DR. PETER G HAWKINS (Orcid ID : 0000-0003-1100-9388)

Article type : Original Article

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

Running Head: Assessment of oropharyngeal cancer survival calculators

Lauren J. Beesley, PhD[1,*]; Peter G. Hawkins, MD, PhD[2,*]; Lahin 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]

[1]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI

[2]Department of Radiation Oncology, University of Michigan Medical School, Ann Arbor, MI

[3]Department of Otolaryngology-Head and Neck Surgery, University of Michigan Medical School, Ann Arbor, MI

*These authors contributed equally to this work.

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; phone: 734-936-3287, fax: 734-763-2215, email: jmgt@umich.edu

Number of pages: 18

Number of figures: 4

Number of tables: 4

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.xxxx/CNCR.31739

Funding Support: This research was supported by National Cancer Institute at the National Institutes of Health grants P50 CA097248 (University of Michigan Head and Neck Specialized Program of Research Excellence), P30 CA46592 (University of Michigan Cancer Center), IU01 CA183848 (Eisbruch), and T32 CA083654 (Cancer Biostatistics Training Program); and by a Young Investigator Award from The American Head and Neck Society (Shuman).

Conflict of Interest Disclosures: The authors declare no conflicts of interest relevant to the present work.

CONDENSED ABSTRACT

Individualized risk calculators for oropharyngeal squamous cell carcinoma demonstrate reasonable predictive accuracy. However, high variability among calculators in predictions for individual patients limits their clinical utility.

ABSTRACT

Background: Accurate, individualized prognostication in 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 in their estimation of five-year overall survival. Treatment outcomes for 856 patients with OPSCC evaluated at a single institution from 2003-2016 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 (ROC) curves, concordance statistics (C-indices), and calibration plots.

Results: Correlation between pairs of calculators varied, with coefficients ranging from 0.63 to 0.90. Only three of six pairs of calculators yielded predictions within 10% of each other for at least 50% of patients. Kaplan-Meier curves of calculator-defined risk groups showed reasonable stratification. Areas under the ROC curve ranged from 0.74 to 0.80, and C-indices from 0.71 to 0.78. Each calculator demonstrated superior discriminatory ability compared to American Joint Committee on Cancer 7th and 8th Editions clinical staging. Among models, the Denmark calculator was best-calibrated to observed outcomes.

*Conclusion*s: Existing calculators exhibited reasonable estimation of survival in OPSCC, but there was considerable variability in predictions for individual patients, which limits clinical utility. Given the increasing role of personalized treatment for OPSCC, further work is needed to improve accuracy and precision, possibly through the identification and incorporation of additional biomarkers.

Author Contributions

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

Key Words: oropharyngeal squamous cell carcinoma, prognosis, prediction, calculator, human papillomavirus

INTRODUCTION

Head and neck cancer (HNC) comprises a diverse group of malignancies that arise from multiple subsites and vary in presentation, treatment, and prognosis.[1] Accurate prognostication in HNC is critical in order to provide effective counseling and individualize optimal treatment. Prognostication in HNC is typically 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 to smoking- and alcohol-related OPSCC, HPV-positivity 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 8th Edition (Ed.) of the American Joint Committee on Cancer (AJCC) *Cancer Staging Manual.*[2, 12, 13] Although the AJCC 8th Ed. staging system for OSPCC accounts for HPV status, its authors explicitly avoided 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 stage. As such, there remains a need for accurate risk stratification that incorporates HPV status as well as other personalized features.

With the goal to provide 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 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 demonstrate impressive accuracy in their respective study cohorts. However, 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 in order 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 Internal Review Board of the University of Michigan. Patient data were extracted from two overlapping datasets at the University of Michigan: 1) a prospectively collected epidemiologic database of HNC patients;[6, 7] and 2) a database of patients treated with radiation therapy (RT) or chemo-radiotherapy (CRT) for OPSCC.[22] Patient data that existed in both datasets were individually checked for agreement, and any conflicting values were resolved by reference to the primary medical record. Patients were diagnosed and treated from 2003 to 2016 per institutional practices, which were consistent with National Comprehensive Cancer Network guidelines. As the selected calculators were designed to be used prior to any intervention, clinical information was analyzed, with pathological data substituted only in cases when 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 five-year overall

survival (OS) prediction, and inclusion of HPV status as determined by detection of HPV deoxyribonucleic acid (DNA) and/or p16.

We identified four calculators: one developed at the MAASTRO Clinic ("MAASTRO"),[20] one based on data from Radiation Therapy Oncology Group (RTOG) trials ("RTOG"),[18] one based on patients treated in eastern Denmark ("Denmark"),[19] and one developed at Erasmus Medical Center (EMC) ("Erasmus").[21, 23] Table 1 summarizes the data sources of each calculator, with additional details provided in Supplemental Tables 1-4. 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 -negative, regardless of p16 results. For the Erasmus calculator, as the method of HPV-status definition has not been reported, we elected to use p16 as is recommended in the AJCC 8th Ed. Cancer Staging Manual.[24] In our dataset, HPV DNA and p16 results were discordant for only 23 (2.7%) patients. These patients were differentially classified as either HPV-positive or -negative depending on the calculator being evaluated. Inputs for each calculator are summarized in Table 2. Additional details are presented in the Supplementary Materials.

Statistical Methods

Predicted five-year OS was computed for each patient using each calculator. Agreement between predictions was assessed using scatterplots and Spearman's correlation coefficients. The ability of each calculator to risk-stratify patients was assessed by dividing subjects into quintiles (five 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 ROC curve (AUCs) at five 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 five-year OS of each group was then plotted against the corresponding Kaplan-Meier estimates of five-year OS.

To handle missing covariates, we used a multiple imputation approach based on a multistate cure model.[28] We used the substantive model compatible (SMC), fully conditional specification (FCS) 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 about this novel methodology are provided in the Supplementary Materials.

RESULTS

Eight-hundred and fifty-six patients were identified for calculator assessment. Patient and disease characteristics are summarized in Table 3. Median follow-up was 61 months. Seventy-four percent and 61% of patients had at least 3 and 5 years of follow-up, respectively, or died before 3 years and 5 years. Figure 1 shows the distribution of predictions from the calculators (on the diagonal), scatterplots showing 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, except for MAASTRO, which tended to predict lower five-year OS. The scatter plots and correlation coefficients revealed variable degrees of association between calculator pairs. The Denmark and Erasmus calculators showed the strongest correlation with each other (rho=0.907), while the RTOG and MAASTRO calculators showed the weakest (rho=0.634). Table 4 lists the percentages of patients for which each pair of calculators yielded predictions within 10% of each other. For only three of the six 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 their 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. While all calculators yielded the expected distribution of survival outcomes, the MAASTRO and RTOG calculators exhibited relatively poorer differentiation of the two lowest-risk groups (Figure 2).

We next generated ROC curves and calculated areas under the ROC curve (AUCs) and C-indices (Figure 3). There was a range of discriminatory ability among the four 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 AJCC 7th Ed. criteria, and 0.72 and 0.68, respectively, using the 8th Ed. (Supplemental Figure 1), indicating inferior discriminatory ability compared to each of the four calculators.

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

DISCUSSION

Utilizing a large patient database, we assessed individualized calculators designed to predict five-year OS in patients with OPSCC. While 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 our 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 no training or validation values have been published for Erasmus. AUCs and C-indices from each calculator were higher than those obtained from AJCC 7th and 8th staging editions. Assessment of absolute predictive accuracy illustrated best calibration for the Denmark calculator.

Although the OPSCC calculators exhibited reasonable accuracy in this 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] Our results suggest that while they may outperform TNM staging, currently available OPSCC calculators still 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, 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 de-intensify treatment in patients with low-risk OPSCC.[33-35] While these approaches consistently incorporate HPV status, they heterogeneously account for other prognostic factors.[36] Our results indicate that individualized risk

calculators improve prognostication beyond AJCC staging, even the 8th Ed. which incorporates HPV status, and may represent a better method for selecting patients for de-intensification. However, the observed variability among calculators suggests that the optimal combination of prognostic variables has not been identified. Further work is needed to improve prognostic calculators to guide individualized treatment.

The variability among calculators was likely a function of differences in the variables included, the manner in which variables were modeled, and the data sources from which variables were identified. For example, performance of the MAASTRO model may have been impaired by inclusion of only patients treated with RT or CRT, and not surgery.[20] However, in a separate analysis of the 692 of our patients treated with RT or CRT, the MAASTRO calculator again exhibited inferior performance in comparison to the other models (Supplemental Figures 2-4). In addition, the MAASTRO calculator was developed using the smallest training cohort and was the only calculator to not 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 RTOG 9003.[18] While 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 patients treated on clinical trials and the exclusion of patients with an ECOG performance status greater than one may have limited generalizability.[38, 39] Also, in contrast to the other calculators, which more commonly modeled variables as ordinal or continuous, the RTOG model utilized 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 (LUMC).[40] This "Leiden" calculator performed poorly in an initial analysis (data not shown), likely because it does not incorporate HPV status. We therefore chose to evaluate the web-based Erasmus calculator, which does include HPV status and is based on a larger, more modern cohort treated at EMC.[21] This was the only calculator we evaluated for which a full description has not been published, although some cohort details have been reported in an analysis of patients with laryngeal cancer.[23] While the online interface is convenient, the lack of corresponding publication describing study patients and model performance limits its clinical utility. 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 OPSCC patients 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, an important prognostic factor in both HPV-related and non-HPV-related OPSCC.[7, 10, 41]

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

While these calculators all include HPV status, there are important differences in how this is defined for each. The MAASTRO calculator defines HPV-positivity based on the presence of HPV DNA, while the RTOG calculator considers only p16. The Denmark calculator considers p16 and HPV DNA separately. The Erasmus calculator's website does not stipulate how HPV status should be determined, and, as 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 consistently performed at all centers. The AJCC 8th Ed. *Cancer Staging Manual* stipulates that p16 staining be preferentially used to define HPV status.[24] Until it is clear that HPV DNA adds prognostic information beyond p16, it is likely 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 is available, caution should be exercised if utilizing these calculators. For staging purposes, in the absence of p16 and HPV status, the AJCC 8th Ed. recommends staging as if the patient were HPV-negative.[24] A similar approach could be taken when using a calculator, although one could also consider calculating an average of predicted outcomes for HPV-positive and 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 resultant prediction. This challenge can arise when any prognostic variable is not available, which may limit the utility of these models.

The current study is similar to our previous efforts to analyze individual predictors of survival in 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 were also important

differences in the datasets used for the larynx and oral cavity compared to OPSCC calculators. Certain of the larynx and oral cavity calculators were developed using the Surveillance, Epidemiology, and End Results (SEER) database, which, while large, lacks details of certain prognostic factors, which limits its usefulness for calculator development.[47] In contrast, the datasets used for the OPSCC calculators were based on institutional and cooperative group databases that contained more detailed information regarding potential prognostic variables. All of these studies, however, revealed substantial variability among calculators that limits their current value.

There are limitations to this study that require consideration. Although our patient dataset was reasonably comprehensive, there was some degree of data missingness (Table 3). As each calculator required different inputs, excluding patients with any missing variable would have substantially reduced the number of evaluable patients. As such, we elected to handle the missing covariates via an SMC-FCS multiple imputation technique.[29] While this methodology is rigorous and theoretically justified, it is based on additional statistical models which, 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 dataset was that it was comprised of patients seen at a single academic institution in the Midwestern United States. Interestingly, calculators derived from US populations were not better-calibrated to our patients than those developed in Europe. Regardless, it remains unclear how generalizable our results may be in relation to other institutions. It is also important to note that we evaluated predictions at only one time point, namely five years. While each calculator also predicts survival at other intervals, five years was chosen for this analysis as it is the only time point shared by all four models. As a result, we are unable to draw conclusions regarding the performance of these calculators in predicting survival prior to or beyond five years.

In conclusion, while existing OPSCC calculators demonstrate reasonable predictive accuracy and superior discriminatory ability compared to AJCC TNM staging, a lack of precision limits their utility 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 OPSCC.

ACKNOWLEDGEMENTS

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.

REFERENCES

1. Harrison LB, Sessions RB, Hong WK. Head and neck cancer: a multidisciplinary approach. Lippincott Williams & Wilkins, 2009.

2. Amin M, Edge S, Greene F, et al. AJCC Cancer Staging Manual, 8th edn. American Joint Committee on Cancer: NewYork, Springer: New York, NY, USA. Google Scholar, 2017.

3. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer. 2018.

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. 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: e173-180.

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. 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. Prediction of survival in patients with HNSCC. Available from URL: http://erasmusmc.thirdwave.nl/model/ [accessed September, 2017].

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 R, 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. Tuttle R, Morris L, Haugen B, et al. AJCC cancer staging manual2017.

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. Biostatistics. 2018.

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.

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: E2214-2220.

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.

FIGURE LEGENDS

Figure 1. Distribution of predicted outcomes (diagonal), scatter plots (below diagonal), and correlation coefficients (above diagonal) for analyzed calculators.

Figure 2. Kaplan-Meier plots of five-year OS in risk-stratified quintile groups of equal patient numbers defined by each model.

Figure 3. ROC curves with associated AUCs and C-indices for each calculator.

AUC = area under the receiver operating characteristic curve. C-index = concordance statistic.

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.

Table 1. Summary of datasets and models for each calculator.

Calculator	Cancers in training dataset	Training dataset	Validation dataset	Reported C-indices and/or AUCs
MAASTRO[20]	OPSCC	168 patients 2000-2011 MAASTRO Clinic	189 patients 2000-2006 VUMC	Training C-index: 0.82 External C-index: 0.73
RTOG[18]	OPSCC	493 patients 2002-2009 Multiple North American centers on RTOG 0129 and 0522	153 patients 1991-1997 Multiple North American centers on RTOG 9003	*Training C-index: 0.76 External C-index: 0.68
Erasmus[21]	Multiple HNC subsites	Cohort size unknown 2006-2013 EMC	None	Not available
Denmark[19]	OPSCC	1,542 patients 2000-2014 Eastern Denmark	None	*Training AUC: 0.8

*These represent uncorrected values; bias-corrected values are available in the corresponding published reports [18, 19].

OPSCC = oropharyngeal squamous cell carcinoma. C-index = concordance statistics. AUC = area under the receiver operating characteristic curve. VUMC = Vrije Universiteit Medical Center. EMC = Erasmus University Medical Center. RT = radiotherapy. CRT = chemo-radiotherapy. AJCC = American Joint Committee on Cancer.

Table 2. Input factors for each calculator.

	MAASTRO	**RTOG**	**Erasmus**	**Denmark**
Age	Not included	Dichotomized (≤ vs. > 50 years)	Continuous	Continuous
Gender	Included	Not included	Included	Not included
Comorbidity	ACE27, dichotomized (none-mild vs. moderate-severe)	Not included	ACE27, ordinal (grades 0-3)	Not included
Performance status	Not included	ECOG (0 vs. 1)	Not included	ECOG, ordinal (0-4)
Smoking status	Pack years, ordinal [none, moderate (1-30 pack-years), or heavy (> 30 pack years)]	Pack years, dichotomized (≤ vs. > 10)	Not included	Pack years, continuous
Education	Not included	Dichotomized (≤ vs. > HS)	Not included	Not included
Determination of HPV status	HPV DNA	p16	Not specified*	HPV DNA and p16
Stage**	Ordinal T- and dichotomized N- (N0-N2a vs. N2b-N3) classifications	Dichotomized T- (T2-3 vs. T4) and N- (N0-2b vs. N2c-3) classifications	Ordinal T-, N-, and M-classifications	Ordinal T- and N-classifications
Hemoglobin	Continuous	Dichotomized (≤ vs. > 13.5 g/dL for men, and ≤ 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)

*The method used to define HPV status for the Erasmus calculator has not been reported. For this analysis of the Erasmus calculator, we defined HPV status using p16. **Stages here refer to AJCC 7th Ed. criteria, although the RTOG calculator allows for use of the 7th or 8th Ed.

ACE27 = adult comorbidity evaluation-27. ECOG = Eastern Cooperative Oncology Group. HPV = human papillomavirus. HS = high school. RT = radiotherapy. CRT = chemo-radiotherapy.

Table 3. Patient characteristics.

Characteristic	**Value**
Age at diagnosis	
Mean years (SD)	58.4 (9.65)
Gender, n (%)	
Male	725 (84.6)
Female	131 (15.3)
Hemoglobin, mean (SD)	13.9 g/dL (1.53)
Unknown	119 (13.9)
ACE-27 comorbidity, n (%)	
None	206 (24.0)
Mild	259 (30.2)
Moderate	112 (13.0)
Severe	44 (5.1)
Unknown	235 (27.4)
Smoking status, n (%)	
Never	283 (33.0)
Former	297 (34.6)
Current	271 (31.6)
Unknown	5 (0.5)
Pack-years	
Mean (SD)	20.3 (25.2)
Median (range)	10 (0-150)
Unknown, n (%)	42 (4.9)
ECOG performance status, n (%)	
0	409 (47.8%)
1	52 (6.0)
2	1 (0.1)
Unknown	394 (46.0)
Maximum level of education, n (%)	
HS or lower	379 (44.2)

> HS	162 (18.9)
Unknown	315 (36.7)
Race, n (%)	
White	348 (40.6)
Black	10 (1.1)
Other	2 (0.2)
Unknown	496 (57.9)
T-classification (AJCC 7), n (%)	
0/is	0
1	195 (22.7)
2	284 (33.1)
3	138 (16.1)
4	236 (27.5)
Unknown	3 (0.3)

Table 3, continued

T-classification (AJCC 8), n (%)	
0/is	0
1	195 (22.7)
2	284 (33.1)
3	138 (16.1)
4	236 (27.5)
Unknown	3 (0.3)
N-classification (AJCC 7), n (%)	
0	110 (12.8)
1	90 (10.5)
2	45 (5.2)
2a	65 (7.5)
2b	327 (38.2)
2c	139 (16.2)
3	80 (9.3)

N-classification (AJCC 8), n (%)	
0	108 (12.6)
1	384 (44.8)
2	177 (20.6)
3	79 (9.2)
Unknown*	108 (12.6)
Clinical stage (AJCC 7), n (%)	
0	0
I	18 (2.1)
II	38 (4.4)
III	98 (11.4)
IV	701 (81.8)
Unknown	1 (0.1)
Clinical stage (AJCC 8), n (%)	
I	275 (32.1)
II	113 (13.2)
III	158 (18.4)
IV	65 (7.5)
Unknown**	245 (28.6)
Viral markers, n (%)	
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 (2.4)
HPV missing/p16-	10 (1.2)
HPV missing/p16 missing	234 (27.3)

Table 3, continued

Treatment modality, n (%)	
CRT	651 (76.0)
RT alone	42 (4.9)
Surgery + adjuvant CRT	33 (3.8)
Surgery + adjuvant RT	35 (4.0)
Surgery alone	23 (2.6)
Chemotherapy alone	14 (1.6)
Palliative, unknown	58 (6.7)

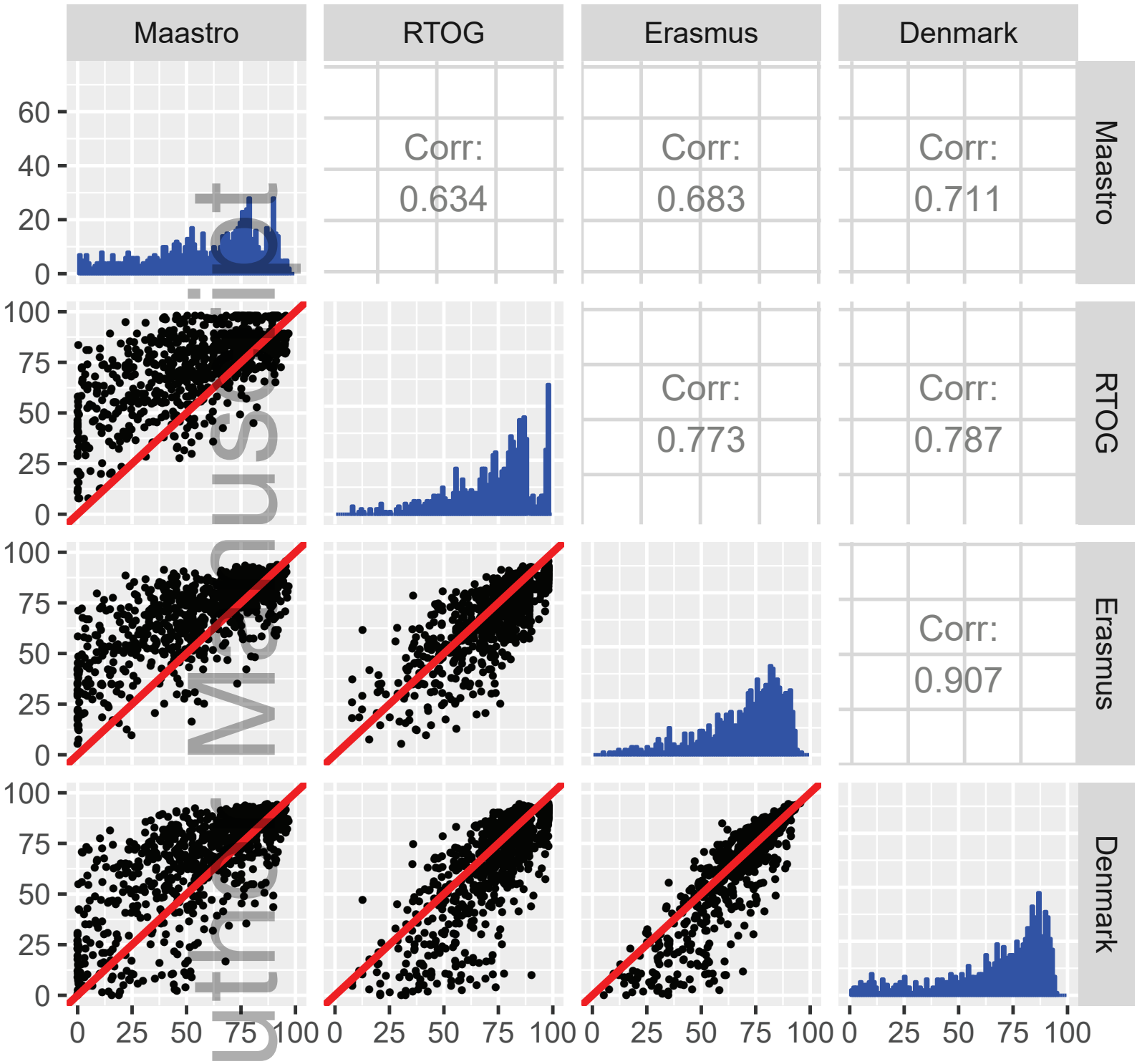
*The AJCC 8 N-classification for these patients was either N1 or N2. **AJCC 8 group stage was unknown in these patients due to unknown N-classification or HPV status.

SD = standard deviation. HS = high-school. RT = radiotherapy. CRT = chemo-radiotherapy.

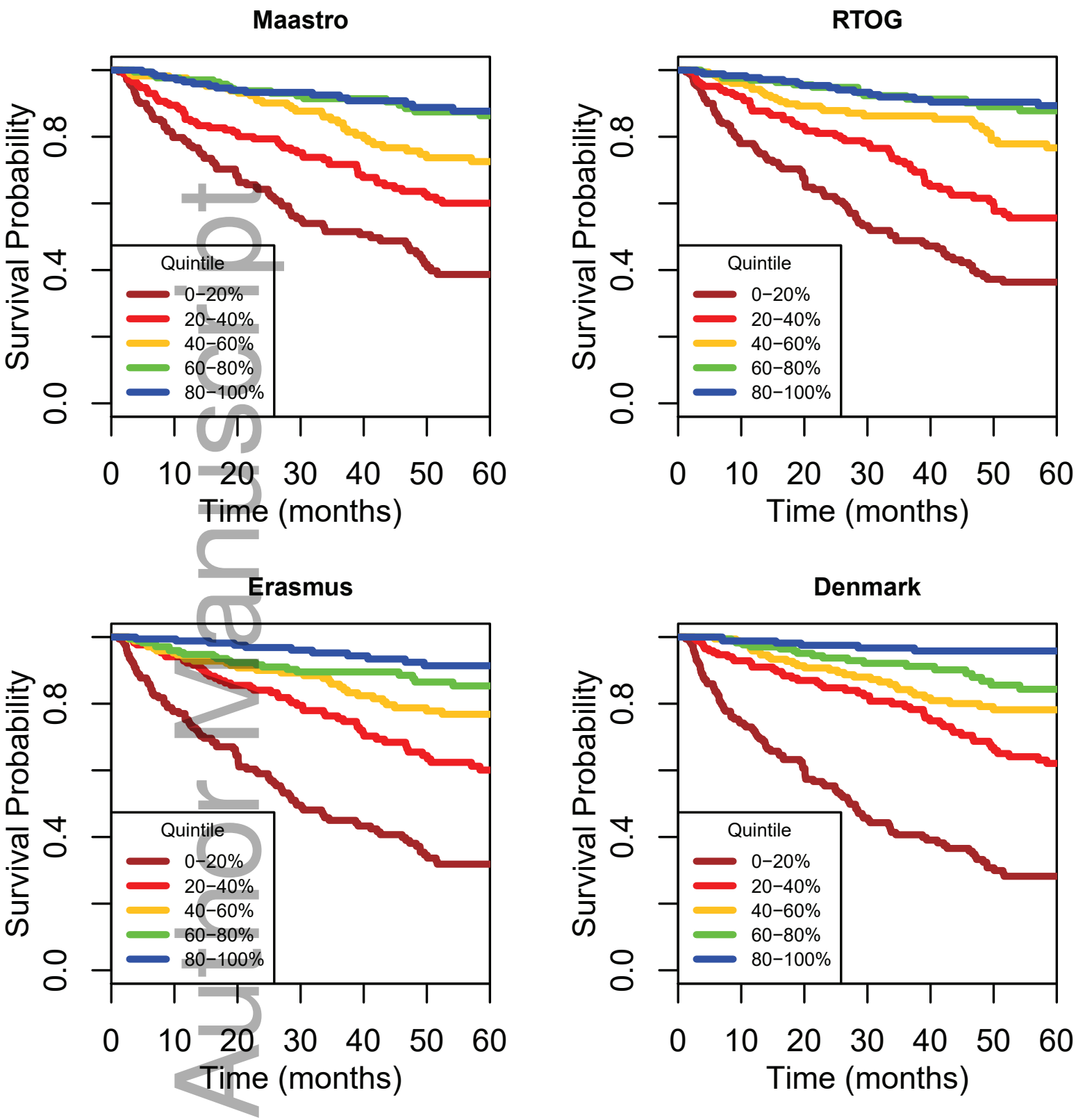
Table 4. Proportion of patients with predicted survival rates within 10% of each other for pairs of calculators.

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

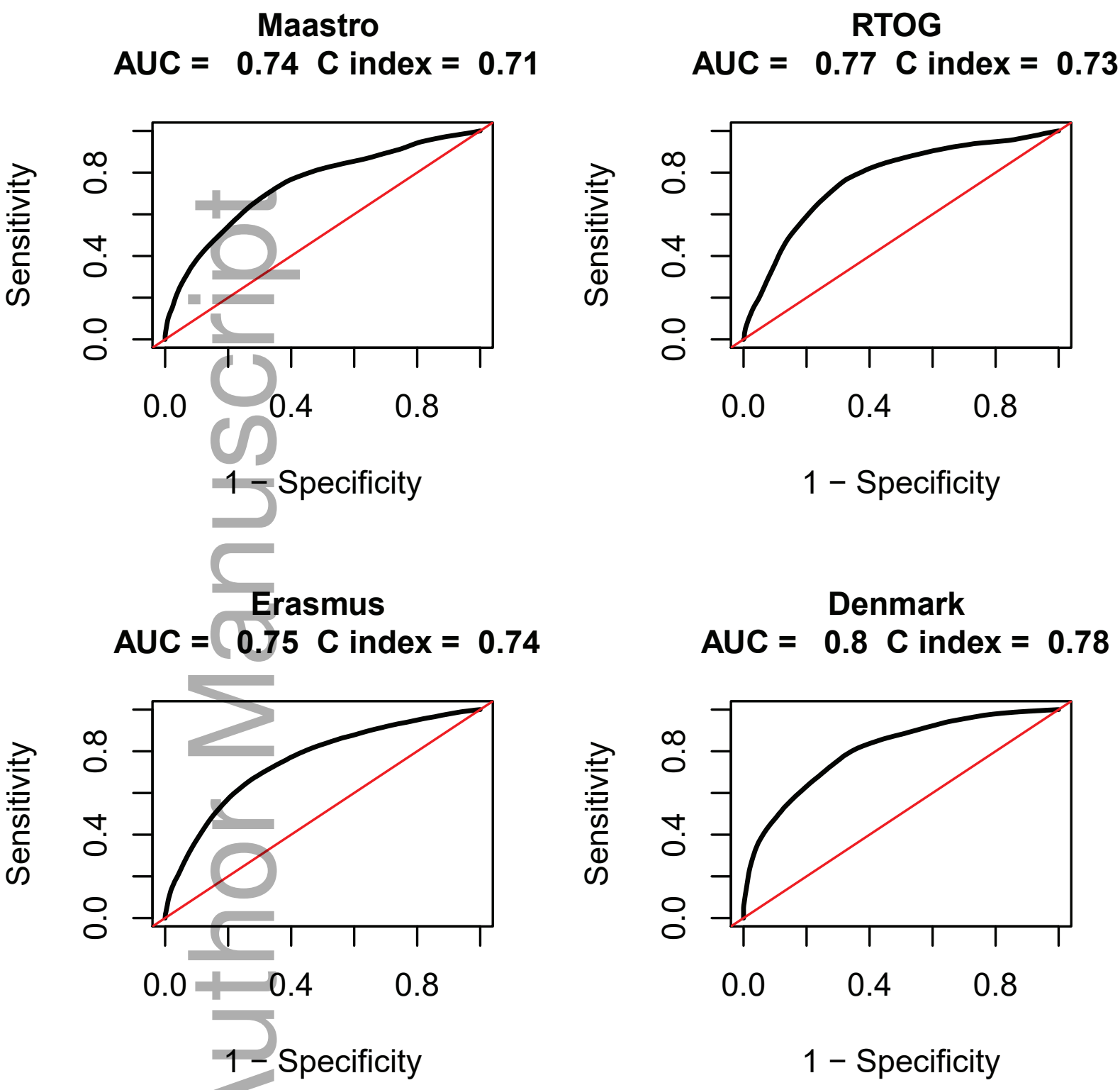
RTOG = Radiation Therapy Oncology Group.



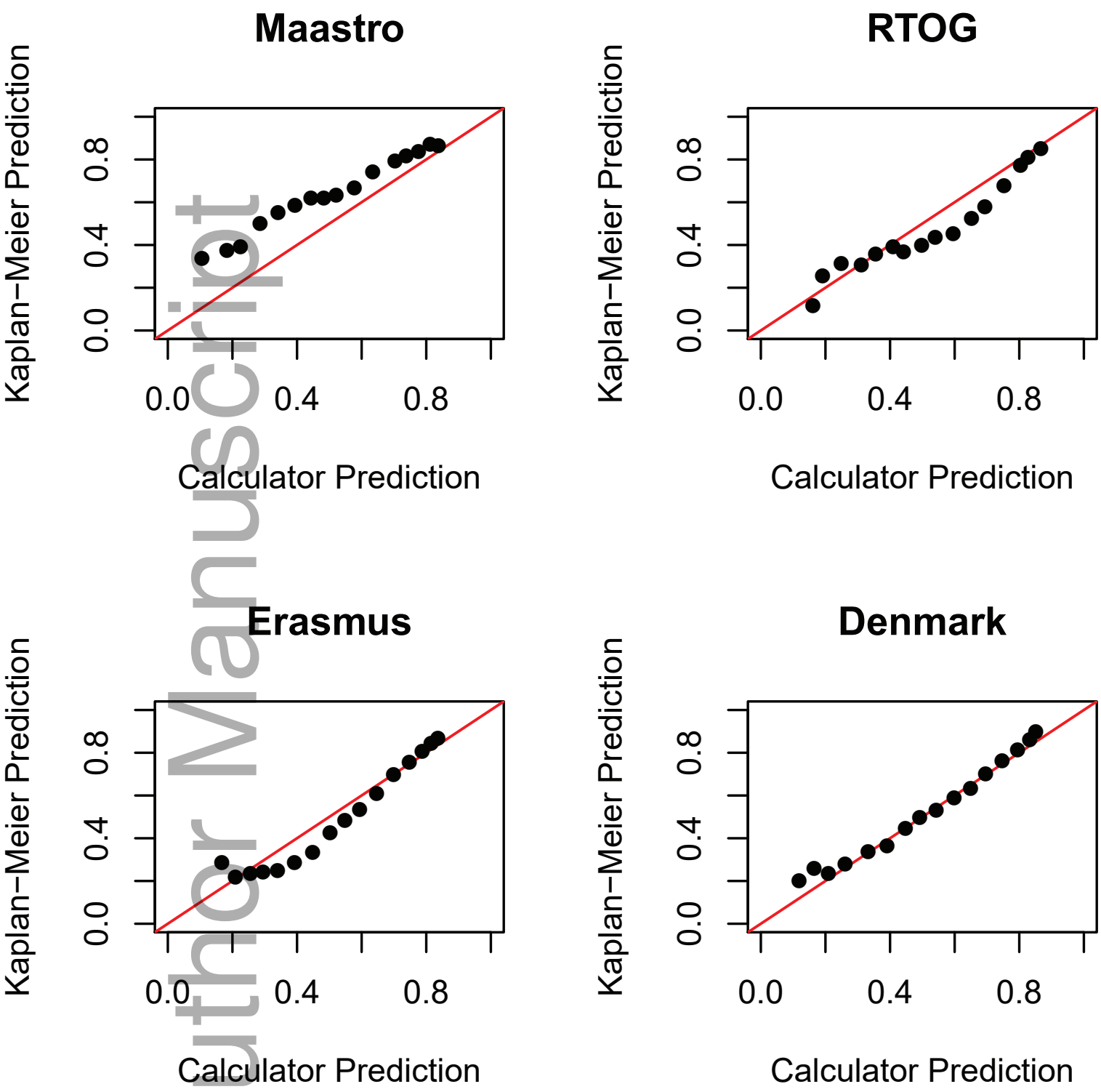
cncr_31739_f1.eps



cncr_31739_f2.eps



cncr_31739_f3.eps



cncr_31739_f4.eps