AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Original Research—Head and Neck Surgery

Otolaryngology–Head and Neck Surgery
2023, Vol. 00(00) 1–10

DOI: 10.1002/ohn.479
http://otojournal.org
WILEY

Individualized Dynamic Prediction Model for Patient-Reported Voice Quality in Early-Stage Glottic Cancer

Maarten C. Dorr, MD[1]*, Eleni-Rosalina Andrinopoulou, PhD[2]*, Aniel Sewnaik, MD, PhD[1], Diako Berzenji, MD[1], Kira S. van Hof, MD[1], Emilie A.C. Dronkers, MD, PhD[1], Simone E. Bernard, MD[1], Arta Hoesseini, MD[1], Dimitirs Rizopoulos, PhD[2], Robert J. Baatenburg de Jong, MD, PhD[1], and Marinella P.J. Offerman, PhD[1]

Abstract

Objective. Early-stage glottic cancer (ESGC) is a malignancy of the head and neck. Besides disease control, preservation and improvement of voice quality are essential. To enable expectation management and well-informed decision-making, patients should be sufficiently counseled with individualized information on expected voice quality. This study aims to develop an individualized dynamic prediction model for patient-reported voice quality. This model should be able to provide individualized predictions at every time point from intake to the end of follow-up.

Study Design. Longitudinal cohort study.

Setting. Tertiary cancer center.

Methods. Patients treated for ESGC were included in this study (N = 294). The Voice Handicap Index was obtained prospectively. The framework of mixed and joint models was used. The prognostic factors used are treatment, age, gender, comorbidity, performance score, smoking, T-stage, and involvement of the anterior commissure. The overall performance of these models was assessed during an internal cross-validation procedure and presentation of absolute errors using box plots.

Results. The mean age in this cohort was 67 years and 81.3% are male. Patients were treated with transoral CO_2 laser microsurgery (57.8%), single vocal cord irradiation up to (24.5), or local radiotherapy (17.5%). The mean follow-up was 43.4 months (SD 21.5). Including more measurements during prediction improves predictive performance. Including more clinical and demographic variables did not provide better predictions. Little differences in predictive performance between models were found.

Conclusion. We developed a dynamic individualized prediction model for patient-reported voice quality. This model has the potential to empower patients and professionals in making well-informed decisions and enables tailor-made counseling.

Keywords

larynx, patient-centered care, prediction, prognostic model, quality of life measures, voice assessment

Received November 11, 2022; accepted July 19, 2023.

Early-stage glottic cancer (ESGC) is a common malignancy of the head and neck area with a good overall clinical outcome in terms of survival and recurrent disease. However, the tumor and its treatment have a significant impact on patient-reported quality of voice.[1-11] Voice and speech are crucial aspects of social communication and interaction and therefore can impact a patient's psychosocial well-being as well. Therefore, providing patients with individualized information on expected voice quality after treatments is important and enable optimal decision-making. Prognostic modeling can be supportive in this process. For survival, prediction models have been developed and used in clinical practice. However,

[1]Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands
[2]Department of Biostatistics, Department of Epidemiology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands

*These authors contributed equally to this article.

Corresponding Author:
Maarten C. Dorr, MD, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.
Email: m.dorr@erasmusmc.nl

prediction modeling for patient-reported outcomes is new and more difficult as structurally collected longitudinal data is scarce, and specific statistical techniques for repeated measurement data are required.[12] Within our institute, a few models for patient-reported outcomes have been developed in urology[13,14] and neurology.[15,16]

Most prediction models in medicine provide static predictions as they have been developed with classic linear, logistic, or Cox regression models.[17-19] During follow-up, however, additional information will become available that might change prognostic estimations of clinical and patient-reported outcomes. Conventional static prediction models are not able to use this updated information to provide new and more adequate predictions. Developing prediction models that are able to combine all available (changing) variables over time requires an extension of the available prediction statistical methods. This could be done by methodological innovations based upon mixed-effects models and joint models for longitudinal and time-to-event data, which have enjoyed a renaissance in recent years in the statistics and biostatistics literature.[12,20,21] Mixed-effect models enable longitudinal analysis by using all available data and account for unbalanced data and correlation between measurements from the same patients.[22] Joint models combine mixed-effect modeling with a time-to-event Cox regression model.

Within our institute, structural collection of patient-reported outcome measurements (PROMs) is embedded in our routine care since 2013 with the Healthcare Monitor.[23] Alongside the use of these PROMs during patient-doctor consultations for improving patient-centered care, this data is used on an aggregated level for obtaining longitudinal insights and developing individualized prediction models.

To our knowledge, no prediction models for longitudinal PROMs and recurrent disease in head and neck cancer are available. With this study, we continue our previous research describing longitudinal trajectories and associated risk factors of patient-reported voice quality in ESGC.[24] In this study, we showed that patient-reported voice quality is heterogeneous and nonlinear, and improved most in the first year of follow-up. Associated risk factors were older age, increased tumor stage, and severe comorbidity. Hence, the goal of our study is to develop a web-based and clinically useful individualized dynamic prediction model for patient-reported voice quality. This model will be dynamic, which means that it can provide new predictions during follow-up at every new consultation, as soon as new information becomes available. By doing this, we will empower patients and professionals to make well-informed decisions and enable tailor-made counseling and customized solutions prior to treatment and during the long period of follow-up.

Methods

Setting and Participants

All patients treated for ESGC (Tcis—T1b, N0M0) with transoral laser microsurgery, local radiotherapy, and single vocal cord irradiation at the Erasmus Medical Center between 2013 and 2018 and participating in the Healthcare Monitor were included in this longitudinal outcome study. The Healthcare Monitor has 95% patient compliance at intake and over 80% during follow-ups up to 5 years. All patients complete questionnaires before every outpatient clinic visit at home or with a tablet at the clinic. In the first year, questionnaires are filled in every 2 months, the second year every 3 months, the third year every 4 months, and every 6 months in year 4. Patients were excluded from this study when they had low-grade dysplasia and were appointed to strict follow-up, had synchronic tumors, a prior head and neck malignancy, had no PROM data available due to insufficient knowledge of the Dutch language or suffering disorders affecting cognitive abilities, or did not provide informed consent on using data for research purposes.

Ethical Considerations

This project was approved by the institutional review board and ethics committee (MEC-2020-0314) from the Erasmus MC. Our study follows the principles of the Declaration of Helsinki. All participating patients provided electronic written informed consent.

Main Outcomes and Measures

In this study, we used the prospectively obtained Dutch version of the Voice Handicap Index (VHI) at previously mentioned time points.[25,26] This is a validated, 30-item, questionnaire that measures the perceived psychosocial voice impairment in daily life.[27] Each item is scored on a 5-point Likert scale (0 = never, 5 = always). The total score is the sum of all scores and ranges from 0 to 120. A higher outcome indicates higher voice impairment.

Treatment modalities in this study are transoral laser microsurgery and radiotherapy.[28-31] The latter can be divided in local radiotherapy with irradiation of the larynx in a total dose of 60 to 66 Gy in 25 to 33 fractions, and single vocal cord irradiation with a mild hypofractioned scheme up to 58.08 Gy in 16 fractions. This resulted in a significant reduction of the radiation dose to the adjacent organs.[29-31] For ESGC, it showed noninferiority compared to local radiotherapy.[30,32,33] Tumor-specific and patient-specific data were retrospectively obtained from Erasmus Medical Center patient records. These variables included: treatment, age (years), gender, ACE-27 comorbidity (0-3), World Health Organization (WHO) performance score (0-4), smoking (yes, no or former), T-stage (Tcis, T1a or T1b), involvement of the

anterior commissure (yes or no). WHO performance score comprises a score for a patient's physical capability of functioning in daily life. Comorbidity was scored at the time of diagnosis by the ACE-27 which is developed specifically for Head and Neck Cancer.[34,35] Time to recurrence was the calculated time from initial treatment up to the occurrence of recurrent disease.

Statistical Analysis

Statistical analyses were performed using R version 4.1.0 (28). Packages that were used are JMbayes2: version 0.3.0 (to apply the joint models)[36]; splines: version 4.2.1 (to assume nonlinear time structure)[37]; lattice: version 0.20.45 (to visualize the data and results)[38]. Descriptive statistics were used to summarize patient, tumor, and treatment characteristics. Means (SD) and medians (Q1-Q3) were used for continuous variables and n (%) for categorical variables. The framework of joint models of longitudinal and time-to-event data was used to obtain dynamic predictions for patient-reported voice quality and recurrent disease. A joint longitudinal model consists of a mixed-effects and a time-to-event submodel. These models can be used when focusing either on the longitudinal outcome (patient-reported outcomes), and we want to correct for nonrandom dropout (due to recurrence), or on the time-to-event outcome (time-to-recurrence) when we want to account for the effect of an endogenous time-dependent covariate (patient-reported outcomes)[12]. An advantage of these models is that the predictions can be updated as more information becomes available.

For the longitudinal submodel, we assumed similar model structures as previously presented. In particular, in previous research, the associations between the outcome and several demographic and clinical variables were investigated.[24] We assumed natural cubic splines (with different degrees of freedom) to capture the nonlinear profiles of the outcome and previously mentioned demographic and clinical variables. By assuming different nonlinear time structures and variations in demographic and clinical variables, we can test whether more included variables also provide better predictions. For the time-to-event submodel, we assume a relative risk submodel with P-splines approximation for the baseline hazard and treatment and age as covariates. The optimal model was selected by means of comparing the predictive performance of the different models. This was done by comparing the predicted and the observed VHI measurement of the testing data set. The overall performance of the longitudinal submodel was assessed by calculating the absolute difference between the predicted and the observed VHI measurement (absolute error). Also, root mean square error (RMSE) per model was calculated. Overfitting-corrected estimates of the predictive performance measures described above were obtained using a cross-validation procedure (internal validation). For long-term clinical relevance, the cross-validation procedure was focused on predictions between 22 and 26 months. The data set was split into 5 subsets, of which 4 were used to fit the model and 1 for obtaining predictions. In smaller data sets, the heterogeneity between these different subsets can be considerable. Hence, to stabilize the results, we have repeated the splitting of the original data set into 5 subsets 100 times. Due to the small number of events, it was not possible to evaluate the risk predictions for the time-to-recurrence event. However, incorporation of the time-to-event model does correct for nonrandom dropout due to recurrent disease. This means that patients with recurrent disease were included; however, it was corrected in the longitudinal prediction analysis. The distribution of the absolute errors is presented using box plots (**Figure 1**). Finally, we illustrate the dynamic longitudinal predictions and the 95% prediction interval for a randomly selected patient.

Results

This prospectively obtained data set consisted of 294 patients treated for ESGC, of which 81.3% of patients are male. Patients endured Tcis (35.0%), T1a (52.7%), and T1b (12.2%) malignancies. Thirty-seven patients (12.6%) had recurrent disease with a mean time to recurrence of 26 months (SD 18.8). The mean follow-up was 43.4 months (SD 21.5) and a total of 2266 measurements were retrieved, with a mean of 8 per patient. The amount of patients during follow-up were: intake (n = 294, 100%), 12 (n = 273, 92.9%), 24 (n = 244, 83.0%), 36 (n = 189, 64.3%), 48 (n = 131, 44.6%), and 60 months (n = 82, 27.9%). For all baseline characteristics, see **Table 1**.

Joint Model Development

Figure 2 depicts all the VHI trajectories and highlights individual patients with varying trajectories. During model development, we used natural cubic splines with 4, 5, and 6 degrees of freedom and variations of the following demographic and clinical variables. These models can be found in **Table 2**. Consequently, the prediction overall performance of these models was validated.

Cross-Validation of Prediction Performance and Model Selection

For the 9 different prediction models, the absolute errors per number of measurements used in the prediction models are plotted in **Figure 1**. All corresponding absolute errors with 25th and 75th percentiles (interquartile range [IQR]) can be found in Appendix 1. As shown in **Figure 1**, increasing the amount of longitudinal VHI measurements decreases the absolute error and provides more trustworthy predictions. For example, the median absolute error for model 1

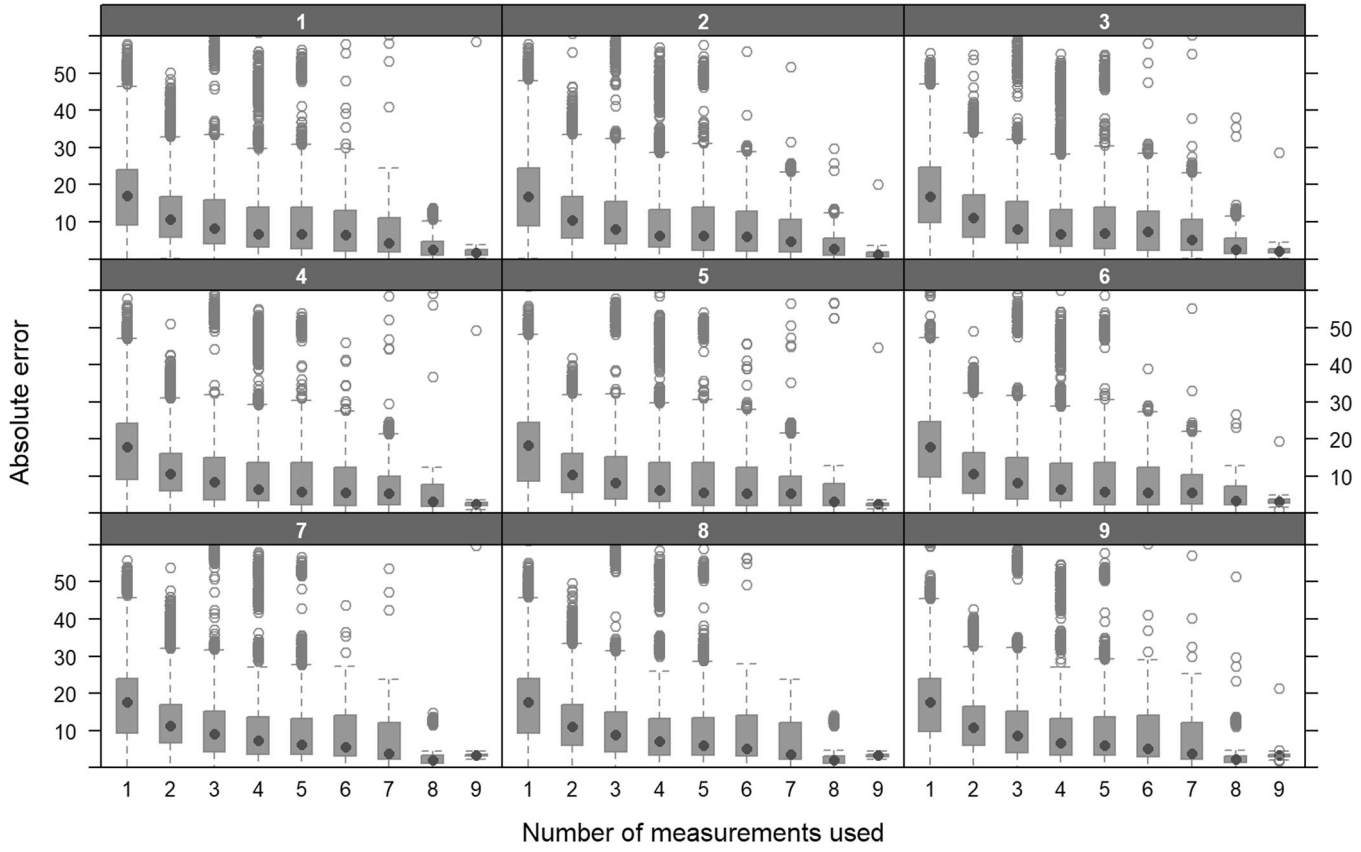

Figure 1. For the 9 different prediction models, the absolute errors per number of measurements used to obtain the predictions. Each panel represents a different model assuming different fixed effects structures

incorporating only 1 VHI measurement is 16.9 (IQR: 9.0-24.1), compared to an error of 1.7 (IQR: 1.0-2.5) when 9 previous measurements are used. When comparing the overall performance of the different models between 22 and 26 months, no clinically significant differences were found in median absolute errors between the models. Including more clinical and demographic variables within these models did not provide better predictions. Furthermore, the time structure did not seem to affect the predictive performance of the models. In addition, the simpler nonlinear structure (4 cubic splines and 3 degrees of freedom) performed similarly to more complex structures. The RMSE measures are 18.6 (1), 17.9 (2), 17.8 (3), 17.6 (4), 18.3 (5), 18.8 (6), 18.1 (7), 18.2 (8), and 17.6 (9).

Overall, the simpler model (model 9) with only treatment can be selected for further analysis and patient-specific prediction visualization. Median absolute errors for this model vary between 17.6 (IQR: 9.7-24.0) when incorporating 1 VHI measurement, and 3.3 (IQR: 2.0-3.6) for 9 incorporated measurements.

Patient-Specific Prediction Visualization

Figure 3 shows an example of a patient-specific longitudinal dynamic prediction trajectory for a randomly chosen patient. This prediction model is able to update predictions at every time point (eg, visit to the outpatient clinic during follow-up) when new information on perceived voice quality becomes available. In this figure, 95% of prediction intervals are visualized in blue. They become narrower and predictions become more accurate when additional measurements are used.

Discussion

In this study, a unique individualized, dynamic prediction model for patient-reported voice quality was developed by using the framework of joint modeling. To our knowledge, this is the first model in head and neck oncology that dynamically predicts longitudinal patient-reported voice quality, which means that individualized predictions can be provided at every time point from intake to the end of follow-up. We propose a clinically applicable model which provides new predictions during follow-up as soon as new information on the VHI becomes available. With this study, we showed the feasibility of an individualized longitudinal prediction model and corresponding graphical outcomes.

Prognostication is considered an important aspect of clinical decision-making. The use of individualized prognostic models in clinical practice enables expectation

Table 1. Baseline Characteristics

Variable	Overall
Patients	294 (100%)
Mean age (SD)	
Gender	
Male	239 (81.3%)
Female	55 (18.7%)
T-stage	
Cis	103 (35.4%)
1a	155 (54.1%)
1b	36 (12.2%)
Comorbidity (ACE-27)	
0	78 (26.5%)
1	129 (43.9%)
2	60 (20.4%)
3	27 (9.2%)
WHO performance status	
0	229 (77.9%)
1	48 (16.3%)
2 + 3	17 (5.8%)
Anterior commissure	
Yes	108 (36.7%)
No	186 (63.3%)
Smoking	
Yes	138 (46.9%)
No	16 (5.4%)
Former	140 (47.6%)
Mean pack years (SD)	34.4 (18.1)

Abbreviations: ACE, Adult Comorbidity Evaluation; WHO, World Health Organization.

Table 2. Created Joint Models Assuming Different Nonlinear Time Structures and Variations of Demographic and Clinical Variables

Mixed-effects submodel 1: interaction between time (natural cubic splines with 6 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, WHO performance score, smoking, T stage and involvement of anterior commissure.*
Mixed-effects submodel 2: interaction between time (natural cubic splines with 6 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, smoking, and T-stage.*
Mixed-effects submodel 3: interaction between time (natural cubic splines with 6 *degrees of freedom*) and treatment.
Mixed-effects submodel 4: interaction between time (natural cubic splines with 5 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, WHO performance score, smoking, T-stage, and involvement of anterior commissure.*
Mixed-effects submodel 5: interaction between time (natural cubic splines with 5 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, smoking, and T-stage.*
Mixed-effects submodel 6: interaction between time (natural cubic splines with 5 *degrees of freedom*) and treatment.
Mixed-effects submodel 7: interaction between time (natural cubic splines with 4 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, WHO performance score, smoking, T-stage, and involvement of anterior commissure.*
Mixed-effects submodel 8: interaction between time (natural cubic splines with 4 *degrees of freedom*) and treatment; *main effects of time, treatment, age, sex, ACE-27 comorbidity, smoking, and T-stage.*
Mixed-effects submodel 9: interaction between time (natural cubic splines with 4 *degrees of freedom*) and treatment.
Time-to-event submodel: main effects of treatment and age.

Every mixed-effects submodel (1-9) was combined with the time-to-event submodel to create a joint model.
Abbreviations: ACE, Adult Comorbidity Evaluation; WHO, World Health Organization.

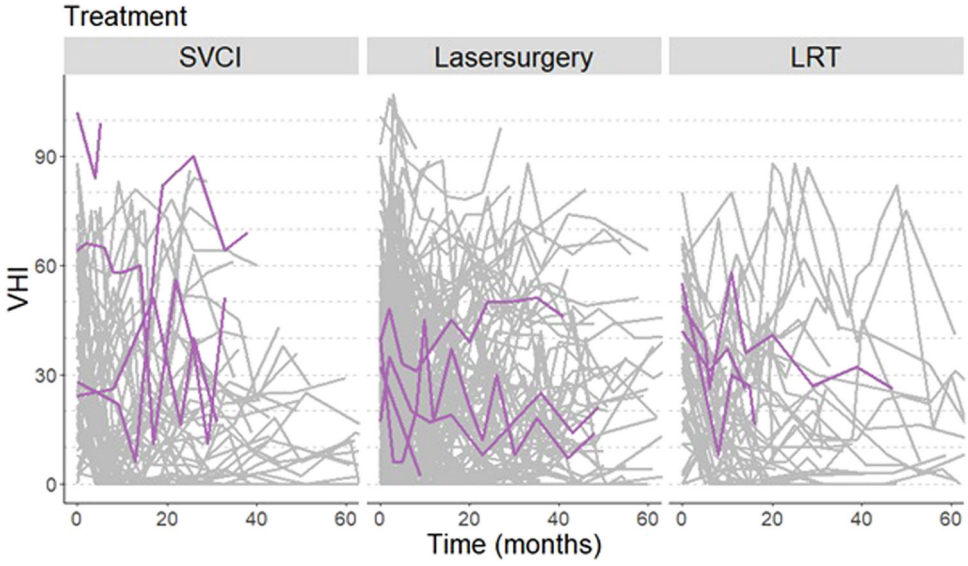

Figure 2. VHI profiles for all 294 patients and highlighting 5 individual patients. This figure shows the variability between patients in longitudinal outcomes. LRT, local radiotherapy; SVCI, single vocal cord irradiation; VHI, Voice Handicap Index

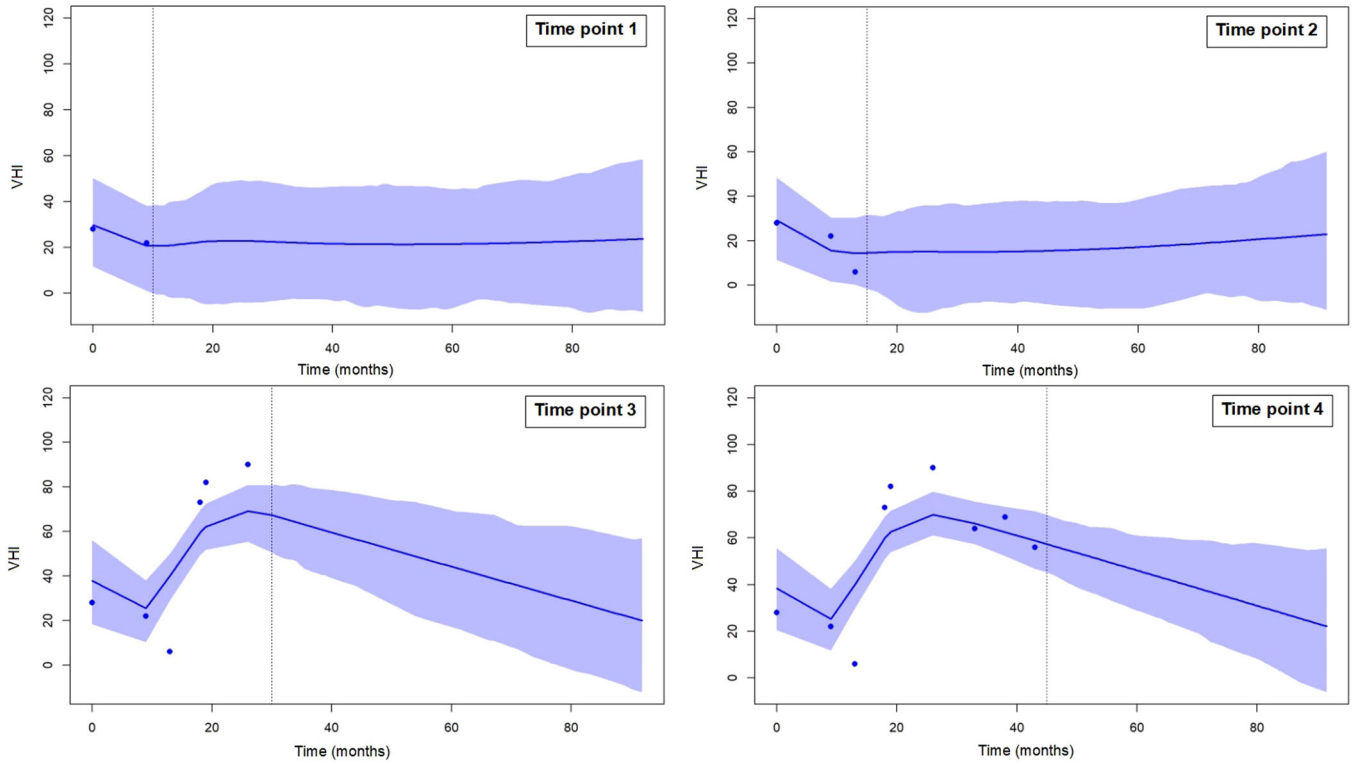

Figure 3. An example of the graphical output of a dynamic prediction model for 1 specific patient. VHI is plotted on the *Y*-axis, and time on the *X*-axis. Every frame is a following point in time. This prediction model is able to update predictions at every time point (eg, visit to the outpatient clinic during follow-up). VHI, Voice Handicap Index

management and, therefore, more personalized counseling and care. Within head and neck oncology, most prediction models focus on clinical binomial outcomes like survival and recurrent disease and use classic linear, logistic, or Cox regression analysis. In contrast to these conventional models, the prediction of longitudinal PROMs requires a different approach. The collection and use of PROMs on an individual level in clinical practice is expanding in all specialties.[39-42] Therefore, we believe that this study is a major step forward within the field of prognostic research and an excellent showcase for the use of PROMs within individualized prediction models.

Both mixed and joint modeling has enjoyed a renaissance in recent years in the statistics and biostatistics literature, which improved the current status quo in prognostic research as it provides more opportunities for longitudinal data than the aforementioned conventional methods.[12,43] These models have shown similar and interesting results and clinically useful models in urology[13,14] and neurology.[15,16]

Unfortunately, we were not able to combine our longitudinal dynamic predictions with the prediction of time-to-event data (recurrent disease) predictions. This was due to the small number of recurrent events, which caused the inability to investigate the predictive performance using a cross-validation procedure. Using a larger, maybe multicenter data set would be beneficial to optimize our prediction model.

During cross-validation, we found that a model with only treatment and nonlinear time assuming no more than 4 degrees of freedom performed just as good as models including more clinical and demographic variables. This is in line with the results of our previous study, in which we identified longitudinal trajectories for patient-reported voice quality for transoral laser microsurgery, local radiotherapy, and single vocal cord irradiation.[24] In that study, we showed that there were few predictive factors for longitudinal patient-reported voice quality, each with a small or negligible effect.

The use of the prediction model that we have developed is able to provide insight into the outcome VHI, which can be used to better inform patients and health care experts on the expectation for a specific patient. However, the used data is not based on a randomized controlled trial and is therefore prone to confounding by indication.[44,45] The current model broadens our possibilities; however, it should not be used as a decision tool. When treatment options are equivalent in ESGC, insights from this model can also be used prior to treatment during shared decision-making in addition to oncological and practical considerations.[46,47]

Clinical Application and Future Perspectives

The dynamic model with treatment and nonlinear time will be integrated into our current PROM-based clinical support system, Healthcare Monitor.[23] By doing this, health care professionals can obtain real-time individualized graphical predictions for patient-reported voice

quality at any given moment during follow-up. At our department, a prognostic model for overall survival (OncologIQ) is already integrated into our electronic health record via Healthcare Monitor.[19,48] By combining quantitative with qualitative prognostic information, we hope to empower patients and professionals to make well-informed and shared decisions and enable tailor-made counseling and customized solutions during follow-up. Based on the methodology of this study, we will continue developing individualized prediction models for other PROMs. For example, we can use domains from other validated and internally used QoL questionnaires (EORTC-QLQ-C30, EORTC-QLQ-HN35, Hospital Anxiety Depression Scale, and Eating Assessment Tool-10). We will also focus on investigating appropriate predictive performance measures using bootstrapping in joint models while assuming different scenarios for the longitudinal and the survival outcomes.[43] Furthermore, we will focus on dashboard development and evaluation together with health care professionals and patients.

Strengths and Limitations

A major strength of our study is the use of relatively new, but appropriate statistical techniques for the prediction of repeated measurements and time-to-event data. We would argue that dynamic prediction modeling should be standard in this field of research as it provides a solution to the need for updated and more precise predictions during follow-up due to changing medical and patient-reported outcomes. Another strength is the amount of included measurements which can be attributed to our institutional routine with the Healthcare Monitor.[23] However, we acknowledge the subjectivity of the VHI and therefore a need for as much data as possible as this would improve the accuracy of predictions.

In this study, a limitation appeared to be the small amount of recurrences for which it was not possible to investigate the predictive performance using a cross-validation procedure for the time-to-recurrence outcome. However, we used the time-to-event model for non-random dropout correction, which enables more fair predictions. Another limitation of our study is the choice of specific (sub)models including specific variables. We have based our models on the results of our previous study, however, the model selection remains arbitrary. Including other variables, or different structures for (non-)linear time can provide other results. Due to missing data, we were not able to include the depth of cordectomy or smoke cessation behavior. Both showed to be important factors for functional outcomes in ESGC.[49,50] In our cohort, 60% underwent TLM, which could cause treatment bias.

Conclusion

In this study, we developed and cross-validated multiple individualized prediction models for longitudinal patient-reported voice quality for patients treated for ESGC. The best-performing joint model was a construct of a mixed-effect model (voice outcome as the function of the interaction of time with treatment) and a time-to-event model (including treatment and age). This dynamic model is able to provide updated predictions during follow-up. We were not able to combine these qualitative predictions with quantitative predictions for recurrent disease due to the small number of events. This model will be integrated into our electronic health record. It has the potential to empower patients and professionals in making well-informed and shared decisions and enable tailor-made counseling and customized solutions during follow-up.

Author Contributions

Maarten C. Dorr, design, conduct, analysis, presentation; **Eleni-Rosalina Andrinopoulou**, analysis, presentation; **Aniel Sewnaik**, design, presentation; **Diako Berzenji**, conduct; **Kira S. van Hof**, analysis; **Emilie A.C. Dronkers**, **Simone E. Bernard**, **Arta Hoesseini**, conduct; **Dimitris Rizopoulos**, analysis, presentation; **Robert J. Baatenburg de Jong**, **Marinella P.J. Offerman**, design, presentation.

Disclosures

Competing interests: The authors hereby declare that there is no conflict of interest related to this study.

Funding source: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

Data can be obtained on request. Requests should be directed toward the data management team of the Head and Neck department of the Erasmus MC Cancer Institute (hoofdhalschirurgie@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and the informed consent of the participants, data cannot be made freely available in a public repository.

ORCID iD

Maarten C. Dorr https://orcid.org/0000-0003-0900-3438

References

1. Ferreira N, Netto E, Fonseca L, et al. Surgery versus radiotherapy: long term outcomes of T1 glottic cancer. *Rep Pract Oncol Radiother*. 2020;25(6):860-866.
2. Abdurehim Y, Hua Z, Yasin Y, Xukurhan A, Imam I, Yuqin F. Transoral laser surgery versus radiotherapy: systematic review and meta-analysis for treatment options of T1a glottic cancer. *Head Neck*. 2012;34(1):23-33.
3. Schrijvers ML, van Riel EL, Langendijk JA, et al. Higher laryngeal preservation rate after CO_2 laser surgery compared with radiotherapy in T1a glottic laryngeal carcinoma. *Head Neck*. 2009;31(6):759-764.
4. Vaculik MF, MacKay CA, Taylor SM, Trites JRB, Hart RD, Rigby MH. Systematic review and meta-analysis of T1

glottic cancer outcomes comparing CO_2 transoral laser microsurgery and radiotherapy. *J Otolaryngol Head Neck Surg*. 2019;48(1):44. doi:10.1186/s40463-019-0367-2
5. Low THH, Yeh D, Zhang T, et al. Evaluating organ preservation outcome as treatment endpoint for T1aN0 glottic cancer. *Laryngoscope*. 2017;127(6):1322-1327.
6. Warner L, Chudasama J, Kelly CG, et al. Radiotherapy versus open surgery versus endolaryngeal surgery (with or without laser) for early laryngeal squamous cell cancer. *Cochrane Database Syst Rev*. 2014;2014(12):002027.
7. Greulich MT, Parker NP, Lee P, Merati AL, Misono S. Voice outcomes following radiation versus laser microsurgery for T1 glottic carcinoma: systematic review and meta-analysis. *Otolaryngol Head Neck Surg*. 2015;152(5):811-819.
8. Guimarães AV, Dedivitis RA, Matos LL, Aires FT, Cernea CR. Comparison between transoral laser surgery and radiotherapy in the treatment of early glottic cancer: a systematic review and meta-analysis. *Sci Rep*. 2018;8(1):11900.
9. Spielmann PM, Majumdar S, Morton RP. Quality of life and functional outcomes in the management of early glottic carcinoma: a systematic review of studies comparing radiotherapy and transoral laser microsurgery. *Clin Otolaryngol*. 2010;35(5):373-382.
10. Du G, Liu C, Yu W, et al. Voice outcomes after laser surgery vs. radiotherapy of early glottic carcinoma: a meta-analysis. *Int J Clin Exp Med*. 2015;8(10):17206-17213.
11. Piazza C, Paderno A, Del Bon F, et al. Long-term oncologic outcomes of 1188 Tis-T2 glottic cancers treated by transoral laser microsurgery. *Otolaryngol Head Neck Surg*. 2021;165:321-328. doi:10.1177/0194599820983727
12. Rizopoulos D. *Joint Models for Longitudinal and Time-to-event Data: With Applications in R*. CRC Press; 2012.
13. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics*. 2019;75(1):153-162.
14. Tomer A, Nieboer D, Roobol MJ, et al. Personalised biopsy schedules based on risk of Gleason upgrading for patients with low-risk prostate cancer on active surveillance. *BJU Int*. 2021;127(1):96-107. doi:10.1111/bju.15136
15. Selles RW, Andrinopoulou ER, Nijland RH, et al. Computerised patient-specific prediction of the recovery profile of upper limb capacity within stroke services: the next step. *J Neurol Neurosurg Psychiatry*. 2021;92(6):574-581.
16. Vliet R, Selles RW, Andrinopoulou ER, et al. Predicting upper limb motor impairment recovery after stroke: a mixture model. *Ann Neurol*. 2020;87(3):383-393.
17. Palazón-Bru A, Mares-García E, López-Bru D, et al. A critical appraisal of the clinical applicability and risk of bias of the predictive models for mortality and recurrence in patients with oropharyngeal cancer: systematic review. *Head Neck*. 2020;42(4):763-773. doi:10.1002/hed.26025
18. Palazón-Bru A, Mares-García E, López-Bru D, Mares-Arambul E, Gil-Guillén VF, Carbonell-Torregrosa MÁ. A systematic review of predictive models for recurrence and mortality in patients with tongue cancer. *Eur J Cancer Care*. 2019;28(6):e13157.
19. Hoesseini A, Van Leeuwen N, Offerman MPJ, et al. Predicting survival in head and neck cancer: external validation and update of the prognostic model OncologIQ in 2189 patients. *Head Neck*. 2021;43:2445-2456.
20. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin*. 2004;14(3):809-834.
21. Andrinopoulou ER, Eilers PHC, Takkenberg JJM, Rizopoulos D. Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*. 2018;74(2):685-693.
22. Detry MA, Ma Y. Analyzing repeated measurements using mixed models. *JAMA*. 2016;315(4):407-408.
23. Dronkers EAC, Baatenburg de Jong RJ, van der Poel EF, Sewnaik A, Offerman MPJ. Keys to successful implementation of routine symptom monitoring in head and neck oncology with "Healthcare Monitor" and patients' perspectives of quality of care. *Head Neck*. 2020;42:3590-3600.
24. Dorr MC. Longitudinal patient-reported voice quality in early-stage glottic cancer. *Otolaryngol Head Neck Surg*. 2022;168:1463-1471.
25. Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, et al. Validation of the Voice Handicap Index by assessing equivalence of European translations. *Folia Phoniatr Logop*. 2008;60(4):173-178. doi:10.1159/000127836
26. Hakkesteegt MM, Wieringa MH, Gerritsma EJ, Feenstra L. Reproducibility of the Dutch version of the Voice Handicap Index. *Folia Phoniatr Logop*. 2006;58(2):132-138. doi:10.1159/000089613
27. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6(3):66-70. doi:10.1044/1058-0360.0603.66
28. Remacle M, Eckel HE, Antonelli A, et al. Endoscopic cordectomy. A proposal for a classification by the Working Committee, European Laryngological Society. *Eur Arch Otrhinolaryngol*. 2000;257(4):227-231.
29. Tans L, Al-Mamgani A, Kwa SLS, et al. Single vocal cord irradiation for early-stage glottic cancer: excellent local control and favorable toxicity profile. *Oral Oncol*. 2022;127:105782.
30. Al-Mamgani A, Kwa SLS, Tans L, et al. Single vocal cord irradiation: image guided intensity modulated hypofractionated radiation therapy for T1a glottic cancer: early clinical results. *Int J Radiat Oncol Biol Phys*. 2015;93(2):337-343.
31. Kocak Uzel E, Figen M, Uzel ÖE. Is single cord irradiation going to be a new standard for T1a glottic carcinoma? *Front Oncol*. 2020;10:1447.
32. Al-Mamgani A, van Rooij PH, Mehilal R, Verduijn GM, Tans L, Kwa SLS. Radiotherapy for T1a glottic cancer: the influence of smoking cessation and fractionation schedule of radiotherapy. *Eur Arch Otrhinolaryngol*. 2014;271(1):125-132. doi:10.1007/s00405-013-2608-8
33. Osman SOS, Astreinidou E, de Boer HCJ, et al. IMRT for image-guided single vocal cord irradiation. *Int J Radiat Oncol Biol Phys*. 2012;82(2):989-997. doi:10.1016/j.ijrobp.2010.12.022

34. Piccirillo JF, Costas I, Claybour P, Borah AJ, Grove L, Jeffe D. The measurement of comorbidity by cancer registries. *J Registry Manage*. 2003;30(1):8-15.
35. Piccirillo JF, Costas I. The impact of comorbidity on outcomes. *ORL J Otorhinolaryngol Relat Spec*. 2004;66(4): 180-185.
36. Rizopoulos D, Papageorgiou G, Afonso PM. JMbayes2: extended joint models for longitudinal and time-to-event data, version 0.3.0. 2022, Accessed June, 2023. https://drizopoulos.github.io/JMbayes2/
37. Wenjie Wang JY. Regression spline functions and classes. 2022, Accessed June, 2023. https://cran.r-project.org/web/packages/splines2/index.html
38. Sarkar D, Andrews F, Wright K, et al. Lattice: trellis graphics for R. 2022, Accessed May, 2023. https://cran.r-project.org/web/packages/lattice/index.html
39. Gibbons C, Porter I, Gonçalves-Bradley DC, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. *Cochrane Database Syst Rev*. 2021;10(10):CD011589. doi:10.1002/14651858.CD011589.pub2
40. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res*. 2008;17(2):179-193.
41. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res*. 2013;22(9):2265-2278.
42. Damman OC, Jani A, Jong BA, et al. The use of PROMs and shared decision-making in medical encounters with patients: an opportunity to deliver value-based health care to patients. *J Eval Clin Pract*. 2020;26(2):524-540.
43. Andrinopoulou ER, Harhay MO, Ratcliffe SJ, Rizopoulos D. Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data. *Int J Epidemiol*. 2021;50(5):1731-1743.
44. Salas M, Hotman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149(11):981-983.
45. Hoesseini A, van Leeuwen N, Sewnaik A, et al. Key aspects of prognostic model development and interpretation from a clinical perspective. *JAMA Otolaryngol Head Neck Surg*. 2021;148:180-186.
46. van Loon Y, Hendriksma M, Langeveld TPM, de Jong MA, Baatenburg de Jong RJ, Sjögren EV. Treatment preferences in patients with early glottic cancer. *Ann Otol Rhinol Laryngol*. 2018;127(3):139-145.
47. Nederlandse Werkgroep Hoofd-Hals Tumoren. *Larynxcarcinoom Landelijke richtlijn*. Nederlandse Werkgroep Hoofd-Hals Tumoren; 2010.
48. Dorr MC, Hoesseini A, Sewnaik A, Hardillo JA, Baatenburg de Jong RJ, Offerman MPJ. Impact of a prognostic model for overall survival on the decision-making process in a head and neck cancer multidisciplinary consultation meeting. *Head Neck*. 2022;44(11):2481-2490.
49. Kujath M, Kerr P, Myers C, et al. Functional outcomes and laryngectomy-free survival after transoral CO_2 laser microsurgery for stage 1 and 2 glottic carcinoma. *J Otolaryngol Head Neck Surg*. 2011;40(suppl 1):S49-S58.
50. Al-Mamgani A, van Rooij PH, Woutersen DP, et al. Radiotherapy for T1-2N0 glottic cancer: a multivariate analysis of predictive factors for the long-term outcome in 1050 patients and a prospective assessment of quality of life and voice handicap index in a subset of 233 patients. *Clin Otolaryngol*. 2013;38(4):306-312.

Appendix 1: Median absolute errors with 25th and 75th percentiles correspond to **Figure 1**

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*	*Model 8*	*Model 9*
1	16.9 (9.0-24.1)	16.7 (8.8-24.5)	16.8 (9.7-24.7)	17.9 (9.1-24.7)	18.3 (8.7-24.5)	17.9 (9.6-24.7)	17.7 (9.3-23.9)	17.7 (9.3-23.9)	17.6 (9.7-24.0)
2	10.6 (5.9-16.7)	10.5 (5.7-16.9)	11.0 (5.9-17.2)	10.6 (6.1-16.0)	10.5 (5.6-16.2)	10.6 (5.4-16.2)	11.2 (6.7-16.9)	11.0 (6.0-17.0)	10.8 (5.9-16.6)
3	8.3 (4.1-15.8)	7.9 (4.1-15.5)	11.0 (4.3-15.5)	8.5 (3.7-15.0)	8.2 (3.8-15.2)	8.3 (3.9-15.0)	9.0 (4.2-15.2)	8.8 (4.2-15.1)	8.5 (4.0-15.3)
4	6.8 (3.3-13.9)	6.4 (3.1-13.3)	8.0 (3.4-15.5)	6.5 (3.3-13.7)	6.2 (3.1-13.8)	6.5 (3.3-13.5)	7.2 (3.6-13.6)	7.0 (3.4-13.2)	6.7 (3.4-13.1)
5	6.7 (2.8-14.0)	6.4 (2.4-13.3)	6.9 (2.8-13.9)	5.8 (2.3-13.6)	5.6 (2.1-13.6)	5.7 (2.3-13.6)	6.2 (3.6-13.3)	6.0 (3.4-13.5)	6.0 (3.4-13.7)
6	6.6 (2.2-13.1)	6.1 (2.2-12.9)	7.3 (2.3-12.8)	5.6 (2.2-12.4)	5.4 (2.1-12.5)	5.5 (2.2-12.4)	5.5 (3.2-14.2)	5.1 (3.0-14.1)	5.1 (2.9-14.1)
7	4.4 (1.9-11.2)	4.7 (2.0-10.6)	5.1 (2.3-10.7)	5.3 (2.4-10.0)	5.4 (2.2-10.0)	5.6 (2.5-10.3)	3.8 (2.3-12.1)	3.7 (2.3-12.1)	3.7 (2.2-12.0)
8	2.6 (1.1-4.8)	2.7 (1.1-5.6)	2.6 (1.5-5.5)	3.1 (2.0-7.8)	3.1 (2.0-7.9)	3.4 (2.2-7.4)	1.9 (1.1-3.3)	2.1 (1.2-3.1)	2.2 (1.4-3.0)
9	1.7 (2.5)	1.1 (0.6-1.9)	2.1 (1.6-2.8)	2.5 (2.1-2.8)	2.4 (2.1-2.8)	2.4 (2.7-3.7)	3.1 (3.0-3.7)	3.3 (2.9-3.6)	3.3 (3.0-3.6)