



UvA-DARE (Digital Academic Repository)

Development and external validation of a risk-prediction model to predict 5-year overall survival in advanced larynx cancer

Petersen, J.F.; Stuiver, M.M.; Timmermans, A.J.; Chen, A.; Zhang, H.; O'Neill, J.P.; Deady, S.; Vander Poorten, V.; Meulemans, J.; Wennerberg, J.; Skroder, C.; Day, A.T.; Koch, W.; van den Brekel, M.W.M.

DOI

[10.1002/lary.26990](https://doi.org/10.1002/lary.26990)

Publication date

2018

Document Version

Final published version

Published in

The Laryngoscope

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Petersen, J. F., Stuiver, M. M., Timmermans, A. J., Chen, A., Zhang, H., O'Neill, J. P., Deady, S., Vander Poorten, V., Meulemans, J., Wennerberg, J., Skroder, C., Day, A. T., Koch, W., & van den Brekel, M. W. M. (2018). Development and external validation of a risk-prediction model to predict 5-year overall survival in advanced larynx cancer. *The Laryngoscope*, 128(5), 1140-1145. <https://doi.org/10.1002/lary.26990>

General rights



It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Development and External Validation of a Risk-Prediction Model to Predict 5-Year Overall Survival in Advanced Larynx Cancer

Japke F. Petersen, MD ; Martijn M. Stuiver, PhD; Adriana J. Timmermans, MD, PhD; Amy Chen, MD, MPH, FACS; Hongzhen Zhang, PhD; James P. O'Neill, MD, PhD; Sandra Deady, PhD; Vincent Vander Poorten, MD, PhD; Jeroen Meulemans, MD; Johan Wennerberg, MD, PhD; Carl Skroder, MD; Andrew T. Day, MD; Wayne Koch, MD, FACS ; Michiel W. M. van den Brekel, MD, PhD

Objectives/Hypothesis: TNM-classification inadequately estimates patient-specific overall survival (OS). We aimed to improve this by developing a risk-prediction model for patients with advanced larynx cancer.

Study Design: Cohort study.

Methods: We developed a risk prediction model to estimate the 5-year OS rate based on a cohort of 3,442 patients with T3T4N0+M0 larynx cancer. The model was internally validated using bootstrapping samples and externally validated on patient data from five external centers (n = 770). The main outcome was performance of the model as tested by discrimination, calibration, and the ability to distinguish risk groups based on tertiles from the derivation dataset. The model performance was compared to a model based on T and N classification only.

Results: We included age, gender, T and N classification, and subsite as prognostic variables in the standard model. After external validation, the standard model had a significantly better fit than a model based on T and N classification alone (C statistic, 0.59 vs. 0.55, $P < .001$). The model was able to distinguish well among three risk groups based on tertiles of the risk score. Adding treatment modality to the model did not decrease the predictive power. As a post hoc analysis, we tested the added value of comorbidity as scored by American Society of Anesthesiologists score in a subsample, which increased the C statistic to 0.68.

Conclusions: A risk prediction model for patients with advanced larynx cancer, consisting of readily available clinical variables, gives more accurate estimations of the estimated 5-year survival rate when compared to a model based on T and N classification alone.

Key Words: Nomogram, risk prediction model, larynx, cancer, total laryngectomy, chemoradiotherapy, radiotherapy.

Level of Evidence: 2c.

Laryngoscope, 128:1140-1145, 2018

From the Department of Head and Neck Surgery and Oncology (J.F.P., M.M.S., A.J.T., M.W.M.V.D.B.), the Netherlands Cancer Institute, Amsterdam, the Netherlands; Department of Clinical Epidemiology, Biostatistics, and Bioinformatics (M.M.S.), Amsterdam Medical Center, Amsterdam, the Netherlands; Department of Otolaryngology-Head and Neck Surgery (A.C., H.Z.), Emory University, Atlanta, Georgia, U.S.A.; Department of Head and Neck Surgery and Oncology (J.P.O.N.), St. James Hospital, Dublin, Ireland; National Cancer Registry (S.D.), Cork, Ireland; Department of Oncology, Head and Neck Oncology Section (V.V.P., J.M.), University Hospitals Leuven, Leuven, Belgium; Department of ENT/Head and Neck Surgery (J.W., C.S.), Lund University Hospital, Lund, Sweden; Department of Head and Neck Surgery and Oncology (A.T.D., W.K.), Johns Hopkins Medical Center, Baltimore, Maryland, U.S.A.; Institute of Phonetic Sciences (M.W.M.V.D.B.), University of Amsterdam, Amsterdam, the Netherlands; and the Department of Oral and Maxillofacial Surgery (M.W.M.V.D.B.), Academic Medical Center, Amsterdam, the Netherlands.

Editor's Note: This Manuscript was accepted for publication October 9, 2017.

Presented as an oral presentation at the International Federation of Otorhinolaryngological Societies World ENT Congress, Paris, France, June 24-28, 2017.

The Netherlands Cancer Institute receives a research grant from ATOS Medical Sweden, which contributes to the existing infrastructure for health-related, quality-of-life research in the Department of Head and Neck Oncology and Surgery.

The authors have no other funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Japke F. Petersen, MD, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands. E-mail: m.petersen@nki.nl

DOI: 10.1002/lary.26990

INTRODUCTION

Larynx cancer is among the most frequently diagnosed head and neck squamous cell cancers (SCC), and approximately 40% of patients present with advanced disease.¹⁻³ The 5-year overall survival (OS) of the advanced (T3T4) tumors varies between 34% and 49%, depending on patient-related factors, tumor-related factors, and treatment.^{3,4} Historically, patients with advanced larynx cancer were treated with a total laryngectomy (TL) with adjuvant radiotherapy (RT). In 1991, the randomized controlled Veterans Affairs (VA) trial demonstrated equal OS for organ preservation (induction chemotherapy [CT] followed by chemoradiotherapy [CRT]) compared to TL plus adjuvant RT.⁵ In 2003, the results of the Radiation Therapy Oncology Group (RTOG) 91-11 study confirmed the added value of CT added to RT; however, large T4N0 larynx cancer patients were excluded.⁶ Furthermore, in a later publication on the VA data, OS for T4N0 patients was significantly higher after TL.⁷ Recently, several other retrospective studies have reported significantly higher OS rates for TL when compared to organ preservation protocols.^{2,4,7-10}

Adequate information regarding the prognosis is crucial in communicating with patients and in clinical decision making. The mixed results regarding the best treatment for

advanced larynx cancer have made the decision process, however, a complex task. Currently, the TNM classification is often used when talking about the estimated prognosis of patients. Although the TNM classification effectively prognosticates at a population level, it works less well on the individual level.^{11,12} Furthermore, the influence of variables such as age and subsite on OS is difficult to assess in the individual patient. Several studies have demonstrated that OS predictions based on a clinical prediction model (CPM) are superior to those made by experienced clinicians.^{11,13–15} The availability of a quantitative prediction model may therefore enhance the quality of the decisional process.

In this study we aimed to develop a CPM to aid decision making in advanced larynx cancer care. We hypothesized that the model would give more accurate predictions on OS than TNM classification alone gives us now. Because of the absence of decisive evidence from randomized controlled trials on the best treatment choice for advanced larynx cancer, a secondary objective of this large observational study was to estimate the effect of treatment on expected survival.

MATERIALS AND METHODS

Derivation Data

We collected patient data from a cohort of the Netherlands Cancer Registry covering all patients who have been diagnosed with advanced SCC of the larynx in the Netherlands (1991–2010). Timmermans et al. recently published the trends in treatment, incidence, and survival of this cohort in which a detailed description is given of the selection and characteristics of the patients.⁴ For the development of the CPM, we included all patients with primary T3T4N0N+M0 SCC of the larynx who were treated with a primary TL, CRT, or primary RT. The derivation dataset initially consisted of 3,794 patients with T3T4N0N+M0 SCC of the larynx diagnosed between 1991 and 2010 in the Netherlands. We excluded patients without follow-up ($n = 7$), patients who had emigrated ($n = 12$), and patients who were not treated with primary RT, CRT, or TL ($n = 333$). Thus, 3,442 patients were included in the study.

Validation Data

External validation of a CPM is crucial to evaluate its performance. We collected data of five independent patient cohorts: 390 patients from an Irish National Cancer Registry, 91 patients from Johns Hopkins, 89 from Emory University Hospital, 100 from Lund Medical Center, and 100 from the University Hospitals Leuven (total = 770). All centers received permission from their institutional review board to participate in this study.

Statistical Analysis

We used descriptive statistics to summarize patient characteristics and compared the pooled validation group and the derivation group by means of the χ^2 or Student t tests. Five-year OS rates were compared by means of the log-rank test, and a multivariable Cox proportional hazard analysis was used to estimate the influence of treatment modality on OS.

Clinical Prediction Model

For the risk-prediction model, we used the Cox proportional hazards model.¹⁶ The model was fully prespecified, with

exception of year of treatment, which was subject to selection based on statistical significance (to control for changes in survival probability due to changes in treatment trends over time if necessary). The predictors included in the model were chosen based on current knowledge, availability, and biological plausibility, and included age (using a restricted cubic spline), gender, subsite within the larynx (*International Classification of Diseases for Oncology, Third Revision*), T classification, and N classification.

Model Performance

We assessed model performance using discrimination and calibration. Discrimination is the ability of a prediction model to distinguish between patients who experience an event from those who do not, and can be measured by means of the C statistic.¹⁶ The C statistic can range from 0.5, which means equal to chance, to 1.0, which means a perfect model. In a Cox proportional hazard model, a C statistic of 0.60 implies that at any point in time, a random patient with an event has a higher risk score than a random patient without an event 60% of the time.^{16,17}

Calibration relates to the agreement between estimated and observed probabilities and is depicted in a calibration plot. In a perfect calibration plot, the lines of the estimated and observed probabilities would follow a 45° line, which implies that the predicted probability is identical to the observed probability.^{16–18}

Internal validation was performed by taking 200 bootstrapping samples. Based on the results of the bootstrap validation, we applied uniform shrinkage to adjust the coefficients. We then performed external validation of the shrunken model and calculated the C statistic and calibration curves.

As a third measure of model performance, we divided the validation data into three risk categories based on tertiles derived from the derivation data. We then plotted the observed Kaplan-Meier (KM) curve of the validation data over the expected KM curve of the derivation set based on the predicted risks, to visually inspect the agreement between observed and expected survival in each of the risk groups.

All models were built using the RMS package in R software.^{19,20}

RESULTS

Derivation and Validation Dataset

The derivation dataset consisted of 3,442 patients. The mean age was 64 years, the majority of patients were male (79%), and the 5-year OS rates were 44% for RT, 45% for CRT, and 49% for TL. All included variables (age, gender, subsite T and N classification, and treatment) had a significant effect on OS ($P < .001$ for all variables except gender: $P < .03$).

Patient characteristics from the derivation and validation dataset can be found in Table I. Patients in the derivation dataset were significantly older than the validation dataset ($P < .001$) and had fewer male patients (79% vs. 85%). In the derivation data, more tumors were located in the supraglottic, and more patients were treated with primary RT (58% vs. 37%) and less with CRT (8% vs. 28%) or primary TL (34% vs. 40%). Furthermore, there were significant differences in T and N classification ($P < .001$) and 5-year OS rates ($P < .001$).

TABLE I.
Patient Characteristics From the Derivation and Validation Datasets.

	Derivation Dataset	Pooled Validation Dataset	Leuven, Belgium, N (%)	NCR Ireland, N (%)	Baltimore, MD, USA, N (%)	Atlanta, GA, USA, N (%)	Lund, Sweden, N (%)
Age, yr, mean (range)	64 (28–100)	62 (16–92)	64 (40–90)	62 (35–85)*	60 (38–92)	59 (16–83)	66 (35–89)
Gender, male, n (%)	2705 (78.6%)	652 (85%)	92 (92%)	339 (86.9%)	72 (79.1%)	64 (74.1%)	85 (85%)
TN classification, n (%)							
T3N0	1237 (35.9%)	282 (36.6%)	34 (34%)	159 (40.8%)	28 (30.7%)	25 (28.1%)	36 (36%)
T3N+	681 (19.8%)	145 (18.8%)	17 (17%)	72 (18.5%)	21 (23.1%)	21 (23.6%)	14 (14%)
T4N0	887 (25.8%)	174(22.6%)	28 (28%)	73 (18.7%)	17 (18.7%)	18 (20.2%)	38 (38%)
T4N+	637 (18.5%)	169 (21.9%)	21 (21%)	86 (22.1%)	25 (27.5%)	25 (28.1%)	12(12%)
Sublocalization, n (%)							
Glottis	1074 (31.2%)	335 (43.5%)	55 (55%)	157 (40.3%)	43 (47.3%)	37 (41.6%)	43 (43%)
Supraglottis	2172 (63.1%)	313 (40.6%)	38 (38%)	147 (37.7%)	45 (49.5%)	36 (40.4%)	47 (47%)
Subglottic	88 (2.6%)	26 (3.4%)	3 (3%)	16 (4.1%)	2 (2.2%)	2 (2.2%)	3 (3%)
Larynx NNO	108 (3.1%)	96 (12.5%)	4 (4%)	70 (17.9%)	1 (1.1%)	14 (15.7)	7 (7%)
Treatment, n (%)							
TL	1168 (33.9%)	311 (40.4%)	54 (54%)	120 (30.8%)	55 (60.4%)	40 (44.9%)	42 (42%)
RT	2009 (58.4%)	281 (36.5%)	15 (15%)	164 (42.1%)	1 (1.1%)	9 (10.1%)	57 (57%)
CRT	265 (7.7%)	213 (27.7%)	31 (31%)	106 (27.2%)	35 (38.5%)	40 (44.9%)	1 (1%)
Total	3442	770	100	390	91	89	100

*After transformation to continuous variable using midpoint of given age category, median = 63.

CRT = chemoradiotherapy; NCR = National Cancer Registry; NNO = not otherwise specified; RT = radiotherapy; TL = total laryngectomy.

Model Performance–Internal Validation

Our main objective was to compare the discriminative power of a multivariable prediction model with a model based on T classification and N classification alone. As a second objective, we evaluated the effect of treatment on OS, for which we added treatment modality as a prognostic variable in a third model containing the same variables as the prediction model. First, internal validation was performed taking bootstrapping samples ($n = 200$). This demonstrated that the prediction model including age, gender, T classification, N classification, and subsite as predictors had significantly better discrimination (C statistic 0.65) than the model based on T and N classification alone (C statistic 0.57) (likelihood ratio test $P < .001$).

Model Performance–External Validation

After external validation on the combined validation dataset ($n = 770$), discrimination proved to be significantly better for the full model (C statistic 0.59, 9% better) compared to the model based on T and N classification alone (C statistic 0.55) (likelihood ratio test $P < .001$). Calibration of the two models is depicted in Figure 1A and B, which shows a slight degree of miscalibration in both models as they do not exactly follow the 45° line. As a third measure of strength, the observed KM curves of the validation sets were plotted over the KM curves estimated from the derivation dataset for the two models (Fig. 2) to test whether a distinction can be made between high-, medium-, and low-risk patients. The

plots show that the models are able to distinguish between the three different risk categories, although OS in the medium- and low-risk groups of the validation set was lower compared to these risk groups in the derivation set.

Influence of Treatment Modality

Treatment modality was significantly related to OS in the validation database ($P < .0001$). The hazard ratio or death adjusted for age, gender, subsite, T classification, and N classification was 1.56 for RT compared to TL ($P < .001$), and 0.95 for CRT compared to TL ($P = .71$). With treatment modality as a prognostic variable added to the prediction model, the C statistic was 0.60.

Exploratory Analysis

Although the prediction model was able to distinguish well between the three risk groups and performed better compared to TNM alone, the C statistic was still relatively low. We hypothesized that adding comorbidity as a prognostic variable might further improve model performance. However, this variable was not recorded in our derivation database because it was retrieved from a national cancer registry. We therefore performed an exploratory post hoc analysis on a subset of the derivation dataset including 181 patients with T3T4N0N+M0 SCC of the larynx, diagnosed and treated with RT, CRT, or TL in the Netherlands Cancer Institute between 1999 and 2008,²¹ for which we were able to collect American

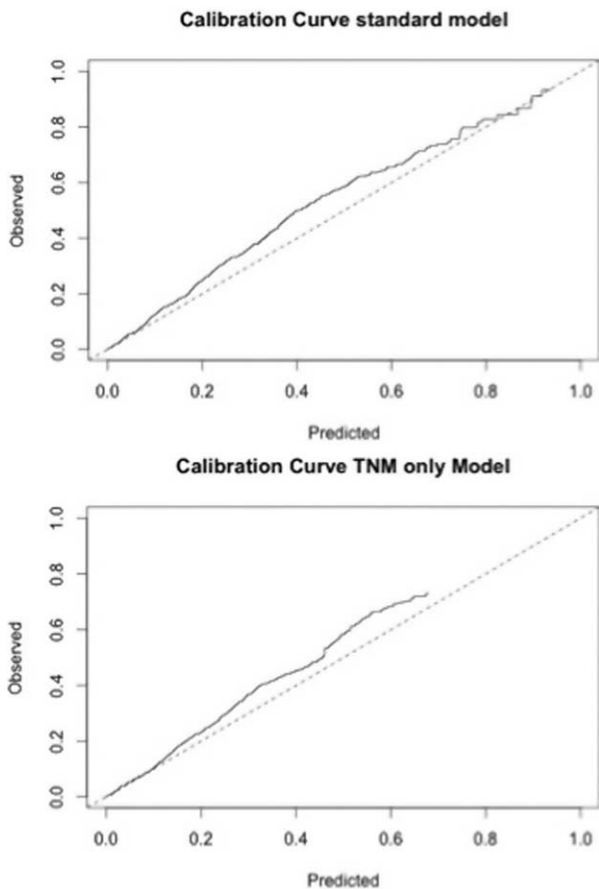


Fig. 1. Calibration curves for the model based on the combined dataset (A) and the model based solely on T and N classification as a prognostic variable (B). A perfect calibration would exactly follow the 45° line (dashed curve).

Society of Anesthesiologists (ASA) scores as a substitute measure for comorbidity. The majority of the external centers had not systematically recorded ASA scores in the patient files; thus, we were unable to perform an external validation on this model. After shrinkage by internal validation the C statistic was 0.68.

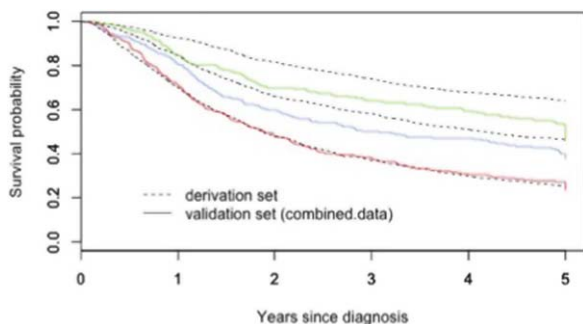


Fig. 2. Kaplan-Meier curve for the low-, medium-, and high-risk groups. The Kaplan-Meier curve of the expected overall survival as estimated by the derivation set (dashed line) and the observed overall survival as seen in the validation set, divided in three risk groups. Green = low risk, blue = intermediate risk, red = high risk on dying. [Color figure can be viewed at www.laryngoscope.com.]

DISCUSSION

The results of our study confirm our hypothesis that a validated multivariable risk prediction model gives more accurate OS predictions for advanced larynx cancer compared to a model based on T and N classification alone. According to estimated and observed KM curves, the model distinguishes adequately between the three risk categories. Yet, with a C statistic of 0.59, the predictive accuracy leaves room for improvement in the context of clinical decision making for individual patients.

As a secondary objective, we aimed to investigate the effect of treatment on expected OS. Estimating the effect of treatment modality in an observational study is troublesome, because this incorporates a bias by indication. However, because a new, large, randomized controlled trial comparing TL with organ preservation strategies may never be performed, we investigated the influence of treatment modality when accounting for the other prognostic variables included in the prediction model. This analysis suggested that survival after TL is better than after CRT or RT, as was suggested by the results published by Timmermans et al.⁴

As also was reported by Timmermans et al. was that the derivation data contained more supraglottic tumors than the validation data. Interestingly, they demonstrated how this distribution was reversed in the T1T2 tumors, in which they found more glottic (78.6%) than supraglottic tumors (19.9%).⁴ The RTOG 91-11 study, with mainly advanced tumors, found a similar rate of supraglottic tumors (69%).^{6,22}

In recent years, several risk-prediction models have been published. In 2001, Baatenburg de Jong et al. developed a risk-prediction model for T1-T4 SCC occurring in all subsites of the head and neck except the esophagus.²³ The model was based on 1,396 patients diagnosed between 1981 and 1998, and included the prognostic predictors of age, gender, tumor site, prior tumor, and TNM classification. In 2013, the model was updated, and the Adult Comorbidity Evaluation-27 was added as a prognostic variable and external validation was performed. After external validation, the model showed a good C statistic of 0.69, but the validation dataset did not include hypopharynx and nasopharynx cancer.²⁴ In their model, the impact of severe comorbidity appeared comparable to the impact of a T4 tumor or N3 neck on OS. We were not able to include comorbidity in our original model, which might explain why our model was less accurate. The exploratory post hoc analysis that included ASA score as an indicator of comorbidity improved the discriminative ability.

Another risk prediction model has been developed by Egelmeier et al., who developed and externally validated a model for T1 to T4 larynx cancer receiving RT, based on a cohort of 994 patients. In concordance with our findings, they reported male gender, older age, higher T status, and nodal involvement to be negative predictive factors for OS. Furthermore, they included hemoglobin level and radiotherapy dose as prognostic factors. The performance of their model ranged from 0.68

to 0.74.²⁵ More recently, another CPM for T3T4 larynx cancer patients was published based on a cohort of 615 patients. In this model, the authors included age, Eastern Cooperative Oncology Group (ECOG) performance status, N classification, and treatment modality, but excluded variables such as T classification, subsite, and smoking status using a stepwise selection procedure. Such a data-driven approach for variable selection results in a model that might not be accurate when used for new patients.²⁶ External validation was not performed, and the authors note that the model needs external validation first and might not be generalizable.²⁷

In the literature, several different patient-specific and tumor-specific factors have been investigated as prognostic variables for head and neck cancer, indicating that factors such as albumin (<4 g/dL), alcohol intake,²⁸ insurance status, race,⁸ tumor volume,²¹ tumor hypoxia,²⁹ and several different biomarkers^{30,31} can have a prognostic influence on overall survival. To help distinguish the actual predictors for OS and create a more accurate CPM, a large prospective cohort should be kept in which multiple parameters are collected or this data could be extracted from electronic patient files. Currently, in the Netherlands, a national prospective audit is being conducted, which in the future could be used to further improve our model.

Next to OS, another frequently used endpoint in clinical studies is larynx preservation. Predicting which patients benefit from organ preservation strategies and which do not could be of great value for avoiding unnecessary toxic treatment with added morbidity after salvage surgery. A well-known model to predict this is the TALK score: a prognostic model developed to facilitate the treatment decision making in larynx preservation.²⁸ TALK is an acronym for T status, Albumin, Alcohol (or liquor) use, and Karnofsky performance score, which were the predictors used. In an external validation on the VA larynx cancer study dataset, a C statistic of 0.57 was obtained for predicting larynx preservation. The TALK score, however, does not indicate which patients suffer from a nonfunctioning larynx after organ preservation, such as those who have a tracheotomy or nasogastric feeding tube in situ. In our derivation cohort, larynx preservation was scored as not having had a laryngectomy after organ preservation. However, information regarding a tracheotomy or feeding tube was missing, due to the fact that it was based on a national cancer registry cohort. We therefore chose not to predict larynx preservation based on these data.

In survival predictions, comorbidity scores can be of great value. However, in our cohort, comorbidity scores were missing. ASA score was available, however, for a subgroup of the derivation dataset. In the ASA score, the burden of comorbidity is incorporated; thus, it could potentially serve as a proxy for an actual comorbidity scale. In 2015, Young et al. compared the ASA score with the ECOG/World Health Organization performance scale as a measure of functional status in a predictive model and demonstrated equal performance in predicting length of stay after cancer surgery.³² In our exploratory post hoc analysis, adding ASA score as a

prognostic variable increased our C statistic to 0.68. We recommend that future studies determine which comorbidity scale might be of most value for prediction of survival outcomes in head and neck cancer, and assess the added value of this scale in a multivariable prediction model.

There are certain limitations to our study. In multivariable prediction modeling, a generally accepted rule of thumb is that a minimum of $m/10$ predictors should be used in a model, where m is the number of uncensored event times (e.g., death).³³ With 2,180 uncensored event times in our cohort, we could have included many more predictors without risking overfitting. However, our choice of predictors was limited to those available in the population-based database. Because the database was anonymized, we were unable to extend our database with variables such as comorbidity, intoxications, tumor volume, race, and insurance status, which might have improved the predictive value of the model for OS.

CONCLUSION

We have developed a ready-to-use prediction model based on a large systematically coded database on advanced-stage larynx cancer. The model gives significantly more accurate predictions on OS than compared to a model based on T and N classification alone. All of the variables included in the model are readily available in clinical practice. Although it should not be used as a replacement for clinical reasoning, it may aid the decision-making process for patients with advanced larynx cancer.

Acknowledgments

The authors thank the registration team of the Netherlands Comprehensive Cancer Organisation (IKNL) for the collection of data for the Netherlands Cancer Registry, as well as IKNL staff for scientific advice, and PALGA (the Dutch nationwide network and registry of histopathology and cytopathology) for their contribution to the national database on which the model was built.

BIBLIOGRAPHY

1. Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer* 2005;114:806–816.
2. Hoffman HT, Porter K, Karnell LH, et al. Laryngeal cancer in the United States: changes in demographics, patterns of care, and survival. *Laryngoscope* 2006;116:1–13.
3. Gatta G, Botta L, Sanchez MJ, Anderson LA, Pierannunzio D, Licitra L. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: the EURO CARE-5 population-based study. *Eur J Cancer* 2015;51:2130–2143.
4. Timmermans AJ, van Dijk BA, Overbeek LI, et al. Trends in treatment and survival of advanced larynx cancer: a 20-year population-based study in the Netherlands. *Head Neck* 2016;38(suppl 1):E1247–E1255.
5. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. The Department of Veterans Affairs Laryngeal Cancer Study Group. *N Engl J Med* 1991;324:1685–1690.
6. Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med* 2003;349:2091–2098.
7. Olsen KD. Reexamining the treatment of advanced laryngeal cancer. *Head Neck* 2010;32:1–7.
8. Chen AY, Halpern M. Factors predictive of survival in advanced laryngeal cancer. *Arch Otolaryngol Head Neck Surg* 2007;133:1270–1276.
9. Dziegielewski PT, O'Connell DA, Klein M, et al. Primary total laryngectomy versus organ preservation for T3/T4a laryngeal cancer: a

- population-based analysis of survival. *J Otolaryngol Head Neck Surg* 2012;41(suppl 1):S56–S64.
10. Megwalu UC, Sikora AG. Survival outcomes in advanced laryngeal cancer. *JAMA Otolaryngol Head Neck Surg* 2014;140:855–860.
 11. Ganly I, Amit M, Kou L, et al. Nomograms for predicting survival and recurrence in patients with adenoid cystic carcinoma. An international collaborative study. *Eur J Cancer* 2015;51:2768–2776.
 12. Burke HB, Henson DE. The American Joint Committee on Cancer. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer* 1993;72:3131–3135.
 13. Specht MC, Kattan MW, Gonen M, Fey J, Van Zee KJ. Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram. *Ann Surg Oncol* 2005;12:654–659.
 14. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002;20:82–88.
 15. Kattan MW, Yu C, Stephenson AJ, Sartor O, Tombal B. Clinicians versus nomogram: predicting future technetium-99m bone scan positivity in patients with rising prostate-specific antigen after radical prostatectomy for prostate cancer. *Urology* 2013;81:956–961.
 16. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173–e180.
 17. Steyerberg ER. *Clinical Prediction Models, A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009:497.
 18. Bonnett LJ, Marson AG, Johnson A, et al. External validation of a prognostic model for seizure recurrence following a first unprovoked seizure and implications for driving. *PLoS One* 2014;9:e99063.
 19. R Core Team. *R: A language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015.
 20. Frank E. Harrell Jr. rms: Regression Modeling Strategies. R package version 5.1-0. Available at: <https://CRAN.R-project.org/package=rms>. Published May 3, 2016.
 21. Timmermans AJ, Lange CA, de Bois JA, et al. Tumor volume as a prognostic factor for local control and overall survival in advanced larynx cancer. *Laryngoscope* 2016;126:E60–E67.
 22. Forastiere AA, Zhang Q, Weber RS, et al. Long-term results of RT09-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol* 2013;31:845–852.
 23. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S. Prediction of survival in patients with head and neck cancer. *Head Neck* 2001;23:718–724.
 24. 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.
 25. Egelmeier 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.
 26. Steyerberg EW. *Clinical Prediction Models*. New York, NY: Springer-Verlag; 2008:528.
 27. Multidisciplinary Larynx Cancer Working Group. Conditional survival analysis of patients with locally advanced laryngeal cancer: construction of a dynamic risk model and clinical nomogram. *Sci Rep* 2017;7:43928.
 28. Sherman EJ, Fisher SG, Kraus DH, et al. TALK score: development and validation of a prognostic model for predicting larynx preservation outcome. *Laryngoscope* 2012;122:1043–1050.
 29. Isa AY, Ward TH, West CM, Slevin NJ, Homer JJ. Hypoxia in head and neck cancer. *Br J Radiol* 2006;79:791–798.
 30. Georgiou A, Gomatos IP, Ferekidis E, et al. Prognostic significance of p53, bax and bcl-2 gene expression in patients with laryngeal carcinoma. *Eur J Surg Oncol* 2001;27:574–580.
 31. Wang R, Guo Y, Ma H, et al. Tumor necrosis factor superfamily member 13 is a novel biomarker for diagnosis and prognosis and promotes cancer cell proliferation in laryngeal squamous cell carcinoma. *Tumour Biol* 2016;37:2635–2645.
 32. Young J, Badgery-Parker T, Dobbins T, et al. Comparison of ECOG/WHO performance status and ASA score as a measure of functional status. *J Pain Symptom Manage* 2015;49:258–264.
 33. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387.