Received: 8 June 2020 | Revised: 8 March 2021 | Accepted: 8 April 2021

DOI: 10.1002/hed.26716

ORIGINAL ARTICLE

WILEY

Predicting survival in head and neck cancer: External validation and update of the prognostic model OncologIQ in 2189 patients

Arta Hoesseini MD, MSc[1] | **Nikki van Leeuwen PhD**[2] | **Marinella P. J. Offerman PhD**[1] | **Jang Zhang BSc**[1] | **Emilie A. C. Dronkers PhD**[1] | **Aniel Sewnaik PhD**[1] | **Hester F. Lingsma PhD**[2] | **Robert Jan Baatenburg de Jong PhD**[1]

[1]Department of Otorhinolaryngology – Head and Neck Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands

[2]Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

Correspondence
Arta Hoesseini, Department of Otorhinolaryngology – Head and Neck Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Dr Molewaterplein 40, 3015 GD Rotterdam, the Netherlands.
Email: a.hoesseini@erasmusmc.nl

Funding information
Zorginstituut Nederland, Grant/Award Number: 2018026704

Abstract

Background: Disclosing prognostic information is necessary to enable patients to make well-informed treatment decisions. OncologIQ is a prognostic model that predicts the overall survival (OS) probability in patients with head and neck squamous cell carcinoma (HNSCC). We aimed to externally validate and update the model with new prognostic factors and translate it to a clinically useful tool.

Methods: A consecutive retrospective sample of 2189 patients eligible for curative treatment of a primary HNSCC were selected. Discriminative performance was determined using the C-statistic.

Results: External validation showed systematic underestimation of OS in the new population, and reasonable discrimination (C-statistic 0.67). Adding smoking, pack years, BMI, weight loss, WHO performance, socioeconomic status, and p16 positivity to the recalibrated multivariable model, improved the internally validated C-statistic to 0.71. The model showed minor optimism and was translated in an online tool (www.oncologiq.nl).

Conclusions: The updated model enables personalized patient counseling during treatment decision consultations.

KEYWORDS
head and neck cancer, prediction, prognosis, shared decision making, survival

1 | INTRODUCTION

Worldwide, head and neck cancer (HNC) accounts for more than 700 000 new cases and 350 000 deaths

Arta Hoesseini and Nikki van Leeuwen contributed equally to this study.

annually.[1] Treatment is often mutilating and interferes with vital functions such as swallowing and breathing.[2] In addition, diagnosis and treatment have an impact on psychosocial functioning, and the incidence of depression is estimated to be as high as 40%.[2-4] Therefore, there is a delicate balance of the trade-offs between quantity and quality of life (QoL) when discussing treatment options. Disclosing prognostic information on both survival probabilities and QoL during these discussions is necessary to enable patients to make well-informed treatment decisions.

Previous research showed that health care providers are not able to accurately forecast cancer patients' life-expectancy and tend to overestimate survival.[5-7] Over the last years, an internally and externally validated prognostic model named "OncologIQ" has been developed. This model estimates the 1- to 10-year overall survival (OS) chances of patients with primary HNC who are eligible for curative treatment.[8-10] While traditionally survival rates are solely based on the TNM-classification of the specific tumor, OncologIQ also includes age, sex, and the Adult Comorbidity Evaluation 27 (ACE-27) as prognostic factors for OS.[8-10] With this model, a more personalized estimate of the OS is calculated based on the average treatment effect. These personalized estimations can be used to counsel patients during treatment decision consultations, aiming to support shared decision making (SDM).

Although many prognostic models are developed, few are actually used in clinical practice or evaluated in an impact study.[11-13] A reason for this could be that patients', physicians', and caregivers' wishes and preferences are not taken into account in the process or that the model is based on outdated data. Although OncologIQ is used by some head and neck oncologic centers in the Netherlands, the patient data included in the original model is outdated. Due to the improvement of OS in HNC in the past years,[14] it is likely that the original model underestimates survival chances. This brings us to the first aim of this study: (1) to explore how OncologIQ performs in a present-day patient population: *external validation*. The second aim of this study was: (2) to explore whether adding new prognostic factors would improve OncologIQ's performance: *model revision*. During our qualitative research on prognostication both patients[15] and doctors' addressed the wish to explore new predictors like smoking and alcohol consumption, in order to further personalize the models' estimations. The third aim was to integrate the updated model in a clinically useful tool. The present study builds on our prior prognostic and qualitative research.[15] Finally, our line of research is described in the discussion which illustrates the steps that were necessary to develop a valued tool for use in clinical practice.

2 | MATERIAL AND METHODS

This study was approved by the ethics committee of the Erasmus Medical Center (MEC number: MEC-2016-751). Patients who were diagnosed with a primary head and neck squamous cell carcinoma, diagnosed and treated at the Erasmus MC Cancer Institute between January 1, 2006 and December 31, 2013 were included.

2.1 | Data collection

2.1.1 | Rotterdam oncological documentation

While the original model is based on the Oncological Documentation of the Leiden University Medical Center, data used for this study were retrieved from the Rotterdam hospital-based cancer registry system: Rotterdam Oncological Documentation (RONCDOC). RONCDOC was built according to the following steps. Data gathered in the Netherlands Cancer Registry (NCR) of patients with HNC were obtained from the Netherlands Comprehensive Cancer Organization (NCCO). Information on every patient with cancer in the Netherlands is recorded in the Netherlands Cancer Registry (NCR) and often directly used for research purposes. In RONCDOC, the consecutive NCR data of patients with HNC were first merged with corresponding clinical data from the patient record files of the Erasmus MC Cancer Institute. These data were manually checked for each patient, and enriched with additional variables from the patient charts according to an extensive data entry manual. When data of the NCR did not correspond with the data in the patient charts, the patient file data was considered leading. All baseline variables were scored at the time of diagnosis. Finally, the data were cleaned by two independent members of the research team. A log was kept of all changes made. The primary endpoint was OS. Follow-up time was last updated on August 16, 2018 by consulting the Municipal Personal Records Database (MPRD). Final day of follow-up time for a patient was defined as the final date that the patient was confirmed to be alive or the date of death.

2.1.2 | Patient inclusion

Patients who were eligible for curative treatment of a primary squamous cell carcinoma of the glottic larynx, supraglottic larynx, hypopharynx, oropharynx, oral cavity, nasopharynx, and lip were included ($n = 2253$). Patients who were lost to follow-up ($n = 5$) or refused curative

treatment ($n = 59$) were excluded. In total 2189 patients were included in the analysis.

2.1.3 | Statistics Netherlands

Statistics Netherlands (SN), also known as *Centraal Bureau voor de Statistiek*, is a governmental institution that gathers microdata on individuals in the Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. The data on education and socioeconomic status (SES) are based on non-public microdata from SN and were merged with the RONCDOC dataset.

2.2 | Definitions of variables

cTNM was staged according to the 7th American Joint Committee on Cancer (AJCC) edition of the TNM classification of malignant tumors.[16] In case of oropharyngeal tumors, immunohistochemical analysis was performed for tumor suppressor protein p16 (cyclin-dependent kinase 2A). p16 positivity was defined as strong and diffuse nuclear and cytoplasmic immunostaining in >70% of the tumor cells. p16 positive tumors were considered HPV positive, conform the 8th AJCC TNM guideline.[17,18] To distinguish between HPV positive and negative tumors, "HPV+" and "HPV−" was added to the oropharynx category. Lymph node category (cN) of these oropharynx tumors was based on the TNM-7. The cumulative quantity of smoking was defined in pack-years (PY) in which one pack year was equal to one packet of 20 cigarettes smoked per day for 1 year. If a patient had stopped smoking for ≥3 months, he or she was considered as a *former smoker*. If a patient had stopped drinking for ≥6 months he or she was considered as a *former drinker*. Number of alcohol units per week were scored according to a standardized list: one unit, or 10 grams, of alcohol is equivalent to 12.5 milliliters of pure ethanol.[19] Weight loss in kilograms (kg) was defined as weight loss in the 6 months before diagnosis. WHO performance status, also known as the Eastern Cooperative Oncology Group (ECOG) score, was scored according to the classification published by Oken et al.[20] Marital status was defined as being married or having a durable relationship versus being single or widowed. SES was categorized as (1) employed (self-employed, employee, managing director, etc.) or being a student, (2) unemployed (including receivers of an unemployment benefit, social assistance benefit, etc.), and (3) retired. Successful completed education was categorized according to a national and international classification: (1) lower; less than primary, primary and lower secondary, (2) intermediate; upper secondary, post-secondary non-tertiary, (3) tertiary; short cycle tertiary, bachelor, master, doctoral.[21,22]

2.3 | Statistical analyses

Statistical analyses were performed using the Statistical Package for the Social Sciences (version 22 and 25), and R statistical software (Vienna, Austria, version 3.4.0 and 3.5.2) using the R packages *mice* and *rms*. Different version of these programs were used due to restrictions in the electronic NS environment. Multiple imputation in R was used to for handling missing data (five iterations) that were missing at random. Education could not be imputed since the missing at random assumption was not plausible.

2.3.1 | External validation

For external validation of the original model, performance was assessed using discrimination and calibration analysis.[23,24] Coefficients of the latest original model are not published and were therefore included in Table S1, Supporting Information. Calibration is the agreement between the predicted probabilities and the observed frequencies, in this case the 5- and 10-year survival. The calibration slope is ideally 1 and reflects whether the effects of the predictors are on average correct.[25] The discriminative performance of OncologIQ was assessed by the Harrell's concordance index (C-statistic).[25] The C-statistic is the most commonly used measure to determine the discriminative performance for binary and time-to-event outcome measures. The C-statistic takes values between 0.5 and 1.0, where 0.5 indicates that the model is not better than chance classification and 1 means perfect discrimination.[26] A C-statistic below 0.6 was considered as poor, a C-statistic over 0.6 as moderate, a C-statistic over 0.7 as good and a C-statistic over 0.8 as strong.[27] Usually, model performance is poorer in external validation compared to the performance in the development data.[13] If this is the case, the model should be updated and adjusted to the conditions in the validation cohort to improve performance.[13,28,29]

2.3.2 | Univariable analyses

The Cox proportional hazard regression model was used to calculate the univariable hazard ratios (HR) of overall mortality (OM) with 95% confidence intervals (CI). The log-rank test was used to test significance. $P < 0.05$ were considered significant.

TABLE 1 Patient characteristics and model performance

Characteristic	Frequency (%)/ median (Q1-Q3)/ mean (SD)	Total no. missing (%)	Univariable HR (95% CI)	*p*	Updated models			
					Recalibrated original model, multivariable HR (95% CI)	*p*	Extended model, multivariable HR (95% CI)	*p*
Total	2189							
Sex (female)	612 (28.0)	0	0.90 (0.79, 1.02)	0.1008	0.85 (0.73, 0.98)	0.0230	0.89 (0.77, 1.03)	0.1129
Age	63.1 (10.8)	0	1.03 (1.02, 1.03)	<0.0001	1.03 (1.03, 1.04)	<0.0001	1.04 (1.03, 1.05)	<0.0001
Localization and HPV (glottic larynx)	528 (29.8)	420 (19.2)[a]	1.0		1.0	<0.0001	1.0	<0.0001
Oral cavity	644 (36.4)		1.63 (1.38, 1.93)	<0.0001	1.38 (1.15, 1.66)	0.0006	1.45 (1.20, 1.75)	0.0001
Oropharynx, HPV−	59 (3.3)		2.75 (2.20, 3.45)	<0.0001	2.10 (1.60, 2.74)	<0.0001	1.89 (1.45, 2.46)	<0.0001
Oropharynx, HPV+	41 (2.3)		1.01 (0.69, 1.47)	0.9779	0.79 (0.52, 1.20)	0.2759	0.88 (0.58, 1.34)	0.5439
Nasopharynx	42 (2.4)		1.54 (1.01, 2.34)	0.0462	1.65 (1.05, 2.58)	0.0291	1.97 (1.24, 3.12)	0.0038
Supraglottic larynx	247 (14.0)		2.24 (1.83, 2.74)	<0.0001	1.55 (1.23, 1.96)	0.0002	1.41 (1.11, 1.79)	0.0043
Hypofarynx	177 (10.0)		2.61 (2.10, 3.26)	<0.0001	1.54 (1.18, 2.01)	0.0015	1.38 (1.06, 1.81)	0.0184
Lip	31 (1.4)		0.91 (0.50, 1.67)	0.7641	1.01 (0.55, 1.85)	0.9787	1.14 (0.62, 2.10)	0.6836
cT (1)	661 (30.2)	0	1.0		1.0	<0.0001	1.0	<0.0001
2	684 (31.2)		1.77 (1.51, 2.08)	<0.0001	1.58 (1.33, 1.88)	<0.0001	1.53 (1.29, 1.82)	<0.0001
3	431 (19.7)		2.62 (2.21, 3.11)	<0.0001	2.05 (1.68, 2.51)	<0.0001	1.82 (1.49, 2.22)	<0.0001
4	413 (18.9)		3.02 (2.55, 3.58)	<0.0001	2.41 (1.98, 2.93)	<0.0001	2.00 (1.64, 2.43)	<0.0001
cN (0)	1446 (66.1)	0	1.0		1.0	<0.0001	1.0	<0.0001
1	250 (11.4)		1.58 (1.33, 1.87)	<0.0001	1.26 (1.04, 1.52)	0.0169	1.24 (1.03, 1.50)	0.0269
2	465 (21.2)		1.68 (1.47, 1.92)	<0.0001	1.43 (1.18, 1.73)	0.0003	1.47 (1.22, 1.78)	<0.0001
3	28 (1.3)		3.54 (2.36, 5.30)	<0.0001	3.56 (2.18, 5.80)	<0.0001	3.78 (2.33, 6.15)	<0.0001
ACE-27 (0)	651 (29.7)	0	1.0		1.0	<0.0001	1.0	0.0007
1	803 (36.7)		1.48 (1.27, 1.72)	<0.0001	1.28 (1.09, 1.49)	0.0020	1.18 (1.01, 1.39)	0.0402
2	519 (23.7)		1.84 (1.57, 2.16)	<0.0001	1.50 (1.27, 1.78)	<0.0001	1.33 (1.12, 1.59)	0.0012
3	216 (9.9)		2.81 (2.31, 3.41)	<0.0001	2.00 (1.62, 2.46)	<0.0001	1.55 (1.23, 1.95)	0.0002
Smoking (no)	265 (12.1)	7 (0.3)	1.0				1.0	0.0128
Former	623 (28.6)		1.44 (1.16, 1.80)	0.0010			1.23 (0.96, 1.56)	0.1004
Current	1294 (59.3)		1.72 (1.40, 2.10)	<0.0001			1.41 (1.10, 1.81)	0.0059
Pack years	33.0 (18.0–46.0)	488 (22.3)	1.01 (1.01, 1.01)	<0.0001			1.004 (1.002, 1.006)	0.0009

TABLE 1 (Continued)

Characteristic	Frequency (%)/ median (Q1-Q3)/ mean (SD)	Total no. missing (%)	Univariable HR (95% CI)	p	Updated models			
					Recalibrated original model, multivariable HR (95% CI)	p	Extended model, multivariable HR (95% CI)	p
Alcohol (no)	422 (19.4)	9 (0.4)	1.0					
Former	207 (9.5)		1.77 (1.43, 2.20)	<0.0001				
Current	1551 (71.1)		1.20 (1.03, 1.40)	0.0169				
No. alcohol units per week	14.0 (2.0–28.0)	214 (9.8)	1.00 (1.00, 1.00)	<0.0001				
Body mass index	24.9 (22.2–27.8)	142 (6.5)	0.96 (0.95, 0.97)	<0.0001			0.98 (0.97, 0.99)	0.0056
Weight loss in the past 6 months (kg)[b]	0.0 (0.0–4.0)	364 (16.6)	1.05 (1.03, 1.06)	<0.0001			1.02 (1.00, 1.03)	0.0404
WHO performance (0)	1316 (63.6)	120 (5.5)	1.0				1.0	<0.0001
1	614 (29.7)		1.93 (1.70, 2.19)	<0.0001			1.32 (1.13, 1.54)	0.0004
2	99 (4.8)		3.36 (2.63, 4.29)	<0.0001			2.05 (1.56, 2.70)	<0.0001
3 and 4	40 (1.9)		3.63 (2.56, 5.14)	<0.0001			2.36 (1.60, 3.47)	<0.0001
Marital status (single)	622 (28.4)	71 (3.2)	1.32 (1.16, 1.50)	<0.0001				
Serum hemoglobin (mmol/L)[c]	9.1 (8.5–9.6)	238 (10.9)	0.83 (0.77, 0.88)	<0.0001				
Socioeconomic status (employed/student)	597 (27.5)	19 (0.9)	1.0				1.0	0.0554
Unemployed	504 (23.2)		1.72 (1.45, 2.03)	<0.0001			1.11 (0.93, 1.33)	0.2466
Retired	1069 (49.3)		1.81 (1.56, 2.09)	<0.0001			0.88 (0.72, 1.09)	0.2363
Education (lower)[d]	212 (51.1)	1774 (81.0)	1.0					
Intermediate	147 (35.4)		0.63 (0.43–0.92)	0.017				
Tertiary	56 (13.5)		0.49 (0.27–0.89)	0.019				
C-statistic						0.69		0.71

[a]Missing data are based on HPV status in oropharynx tumors.
[b]1 kilogram (kg) = 2.20 pounds.
[c]Hemoglobin of 9.1 mmol/L = 14.7 g/dL.
[d]Complete case analysis: due to the high % missing data multiple imputation was not possible.

2.3.3 | Model revision (update) and presentation

The model was updated by testing whether the new variables improved model performance. The backward selection method was used for variable selection. Variables were excluded one by one until all variables left in the new model had a $P < 0.10$ (two-sided tests). We used the variables included in the original model (sex, age, tumor localization, T, N, ACE-27) as fixed predictors based on our previous research and for face-validity.[8-10] In the original development set 17 patients with M1 disease were included.[9] However, the current update focuses solely on curative patients; therefore, M status was not included in the update. New predictors to be tested in the stepwise backward selection were smoking, pack years (PY), alcohol consumption, alcohol units per week, body mass index (BMI), weight loss in kg, WHO performance, marital status, serum hemoglobin (Hb), SES. Hereafter, year of diagnosis was added to the extended model to test whether it affected OS. The model was tested for prognostic accuracy using the C-statistic. The difference in survival probabilities for individual patients in the original and the updated model were plotted. Further, the updated model was internally validated using bootstrapping techniques and the shrinkage factor was calculated. The regression coefficients should be multiplied by the shrinkage value to provide more reliable predictions for new patients. A shrinkage factor close to 1 means minor optimism of the model. In other words: a minor difference between the true performance of the model in the underlying population and the apparent performance in the sample.[26,30] The prognostic model was validated, updated, and reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement.[31] The model was integrated in an online tool using shiny R software.

3 | RESULTS

3.1 | Patient population

Baseline characteristics and missing data are summarized in Table 1. Out of the 2189 included patients, 1208 (55.2%) died over 10-year follow-up. The median follow-up time was 5 years and 5 months. Of the 520 oropharynx tumors, in 420 p16 positivity was not tested. In the 100 tested cases, 41 (41%) had p16 positive tumors (see also Table 1). This is comparable with a recent study that included 1204 patients with oropharyngeal squamous cell carcinoma in the Netherlands of which 32% had a p16positive tumor.[32]

3.2 | External validation

Figure 1 shows the calibration plots of the external validation. Predicted probabilities for OS of the original model were systematically lower than the observed frequencies in the new cohort. The calibration slope was 0.88 (95% CI, 0.80–0.98). Previous published discrimination of the original model in the development data showed a C-statistic of 0.73.[9] Discrimination of the original model was tested in the new RONCDOC dataset by applying the coefficients as shown in Table S1. This resulted in a lower discriminative performance with a C-statistic of 0.67.

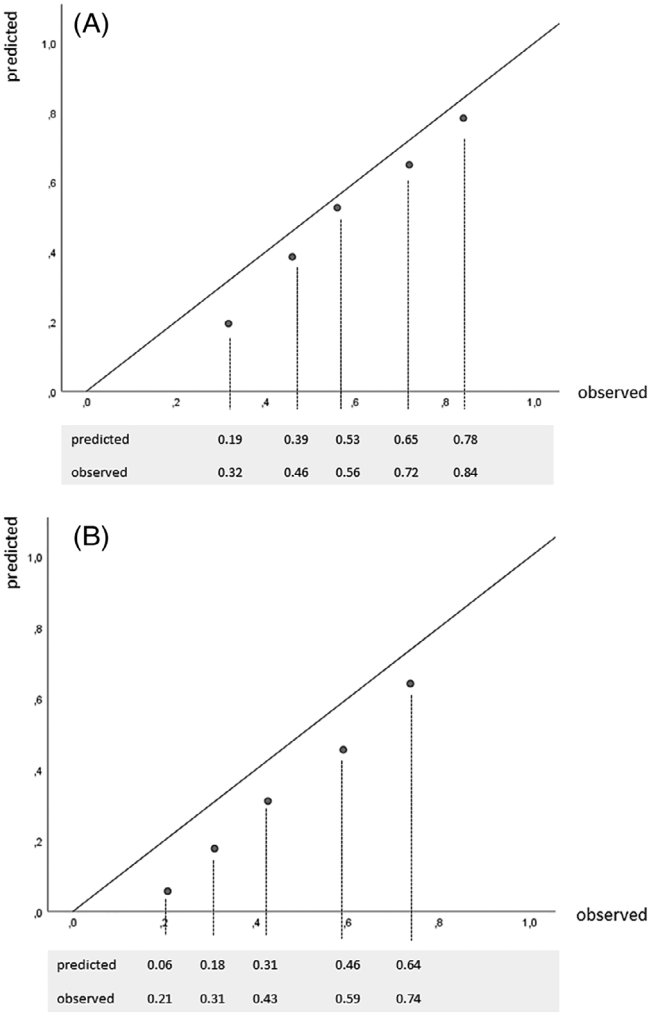

FIGURE 1 Calibration plots: predicted probabilities and observed frequencies in five quintiles at 5 and 10 years after diagnosis. (A) 5-year survival probability. (B) 10-year survival probability

3.3 | Univariable analysis

All candidate predictors showed a univariable significant effect on OM, except for sex (Table 1). Having an HPV+ oropharynx tumor was related to a better survival probability compared to having an HPV− oropharynx tumor. Current smokers had a higher probability of mortality compared to former smokers. Former drinkers had a significant worse survival probability ($p = 0.000$) compared to current drinkers, while the median alcohol consumption per week in the former group was higher (56 units, IQR 28–105) than in de current drinkers (21 units, IQR 11–35). Retired and unemployed patients had a lower survival probability than employed patients. Having finished intermediate or primary education was associated with better survival probability compared to patients who finished lower education.

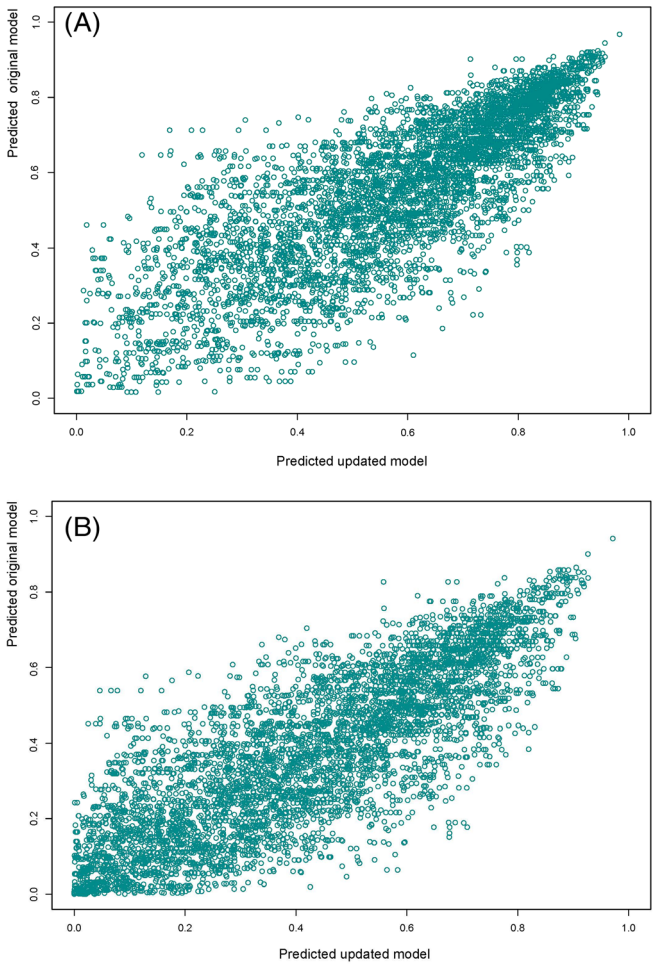

FIGURE 2 The difference in survival probabilities for individual patients in the original and the updated model, at (A) 5 years and (B) 10 years. (A) 5-year survival probability. (B) 10-year survival probability. Each dot represents an individual patient. On the *x*-axis the predicted survival probabilities of the updated model are shown, on the *y*-axis the predicted survival probabilities of the original model are shown. For most patients, both predicted probabilities are similar. However, on an individual level, the updated model shows reasonably different estimations, compared to the original model. This indicates the added value of the updated model at an individual level. The maximal difference in individually estimated probabilities ranged from −0.51 to 0.54 for 5-year survival, and −0.53 to 0.49 for 10-year survival [Color figure can be viewed at wileyonlinelibrary.com]

3.4 | Model revision (update)

The coefficients of the variables in the original model were recalibrated per variable and are shown in Table 1. After recalibrating the original model, the discriminative performance was improved from a C-statistic of 0.67–0.69. Using a stepwise backward selection, the final multivariable model included six new variables. The new prognostic variables smoking, PY, BMI, weight loss, WHO performance, and SES were significant predictors for OS in the multivariable analysis. The HRs for the new variables in the model can be found in Table 1. Adding these prognostic factors improved the discriminative performance of the model from a C-statistic of 0.69 to 0.71. No statistically significant effect of year of diagnosis was found, thus is was not necessary to correct for it. Figure 2 shows the difference in survival probability on an individual level between the original and the updated model. The internally validated C-statistic was 0.71. The model showed only minor optimism, with a shrinkage factor of 0.94.

3.5 | Model presentation (online tool)

The final extended model was translated into the existing online tool.[33] In this updated version smoking, PY, BMI, weight loss, WHO performance and SES were added. As patients[15] and doctors expressed different presentation wishes during our qualitative research, two separate pages were made: one with a survival graph (doctors) and one with a pie chart (patients). Figure 3 shows a screenshot of the online prediction model. The coefficients for the full model equation can be found in Table 2. The updated coefficients were multiplied with the shrinkage factor. Shrinked coefficients are also shown in Table 2.

4 | DISCUSSION

This study aimed to optimize the prognostic model OncologIQ. In total 2189 patients with HNC were

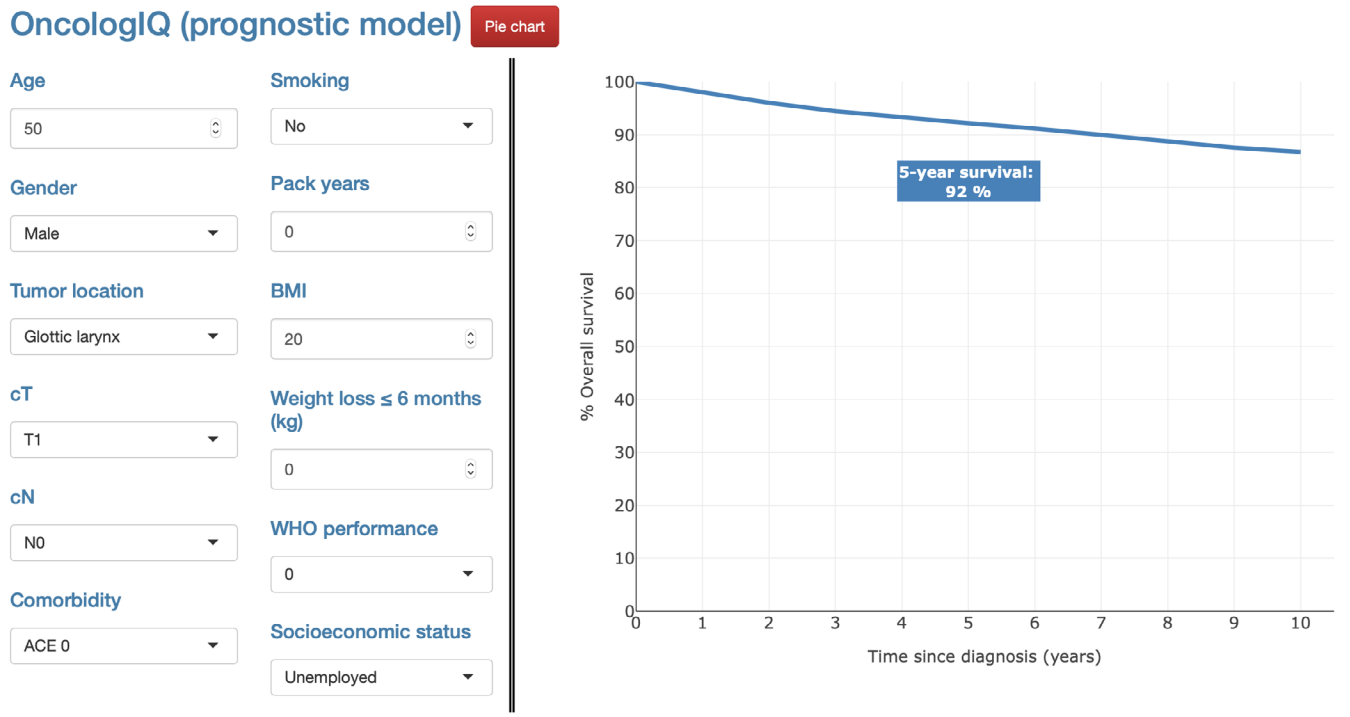

FIGURE 3 Screenshot of the updated model www.oncologiq.nl [Color figure can be viewed at wileyonlinelibrary.com]

included to externally validate and update the original model. In summary, the original model showed systematically underestimation of overall survival in the external validation. Consequently, it was recalibrated. In the updated model the new prognostic variables smoking, PY, BMI, weight loss, WHO performance, and SES were added, which improved its discriminative performance from moderate to good (C-statistic 0.71). The final extended model was integrated in a clinically useful tool, which enables up-to-date personalized prognostic counseling.

4.1 | Predictors of survival after HNC diagnosis

Univariable analysis showed statistically significant relations between all new candidate predictors and OM (see Table 1). For example, alcohol consumption at diagnosis showed to be harmful, while earlier studies among patients with HNC showed conflicting results.[34] Having no partner was associated with worse survival, while a higher serum Hb and a higher BMI showed to be protective. These findings agree with earlier research[35,36] and should be taken into account during counseling. Former drinkers had a lower survival probability compared to current drinkers. This was explained by the significant higher reported alcohol intake in the former drinker group. Although univariable analysis showed a significant relation between OM and all new candidate predictors, in multivariable analysis some of these factors had no extra prognostic value over the factors that were added to the final model. An explanation is that specific predictors, for example, alcohol consumption and smoking, are highly correlated to each other. Therefore the effect of drinking alcohol is implicitly included in the effect of smoking on survival and thus does not improve survival prediction. The discriminative performance of our updated model is comparable with other prediction models that are widely used in clinical practice. For example, the Framingham risk model and the APACHE scores with a C-statistic of little over 0.70.[37,38] Likely reasons for their wide use are the relevance of their predicted outcomes (cardiovascular disease and mortality) and their face validity.[39] The same applies to our model. OncologIQ has high face validity given all the included variables. Furthermore, the prediction of long-term survival chances is highly relevant in HNC care. In addition, the updated model provides substantial different estimated survival probabilities on an individual level (see Figure 2), underlining the necessity to update the original model.

4.2 | Model presentation and communication

In accordance with a recent published guide on how to present prediction models in clinical practice,[40] we

TABLE 2 Coefficients of the full model equation

Characteristic	Coefficients	Shrinked coefficients
Sex (female)	−0.1169	−0.1098
Age	0.0429	0.0403
Localization and HPV (glottic larynx)		
Oral cavity	0.3688	0.3464
Oropharynx, HPV−	0.6370	0.5983
Oropharynx, HPV+	−0.1307	−0.1228
Nasopharynx	0.6774	0.6363
Supraglottic larynx	0.3443	0.3234
Hypofarynx	0.3234	0.3038
Lip	0.1273	0.1196
cT (1)		
2	0.4258	0.3999
3	0.5978	0.5615
4	0.6917	0.6497
cN (0)		
1	0.2162	0.2031
2	0.3883	0.3647
3	1.3310	1.2505
ACE-27 (0)		
1	0.1669	0.1568
2	0.2872	0.2698
3	0.4386	0.4120
Smoking (no)		
Former	0.2034	0.1910
Current	0.3462	0.3252
Pack years	0.0038	0.0036
Body mass index	−0.0203	−0.0191
Weight loss in the past 6 months (kg)	0.0151	0.0142
WHO performance (0)		
1	0.2795	0.2625
2	0.7189	0.6752
3 and 4	0.8567	0.8047
Socioeconomic status (employed/student)		
Unemployed	0.1050	0.0986
Retired	−0.1248	−0.1172
Baseline survival		
5 year	0.58	
10 year	0.39	

engaged patients and physicians to examine the best visual format. Patients preferred a pie chart presenting the 5-year survival rate,[15] while physicians preferred to see the individual 1- to 10-year survival graph (unpublished data). This resulted in a digital tool with separate visual formats. Both visual formats were updated, according to the new extended model. During our qualitative research both patients and doctors asked about the possibility to include treatment in the prognostic model.[15] Although it would be tempting to include treatment modalities in our analyses, this would be invalid to estimate in our observational data. Estimating treatment effectiveness in observational data would provide invalid estimates due to confounding by indication.[41] The golden standard for estimating the effect of treatment is using randomized controlled trial data. Since these data are lacking, we could not include treatment in OncologIQ. The models' survival probabilities are based on the average treatment effectiveness in our population. In the former model, the benefit of chemotherapy was added by an adaptation method.[42] Reason for this was that the development set of the original model hardly included cases in which chemotherapy was added to locoregional treatment, while in the course of time many studies proved the benefit of it.[43,44] During the time patients were included in the RONCDOC database, chemotherapy was part of the standard treatment protocol. Therefore, this hazard is no longer included in the visualization to prevent overestimation of the beneficial effect of chemotherapy. The next step would be to use the model in clinical practice. However, there is little research on how to communicate prognostic model estimates during patient counseling. By organizing focus groups, we recently identified HNC patient preferences on receiving prognostic information.[15] Based on the results, a clinical practice guideline has been developed to support physicians in sharing individualized prognostic information.[15]

4.3 | Strengths and limitations

A major strength of this study was the large number of consecutive patients with HNC ($n = 2189$) included in our study and the high number of events ($n = 1208$) that determines high statistical power. In addition, only five patients were lost to follow-up, as we were able to link our dataset to the MPRD. Given the consecutive design of the RONCDOC cohort, selection bias is eliminated. While data from the NCR is often directly used for analysis, we embedded an extra validation check by comparing the NCR data with the data registered in the patient record files, resulting in a high quality dataset.

A limitation of our study is the presence of missing values of some baseline characteristics. However, the percentage of missing values was limited except for

education and the HPV status of the oropharyngeal tumors. To tackle this issue we used multiple imputation, which is the preferred method over a complete case analysis.[45-47] Also, all tumors were staged according to the TNM-7. To distinguish between HPV positive and negative tumors, "HPV+" and "HPV−" was added to the oropharynx category based on p16 positivity. However, cN and cT classification of these oropharynx tumors was still based on the TNM-7. This implies that the TNM-7 N2a (metastasis in a single ipsilateral lymph node >3 and ≤6 cm) and N2b (metastasis in multiple ipsilateral lymph nodes ≤6 cm) should not be entered as N1 in OncologIQ but as N2. Future research will focus on staging according to the TNM-8. Furthermore, this model has been developed and externally validated using patient data from the Netherlands and the United States.[8-10] As etiologic cause and prognosis may differ in other regions in the world, this model may not be generalizable for all patients with HNC globally. However, this model can be used to externally validate and update in a new and possible different population.

4.4 | Clinical implications and future perspectives

Our overall line of research illustrates a perfect example of the consecutive steps, necessary to develop a valued and clinically useful model[13] that is tailored to patients' and physicians' needs.

- Step 1: Model development and validation.[8-10]
- Step 2: Qualitative research: prognostic disclosure during treatment discussions.[48]
- Step 3: Qualitative research: patients', caregivers', and doctors' preferences on sharing prognostic information.[15,34]
- Step 4: External validation and update.
- Step 5: Evaluation of the clinical impact of the model.

After developing and validating the original model (step 1), the current disclosure of prognostic information was measured (step 2). We used the results of our qualitative research (steps 2 and 3) as the foundation for the current study. The updated model can be used to support SDM during treatment decision consultations. While many prognostic models in cancer care have been developed, few are used in clinical practice or evaluated in an impact study.[11,13] Therefore, it often remains unclear what the impact of model implementation is on both patients and health care providers.[13] A current trial in our institution is evaluating its clinical impact by measuring decisional conflict[49] among patients with HNC (step 5). In this prospective clinical trial with sequential cohorts, patients who received counseling with and without OncologIQ are compared. Also, a pilot study has been done to examine the effect of the use of OncologIQ in our tumor board meetings. In addition, future external validation of the updated model with the new prognostic variables would be appreciated. Information on survival is not a standalone concept and should be combined with information on QoL. Due to the implementation of our Healthcare Monitor[50] we are currently obtaining QoL data. Future research will focus on modeling QoL together with survival.

5 | CONCLUSIONS

This study enabled the external validation and recalibration of the prognostic survival model OncologIQ for patients with primary HNC, eligible for curative treatment. Recalibration was necessary due to systematic underestimation of OS in the new patient population. Adding the new prognostic variables smoking, PY, BMI, weights loss, WHO performance, and SES improved OncologIQ's discriminative performance. The model was translated in a clinically useful tool, enabling up-to-date personalized patient counseling during treatment decision consultations. Current research in our institution is evaluating its clinical impact.

ACKNOWLEDGMENTS

We would like to thank all the medical students that joined the RONCDOC research team. We also would like to thank L. M. Lawant, E. F. van der Poel, and T. Grevelink for enabling the visual formats of OncologIQ. This research partly received funding from the National Health Care Institute (Zorginstituut Nederland).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Arta Hoesseini https://orcid.org/0000-0001-8049-6701
Emilie A. C. Dronkers https://orcid.org/0000-0003-1238-0613

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
2. De Boer MF, McCormick LK, Pruyn JF, Ryckman RM, van den Borne BW. Physical and psychosocial correlates of head and neck cancer: a review of the literature. *Otolaryngol Head Neck Surg*. 1999;120(3):427-436.
3. Barber B, Dergousoff J, Slater L, et al. Depression and survival in patients with head and neck cancer: a systematic review. *JAMA Otolaryngol Head Neck Surg*. 2016;142(3):284-288.
4. Sehlen S, Lenk M, Herschbach P, et al. Depressive symptoms during and after radiotherapy for head and neck cancer. *Head Neck*. 2003;25(12):1004-1118.
5. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ*. 2003;327(7408):195-198.
6. White N, Reid F, Harris A, Harries P, Stone P. A systematic review of predictions of survival in palliative care: How accurate are clinicians and who are the experts? *PLoS One*. 2016;11 (8):e0161407.
7. Hoesseini A, Offerman MPJ, van de Wall-Neecke BJ, Sewnaik A, Wieringa MH, Baatenburg de Jong RJ. Physicians' clinical prediction of survival in head and neck cancer patients in the palliative phase. *BMC Palliative Care*. 2020; 19:176.
8. 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(9):718-724.
9. 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(6):728-736.
10. Datema FR, Ferrier MB, Vergouwe Y, et al. Update and external validation of a head and neck cancer prognostic model. *Head Neck*. 2013;35(9):1232-1237.
11. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.
12. Simmons CPL, McMillan DC, McWilliams K, et al. Prognostic tools in patients with advanced cancer: a systematic review. *J Pain Symptom Manage*. 2017;53(5):962-970.e10.
13. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698.
14. Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. *Oncologist*. 2010;15(9):994-1001.
15. Hoesseini A, Dronkers EAC, Sewnaik A, Hardillo JAU, Baatenburg de Jong RJ, Offerman MPJ. Head and neck cancer patients' preferences for individualized prognostic information: a focus group study. *BMC Cancer*. 2020;20(1):399.
16. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471-1474.
17. Brierley JD, Gospodarowicz MK, Wittekind C. *UICC TNM Classification of Malignant Tumours*. 8th ed. Hoboken, NJ: Wiley-Blackwell; 2017.
18. Amin MB, Edge SB. *AJCC Cancer Staging Manual*. 8th ed. New York: Springer; 2017.
19. UK Parliament. *House of Commons SaTC. Alcohol Guidelines, Eleventh Report of Session 2010–12*. London: Author; 2011.
20. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol*. 1982;5(6):649-655.
21. International standard classification of education: ISCED 2011: UNESCO Institute for Statistics Montreal; 2012.
22. Standaard onderwijsindeling 2016: Centraal Bureau voor de Statistiek (CBS); 2017.
23. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6): 515-524.
24. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1): 128-138.
25. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Switzerland: Springer; 2015.
26. Steyerberg EW. *Clinical Prediction Models*. Switzerland: Springer; 2009.
27. Klompenhouwer AJ, Alblas M, van Rosmalen BV, et al. Development and validation of a model to predict regression of large size hepatocellular adenoma. *Am J Gastroenterol*. 2019;114(8): 1292-1298.
28. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2286.
29. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76-86.
30. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148-158.
32. Nauta IH, Rietbergen MM, van Bokhoven A, et al. Evaluation of the eighth TNM classification on p16-positive oropharyngeal squamous cell carcinomas in the Netherlands and the importance of additional HPV DNA testing. *Ann Oncol*. 2018;29(5): 1273-1279.
33. Hoesseini A. OncologIQ (prognostic model). http://www.oncologiq.nl/
34. Beynon RA, Lang S, Schimansky S, et al. Tobacco smoking and alcohol drinking at diagnosis of head and neck cancer and all-cause mortality: results from head and neck 5000, a prospective observational cohort of people with head and neck cancer. *Int J Cancer*. 2018;143(5):1114-1127.
35. Aizer AA, Chen MH, McCarthy EP, et al. Marital status and survival in patients with cancer. *J Clin Oncol*. 2013;31(31):3869-3876.
36. 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.
37. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619-1636.
38. Liao Y, McGee DL, Cooper RS, Sutkowski MB. How generalizable are coronary risk prediction models? Comparison of Framingham and two national cohorts. *Am Heart J*. 1999;137(5):837-845.
39. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
40. Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. *BMJ*. 2019;365:l737.
41. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63(1):64-74.
42. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med*. 2000;19(2):141-160.
43. Pignon JP, Bourhis J, Domenge C, Designe L. Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. Meta-analysis of chemotherapy on head and neck cancer. *Lancet*. 2000;355(9208):949-955.
44. Pignon JP, le Maitre A, Maillard E, Bourhis J, Group M-NC. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. *Radiother Oncol*. 2009;92(1):4-14.
45. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. *J Clin Epidemiol*. 2007;60(9):979.
46. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087-1091.
47. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. 2006;59(10):1102-1109.
48. Dronkers EAC, Hoesseini A, de Boer MF, Offerman MPJ. Communication of prognosis in head and neck cancer patients; a descriptive qualitative analysis. *Oral Oncol*. 2018;84:76-81.
49. O'Connor AM. *User Manual-Decisional Conflict Scale*. Vol 1993. Ottawa: Ottawa Hospital Research Institute; 1993.
50. 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: 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. https://doi.org/10.1002/hed.26716