27. Hoban CW, Beesley LJ, Bellile EL, et al. Individualized outcome prognostication for patients with laryngeal cancer. *Cancer*. 2018;124:706-716.
28. Beesley LJ, Taylor JMG. EM algorithms for fitting multistate cure models [published online ahead of print March 23, 2018]. *Biostatistics*. doi: https://doi.org/10.1093/biostatistics/kxy011.
29. Bartlett JW, Seaman SR, White IR, Carpenter JR. Alzheimer's Disease Neuroimaging I. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res*. 2015;24:462-487.

30. Kehl KL, Landrum MB, Arora NK, et al. Association of actual and preferred decision roles with patient-reported quality of care: shared decision making in cancer care. *JAMA Oncol*. 2015;1:50-58.
31. Kane HL, Halpern MT, Squiers LB, Treiman KA, McCormack LA. Implementing and evaluating shared decision making in oncology practice. *CA Cancer J Clin*. 2014;64:377-388.
32. Stacey D, Paquet L, Samant R. Exploring cancer treatment decision-making by patients: a descriptive study. *Curr Oncol*. 2010;17:85-93.
33. Marur S, Li S, Cmelak AJ, et al. E1308: phase II trial of induction chemotherapy followed by reduced-dose radiation and weekly cetuximab in patients with HPV-associated resectable squamous cell carcinoma of the oropharynx–ECOG-ACRIN Cancer Research Group. *J Clin Oncol*. 2017;35:490-497.
34. Masterson L, Moualed D, Liu ZW, et al. De-escalation treatment protocols for human papillomavirus–associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. *Eur J Cancer*. 2014;50:2636-2648.
35. Chera BS, Amdur RJ, Tepper JE, et al. Mature results of a prospective study of deintensified chemoradiotherapy for low-risk human papillomavirus–associated oropharyngeal squamous cell carcinoma. *Cancer*. 2018;124:2347-2354.
36. Chera BS, Amdur RJ. Current status and future directions of treatment deintensification in human papilloma virus–associated oropharyngeal squamous cell carcinoma. *Semin Radiat Oncol*. 2018;28:27-34.
37. Syrigos KN, Karachalios D, Karapanagiotou EM, Nutting CM, Manolopoulos L, Harrington KJ. Head and neck cancer in the elderly: an overview on the treatment modalities. *Cancer Treat Rev*. 2009;35:237-245.
38. Braunholtz DA, Edwards SJ, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". *J Clin Epidemiol*. 2001;54:217-224.
39. Fossa SD, Skovlund E. Selection of patients may limit the generalizability of results from cancer trials. *Acta Oncol*. 2002;41:131-137.
40. 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.
41. van Imhoff LC, Kranenburg GG, Macco S, et al. Prognostic value of continued smoking on survival and recurrence rates in patients with head and neck cancer: a systematic review. *Head Neck*. 2016;38(suppl 1):E2214-E2220.
42. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8:283-298.
43. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5:1315-1316.
44. Reimers N, Kasper HU, Weissenborn SJ, et al. Combined analysis of HPV-DNA, p16 and EGFR expression to predict prognosis in oropharyngeal cancer. *Int J Cancer*. 2007;120:1731-1738.
45. Michaelsen SH, Larsen CG, von Buchwald C. Human papillomavirus shows highly variable prevalence in esophageal squamous cell carcinoma and no significant correlation to p16INK4a overexpression: a systematic review. *J Thorac Oncol*. 2014;9:865-871.
46. Prince V, Bellile EL, Sun Y, et al. Individualized risk prediction of outcomes for oral cavity cancer patients. *Oral Oncol*. 2016;63:66-73.
47. Yu JB, Gross CP, Wilson LD, Smith BD. NCI SEER public-use data: applications and limitations in oncology research. *Oncology (Williston Park)*. 2009;23:288-295.
48. Williams PM, Lively TG, Jessup JM, Conley BA. Bridging the gap: moving predictive and prognostic assays from research to clinical use. *Clin Cancer Res*. 2012;18:1531-1539.
49. Bernstein JM, Homer JJ, West CM. Dynamic contrast-enhanced magnetic resonance imaging biomarkers in head and neck cancer: potential to guide treatment? *A systematic review. Oral Oncol*. 2014;50:963-970.