Survival Prediction in Head and Neck Cancer: Impact of Tumor and Patient Specific Characteristics

Frank R. Datema

Survival Prediction in Head and Neck Cancer:
Impact of Tumor and Patient Specific Characteristics

Author : F. R. Datema
Lay-out : F. R. Datema
Printing : Ridderprint Offsetdrukkerij BV, Ridderkerk
Cover : F. R. Datema

Publication of this thesis was financially supported by:
RS Datema, J-ACHT, Atos Medical BV, Stallergenes BV, Nutricia Advanced Medical Nutrition, GlaxoSmithKline, Medoc Huisartsenpraktijk NV, Carl Zeiss BV, Daleco Pharma BV, Olympus Nederland BV

Survival Prediction in Head and Neck Cancer: Impact of Tumor and Patient Specific Characteristics

Voorspellen van Overleving in de Hoofd-Hals Oncologie: Invloed van Tumor en Patiënt Specifieke Kenmerken

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van Rector Magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 22 juni 2012 om 11.30

door
Frank Roelf Datema
geboren te Breda

Promotor: Prof.dr. R.J. Baatenburg de Jong

Co-promotor: dr. H.M. Blom

Leescommissie: Prof.dr. E.B. Wolvius
Prof.dr. J.J.B. van Lanschot
Prof.dr. E.W. Steyerberg

CONTENTS

Chapter 1	General introduction and outline of the thesis	**9**
Chapter 2	Comorbidity as the 8th prognostic variable for overall survival estimation in head and neck cancer	**19**
Chapter 3	Internal and external validation of the updated prognostic head and neck cancer model	**37**
Chapter 4	Severe malnutrition as a potential 9th prognostic variable for overall survival estimation in head and neck cancer	**55**
Chapter 5	Survival visualization	**69**
Chapter 6	Incidence and prediction of major cardiovascular complications in extensive head and neck surgery	**73**
Chapter 7	Random Survival Forests versus Cox Proportional Hazards Regression	**93**
Chapter 8	General Discussion	**113**
Chapter 9	Summary / Samenvatting	**123**

List of abbreviations **133**

Dankwoord **134**

Curriculum Vitae **137**

List of publications **138**

aan mijn ouders

voor Jeske, Julia en Veerle

General introduction and outline of the thesis

Chapter One

1.1 Head and neck carcinoma: epidemiology, etiology, and prognosis

Head and neck cancer accounts for almost 5% of all malignant tumors in the Netherlands. The most up-to-date Dutch Cancer Registry (NCR) database from 2009 reported 2878 new patients with an invasive carcinoma of the lip, oral cavity, pharynx and larynx (general incidence 17:100.000). In this thesis we focus on head and neck *squamous cell* carcinoma (HNSCC).

Head and neck squamous cell carcinomas originate from the mucosal lining of the upper aero-digestive tract. *Tobacco and alcohol* are irritants to this mucosal lining and therefore form major risk factors for the genesis of malignant epithelial tumors. Other reported etiological factors are malnutrition, viral factors (Epstein Barr virus and Human Papilloma virus), genetic predispositions and occupational hazards.

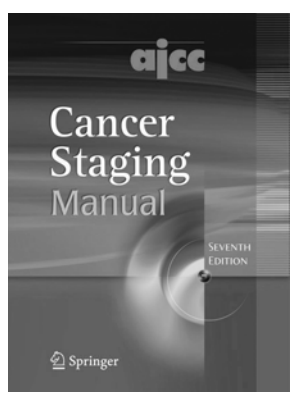
When diagnosed with cancer, the patient's life changes dramatically. Uncertainty about future life expectancy, quality of life and (side) effects of upcoming treatment can form a physical and emotional challenge. Accurate information on what to expect from the course of disease (modified by treatment) and from the likely outcome of disease, can help patients and their loved-ones to cope and prepare and to balance the burden of treatment against the possible gain in life expectancy and quality of life. Furthermore, an individualized treatment can only be the result of an accurate prognosis. Over- and underestimation of survival may result in under- and overtreatment.

In general, clinicians are very capable in providing clear information about the disease and short-term and long-term (side) effects of oncological treatment. When the patient asks for his prognosis, it can become difficult. The easy answer to the question is that the 5-year survival rate of HNSCC patients (the percentage of patients that survive at least five years after cancer is detected) is approximately 50%. This percentage however can greatly vary depending on the impact and interaction of prognostic factors. Examples are: the age of the patient, the location of the primary tumor, the size of the tumor, presence of loco regional and/or distant metastasis, and the patient's general health status.
For example: *the otherwise healthy 42-year-old T1N0M0-glottic larynx carcinoma patient probably has a better prognosis than the insulin dependent 80-year-old T4N2M0-oropharynx carcinoma patient*.

Because the *patient demands an accurate prognosis estimate*, the *clinician is challenged* to recognize relevant prognostic factors in the patient and to determine their impact and interaction on prognosis. Unfortunately we lack valid instruments that integrate prognostic factors and their corresponding p-values and hazard ratios into an individualized prognosis. So the clinician is left with an expanding list of prognostic factors and is not adequately equipped to apply this prognostic knowledge in everyday practice. Fortunately there are *statistical survival analyses*. These analyses, allow the identification of prognostic factors from historical patients with similar patient and tumor specific characteristics, and can test their strength, significance and independence towards overall survival. When several criteria are met prognostic factors can be combined into a multivariable survival model that predicts the prognosis of an individual patient. In this thesis, the development of several *prognostic HNSCC survival models* are presented and statistical output is translated into clinical relevance and applicability. To equip the clinician, we developed free, on-online software (also known as OncologIQ) that calculates an individual overall survival estimate of the newly diagnosed HNSCC patient based on eight prognostic factors. We believe that this model can complement medical craftsmanship when communicating prognosis to the patient or when the clinician needs to choose between suitable treatment options.

1.2 The TNM-classification as a cornerstone in outcome prediction

Up to the 1950s many incompatible cancer staging systems were in use. With the rise of effective therapies against cancer, the need for an universal staging system emerged. This staging system needed to be capable to assess an accurate prognosis, because the selection of therapy usually depends on estimated outcome. As a response to this need, Pierre Denoix (former chairman of the Union for International Cancer Control (UICC) and former chairman at the Institute Gustave-Roussy in France), developed the TNM-classification.

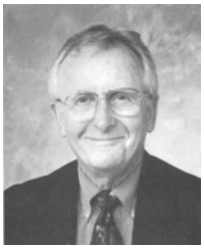


The TNM-classification incorporates the *size* and extent of the primary tumor (T-classification), its *regional lymphatic involvement* (N-classification) and the presence of *distant metastasis* (M-classification) to stage the progression of cancer.

In 1987, the UICC and American Joint Committee on Cancer (AJCC) joined hands and unified their staging systems into a single cancer staging system (the 4th TNM edition). Since then, The TNM-classification has been revised and updated several times. The most up to date published TNM-classification of Malignant Tumors is the 7th edition (2010) and forms *the globally accepted method of choice* to describe the anatomical extent of cancer.

Because the TNM-classification is a globally accepted uniform staging system, it became a very suited instrument to facilitate treatment planning, uniform evaluation of treatment results, patient selection for clinical trials, and to communicate prognosis to patients. It can be said that the TNM-classification has been the *cornerstone of cancer outcome prediction* for many years. In several chapters of this thesis the importance and impact of the TNM-classification on overall survival is acknowledged and described.

With its revisions and updates, the TNM-classification evolved into an excellent *descriptive instrument* for tumor-specific characteristics. As a predictive instrument, the TNM-classification showed *limitations*, especially in prognosis prediction of the individual cancer patient. These limitations were well described in 1991 by Byron J. Bailey, Chairman of the Committee to study the TNM-classification of the Laryngeal Cancer Association:



''Physicians are focused on optimal treatment while patients are interested in their prognosis, and the TNM is not designed to provide answers to either set of questions. At the present time, the TNM system is neither a roadmap for patient management nor is it a crystal ball with the answers sought by patients.''

This comment is a good illustration of the need for a prognostic instrument that goes further than the anatomical extent of cancer alone. Combining the TNM-classification with other relevant oncological and non-oncological prognostic factors into a model, allows the estimation of a more accurate individual prognosis prediction of the newly diagnosed HNSCC patient.

1.3 AJCC recommendations for prognostic factors and for an enhanced prognostic system (model)

All variables are potentially prognostic, but few variables actually have an independent prognostic value. In 1993 the AJCC communicated their vision on prognostic factors and formed criteria for an enhanced prognostic system (model). Most of these criteria form the basis for the research that is presented in this thesis and are therefore worth mentioning.

AJCC criteria for prognostic factors

1. The prognostic factor must be *significant*, meaning that it rarely occurs by chance.
2. The prognostic factor must be *independent*, meaning that it retains its prognostic value when new prognostic factors are added
3. The prognostic factor must be *clinically important*, meaning that it can change patient management and thereby change outcome.

Note: In the prognostic research presented in this thesis, the significance of a prognostic factor is tested with Kaplan Meier Curves and the log-rank test. Following these univariate analyses, independence of each prognostic factor was tested with a multivariate Cox regression analysis. Prognostic factors discussed in this thesis are clinically important and some results from this thesis have strengthened certain management changes in the head and neck oncology department of the Erasmus Medical Centre. For example, in 2010 an internal physician joined the oncological staff to identify and optimize comorbidity before, during and after treatment to reduce complication rates and to optimize the overall survival probability of our HNSCC patients.

AJCC criteria for a prognostic system (prognostic model)

1. The system is easy for physicians to use.
2. Provides predictions for all types of cancer.
3. Provides the most accurate relapse and survival predictions at diagnosis and for every year lived for each patient.
4. Provides group survival curves, where the grouping can be by any variable, including outcome and therapy.
5. Accommodates missing data and censored patients and is tolerant of noisy and biased data.
6. Makes no a priori assumptions regarding the type of data, the distribution of the variables, or the relationships among the variables.
7. Tests putative prognostic factors for significance, independence, and clinical relevance.
8. Accommodates treatment information in the evaluation of prognostic factors.
9. Accommodates new prognostic factors without changing the model.
10. Accommodates emerging diagnostic techniques: not only molecular tests, but also new imaging modalities.
11. Provides information regarding the importance of each predictive variable.
12. Is automatic, that is, the model's output does not depend on the operator.

Note: Prognostic models that are presented in this thesis comply with criterion 1, 3-7, 9, 11 and 12. Criterion 2 cannot be upheld since the models are only suited for primary HNSCC patients. Criterion 8 cannot be upheld since treatment is not considered a prognostic factor. Criterion 10 cannot be upheld since the models do not incorporate diagnostic techniques. The TNM-classification however is partially derived from diagnostic (imaging) results.

1.4 Sources of data and prognostic model covariables

In this thesis, three versions of our HNSCC prognostic model are presented. The first version is an extended version of the original model of Baatenburg de Jong et. al. (2001) to which comorbidity is added as an 8th prognostic factor. The second version is an updated and externally validated model which has more recent follow-up data reaching until January 2010. The third model is created with an alternative statistical approach called Random Survival Forests. The end-point of all models is overall survival (death of all causes). The models are fitted on the historical data of 1371 primary HNSCC patients, diagnosed and treated at the Leiden University Medical Centre (LUMC) between 1981 and 1999. Follow-up is complete for 97.5%.

Covariables	Subcategories
1. Age at diagnosis	*continuous*
2. Sex	*male, female*
3. T-classification	*T1, T2, T3, T4*
4. N-classification	*N0, N1, N2, N3*
5. M-classification	*M0, M1*
6. Tumor location	*lip, oral cavity, oropharynx, nasopharynx, hypopharynx, glottic larynx and supraglottic larynx*
7. Prior tumors	*all preceding malignancies except basal cell or squamous cell carcinoma of the skin*
8. Comorbidity	*Adult Comorbidity Evaluation (ACE27) grade 0, grade 1, grade 2, grade 3*
9. Malnutrition	*Weight loss < 5%, weight loss 5-10%, weight loss > 10%*

Note: Malnutrition is considered a potential 9th predictor for our model. Since data on weight loss are only available for a subset of the baseline population, malnutrition cannot be added to the model.

1.5 Outline of this thesis

In 2001 Baatenburg de Jong et. al. presented a 7-variable-prognostic Cox regression model (TNM-classification, tumor location, age at diagnosis, prior tumors and sex). In *chapter two* of this thesis, his prognostic model is enhanced with comorbidity as an 8th predictor. The significance and independent impact of comorbidity on overall survival and short-term mortality is investigated and discussed. The new model is internally validated.

In *chapter three* the 8-variable-prognostic model is updated with very recent follow-up and survival data reaching until January 2010. The impact of this update on model performance is discussed and more importantly, an external validation with a secondary dataset from the United States of America is performed. External validation is the most stringent test for a model and is essential before implementing prediction models into clinical practice.

In *chapter four* we discuss the impact of severe malnutrition, defined as a weight loss of more than 10% in the six months preceding cancer diagnosis, on overall survival and short term mortality. Severe malnutrition is a potential 9th predictor for a prognostic head and neck cancer model.

In *chapter five* we explain how our most recent model can be accessed on-line and used for free during daily practice. With a very user-friendly interface it is possible to create an individualized 5-year survival chart for your newly diagnosed HNSCC patient, without extensive statistical knowledge. We believe that this on-line visual interpretation of our model can complement medical craftsmanship in communicating prognosis to the patient.

A substantial amount of tobacco and alcohol induced comorbidity was found in our study population. Some of these illnesses are known risk-factors for major cardiovascular complication occurrence during and after extensive head and neck surgery. Since major cardiovascular complications form an elevated mortality risk, they are worth preventing. For this purpose, a new risk stratification tool (the modified Lee Cardiac Risk Index) is introduced for head and neck oncology in *chapter six*.

In *chapter seven* an alternative modeling approach, called Random Survival Forests (RSF) is compared to the most generally used modeling technique, Cox Regression. RSF is known to deliver accurate models and is more easily automated than Cox models (criterion 12 in paragraph 1.3). RSF model performance is compared to Cox model performance. Similarities, differences, limitations and advantages are investigated and discussed.

Comorbidity as the 8th prognostic variable for overall survival estimation in head and neck cancer

Datema FR, Ferrier MB, vd Schroeff MP, Baatenburg de Jong RJ. The impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. Head Neck 2010; 32: 728-36.

Chapter Two

Abstract

Background In 2001, we presented a Cox regression model that is able to predict survival of the newly diagnosed patient with head and neck squamous cell carcinoma (HNSCC). This model is based on the TNM classification and other important clinical variables such as age at diagnosis, sex, primary tumor site, and prior malignancies. We aim to improve this model by including comorbidity as an extra prognostic variable. Accurate prediction of the prognosis of the newly diagnosed patient with head and neck cancer can assist the physician in patient counselling, clinical decision-making, and quality maintenance.

Methods All patients with HNSCC of the oral cavity, pharynx, and larynx diagnosed in the Leiden University Medical Centre between 1981 and 1998 were included. From these 1371 patients, data on primary tumor site, age at diagnosis, sex, TNM classification, and prior malignancies were already available. Comorbidity data were collected retrospectively according to the ACE27 manual. The univariate prognostic value of each variable on overall survival was studied with Kaplan–Meier curves and the log-rank test. The Cox regression model was used to investigate the impact of these variables on overall survival simultaneously. Furthermore, univariate analyses were performed to investigate the impact of comorbidity severity on short-term mortality and to investigate the impact of organ-specific-comorbidity on short-term mortality.

Results Comorbidity was present in 36.4% of our patients. Mild decompensation was seen in 17.4%, moderate decompensation in 13.5%, and severe decompensation in 5.5%. Most frequently observed ailments were cardiovascular, respiratory, and gastrointestinal. In univariate analyses, all prognostic variables, including comorbidity, contributed significantly to overall survival. Their contribution (except sex) remained significant in the multivariate Cox model. Internal validation of this model showed a concordance index of 0.73, indicating a good predictive value. Short-term mortality was seen in 5.7% of our patients. Cardiovascular comorbidity, respiratory comorbidity, gastrointestinal comorbidity, and diabetes showed a significant relationship with short-term mortality.

Conclusions Comorbidity impacts overall survival of the newly diagnosed patient with HNSCC. There is a clear distinction between the impact of the four ACE27 severity grades. The impact of an ACE27 grade 3 is comparable to the impact of a T4 tumor or an N2 neck. Comorbidity impacts short-term mortality as well. Especially cardiovascular comorbidity, respiratory comorbidity, gastrointestinal comorbidity, and diabetes show a strong relationship.

Introduction

The TNM-classification has been the cornerstone of cancer staging and outcome prediction in patients with head and neck squamous cell carcinoma (HNSCC) for many years. Since its introduction, the TNM-classification has been revised and updated several times and has evolved into an excellent descriptive instrument for tumor-specific characteristics.

The anatomic extent of cancer alone, however, is not the most accurate way to predict the outcome of an individual patient with cancer at initial presentation. A more accurate prognosis can be achieved by combining the TNM-classification with patient-specific variables that impact survival as well [1]. In 2001, this principle was illustrated by the presentation of a validated, Cox-regression model that included the prognostic variables: age, sex, primary tumor site, TNM-classification, and prior malignancies [2]. The suggestion to include other prognostic variables to improve the predictive value of the model was made.

A factor that has high potential as a prognostic variable is comorbidity. Comorbidity is described as *''the presence of one or more medical ailments, next to the primary tumor but not caused by the primary tumor.''* Risk factors for the development of HNSCC, such as smoking and alcohol abuse, contribute to other diseases as well (eg, cardiovascular, pulmonary, or hepatic diseases). Therefore, comorbidity is to be expected in the patient with HNSCC. Previously reported incidences of comorbidity in patients with HNSCC ranged from 30% to 55% [3–5]. In some cases, the patient's comorbidity conditions form an even greater risk on mortality than the primary tumor [3–6].

There are several validated instruments designed to code and quantify comorbidity in patients: the cumulative illness rating scale (CIRS), the Kaplan Feinstein Comorbidity Index (KFI), the Charlson Comorbidity Index (CCI) and the Index of Coexistent Disease (ICED) [7-10]. In a comparative study of these 4 instruments, the KFI was the most successful in stratifying patients with head and neck cancer [11].

In 1999, the adult comorbidity evaluation (ACE27), also known as the modified KFI, was introduced. The ACE27 has been revised several times and has proven to be a valid tool to code and quantify the presence of comorbidity in HNSCC patients. [4] In this retrospective study, the ACE27 was used to code and quantify comorbidity in 1371 patients with HNSCC. The first aim of this study was to examine the quality of our retrospective comorbidity data by means of calculating the intra-observer and inter-observer variability. The second aim of this study was to gain insights into the prevalence and severity of comorbidity in this (Dutch) head and neck cancer patient dataset. The third aim of this study was to investigate the univariate impact of comorbidity on overall survival of patients with head and neck cancer. The univariate impact of previously identified variables (primary tumor site, age at diagnosis, sex, prior malignancies, and TNM classification) were investigated as well.

The fourth aim of this study was to investigate the multivariate impact on overall survival of these variables simultaneously. The fifth aim of this study was to investigate and describe the univariate impact of comorbidity and organ system–specific comorbidity on overall survival and on short-term mortality (mortality within 6 months after diagnosing the primary HNSCC).

Materials and Methods

For this study, the data of 1662 patients diagnosed with primary HNSCC at the Leiden University Medical Centre (LUMC) between 1981 and 1998 was available.

Patients with oesophageal cancer and subglottic cancer (n = 218) were excluded because the prognosis of these patients is poor and the number of incomplete TNM classifications in this group was relatively large. Patients with carcinoma in situ (n = 51; including 3 patients with oesophageal cancer) were excluded because the prognosis of these patients is exceptionally good. Another 25 patients were excluded because it was not possible to designate a proper TNM. The final study population consisted of 1371 patients with histologically proven squamous cell carcinoma of the lip, oral cavity, oropharynx, nasopharynx, hypopharynx, glottic-larynx, and supraglottic-larynx (Table 1).

Prior malignancies were defined as all preceding malignant tumors except for basal cell and squamous cell carcinoma of the skin. Based on the therapeutic nil hypotheses, the type of treatment were not considered a prognostic factor for our model in this study (see discussion) [12–15].

Study Design

Patient data on age, sex, primary tumor site, prior malignancies, and TNM classification were available from the hospital-based cancer registry system (ONCDOC). The ONCDOC was established in 1969 and contains patient, treatment, and follow-up data for each patient with cancer diagnosed in the LUMC. Trained oncologic data managers store these data and safeguard an adequate follow-up by contacting the general practitioner and/or Registry of Births, Deaths, and Marriages when patients are lost. As a result, the percentage lost to follow-up for this study population was only 2.5% and the mean follow-up time was 12.3 years. Another task of ONCDOC is to monitor the registered TNM classification of each tumor and to raise discussion when discrepancies exist. Disease was always staged or restaged according to the Union Internationale Contre le Cancer (UICC) manual, which was up-to-date then. Data on comorbidity were collected retrospectively from the patient's medical chart. The presence and severity of comorbidity in the timeline before diagnosing the primary HNSCC tumor was coded according to the ACE27 manual with 1 exception. [16] Prior malignancies were not scored as a comorbidity condition because this factor was already a variable in our existing predictive model. The impact of prior malignancies on overall survival would be unjust when scored twice.

The patient's overall comorbidity severity score was defined according to the highest ranked single ailment (coded as grade 1: mild decompensation; grade 2: moderate decompensation; or grade 3: severe decompensation), except when 2 or more grade 2 ailments occurred in different organ systems. In this case, the overall comorbidity severity score was designated as a grade 3. For example, a patient with chronic heart failure more than 6 months previously and portal hypertension without compensation (two grade 2 ailments) would have an overall comorbidity severity score of grade 3.

To investigate the integrity and quality of our comorbidity data, the inter-observer variability and intra-observer variability were calculated. For this purpose, a second researcher coded comorbidity on 60 randomly selected patients, and the initial researcher coded 20 randomly selected patients twice.

Table 1. Demographic and tumor data of baseline study population (n = 1371).

Characteristics	No. (%)	Characteristics	No. (%)
(Sub)sites		Age categories	
Lip	123 (9.0%)	< 50y	182 (13.3%)
Oral cavity	280 (20.4%)	50-59y	367 (26.8%)
Oropharynx	152 (11.1%)	60-69y	427 (31.1%)
Nasopharynx	41 (3.0%)	≥ 70y	395 (28.8%)
Hypopharynx	137 (10.0%)	Sex	
Larynx-glottic	442 (32.2%)	Male	1088 (79.4%)
Larynx-supraglottic	196 (14.3%)	Female	283 (20.6%)
T-classification		Year of diagnosis	
T1	516 (37.6%)	1981-85	296 (21.6%)
T2	369 (26.9%)	1986-90	359 (26.2%)
T3	208 (15.2%)	1991-95	416 (30.3%)
T4	278 (20.3%)	1996-98	300 (21.9%)
N-classification		Prior malignancies	
N0	964 (70.3%)	Yes	133 (9.7%)
N1	145 (10.6%)	No	1238 (90.3%)
N2	180 (13.1%)	Treatment	
N3	82 (6.0%)	Radiotherapy	798 (58.2%)
M-classification		Chemotherapy	8 (0.6%)
M0	1354 (98.8%)	Surgery	205 (15.0%)
M1	17 (1.2%)	Surgery + PORT	250 (18.2%)
		Otherwise *	110 (8.0%)

**Otherwise: cryogenic or laser therapy; PORT: post operative radio therapy.*

Statistical analyses

To investigate the quality of our retrospective ACE27 comorbidity data, the inter-observer and intra-observer variability was determined by calculating kappa. A kappa value above 0.80 was interpreted as an almost perfect level of agreement. [17]

To investigate the univariate impact of the variables: primary tumor site, age at diagnosis, sex, prior malignancies, comorbidity, and TNM classification on overall survival, Kaplan–Meier curves and the log-rank test were used. The endpoint for overall survival was death (of all causes).

The Cox regression model was used to investigate the impact of all mentioned variables on overall survival simultaneously. The Cox regression model can be written as:

$$H(t,\mathbf{X}) = h_0(t) * exp^{(\beta 1X1 + \beta 2X2 + .. + \beta nXn)}$$

The reference hazard $h_0(t)$ is also known as the baseline hazard and is calculated in the Cox model. The baseline hazard reflects the risk of dying for the individual patient at certain point in time when all variables are equal to zero. Therefore, H(t,**X**) is the resultant hazard or cumulative hazard and is based on the impact of each model variable of the respected individual ($\beta_n X_n$) multiplied by the baseline hazard. The X is the covariate vector and β is the regression coefficient. In this study, the regression coefficients of each categorical variable reflect the additional risk for death (of all causes) adjusted for their reference category to which the risk is set to 1.0. The reference category is the category with the best prognosis. Since age is not a categorical but a continuous variable, the mean population age (62.6 years) was set as the reference.

The predictive accuracy of the Cox regression model was determined by calculating the concordance index (C-index). A C-index of 0.5 indicates that outcomes are completely random, whereas a C-index of 1.0 indicates that the model is a perfect predictor. A bootstrap procedure with 500 samples was used for internal validation by computation of an unbiased estimate of the C-index [18].

In addition to comorbidity as a single entity, the same statistical approaches were followed in an attempt to identify the impact of organ-system-specific comorbidity on overall survival. To investigate the impact of the variables: primary tumor site, age at diagnosis, sex, prior malignancies, comorbidity, and TNM classification on short-term-mortality, backward and forward binary logistic regression analyses were performed. The endpoint for short-term mortality was death (of all causes) within 6 months after diagnosing the primary tumor.

The impact of organ system–specific comorbidity (categorized as described in the ACE27 manual) on short-term mortality was investigated with backward and forward binary logistic regression analyses. In all analyses, only *p* values below .05 were considered statistically significant. Calculations were performed in SPSS for Windows (version 16.0) and R (version 2.7.0) (SPSS Inc, Chicago, IL).

Results

Inter-observer and intra-observer variability

The kappa value for inter-observer variability of the ACE27 data was 0.89 ($p < .01$), and the kappa value for intra-observer variability was 0.90 ($p < .01$). These findings indicate an almost perfect level of agreement and conclude that the ACE27 is an instrument that leaves little space for inconsistent data.

Prevalence, type, and severity of comorbidity conditions

From the 1371 reviewed patient charts, 87 charts (6.5%) did not contain enough information to designate a proper ACE27 score. From the 1282 charts with information, 500 patients had comorbidity (36.4%). Some patients had multiple comorbid conditions, resulting in a total of 835 scored comorbidity ailments. Cardiovascular disease, respiratory disease, and gastrointestinal disease were observed most frequently. No patient had a body mass index (BMI) above 38 (a grade 2 comorbidity), and no data regarding the presence of acquired immune deficiency syndrome (AIDS) or human immuno-sufficiency virus (HIV) positivity was encountered. Further results regarding prevalence and severity of comorbidity are shown in Table 2.

Table 2. Presence and severity of comorbidity conditions in the study population (n = 1371).

	Grade 1: mild	Grade 2: moderate	Grade 3: severe
Overall ACE27 score	239 (17.4%)	185 (13.5%)	76 (5.5%)
Specific ACE27 categories			
Cardiovascular disease	276 (20.1%)	155 (11.3%)	8 (0.6%)
Respiratory disease	38 (2.8%)	33 (2.4%)	13 (0.9%)
Gastro-intestinal disease	45 (3.3%)	53 (3.9%)	0 (0.0%)
Renal disease	9 (0.7%)	2 (0.1%)	0 (0.0%)
Diabetes	36 (2.6%)	16 (1.2%)	2 (0.1%)
Neurological disease	38 (2.8%)	17 (1.3%)	1 (0.1%)
Pschycologic disease	7 (0.5%)	1 (0.1%)	0 (0.0%)
Rheumatologic disease	13 (0.9%)	2 (0.1%)	0 (0.0%)
Substance abuse	44 (3.2%)	17 (1.2%)	9 (0.7%)

**n = 89 (6.5%), missing values due to absence of medical data.*

Impact of comorbidity on overall survival

In univariate analyses, all variables, except sex (p = .19) showed a significant relationship with overall survival. There was a clear distinction between the 4 ACE27 comorbidity severity grades regarding their impact on overall survival. For example, the 2-year survival probability of a patient without comorbidity was 75.0%, against 67.0% for a patient with grade 1 comorbidity. A 63.0% survival probability was seen in a patient with grade 2 comorbidity and 46.0% in a patient with grade 3 comorbidity. The univariate impact of grade 3 comorbidity is comparable to the impact of a tumor classified as T4 or a neck classified as N2. In 1-year and 5-year survival probabilities, comparable results were found (Table 3).

The next step was to examine how these variables performed simultaneously in a multivariate Cox-regression analysis. All variables, except sex (p = .43) remained significant in the multivariate Cox model. The regression coefficients (exp[β]) and 95% confidence intervals of each variable are displayed in Table 4. In the multivariate analysis, the impact of grade 3 comorbidity on overall survival remained comparable to a tumor classified as T4 or a neck classified as N2.

Table 3. Univariate analysis for all variables – impact on overall survival

Predictor	Subcategory	1-year survival probability	2-year survival probability	5-year survival probability	P-value log rank
Site	Lip	0.97	0.93	0.75	< 0.01
	Oral cavity	0.75	0.54	0.38	
	Oropharynx	0.78	0.58	0.39	
	Nasopharynx	0.81	0.68	0.56	
	Hypopharynx	0.69	0.45	0.33	
	Larynx glottic	0.92	0.85	0.70	
	Larynx supraglottic	0.86	0.73	0.50	
Sex	Male	0.85	0.72	0.54	0.19
	Female	0.80	0.65	0.49	
Age	< 50 years	0.92	0.81	0.67	< 0.01
	50-59 years	0.86	0.73	0.60	
	60-69 years	0.86	0.71	0.53	
	≥ 70 years	0.77	0.62	0.41	
T-stage	T1	0.94	0.89	0.74	< 0.01
	T2	0.89	0.74	0.54	
	T3	0.72	0.53	0.37	
	T4	0.67	0.43	0.27	
N-stage	N0	0.91	0.81	0.65	< 0.01
	N1	0.78	0.51	0.32	
	N2	0.67	0.43	0.28	
	N3	0.49	0.30	0.12	
M-stage	M0	0.85	0.71	0.54	< 0.01
	M1	0.15	0.08	0.00	
Prior mal*	Yes	0.85	0.62	0.38	< 0.01
	No	0.84	0.71	0.55	
ACE –27	Grade 0	0.87	0.75	0.58	< 0.01
	Grade 1	0.81	0.67	0.53	
	Grade 2	0.79	0.63	0.44	
	Grade 3	0.64	0.46	0.25	

Abbreviations; Prior mal: prior malignancies in the timeline before diagnosing the primary head and neck tumor

Table 4. Multivariate Cox-regression analysis – impact on overall survival

Predictor	Subcategory	Regression Coefficient (exp β)	P-value	95% Confidence interval
Site	Lip	1.000	<0.01	
	Oral cavity	1.580		1.18 – 2.11
	Oropharynx	1.597		1.16 – 2.21
	Nasopharynx	1.191		0.73 – 1.94
	Hypopharynx	1.859		1.33 – 2.60
	Larynx glottic	1.021		0.78 – 1.33
	Larynx supraglottic	1.304		0.96 – 1.77
Sex	Male	1.000	0.43	
	Female	0.939		0.80 – 1.10
Age	Mean age [62.6 years]	1.000	<0.01	
	Above*	1.040		1.03 – 1.04
T-stage	T1	1.000	<0.01	
	T2	1.325		1.12 – 1.57
	T3	1.574		1.28 – 1.94
	T4	2.002		1.64 – 2.45
N-stage	N0	1.000	<0.01	
	N1	1.452		1.18 – 1.79
	N2	1.899		1.55 – 2.32
	N3	2.440		1.88 – 3.16
M-stage	M0	1.000	<0.01	
	M1	6.398		3.72 – 11.00
Prior Malignancies	No	1.000	<0.01	
	Yes	1.723		1.41 – 2.11
ACE –27	Grade 0	1.000	<0.01	
	Grade 1	1.043		0.88 – 1.24
	Grade 2	1.379		1.15 – 1.65
	Grade 3	2.229		1.73 – 2.87

** For each year above mean age, 0.039 needs to be added to the regression coefficient and vice versa when younger.*

Impact of organ system-specific comorbidity on overall survival
In univariate analysis, comorbidity from the cardiovascular system, respiratory system, gastrointestinal system, endocrine (diabetes) system, neurological system, and substance abuse system showed a significant relationship with overall survival. Comorbidity from the rheumatologic system showed no significant impact (p = .16). Immunological disease (AIDS/HIV), obesity (BMI >38), psychiatric, and renal disease were excluded from this analysis because each system contained less than 11 patients and significant results were, therefore, not expected.

In the multivariate regression analysis, comorbidity from the cardiovascular system, respiratory system, and substance abuse remained significant variables with an impact on overall survival, next to the variables: age at diagnosis, primary tumor site, TNM classification, and prior tumors. Grade 3 comorbidity within these specific organ systems had a regression coefficient of approximately 1.8 or higher (Table 5).

Figure 1. Expected survival for a 52-year-old male with a T3N2M0 glottic larynx carcinoma, without prior malignancies and with comorbidity grade 2.
The expected two- and five-year overall survival rates are 73.6% and 55.3%, respectively.

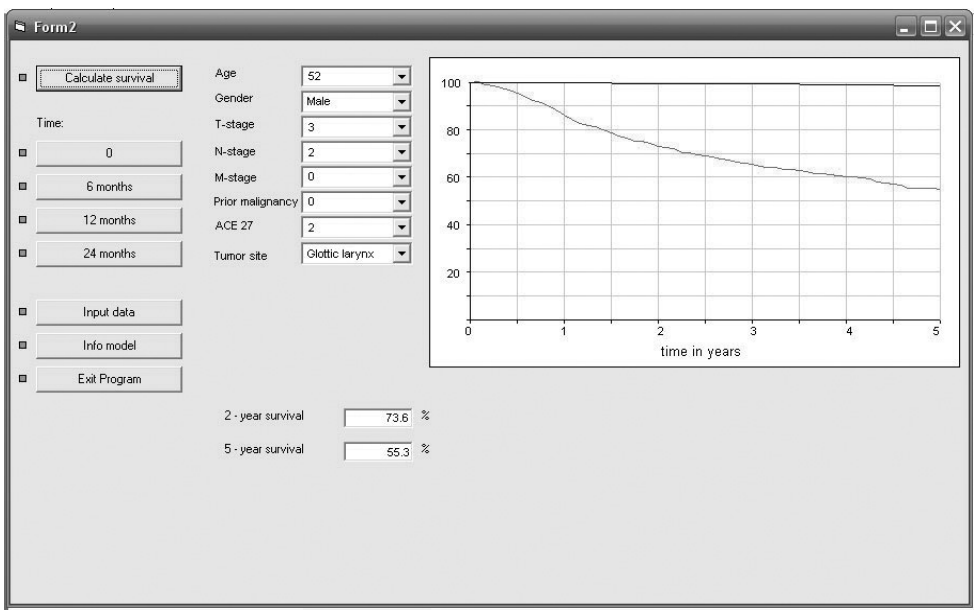


Table 5. Multivariate Cox-regression analysis: specific ACE subcategories – impact on overall survival

Predictor	Subcategory	Relative Risk (Exp(ß))	P-value
Site	Lip	1.000	<0.01
	Oral cavity	1.549	
	Oropharynx	1.589	
	Nasopharynx	1.249	
	Hypopharynx	1.929	
	Larynx glottic	1.038	
	Larynx supraglottic	1.323	
Age	Mean (62.2 years)	1.000	<0.01
	Above*	1.042	
T-stage	T1	1.000	<0.01
	T2	1.334	
	T3	1.533	
	T4	1.902	
N-stage	N0	1.000	<0.01
	N1	1.482	
	N2	1.797	
	N3	2.524	
M-stage	M0	1.000	<0.01
	M1	6.120	
Prior Malignancies	No	1.000	<0.01
	Yes	1.652	
ACE -27 cardiovascular	Grade 0	1.000	<0.01
	Grade 1	0.922	
	Grade 2	1.313	
	Grade 3	1.839	
ACE27 respiratory	Grade 0	1.000	<0.01
	Grade 1	1.354	
	Grade 2	1.887	
	Grade 3	1.773	
ACE27 substance abuse	Grade 0	1.000	<0.01
	Grade 1	1.966	
	Grade 2	3.952	
	Grade 3	3.232	

** For each year above mean age, 0.041 needs to be added to the relative risk and vice versa when younger*

Impact of comorbidity and organ system-specific comorbidity on short-term mortality. From 1371 reviewed patients, 78 patients (5.7%) died within 6 months after diagnosing the primary head and neck tumor. In univariate analysis, the prognostic variables: age at diagnosis, primary tumor site, TNM classification, and comorbidity had significant impact on short-term mortality. Sex (p = .09) and prior tumors (p = .89) showed no significant impact. Having established the univariate relationship between comorbidity and short-term mortality, the next step was to see if it was possible to investigate the univariate impact of organ-system specific comorbidity on short-term mortality. The organ systems, as described in the ACE27 manual, were chosen as categorical variables. Univariate analysis showed a (borderline) significant contribution from cardiovascular disease (p = .05), respiratory disease (p < .01), gastrointestinal disease (p = .04), and diabetes (p = .03). The most frequently observed comorbid conditions within these 4 organ systems were: myocardial infarct grade 2 (11.5%), diastolic blood pressure grade 1 (7.7%), chronic obstructive pulmonary disease (COPD) grade 2 (9.0%), ulcers grade 2 (6.4%), and diabetes grade 1 (9.0%). Multivariate analyses were not performed because the subgroups for these analyses contained too little data for a relatively short timeline.

Prognostic model update and internal validation

Based on the regression coefficients and the baseline hazard, which is calculated in the Cox model, the expected survival for a new patient with HNSCC can be calculated by use of the formula shown in the statistical section. Based on this formula, we built a user-friendly interface for survival calculation, and visual graphic output is available through specially designed, dedicated software (Figure 1). The C-index, based on 500 bootstrap samples, was 0.73. This indicates a good predictive value of our Cox regression model. Furthermore, this C-index means an improvement in accuracy of prognosis prediction over our previous model.

Discussion

This study shows that comorbidity impacts overall survival and short-term mortality of the patient with newly diagnosed HNSCC. There is a clear distinction between the impact of the four ACE27 comorbidity severity grades. Next to comorbidity, the variables prior tumors, primary tumor site, TNM classification, and age at diagnosis remain significant predictive covariates in the Cox regression model.

The limitations of the TNM classification as a predictive instrument were well described by Byron J Bailey, Chairman of the Committee to study the TNM classification of the Laryngeal Cancer Association [19]: ‘‘Physicians are focused on optimal treatment while patients are interested in their prognosis, and the TNM is not designed to provide answers to either set of questions. At the present time, the TNM system is neither a roadmap for patient management nor is it a crystal ball with the answers sought by patients.’’

To include treatment modalities as a prognostic variable in our model was not a part of this study. First, in any index for prognostic stratification, choices between treatment options that are not under control of the investigator will influence outcome. This systematic error cannot be eliminated and is the reason this study is based on the therapeutic nil hypotheses. Second, the choices between treatment options are partially determined by variables that are already included in our model (e.g., TNM classification, age at diagnosis, and comorbidity). The inclusion of treatment as an extra prognostic variable would, therefore, introduce a significant bias.

In this study, comorbidity was scored according to the ACE27 manual. The kappa values found in this study indicate that the ACE27 is an easy to use instrument that leaves little space for inconsistent data. Our kappa values are in agreement with kappa values found in the study of Paleri et. al. (0.81–1.00) [20]. From the 1371reviewed patient charts, 87 charts (6.5%) did not contain enough information to designate a proper ACE27 score. A comparable percentage of patients without a comorbidity score were found in a study of Singh et. al. using the KFI [21]. Comorbidity was present in 36.4% of our patients. This percentage is consistent with findings in other studies where comorbidity among patients with HNSCC ranged from 30% to 55% [3-5].

Previous studies already described a few limitations of the ACE27. For example, whether a patient is HIV positive or not contributes significantly to survival. This information, however, will not always be found in the patient's medical chart due to privacy reasons, as was the case in this study. During data collection, we found other ailments that could possibly affect survival but cannot be scored in the ACE27. Two examples are valvular heart disease and chronic anaemia. Valvular heart disease can only be scored as ACE27 comorbidity, when it results in congestive heart failure or arrhythmia. In our medical charts, valvular heart disease was seen, but in some cases without reporting the consequence of this disease. For these patients, this could have resulted in a lower ACE27 score. Chronic anaemia is a condition that can affect treatment response and can aid or introduce post-treatment morbidity, 2 conditions with a possible influence on overall survival. Besides missing comorbid conditions in the ACE27, we were unable to code comorbidity in the obese category. A BMI above 38 is not frequently seen in the general Dutch population because Dutch people are relatively tall. Especially in an oncologic population in which weight is frequently reduced, a BMI above 38 is highly unlikely. Cardiovascular comorbidity, respiratory comorbidity, and substance abuse impact overall survival. The clear distinction between the impact of the four ACE27 severity grades within these organ systems was less discriminative than was encountered in overall comorbidity. This was a direct result of the distribution of patients with comorbidity in these specific organ systems. The difference in impact between absent comorbidity and present comorbidity on overall survival, however, remained, especially in grade 2 and grade 3 ailments. Cardiovascular comorbidity, respiratory comorbidity, diabetes, and gastrointestinal comorbidity showed a significant impact on short-term mortality.

Our findings form a stronger foundation for the implementation of identification and optimization strategies toward comorbid conditions. In the Erasmus Medical Centre (Rotterdam) a recent proposal was presented to perform pulmonary function testing in all patients with HNSCC that are planned for major surgery. This allows the pulmonary physician to optimize the pulmonary status of the patient preoperatively. It is our theory that this change in preoperative management could lead to a better survival prognosis (and fewer complications). Further optimization strategies towards comorbidity are currently being explored, discussed, and implemented.

Conclusion

In order to establish optimal treatment for head and neck cancer, clinicians need to be aware of all the relevant factors that determine course and prognosis. It is possible to predict survival in a new patient with HNSCC based on historic results from a dataset analyzed with the Cox regression model. Our predictive model contained the prognostic variables age at diagnosis, sex, TNM classification, prior malignancies, and primary tumor site and is now extended with the prognostic variable comorbidity. This update allows the model to predict a more accurate and individualized survival estimate.

The impact of comorbidity on the overall survival of a patient with HNSCC is comparable to the impact of tumor size (T) and nodular metastasis (N). The results of the Cox regression may be used in patient counselling, clinical decision-making, and quality maintenance.

Comorbidity impacts short-term mortality as well. A differentiation into organ system-specific comorbidity showed a significant impact from cardiovascular disease, respiratory disease, gastrointestinal disease, and diabetes. These findings motivate us to have a greater sense of awareness towards these diseases in the pre-treatment time-period. Furthermore they form a stronger foundation for the implementation of identification and optimization strategies towards comorbidity in patients with head and neck cancer.

References

1. Piccirillo JF, Feinstein AR. Clinical Symptoms and comorbidity: significance for the prognostic classification of cancer. Cancer. 1996; 77: 834-42.
2. 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: 718-24.
3. Piccirillo JF, Lacy PD, Basu A, Spitznagel L. Development of a new head and neck cancer-specific comorbidity Index. Arch Otolaryngol Head Neck Surg. 2002; 128: 1172-9.
4. Piccirillo JF, Costas I, The impact of comorbidity on outcomes. ORL J Otorhinolaryngol Relat Spec. 2004; 66: 180-5.
5. Piccirillo JF, Tierney RM, Costas I, Grove L. Spitznagel EL Jr. Prognostic importance of comorbidity in a hospital-based cancer registry. JAMA. 2004; 291: 2441-7.
6. Ferrier MB, Spuesens EB, Le Cessie S, Baatenburg de Jong RJ. Comorbidity as a major risk factor for mortality and complications in head and neck Surgery. Arch Otolaryngol Head Neck Surg. 2005; 131: 27-32.
7. Linn BS, Linn MW, Gurel L. Cumulative Illness Rating Scale. J Am Geriatr Soc. 1968; 16: 622-6.
8. Kaplan MH, Feinstein AR. The importance of classifying initial co-morbidity in evaluating the outcome of diabetes mellitus. J Chronic Dis. 1974; 27: 387-404.
9. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A New Method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987: 40: 373-83.
10. Cleary PD, Greenfield SG, Mulley AG, Pauker SG. Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. JAMA 1991; 266: 73–9.
11. Hall SF, Rochon PA, Streiner DL, Paszat LF, Groome PA, Rohland SL. Measuring comorbidity in patients with head and neck cancer. Laryngoscope. 2002; 112: 1988-96.
12. Feinstein AR. Clinical biostatistics XIV. Clin. Pharmacol. Ther. 1972; 13: 285-97.
13. Feinstein AR. Clinical biostatistics XV. Clin. Pharmacol. Ther. 1972; 13: 442-57.
14. Feinstein AR. Clinical biostatistics XVI. Clin. Pharmacol. Ther. 1972; 13: 609-24.
15. Feinstein AR. Clinical biostatistics XVII. Clin. Pharmacol. Ther. 1972; 13: 755-68.
16. ACE27 Comorbidity Calculator: available on-line at: http://oto.wustl.edu/clinepi/calc.html
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
18. Harrell F. Regression modeling strategies, Springer-report New York 2001.
19. Bailey BJ. Beyond the 'new' TNM classification. Arch Otolaryngol. Head Neck Surg. 1991; 117: 481-3.
20. Paleri V, Wight RG, Davies GR. Impact of comorbidity on the outcome of laryngeal squamous cancer. Head Neck. 2003; 25: 1019-26
21. Sing B, Bhaya M, Zimbler M, et. al. Impact of comorbidity on outcome of young patients with head and neck squamous cell carcinoma. Head Neck. 1998; 20: 1-7.

Internal and external validation of the updated prognostic head and neck cancer model

Chapter Three

Datema FR, Ferrier MB, Vergouwe Y, Moya A, Molenaar J, Piccirillo JF, Baatenburg de Jong RJ. Update and external validation of a head and neck cancer prognostic model. Awaiting final decision after revisions. Head Neck

Abstract

Purpose Update and external validation of a prognostic model that is able to predict the survival probability of newly diagnosed head and neck cancer patients.

Materials and methods The Leiden prognostic model is based on historical data of 1371 primary head and neck cancer patients, diagnosed and treated in the Leiden University Medical Center, between 1981 and 1999. The model contains the predictors age, sex, tumor site, TNM-classification, prior tumors and comorbidity. We updated the model with follow-up data until January 2010. The updated model was then externally validated in 598 head and neck cancer patients from the Siteman Cancer Center/Barnes-Jewish Hospital, St. Louis, Missouri, USA.

Results Median follow-up was 5.5 years (range 0-25.5). Only 2.5% of patients were lost to follow-up. During follow-up 1099 (80.2%) passed away. Discrimination of the updated prognostic model was good with a C-index of 0.73 after internal validation. The discrimination was slightly lower in the external validation set (C- index 0.69). The predicted 2-year and 5-year survival was satisfactory with some slight deviations from the perfect calibration line.

Conclusions We used recent follow-up information to update the Leiden prognostic model for newly diagnosed head and neck cancer patients. The model showed acceptably good calibration and discrimination results in internal and external validation procedures.

Introduction

An accurate prognostic assessment of the newly diagnosed head and neck cancer patient can assist the treating physician in clinical decision making, patient information about the likely outcome of disease and treatment evaluation. Physicians are often worried about the accuracy of their assessment, because it is very difficult to determine the impact and interaction of applicable factors on prognosis [1]. To aid the physician, statistical survival analyses are mandatory.

We recently developed a prognostic Cox model that combines tumor specific predictors (TNM-classification, tumor location) with patient specific predictors (age, sex, prior tumors, and comorbidity) [2]. The prognostic model is based on the historical data of 1371 Dutch primary head and neck squamous cell carcinoma (HNSCC) patients and was internally validated.

In this paper we describe the update of the Leiden prognostic model with new follow-up and survival data, reaching until January 2010. Furthermore, the predictive accuracy of the model is evaluated with an external validation.

The goal of the external validation is to study the performance of the prognostic model for patients from a different population. The external validation population should be similar to the development population in terms of index disease, but different from the development population in terms of geographic location, historical time period, or other important ways. For the purposes of assessing the external validation or generalizability of the prognostic model, a dataset containing 598 primary HNSCC patients, diagnosed and treated at the Siteman Cancer Center/Barnes-Jewish Hospital, St. Louis, Missouri, USA from 1995 to 2000 was used.

To our knowledge it is the first time that a long-term-follow-up prognostic head and neck cancer model is externally validated. If the Leiden prognostic model shows a good predictive performance in American data, we will be more confident about the accuracy and transportability of the model.

Materials and methods

Sources of data

The Leiden prognostic model is based on historical data of 1371 consecutive primary HNSCC patients diagnosed and treated at the Leiden University Medical Centre (LUMC) between 1981 and 1999. All tumors are histologically confirmed squamous cell carcinomas of the lip, oral cavity, oropharynx, nasopharynx, hypopharynx, glottic larynx or supraglottic larynx. Data on age at diagnosis, sex, prior tumors, tumor location and the TNM-classification were available from the hospital-based cancer registry system (ONCDOC). The ONCDOC, established in 1969, contains patient, treatment, and follow-up data of all oncological patients diagnosed and treated at the LUMC. Trained oncologic data managers store these data and safeguard an adequate follow-up by contacting the general practitioner and/or Registry of Births, Deaths, and Marriages when patients are lost. Between 1990 and 1994 twenty-seven patients from Suriname were treated in the LUMC. These patients returned to their home country afterwards and were therefore not followed by ONCDOC. Correcting for these patients, our percentage lost to follow-up was only 0.5%. Follow-up data for this study reaches until January 2010. Disease was always staged or restaged according to the most up-to-date Union Internationale Contre le Cancer (UICC) manual. Comorbidity (present before primary tumor diagnosis) was coded according to the Adult Comorbidity Evaluation (ACE27) manual [3]. Prior tumors were defined as “all preceding malignant tumors, except for basal cell or squamous cell carcinomas of the skin”. All model covariables are considered categorical, except age which is a continuous variable. The reference categories for categorical covariables are set as the category with best prognosis. For example, the reference category for tumor location is the lip and the reference category for tumor size (T-classification) is T1.

The secondary dataset for external validation consist out of 598 primary HNSCC patients, diagnosed and treated at the Siteman Cancer Center/Barnes-Jewish Hospital between 1995 and 2000. All tumors are histologically confirmed squamous cell carcinomas of the lip, oral cavity, oropharynx, glottic larynx or supraglottic larynx. Data on age at diagnosis, sex, prior malignancies, tumor location, TNM-classification and comorbidity are complete. Follow-up for all American patients reaches until November 2009.

Statistical analyses

Update of the prognostic model

Information on survival status and follow up time for the individual Dutch patients was updated until January 2010. For 89 patients information on comorbidity was missing. Missing values were imputed once, allowing all observed values to be analysed. We used the MICE algorithm, which works with R software. With these data a new Cox regression model was fitted, that contained the same predictors as the original model. Regression coefficients (ß), 95% confidence intervals and covariable importance values (Z-value) are reported. The Z-value is the ratio of the regression coefficient to its standard error. A covariable with a high Z-value corresponds to an important predictor for the model.

Model performance measures

The accuracy of the prognostic model was evaluated by means of calibration and discrimination. Calibration defines if predictions correspond to the observed outcomes. This is illustrated with calibration curves. Calibration was studied for 2-year and 5-year survival. Patients were divided in eight groups based on the predicted risks. For each group a Kaplan Meier curve was constructed to assess the observed survival probability, that was plotted against the survival probability calculated by the prognostic model. Discrimination defines the accuracy of the prognostic model in distinguishing between patients who survive and who die. The Concordance index (C-index) is commonly used to quantify discrimination. The C-index estimates the probability that in a randomly selected pair of patients, the patient who dies first has the worst predicted outcome. A C-index of 0.5 corresponds to a model doing no better than random guessing. A value of 1.0 corresponds to a perfectly discriminating model.

Internal and external validation

Over-fitting is a common problem in prognostic modelling. An over-fitted model results in low predictions being too low and high predictions, being too high. When a model is over-fitted its predictive accuracy may be quite good when it is applied on the development dataset, yet when the model is applied to a new dataset performance will be poor [4]. The extent of over-fitting was estimated in an internal validation procedure using 100 bootstrap samples. A shrinkage factor was calculated and used to shrink the regression coefficients to obtain well calibrated predictions for new patients [5-8]. The bootstrap procedure also yielded an 'optimism corrected' C-index, which reflects the discriminative capability of our updated prognostic model in new, similar patients.

For external validation, the prognostic model with shrunken regression coefficients was applied on the American data. A graphical representation of the complete external validation approach is given in figure 1. For each American patient a prognostic index was calculated based on the regression coefficients from the updated Leiden model. The Leiden tertile cutoff values were used to divide the American cohort into three risk groups (high-intermediate-low). Observed survival assessed as Kaplan Meier plots were compared with the predicted survival of each risk group.

SPSS for Windows © version 17.0 and R version 2.11.1 were used to perform all statistical and explorative analyses. Only p-values < .05 were considered statistically significant.

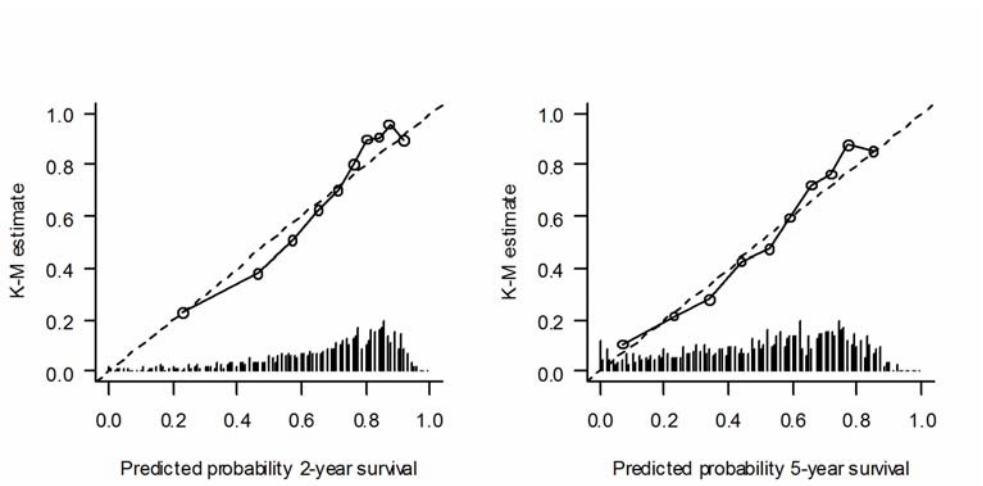
Results

Update of the prognostic model

In table 1 the updated Cox model is presented. All predictors (except sex) have a significant independent impact on overall survival. These results are similar to the original prognostic model [1]. Looking at the Z-values, age at diagnoses is the most important predictor. Other highly important predictors are T-classification, N-classification, M-classification and comorbidity.

In figure 1 calibration of the model is shown. In general, predictions and observed outcomes for 2 and 5-year survival showed a close proximity to the perfect calibration slope (dashed line), which may be expected in development data. Internal validation of the model resulted in a shrinkage factor of 0.95, indicating relatively minor over-fitting. The 'optimism corrected' C-index was 0.73.

Figure 1 Validation plot of the Dutch prognostic model for 2-year survival and 5-year survival.



For each percentile of predicted probability, the average predicted probability is plotted against the Kaplan-Meier estimate. Distribution of the predicted probabilities is indicated with vertical lines at the bottom. The dashed line is the perfect calibration slope.

Table 1. Results of the updated prognostic model

Covariable		ß	Exp (ß)	Z-value	P-value	95% CI
Age		0.04	1.04	13.33	< 0.01	1.03 - 1.05
Gender	Male [RC]	--	1.00	--	--	--
	Female	0.08	0.92	0.98	0.33	0.79- 1.08*
Tumor Location	Lip [RC]	--	1.00	--	--	--
	Hypopharynx	0.62	1.85	3.75	< 0.01	1.35 - 2.54
	Oral Cavity	0.41	1.51	2.93	< 0.01	1.14 - 1.99
	Oropharynx	0.47	1.60	2.94	< 0.01	1.17 - 2.17
	Glottic larynx	0.04	1.04	0.28	0.78	0.81 - 1.33*
	Supraglottic larynx	0.27	1.31	1.82	0.07	0.98 - 1.77*
	Nasopharynx	0.18	1.20	0.75	0.46	0.74 - 1.94*
T-classification	T1 [RC]	--	1.00	--	--	--
	T2	0.25	1.28	2.91	< 0.01	1.09 - 1.51
	T3	0.42	1.52	3.99	< 0.01	1.24 - 1.86
	T4	0.67	1.95	6.64	< 0.01	1.60 - 2.37
N-classification	N0 [RC]	--	1.00	--	--	--
	N1	0.37	1.45	3.53	< 0.01	1.18 - 1.78
	N2	0.61	1.85	6.07	< 0.01	1.52 - 2.26
	N3	0.90	2.45	6.76	< 0.01	1.89 - 3.17
M-classification	M0 [RC]	--	1.00	--	--	--
	M1	1.85	6.36	6.70	< 0.01	3.70 -10.93
Prior tumors	No [RC]	--	1.00	--	--	--
	Yes	0.50	1.64	5.02	< 0.01	1.34 - 2.00
Comorbidity	ACE27 Grade 0 [RC]	--	1.00	--	--	--
	ACE27 Grade 1	0.06	1.07	0.86	0.45	0.91 -1.25*
	ACE27 Grade 2	0.34	1.39	3.96	< 0.01	1.18 - 1.66
	ACE27 Grade 3	0.79	2.21	6.38	< 0.01	1.73 - 2.80

*Abbreviations; RC: reference category; ß: unshrunken regression coefficient; Exp(ß): multiplicative factor; Z-value: ratio of regression coefficient to its standard error; 95% CI: 95% confidence interval, *: Not statistically significant.*

Comparison between Dutch and USA cohorts

In table 2, the patient, tumor and follow-up data of the Leiden and American cohort are displayed. In general, the two datasets were fairly comparable. The American diagnostic time frame is more recent (1995-2000) than the Dutch (1981-1999) and median follow up is shorter (3.9 years and 5.5 years respectively). During follow-up, 80.2% of the Dutch patients passed away and in the American cohort this percentage was 68.9%. Regarding predictor distribution, some remarkable discrepancies are present. The majority (41.5%) of American patients had an oral cavity tumor where this tumor location only represents 20.4% of the Leiden population. In the American dataset nasopharynx and hypopharynx carcinomas are not present. In the Leiden population these tumor locations are present in 3.0% and 10.0% respectively.

Table 2. Demographic and follow-up data of the two study populations

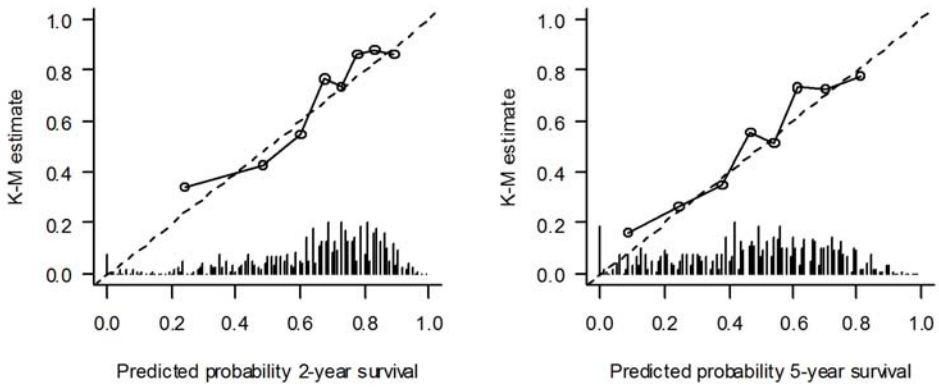
Covariable		Explanation	Dutch population [N = 1371]	USA population [N = 598]
Age	-	Mean ± SD	62.6 ± 12.0	61.6 ± 12.4
Gender	Female	N [%]	283 [20.6%]	171 [28.6%]
	Male		1088 [79.4%]	427 [71.4%]
Tumor location	Lip	N [%]	123 [9.0%]	22 [3.7%]
	Oral cavity		280 [20.4%]	248 [41.5%]
	Oropharynx		152 [11.1%]	91 [15.2%]
	Nasopharynx		41 [3.0%]	-
	Hypopharynx		137 [10.0%]	-
	Glottic larynx		442 [32.2%]	137 [22.9%]
	Supraglottic larynx		196 [14.3%]	100 [16.7%]
T-classification	T1	N [%]	516 [37.6%]	217 [36.3%]
	T2		369 [26.9%]	187 [31.3%]
	T3		208 [15.2%]	101 [16.9%]
	T4		278 [20.3%]	93 [15.6%]
N-classification	N0		964 [70.3%]	381 [63.7%]
	N1		145 [10.6%]	82 [13.7%]
	N2		180 [13.1%]	118 [19.7%]
	N3		82 [6.0%]	17 [2.8%]
M-classification	M0	N [%]	1354 [98.8%]	585 [97.8%]
	M1		17 [1.2%]	13 [2.2%]
Prior tumors	No	N [%]	1238 [90.3%]	502 [83.9%]
	Yes		133 [9.7%]	96 [16.1%]
Comorbidity*	ACE27 Grade 0	N [%]	845 [61.6%]	256 [42.8%]
	ACE27 Grade 1		251 [18.3%]	182 [30.4%]
	ACE27 Grade 2		193 [14.1%]	118 [19.7%]
	ACE27 Grade 3		82 [6.0%]	42 [7.0%]
Diagnostic time frame		Years	1981-1998	1995-2000
Follow-up time		Years mean/median	7.5 / 5.5	5.4 / 3.9
Status at last follow-	Dead	N [%]	1099 [80.2%]	412 [68.9%]
	Alive		272 [19.8%]	186 [31.1%]

*Abbreviations; SD: standard deviation; N: number of patients; N[%]: percentage of total number of patients; * by imputation, the missing comorbidity data of 89 patients was added to the data*

External validation

In figure 2 calibration of the updated model in American patients is shown. In general, predictions and observed outcomes for 2 and 5-year survival showed good agreement with some deviations from the perfect calibration slope (dashed line). Overall, predicted and observed survival estimates of the whole American population were 0.66 versus 0.68 and 0.48 versus 0.51 respectively for two- and five-year survival.

Figure 2 Validation plot of the prognostic model with USA data for 2-year survival and 5-year survival.



For each percentile of predicted probability, the average predicted probability is plotted against the Kaplan-Meier estimate. Distribution of the predicted probabilities is indicated with vertical lines at the bottom. The dashed line is the perfect calibration slope.

In figure 3 the calculated survival curves of three American risk groups are compared to observed Kaplan Meier survival curves. The Leiden tertile cut-off values distributed American patients into a low-risk group (N = 142), an intermediate-risk group (N =238) and a high-risk group (N = 218). The figure shows that our prognostic model is capable of discriminating between American patients with a good, intermediate, and a poor prognosis. This result is confirmed by the C- statistic of 0.69.

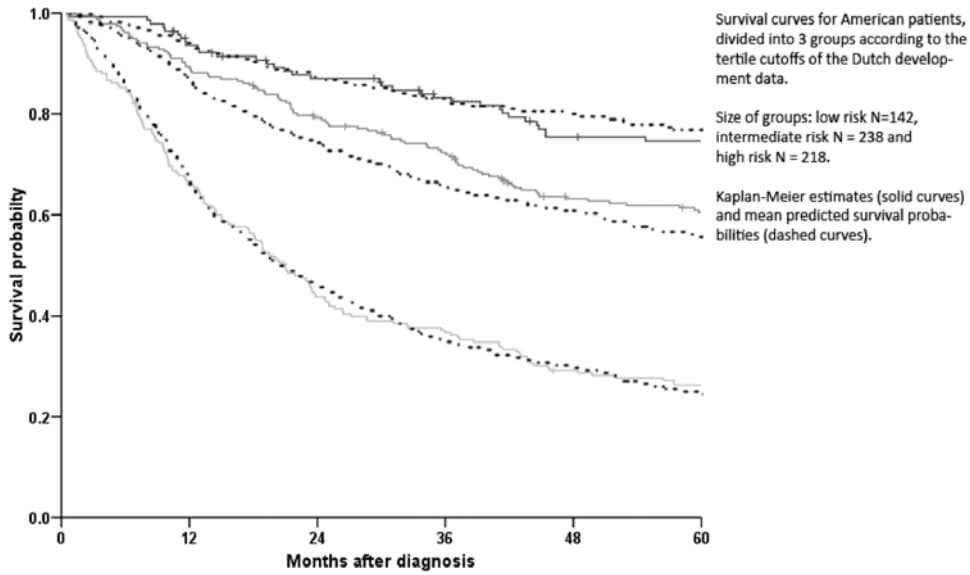


Figure 3 Graphical presentation of external validation performance

Discussion

In this study we described the internal validation of an eight-variable prognostic Cox regression model, developed from the survival experience of 1371 consecutive primary Dutch HNSCC patients. Following internal validation, an external validation procedure was performed in a cohort of similar patients treated in an American teaching hospital and national cancer center.

When we take a closer look at the prognostic factors in the model, age, T-classification, N-classification, M-classification, and (moderate to severe) comorbidity are the most important predictors. These findings are in agreement with clinical experience and results from prior research [3, 9-10, 12-16]. Sex did not influence prognosis. This is possibly explained by the knowledge that the majority of our patients were younger than 65 years. The average life expectancy for Dutch men and women younger than 65 is similar [17]. Mild comorbidity did not influence prognosis as well. This is explained by the fact that mild comorbidity is usually a historical medical event without detrimental effects. When necessary, mild conditions are usually treated with therapeutic or prophylactic medication and then form little mortality risk to the patient. The impact of severe comorbidity on overall survival however is comparable to the impact of a T4-tumor or N3 neck, stating the importance of recognizing comorbidity in the HNSCC patient. Prior malignancies have an independent impact on prognosis as well. There are however some considerations: the model does not distinguish patients surviving an aggressive tumor (with potentially detrimental effects on general health) from patients surviving a mild tumor. Second, the time-interval between the prior tumor and the primary HNSCC is not taken into account. It is clear that for example, the impact of a lung tumor 6 months prior to HNSCC diagnosis is different from the impact of a lung tumor 8 years prior to HNSCC diagnosis. This problem was previously addressed by Goeman et. Al [18]. Despite such limitations, an internal validation procedure of our updated model showed acceptably good results and model performance improvement (C-index of 0.73 versus 0.71).

We believe that a more accurate model performance is only possible when we expand the development study population, preferably until we have included a large group of patients from a time-frame with similar diagnostic and therapeutic options as today. This would partially counter the “out-of-date principle” of the model. Another possibility to improve model performance is to include additional prognostic covariables such as for example severe malnutrition or oncological biomarkers as developed for HPV [11].

In the published literature, prognostic model performance is frequently reported with results from an internal validation procedure. The most stringent test for a model however is a test of generalizability or external validation. External validations are much less frequently reported since the procedure requires an independent group of patients with similar characteristics as the development study population. These databases are hard to come by. In this paper we described how an external validation of our model was performed on a dataset of patients receiving care at an America cancer center.

At first glance, the Leiden and American datasets may seem quite similar. However, upon closer inspection, the two datasets reveal some discrepancies that require attention. An important finding is that the American dataset does not include patients with nasopharynx or hypopharynx carcinomas. In the Leiden dataset, patients with tumors in these locations are present in 3.0% and 10.0%, respectively. Another discrepancy is formed by the inclusion period and the median follow up time. Dutch patients were included from 1981 to 1998 with a median follow-up time of 5.5 years. American patients were included from 1995 to 2000 with a mean follow-up time of 3.9 years. American patients are from a more recent time frame (with more up-to date treatment regimes and modalities), do not include hypopharynx tumors (with a generally worse prognosis), and have a shorter median follow-up time. Despite these differences, good results were found in the external validation procedure. A C-index derived from external data of 0.69, compared to a C-index of 0.73 derived from internal data is a good result. Furthermore, the model appeared very capable in discriminating between American patients with a good, intermediate, and bad prognosis. Similar internal and external validation results were found in a study of Campbell et. al. who found a C-index of 0.745 and 0.697 after internal and external validation of a British breast cancer model [19].

External validation is essential before implementing prediction models in clinical practice. The results of this study make us feel confident about the clinical applicability of the Leiden prognostic model. But when is a prognostic model good enough? In 1993, the American Joint Committee on Cancer (AJCC) communicated their vision on this question and proposed criteria for prognostic factors and a prognostic system [20]. All predictors in our model, except gender, are significant, independent and clinically important and therefore comply with the AJCC prognostic factor criteria. Out of twelve criteria for a prognostic system, the most important 9 criteria are upheld by the Leiden prognostic model. E.g., treatment (AJCC) is not a prognostic factor in our model, because in any index of prognostic stratification, choices between treatment options that are not under control of the investigator will influence outcome. This systematic error cannot be eliminated.

Furthermore, choices between treatment modalities are partially determined by variables that are already included in our model (e.g, TNM- classification, age at diagnosis, and comorbidity). The inclusion of treatment as an extra prognostic variable would, therefore, introduce bias and confounding. Emerging diagnostic techniques such as molecular testing and new imaging techniques (AJCC) are not incorporated as predictors in our model. The TNM-classification however is partially derived from diagnostic imaging results.

Conclusion

External validation is essential before implementing prediction models into clinical practice. The Leiden prognostic model showed acceptably good calibration and discrimination results in an external validation procedure. Therefore, we recommend the use of this prognostic model for HNSCC patients receiving care at medical centers in developed countries.

References

1. Justice AC, Covinsky KE, Berlin JA. Assessing the Generalizability of Prognostic Information. Ann Intern Med. 1999; 130: 515-524.
2. Datema FR, Moya A, Krause P et. al. Novel Head and Neck Cancer Survival Analysis Approach: Random Survival Forests versus Cox Proportional Hazards Regression. Head Neck 2012; 34: 50-8.
3. Piccirillo FJ, Tierney RM, Costas I, Grove L, Spitznagel EL Jr. Prognostic Importance of Comorbidity in a Hospital-Based Cancer Registry. JAMA 2004; 291:2441-7.
4. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. J Clin Epidemiol. 2003; 56: 441-7.
5. Royston P, Parmar MKB, Altman DG. External validation and updating of a prognostic survival model. Oxford Research report No. 307. 2010.
6. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. Br J Cancer 2003; 89: 605-11.
7. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis part IV: Further concepts and methods in survival analysis. Br J Cancer 2003; 89: 781-786.
8. 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.
9. Datema FR, Ferrier MB, vd Schroeff MP, Baatenburg de Jong RJ. The impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. Head Neck 2010; 32: 728-36.
10. 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: 718-24
11. Datema FR, Ferrier MB, Baatenburg de Jong RJ. Impact of severe malnutrition on short-term mortality and overall survival in head and neck cancer. Oral. Oncol. 2011; 47:910-4.
12. Piccirillo JF, Wells CK, Sasaki CT, Feinstein AR. New clinical severity staging system for cancer of the larynx. Five-year survival rates. Annals of Otology, Rhinology & Laryngology 1994; 103: 83-92.
13. Pugliano FA, Piccirillo JF, Zequeira MR, Emami B, Perez CA, Simpson JR, Fredrickson JM. Clinical-severity staging system for oropharyngeal cancer: five-year survival rates. Archives of Otolaryngology - Head & Neck Surgery 1997; 123: 1118-24.
14. Pugliano FA, Piccirillo JF, Zequeira MR, Fredrickson JM, Perez CA, Simpson JR. Clinical-severity staging system for oral cavity cancer: 5-year survival rates. Otolaryngology - Head & Neck Surgery 1999; 120: 38-45
15. Piccirillo JF. Importance of comorbidity in head and neck cancer. Laryngoscope. 2000; 110: 593-602.
16. Schroeff vd MP, Steyerberg EW, Wieringa MH, Langeveld TP, Molenaar J, Baatenburg de Jong RJ. Prognosis: A variable parameter. Dynamic prognostic modeling in head and neck squamous cell carcinoma. Head Neck 2011 (epub ahead of print).
17. Poos MJJC (RIVM). What is the life expectancy in the Netherlands? Public Health, Future exploration, National Institute of Health 2008. www.nationaalkompas.nl

18. Goeman JJ, Le Cessie S, Baatenburg de Jong RJ, van de Geer SA. Predicting survival using disease history: a model combining relative survival and frailty. Statistica Neerlandica 2004; 58: 21-34.
19. Campbell HE, Gray AM, Harris AL, Briggs AH, Taylor MA. Estimation and external validation of a new prognostic model for predicting recurrence-free survival for early breast cancer patients in the UK. Br J Cancer. 2010; 103: 776-86.
20. Burk HB, Henson DE. The American Joint Committee on Cancer. Criteria for prognostic factors and for an enhanced prognostic system. Cancer. 1993; 72: 3131-5.

Severe malnutrition as a potential 9th prognostic variable for overall survival estimation in head and neck cancer

Datema FR, Ferrier MB, Baatenburg de Jong RJ. Impact of severe malnutrition on short-term mortality and overall survival in head and neck cancer. Oral Oncol. 2011; 47: 910-4.

Chapter Four

Abstract

Background Basic patient and tumor characteristics impact overall survival of head and neck squamous cell carcinoma patients. Severe malnutrition, defined as weight loss > 10% in the 6 months preceding primary tumor diagnosis, impact overall survival as well. Little attention has been paid to the interaction between severe malnutrition and other relevant prognostic covariables. This study investigates the impact of malnutrition on short-term mortality and overall survival, together with the covariables age, tumor site, gender, TNM-classification, comorbidity and prior tumors.

Methods 383 consecutive primary HNSCC patients, diagnosed and treated between 1995 and 1999 were followed until January 2010. Impact of covariables on short-term mortality and overall survival was studied univariately with Kaplan-Meier curves and the log-rank test. Cox-regression and binary logistic regression were used for multivariate analyses.

Results 28 (7.3%) patients were severely malnourished. All covariables, except gender and prior tumors had significant impact on overall survival. The relative risk of severe malnutrition was 1.8 and is comparable to the impact of a T2 tumor, a N1 neck or moderate comorbidity. A univariate relationship between severe malnutrition and short-term mortality was established.

Conclusions Severe malnutrition has major and independent impact on overall survival, with a relative risk comparable to a tumor sized T2, a neck staged N1 or moderate comorbidity. Early and continuous intervention of malnutrition is mandatory.

Introduction

Cancer cachexia, chewing and swallowing impairments caused by the local tumour or by side effects from oncological treatment can result in malnourishment of head and neck cancer patients. A malnourished patient is at risk for increased morbidity and mortality. Fortunately it is possible to diagnose and treat malnutrition, for example with (par) enteral feeding [1].

Despite adequate diagnosis and intervention strategies there are patients who remain malnourished during and after cancer treatment. These patients are at risk for adverse events, such as body tissue catabolism and wound healing disorders. These patients sometimes cannot tolerate optimal treatment. Adverse events and suboptimal treatment are two risk factors for a decreased overall survival as well. Furthermore, adverse events are associated with a decrease in quality of life, which has a proven negative impact on overall survival. These detrimental effects of malnutrition are reported in previous research [1-7].

Few studies are performed that quantify the impact of malnutrition on overall survival of head and neck squamous cell carcinoma (HNSCC) patients, together with other prognostically relevant predictors.

Recently we developed a Cox regression model that contains eight predictors with an independent impact on overall survival of newly diagnosed HNSCC patients [8]. The model combines tumour specific characteristics (TNM-classification and tumour location) with patient specific characteristics (age at diagnosis, gender, comorbidity and prior tumors) to assess an overall survival probability. The model is based on the historical data of 1371 consecutive primary Dutch HNSCC patients and is validated internally and externally (C-index 0.73 and 0.69 respectively). Malnutrition could be a potential ninth predictor for this model.

This study investigates the univariate and multivariate impact of malnutrition on overall survival using a subset (n = 383) of the original 1371 patients. Furthermore, analyses were performed to identify a possible relationship between severe malnutrition and short-term mortality.

Prognostic models can assist the head and neck oncologist in a more accurate prognosis prediction of newly diagnosed patients which is useful in patient counselling, clinical decision-making and quality maintenance.

Materials and methods

Sources of data

For this study the data of 383 consecutive, primary HNSCC patients were available. All patients were diagnosed and treated at the Leiden University Medical Centre (LUMC) between January 1995 and December 1998. All tumours are histologically confirmed squamous cell carcinoma of the lip, oral cavity, oropharynx, nasopharynx, hypopharynx, glottic larynx or supraglottic larynx.

Patient data on age, gender, primary tumour site, prior malignancies and TNM-classification were available from the hospital based cancer registry system (ONCDOC). ONCDOC was established in 1969 and contains patient, treatment and follow-up data of each cancer patient diagnosed in the LUMC. Trained oncological data managers store these data and safeguard an adequate follow-up by contacting the general practitioner and/or Registry of Births, Deaths and Marriages when patients are lost. As a result, no patients were lost to follow-up until January 2010.

Age at diagnosis was considered a continuous predictor and all other predictors were considered categorical. The TNM-classification was staged according to the 5^{th} edition of the Union Internationale Contre le Cancer (UICC) manual. Prior tumours were defined as all preceding malignant tumours except for basal cell and squamous cell carcinoma of the skin. Comorbidity was coded according to the Adult Comorbidity Evaluation (ACE27) where grade 0 corresponds to no comorbidity, grade 1 to mild comorbidity, grade 2 to moderate comorbidity and grade 3 to severe comorbidity [8].

Data on weight loss were collected and recorded into the medical chart by a trained dietician that especially joined the oncological head and neck team during the above mentioned time-frame. The patient's weight at cancer diagnosis was measured and compared to the patient's weight 6 months prior to diagnosis (these data were obtained from anamnesis or from prior medical documentation when available). Based on the weight difference, the dietician calculated a percentage of estimated weight loss in the 6 months preceding cancer diagnosis. For this study three variables were analysed: *severe malnutrition*, defined as > 10% weight loss, *moderate malnutrition*, defined as a weight loss of 5 to 10% and *no malnutrition*, defined as a weight loss of less than 5%.

Based on the therapeutic nil hypothesis, the type of treatment was not considered a prognostic factor (see discussion) [9-12]. Treatment modalities are however reported in table 1.

Table 1. Demographic and tumour data of baseline study population

	Total population N = 383	No malnutrition N = 335	Moderate malnutrition N = 20	Severe malnutrition N = 28
	No. [%]	No. [%]	No. [%]	No. [%]
Tumour location				
Lip	29 [7.6%]	29 [7.6%]	0 [0.0%]	0 [0.0%]
Oral cavity	94 [24.5%]	82 [21.4%]	4 [1.0%]	8 [2.1%]
Oropharynx	50 [13.1%]	42 [11.0%]	4 [1.0%]	4 [1.0%]
Nasopharynx	8 [2.1%]	5 [1.3%]	1 [0.3%]	2 [0.5%]
Hypopharynx	42 [11.0%]	33 [8.6%]	3 [0.8%]	6 [1.6%]
Glottic larynx	119 [31.1%]	110 [28.7%]	3 [0.8%]	6 [1.6%]
Supraglottic larynx	41 [10.7%]	34 [8.9%]	5 [1.3%]	2 [0.5%]
T-stage				
T1	135 [35.2%]	129 [33.7%]	3 [0.8%]	3 [0.8%]
T2	95 [24.8%]	83 [21.7%]	9 [2.3%]	3 [0.8%]
T3	48 [12.5%]	39 [10.2%]	3 [0.8%]	6 [1.6%]
T4	101 [26.4%]	80 [20.9%]	5 [1.3%]	16 [4.2%]
N-stage				
N0	268 [70.0%]	239 [62.4%]	13 [3.4%]	16 [4.2%]
N1	27 [7.0%]	22 [5.7%]	1 [0.3%]	4 [1.0%]
N2	74 [19.3%]	60 [15.7%]	66 [1.6%]	8 [2.1%]
N3	14 [3.7%]	14 [3.7%]	0 [0.0%]	0 [0.0%]
M-stage				
M0	377 [98.4%]	330 [86.2%]	20 [5.2%]	27 [7.0%]
M1	6 [1.6%]	5 [1.3%]	0 [0.0%]	1 [0.3%]
Gender				
Female	94 [24.5%]	82 [21.4%]	5 [1.3%]	7 [1.8%]
Male	289 [75.5%]	253 [66.1%]	15 [3.9%]	21 [5.5%]
Comorbidity				
ACE27 Grade 0	182 [47.5%]	157 [41.0%]	14 [3.7%]	11 [2.9%]
ACE27 Grade 1	92 [24.0%]	80 [20.9%]	4 [1.0%]	8 [2.1%]
ACE27 Grade 2	57 [14.9%]	52 [13.6%]	2 [0.5%]	3 [0.8%]
ACE27 Grade 3	30 [7.8%]	24 [6.3%]	0 [0.0%]	6 [1.6%]
No data	22 [5.7%]	22 [5.7%]	0 [0.0%]	0 [0.0%]

Malnutrition				
Prior Tumours				
No	339 [88.5%]	297 [77.5%]	19 [5.0%]	23 [6.0%]
Yes	44 [11.5%]	38 [9.9%]	1 [0.3%]	5 [1.3%]
Treatment type				
Surgery	78 [20.4%]	77 [20.1%]	1 [0.3%]	0 [0.0%]
Surgery followed by RTx	106 [27.7%]	90 [23.5%]	6 [1.6%]	10 [2.6%]
Primary RTx	165 [43.1%]	144 [37.6%]	10 [2.6%]	11 [2.9%]
Chemotherapy	8 [2.1%]	6 [1.6%]	0 [0.0%]	2 [0.5%]
Chemo radiation	23 [6.0%]	15 [3.9%]	3 [0.8%]	5 [1.3%]

Abbreviations; ACE27: Adult Comorbidity Evaluation 27, RTx: radiotherapy

Statistical analyses and endpoint definition

To investigate the univariate impact of the nine predictors on *overall survival*, Kaplan Meier Curves and the Log Rank test were used. The endpoint for overall survival was death of all causes.

Cox regression was used to investigate the impact of all predictors on overall survival simultaneously. Only variables with a statistically significant impact in univariate analysis were included in the Cox model. For each categorical predictor, the reference category was set as the category with best prognosis. For example: T1 was the reference category for T-classification and N0 the reference category for N-classification.

To investigate the univariate impact of the nine predictors on *short-term mortality,* Kaplan Meier curves and the Log Rank test were used. For malnutrition two subcategories were used: severe malnutrition and no malnutrition (see results). Binary logistic regression analysis was used for multivariate analysis.

All calculations were performed in SPSS for Windows © (version 17.0). Only p-values < .05 were considered statistically significant.

Results

Explorative data analyses

The mean age at diagnosis was 62.9 years with a standard deviation of 11.8 years. The majority of patients were male (75.5%). Severe malnutrition was encountered in 28 patients (7.3%). From these 28 patients, a majority had a T4 stage tumour (57.1%) and 1 patient (3.6%) had distant metastasis. Moderate malnutrition was encountered in 20 patients (5.2%).

A minority of patients had a preceding malignancy (11.5%). Comorbidity was found in 179 patients (46.7%) of whom 7.8% had severe comorbidity. Most tumours were located in the glottic larynx (31.1%) and oral cavity (24.5%). Further distribution of tumour locations, TNM classification and treatment modalities can be found in table 1.

The mean follow-up time was 6.4 years and the median follow-up time was 6.2 years. None of the patients were lost to follow-up. During follow-up, 270 patients (70.5%) eventually died. The remaining 113 patients (29.5%) were alive in December 2009. Short-term mortality was encountered in 37 patients (9.7%).

Univariate analysis for overall survival

In univariate analysis, all variables except gender ($p = .43$) and prior tumours ($p = .20$), showed a significant impact on overall survival. From the moment of primary tumour diagnosis until 10 years after, a clear distinction between the overall survival of patients with severe malnutrition and patients without malnutrition was present. Moderate malnutrition showed a clear decrease in overall survival probability, starting two years after primary tumour diagnosis. The two-year survival probability of severely malnourished patients was 0.50, compared to 0.70 for patients with moderate malnutrition and no malnutrition ($p < .01$). The 5-year survival probability of severely malnourished patients was 0.36 compared to 0.45 and 0.57 for moderate malnutrition and no malnutrition respectively ($p < .01$). After 10 years, patients with moderate malnutrition showed almost comparable survival as patients without malnutrition. The relative risk of severe malnutrition is comparable to the relative risk of a tumor classified as T3 or a neck classified as N1.

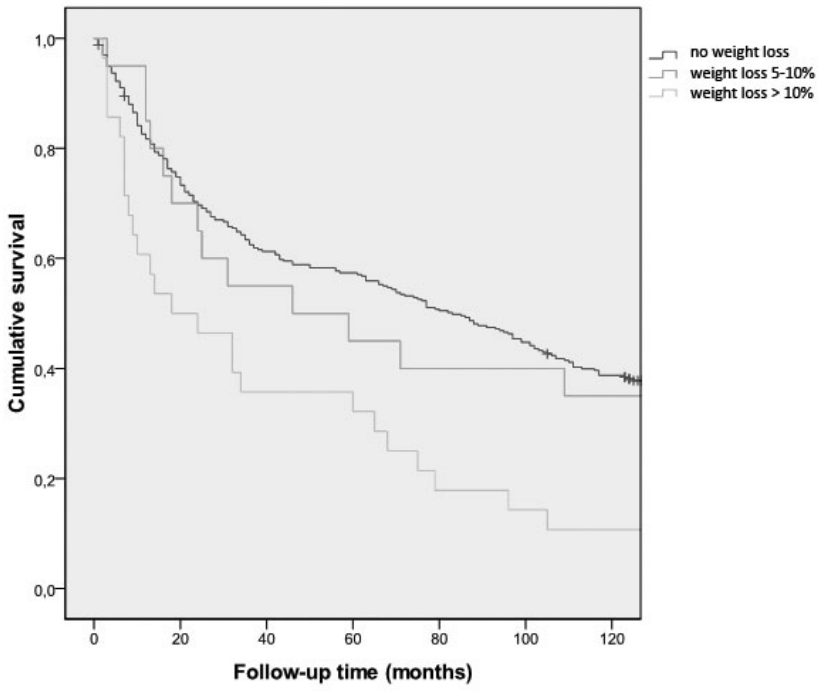


Figure 1 Kaplan Meier Curves showing univariate impact of malnutrition on overall survival

Multivariate analysis for overall survival

In multivariate Cox regression analysis, all variables except gender and prior tumors were included. All included predictors, except tumor location showed a significant impact on overall survival. The independent impact of severe malnutrition on overall survival is comparable to a tumor classified as T2, a neck classified as N1 or moderate comorbidity (ACE27 Grade 2). The hazard ratios, p-value and 95% confidence interval for all covariables are displayed in table 2.

Table 2. Results of the Cox Model

Covariate		B	Exp (B)	P-value	95% CI
Age		0.05	1.05	< .01	1.03 - 1.06
Tumor Location	Lip [RC]	--	1.00	--	--
	Oral Cavity	0.28	1.33	0.38	0.59 - 3.02*
	Oropharynx	0.46	1.57	0.50	0.66 - 3.70*
	Nasopharynx	0.13	1.14	0.31	0.34 - 3.82*
	Hypopharynx	0.63	1.86	0.17	0.77 - 4.48*
	Glottic larynx	0.06	1.05	0.83	0.48 - 2.32*
	Supraglottic larynx	0.49	1.55	0.32	0.65 - 3.70*
T-classification	T1 [RC]	--	1.00	< .01	--
	T2	0.46	1.59		1.08 - 2.33
	T3	0.81	2.33		1.48 - 3.67
	T4	0.59	1.82		1.21 - 2.74
N-classification	N0 [RC]	--	1.00	< .01	--
	N1	0.66	1.91		1.16 - 3.14
	N2	0.82	2.21		1.59 - 3.13
	N3	1.24	3.40		1.82 - 6.34
M-classification	M0 [RC]	--	1.00	< .01	--
	M1	1.54	4.76		1.66 - 13.66
Malnutrition	No [RC]	--	1.00	.03	--
	5-10% weight loss	- 0.16	0.86		0.49 - 1.50
	> 10% weight loss	0.60	1.82		1.15 - 2.87
Comorbidity	ACE27 Grade 0 [RC]	--	1.00	< .01	--
	ACE27 Grade 1	-0.06	0.94		0.68 - 1.30
	ACE27 Grade 2	0.44	1.56		1.08 - 2.25
	ACE27 Grade 3	0.91	2.49		1.60 - 3.88

*Abbreviations; RC: reference category; Exp(B): multiplicative factor on the hazard; 95% CI: 95% confidence interval, *: Not statistically significant.*

Univariate analysis for short-term mortality

Since moderate malnutrition had no significant impact on overall survival in the first 24 months following primary tumor diagnosis, a significant impact form this group on short-term mortality was not expected. In univariate analysis patients with moderate malnutrition were therefore excluded. All covariables except gender (p = .98), prior tumors (p = .89) and age (p = .32), showed a significant relationship with short-term mortality. The 6 month survival probability of a severely malnourished patient was significantly less (p = .03) than for a patient without malnutrition (86% and 92% respectively).

Multivariate analysis for short-term mortality

In multivariate analysis, severe malnutrition did not impact short-term mortality (p = .30). Expected important predictors such as M1 classification, T4 classification and severe comorbidity did show significant impact on short-term mortality.

Discussion

In this study a significant and independent impact of severe malnutrition on overall survival of primary head and neck cancer patients was established. The relative risk of dying for a severely malnourished patient is 1.8 times higher than for patients without malnutrition. Kaplan Meier curves show a clear distinction between the overall survival probability of severely malnourished patients and patients without malnutrition. This distinction remains even 10 years after primary tumor diagnosis. Patients with moderate malnutrition show a decreased overall survival probability as well, starting approximately two years after primary tumor diagnosis. Since we have no specific causes of death for all patients in this study, we can only speculate why this decrease starts after 24 months. Perhaps delayed radiation effects or tumor recurrence can be the explanation.

Our findings emphasize the importance of identification and optimal treatment of malnourishment before, during but also after cancer treatment. Regarding a preventive strategy towards malnutrition we are in agreement with Meuric et. al, who presented a "Good clinical practice in nutritional management of head and neck cancer patients" [15].

Severe malnutrition showed a significant impact on short-term mortality in univariate analysis (additional risk of 6%). A multivariate relationship however could not be established. This is likely explained by the low number of events (9.7%) making it difficult to establish significance. Another explanation could be that most patients (66.7%) who died within six months had distant metastasis.

In our consecutive cohort of 383 patients, a malnutrition incidence of 12.5% (n = 48) was found. Literature reports an incidence ranging from 30 to 50% in head and neck cancer [1-2, 13-14]. The lower incidence in this study can be explained by the definitions for malnutrition that we used, the absence of esophageal cancer as a primary tumor site and the fact that our patients were consecutive (meaning that no selection or inclusion criteria were used other than a primary HNSCC). Three recent studies that use similar definitions for moderate and severe malnutrition, report a more comparable incidence of 16 to 19% [3-4, 16]. Despite a variation in incidence, all studies conclude that malnutrition impacts the overall survival of head and neck cancer patients. Little attention however has been paid to the interaction between malnutrition and other predictors.

Data on weight loss were gathered by a trained dietician that joined the oncological team for this study. To calculate a weight loss percentage, she depended on anamnesis and (when available) on weight data recorded in prior medical documentation. This methodology comes with a possible uncertainty regarding the true percentage of weight loss. There was however no other method available to get more precise results, especially since there is no way of knowing which persons will develop a head and neck tumor in the next six months.

Despite these limitations, results in this study were illustrative and significant and we feel confident about the potential of (severe) malnutrition as a predictor for our existing prognostic model. With extension of model predictors we attempt to get a more accurate model performance. This is important because even for the very experienced head and neck surgeons it remains a difficult task to assess the overall survival probability of a newly diagnosed HNSCC patient. This difficulty exists because overall survival is determined by the impact and interaction of multiple covariables. Statistical survival analyses, such as performed in this study, can aid the physician.

To include treatment modalities as an extra prognostic variable in our prognostic model was not a part of this study. First, in any index of prognostic stratification, choices between treatment options that are not under control of the investigator will influence outcome. This systematic error cannot be eliminated and is the reason this study is based on the therapeutic nil hypotheses. Second, choices between treatment modalities are partially determined by variables that are already included in our model (eg, TNM-classification, age at diagnosis, and comorbidity). The inclusion of treatment as an extra prognostic variable would, therefore, introduce bias and confounding.

Conclusion

Severe malnutrition has a significant and independent impact on overall survival of primary head and neck cancer patients. There is a clear distinction between patients with and without severe malnutrition from moment of diagnosis until 10 years after. This emphasizes the importance of identification and optimal treatment of malnutrition before, during and after cancer treatment.
A univariate relationship between severe malnutrition and short-term mortality was established but a multivariate relationship could not be found.

References

1. Reilly JJ. Does nutrition management benefit the head and neck cancer patient? Oncology (Williston Park) 1990; 4: 105-15.
2. Brookes GB. Nutritional status – a prognostic indicator in head and neck cancer. Otolaryngol Head Neck Surg 1985; 93: 69-74.
3. Jager-Wittenaar H, Dijkstra PU, Vissink A, van der Laan BF, van Oort RP, Roodenburg JL. Malnutrition and quality of life in patients treated for oral or oropharyngeal cancer. Head Neck 2010; 7: epub ahead of print.
4. Jager-Wittenaar H, Dijkstra PU, Vissink A, van Oort RP, van der laan BF, Roodenburg JL. Malnutrition in patients treated for oral or oropharyngeal cancer – prevalence and relationship with oral symptoms: an explorative study. Support Care Cancer 2010; 16: epub ahead of print.
5. Hammerlid E, Wirblad B, Sandin C et. al. Malnutrition and food intake in relation to quality of life in head and neck cancer patients. Head Neck 1998; 20: 540-8.
6. Van Bokhorst-van der Schueren MA, van Leeuwen PA, Sauerwein HP, Kuik DJ, Snow GB, Quak JJ. Assessment of malnutrition parameters in head and neck cancer and their relationship to postoperative complications. Head Neck 1997; 19: 419-25.
7. Oskam IM, Verdonck-de Leeuw IM, Aaronson NL et. al. Quality of life as a predictor of survival: a prospective study on patients treated with combined surgery and radiotherapy for advanced oral and oropharyngeal cancer. Radiother Onco 2010; 97: 258-62.
8. Datema FR, Ferrier MB, van der Schroeff MP, Baatenburg de Jong RJ. Impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. Head Neck 2010; 32: 728-36.
9. Feinstein AR. Clinical biostatistics. XIV. The purposes of prognostic stratification. Clin Pharmacol Ther 1972; 13: 285-297.
10. Feinstein AR. Clinical biostatistics. XV. The purposes of prognostic stratification. I. Clin Pharmacol Ther 1972; 13: 442-457.
11. Feinstein AR. Clinical biostatistics. XVI. The purposes of prognostic stratification. 2. Clin Pharmacol Ther 1972; 13: 609-624.
12. Feinstein AR. Clinical biostatistics. XVII. Synchronous partition and bivariate evaluation in predictive stratification. Clin Pharmacol Ther 1972; 13: 755-768.
13. Hussain M, Kish JA, Crane L, et. al. The role of infection in the morbidity and mortality of patients with head and neck cancer undergoing multimodality treatment. Cancer 1991; 67: 716-21.
14. Sako K, Loré JM, Kaufman S, Razack MS, Bakamjjian V, Reese P. Parenteral hyperalimentation in surgical patients with head and neck cancer: a randomized study. J Surg Oncol 1981; 16: 391-402.
15. Meuric J, Garabige V, Blanc-Vincent MP, Lallemand Y, Bachmann P. Good clinical practice in nutritional management of head and neck cancer patients. Bull Cancer 1999; 86: 843-54.
16. Jager-Wittenaar H, Dijkstra PU, Vissink A, van der Laan BF, van Oort RP, Roodenburg JL. Critical weight loss in head and neck cancer – prevalence and risk factors at diagnosis: an explorative study. Support Care Cancer 2007; 15: 1045-50.

Survival visualization

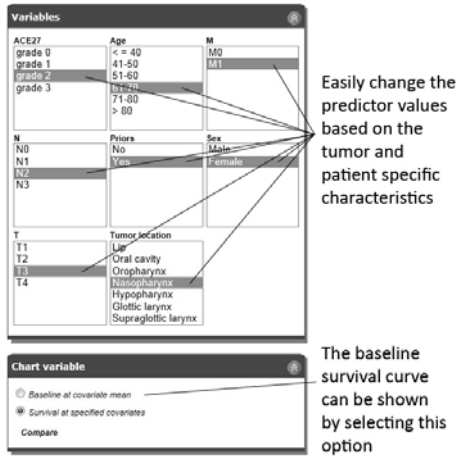
Chapter Five

Introduction

Our most recent prognostic model (chapter three) showed acceptably good results in both internal and external validation procedures. We believe that the model can be used in medical centers located in developed countries. With use of specific software, designed by Henk Jan van der Wijk (department of medical statistics and bioinformatics at the LUMC), it was possible to create an interactive, on-line version of our model. The on-line software generates an individual 5-year survival chart for your newly diagnosed HNSCC patient, without the need for extensive statistical and mathematical knowledge. The model can be accessed and used for free at: www.oncologiq.nl

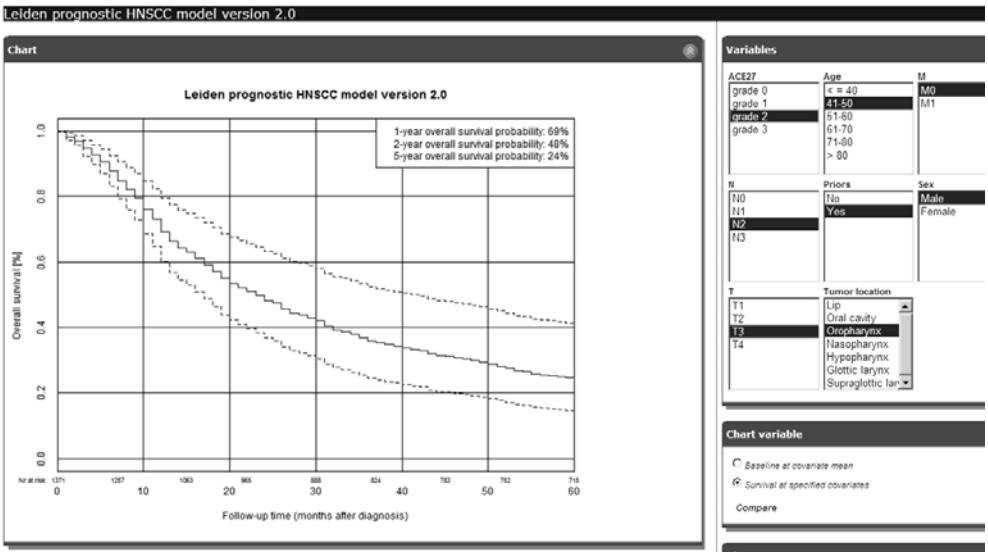
Short tutorial for use of the on-line model

The software provides a user-friendly interface with a *variables menu* on the right side of the screen. In this menu, the eight covariables of our prognostic model are present with corresponding subvariables. By selecting the applicable covariables of your primary HNSCC patients, *an individual 5-year overall survival chart* is generated and presented on the left side of the screen. The X-axis presents the follow-up time in months after diagnosis and the Y-axis the overall survival probability in percentages. To make it easier, the precalculated 1-year, 2-year and 5-year overall survival probability is given.
By manually changing the subvariables in the variables menu, the 5-year survival chart will change with it immediately. It is therefore possible to view the effect of, for example: aging, tumor stage progression and comorbidity progression (but also regression!). In the *chart variable section*, it is possible to show the baseline survival function curve of the entire population.

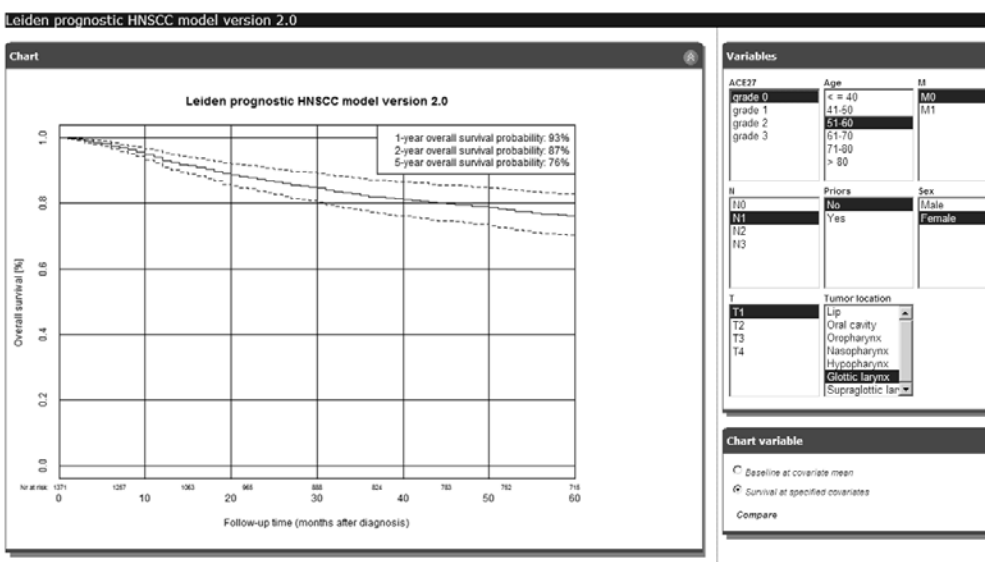


Examples

Survival curve for a 47-year old male with a T3N2M0 oropharynx carcinoma, with a prior tumor and ACE27 score 2 comorbidity. The 1-year and 5-year overall survival probability for this patient is approximately 69% and 24% respectively.



Survival curve for a 51-year old female with a T1N1M0 glottic larynx carcinoma. She had no prior tumors or comorbidity. The 1-year and 5-year overall survival probability for this patient is approximately 93% and 76% respectively.



Software modifications / limitations

In all prognostic models presented in this thesis, age is considered a continuous variable where the mean population age (62.2 years) serves as a reference with a relative risk of 1.0. The relative risk increases with 4% for each year above mean age and decreases with 4% for each year below mean age. At the moment, the on-line software only allows the use of categorical variables. We therefore modified our model development database to create six age-categories and new univariate and multivariate analyses were performed. For illustrative purposes the relative risk of each age category (Cox regression) is shown in table 1. The six age categories show an increasing impact on overall survival. The relative risks of the other model covariables were hardly affected by this exercise. For example, the most extreme relative risk change was seen in M1-stage (6.36 to 6.23). We therefore believe that the online model is representable for the most up-to-date model in this thesis (chapter three).

Table 1. Hazard ratios for age (as a categorical variable)

Covariable		N	Exp (β)	P-value
Age	≤ 40 years	37	1.00 [RC]	< .01
	41-50 years	175	1.20	
	51-60 years	382	1.42	
	61-70 years	424	2.12	
	71-80 years	270	2.97	
	≥ 80 years	83	6.40	

Abbreviations: RC: reference category; Exp (β): hazard ratio; N: number of patients

We welcome suggestions from other users to improve the program.

Incidence and prediction of major cardiovascular complications in extensive head and neck surgery

Datema FR, Poldermans D, Baatenburg de Jong RJ. Incidence and prediction of major cardiovascular complications in head and neck surgery. Head Neck 2010; 32: 1485-93.

Chapter Six

Abstract

Background Patients with head and neck squamous cell carcinoma (HNSCC) usually have a history of tobacco and alcohol abuse. These 2 intoxications not only are main oncologic risk factors but also show a strong causal relationship with certain comorbid conditions. Examples are coronary artery disease, stroke, renal dysfunction, and heart failure, which are all proven major risk factors for an adverse postoperative outcome after stressful non-cardiac surgery. Preoperative identification of these conditions could lead to preventive measures in patients with HNSCC that undergo extensive surgery. Preventing morbidity and mortality is of medical and economical importance.

Methods All comorbidity of 135 consecutive patients with HNSCC that underwent extensive oncologic and reconstructive surgery as the first form of treatment between 2001 and 2007 was investigated. Based on these data, a Lee Cardiac Risk Index (LCRI) Score and an overall Adult Comorbidity Evaluation (ACE-27) severity score were calculated. The predictive value of these scores and the American Society of Anesthesiologists' (ASA) classification toward major cardiovascular complication development were investigated. Major cardiovascular complications were defined as: cardiac death, nonfatal myocardial infarction, heart failure, and cardiac arrhythmias. The impact of these complications on duration of hospitalization, medical costs, and short-term mortality (defined as death within 6 months after primary tumor diagnosis) were investigated as well. The cardioprotective effect of preoperatively prescribed beta blockers and statins are discussed.

Results Twenty-two patients developed 23 major cardiovascular complications (16.3%). In univariate and multivariate analyses, a higher LCRI score was associated with an increased risk for major cardiovascular complications, as was an age >70 years (all values of $p < .01$). The area under the receiver operating characteristics (ROC) curve (AUC) for the multivariate model was 0.84, indicating a good prognostic value. In univariate and multivariate analysis, a higher ACE-27 score was associated with an increased risk for major cardiovascular complications, as was as age >70 years (all values of $p < .01$). The AUC for this model was 0.84, indicating a performance similar to that of the LCRI score model. No statistically significant results were found for the ASA scores ($p = .38$).

Preoperative beta-blocker use showed a significant cardioprotective function in univariate analysis, whereas statins did not. The mean duration of hospitalization was prolonged by 7 days in patients with a major cardiovascular complication. In economic terms, this means a cost increase of at least 3500 Euros. None of the patients died during admission because of a major cardiovascular complication. The short-term mortality rate was 11.1%, but no specific cardiovascular cause of death was reported in these patients.

Conclusions Prevention of major complication occurrence after extensive HNSCC surgery is of medical and economic importance. Our results show that the ACE-27 and the LCRI are suitable instruments for preoperative major cardiovascular complication risk assessment. Addition of the variable age >70 years shows an improvement in predictive value of both instruments. Because of its simplicity we advise the implementation of the LCRI into preoperative HNSCC screening protocols. We advise the exploration of low-dose long-acting beta blockers as a preventive treatment strategy.

Introduction

Patients with head and neck squamous cell carcinoma (HNSCC) are prone to develop significant comorbidity mainly because of the high incidence of tobacco and alcohol abuse, which are 2 main risk factors for HNSCC development.

Comorbidity can form a greater risk for morbidity and mortality than the primary tumor and therefore requires attention in the course of the disease of a patient with HNSCC [1-4] Examples of comorbid conditions that show a causal relationship with alcohol and tobacco abuse are ailments from the cardiovascular, neurologic, and renal systems. Coronary artery disease, stroke, renal dysfunction, and heart failure are proven major risk factors for adverse postoperative outcome after stressful non-cardiac surgery [5]. Preoperative identification of these conditions could lead to preventive measures to reduce perioperative morbidity and mortality (eg, cardiovascular protective medication and intensified perioperative management, including prolonged postoperative stay at the intensive care unit).

Several instruments are at hand to stratify the risk for perioperative morbidity and mortality in patients that require surgery. An instrument that is commonly used for this purpose is the American Society of Anaesthesiologists (ASA) Physical Status Classification System (ASA classification). A second instrument that has a proven predictive value toward complication development, short-term mortality, and overall survival of the patient with HNSCC is the Adult Comorbidity Evaluation (ACE-27) [2, 4]. The ACE-27 is an extensive but thorough instrument that codes both the presence and the severity of organ system-specific comorbidity. A third instrument with a high potential predictive value toward major perioperative cardiovascular complication development is the Lee Cardiac Risk Index (LCRI). The LCRI is a multivalidated and simple 6-item index that can easily be implemented as part of a preoperative risk stratification protocol [5, 6]. The predictive value of the LCRI for a patient population with primary HNSCC, however, is not investigated in prior research.

This retrospective study was performed to gain further insights into the type and severity of comorbid conditions of patients with primary HNSCC requiring extensive oncologic and reconstructive surgery. The collected comorbidity data were used to define an overall ACE-27 and LCRI score. The predictive value toward major perioperative cardiovascular complication occurrence of the ACE-27, LCRI, and ASA classification was researched. Major cardiovascular complications were defined as: cardiac death, nonfatal myocardial infarction, heart failure, and cardiac arrhythmias [5]. A comparison between predictive performances was made to investigate which instrument is most suited to be integrated into a preoperative risk stratification protocol.

Materials and methods

Hospital Setting and Study Population

The Erasmus Medical Centre is a university hospital that serves a population of approximately 3 million people in the south-western area of The Netherlands. It acts as a tertiary referral centre for approximately 30 affiliated hospitals.

Our study population consists of 135 consecutive patients, diagnosed with a primary squamous cell carcinoma of the oral cavity or oropharynx between September 2001 and December 2007. All patients had extensive oncologic and reconstructive surgery as the first line of treatment.

Sources of Data

The data, assessed at preoperative screening, for this retrospective study originated from our computerized hospital information system. Information was collected by the first author of this article, from referral letters, outpatient clinic letters, clinical discharge letters, anaesthesiology reports, surgical reports, complication registration system, laboratory database, chest X-ray reports, electrocardiogram (ECG) database, and pathology reports.

Covariates Collected: Lee Cardiac Risk Index

The Lee Cardiac Risk Index (LCRI) consists of 6 items that define an overall Lee Index score: Lee I: 0 risk variables; Lee II: 1 risk variable; Lee III: 2 risk variables; and Lee IV: >2 risk variables. Five risk factors are comorbid conditions: a history of ischemic heart disease (angina pectoris and/or myocardial infarction), heart failure, history of cerebrovascular disease, insulin-dependent diabetes, and kidney failure (preoperative serum creatinine >2 mg/dL). The sixth risk factor is a high-risk type of surgery [5, 6]. The variables of the LCRI were scored when several criteria were met. Diabetes mellitus was coded in the patient who used insulin or oral medication and in the patient with fastening blood glucose levels >7.0 mmol/L or random blood glucose levels >11.0 mmol/L. Myocardial infarction was coded when the patient had a history of myocardial infarction or when preoperative ECG results showed signs of an old myocardial infarction. Angina pectoris was coded in the presence of typical chest pain complaints.

Patients with a history of bypass surgery or angioplasty were not coded when typical angina pectoris cardiac complaints were absent. Heart failure was coded when patients complained of dyspnea during exercise or dyspnea in rest, supported by physical examination abnormalities suggestive for heart failure. Results from the golden standard (preoperative echocardiography) were not available since this test is not a routine part of preoperative testing in our hospital. Cerebrovascular accidents were coded in patients with a history of transient ischemic attack or cerebrovascular ischemia or haemorrhage.

Renal disease was coded as defined by Lee when the preoperative serum creatinine level was >160 µmol/L or 2.0 mg/dL. In addition to these LCRI variables, the patient's age at diagnosis was calculated and cardiac medication use (beta blockers and statins) was coded. The hospital protocol considering chronic perioperative beta-blocker therapy recommends a resting heart rate between 60 and 70 beats per minute, with dose adjustments prior to surgery to achieve this target heart rate. Statins are prescribed and dosed according to national guidelines [6].
All patients received the same form of surgical treatment. Type of treatment thus has no discriminative value in our current risk analyses. Furthermore, head and neck surgery is defined as surgery with an intermediate to high risk and is therefore not a contributing factor to the Lee Index Score [7].

Covariates Collected: Adult Comorbidity Evaluation (ACE-27)
An overall comorbidity severity grade was designated according to the ACE-27 manual. The overall comorbidity severity grade is defined by the highest-ranked single ailment (coded as grade 1: mild decompensation; grade 2: moderate decompensation; or grade 3: severe decompensation) except when 2 or more grade 2 ailments occur in different organ systems. In this case, the overall comorbidity severity score is grade 3. For example, a patient with chronic heart failure >6 months ago and portal hypertension without compensation (2 grade 2 ailments) would have an overall comorbidity severity score of grade 3. The first author of this report, who has substantial experience with comorbidity coding, designated all ACE-27 scores.

Covariates Collected: American Society of Anesthesiologists' Classification of Physical Status (ASA)

The ASA classification has been revised several times and has evolved into a widely used and commonly accepted risk stratification instrument. The most recent ASA classification is divided into 5 risk classes. Class I is designated to a normal healthy patient; Class II patients have mild systemic disease; Class III patients have severe systemic disease; Class IV patients have a life-threatening systemic disease; and Class V is designated to the moribund patient who is unlikely to survive the operation. The ASA class is routinely determined by the anesthesiologist in a preoperative anesthesiologic intake and recorded in the patients' medical chart. In most cases the ASA class data were therefore readily available to us for analysis.

End-Point Definition

The primary end point for this study was the occurrence of major cardiovascular complications from the timeframe between hospital admission and hospital discharge. Major cardiovascular complications were defined as: cardiac death, nonfatal myocardial infarction, heart failure, and cardiac arrhythmias. Myocardial infarction was coded when ECG reports and positive serum troponine T levels could confirm the diagnosis. Heart failure was coded when the radiologists confirmed the clinical diagnosis on chest X-rays. Ventricular fibrillation was coded with an ECG-confirmed diagnosis. Primary cardiac arrest and complete heart block did not occur.

The secondary end points for this study were duration of hospitalization, medical costs, and short-term mortality (defined as death within 6 months after primary tumor diagnosis).

Statistical Analysis

The Pearson chi-square test was used for univariate impact analysis of the categorical variables: age >70 years, LCRI, ASA class, and overall ACE-27 grade, on major cardiovascular complication occurrence. A threshold of $p < .05$ in univariate analysis was set for entering a multivariate analysis.

Multivariate analyses were performed with bivariate logistic regression (backward and forward). Only values of $p < .05$ were considered to be statistically significant.

Performances of the LCRI risk model and the ACE-27 risk model were determined by calculating the area under the receiver operating characteristics (ROC) curve (AUC), which indicates how well a model orders patients with respect to their outcome (where 0.5 indicates no predictive value and 1.0 indicates perfect performance).

All calculations and descriptive statistics were performed in SPSS for Windows (version 16.0; Chicago, IL).

Results

Baseline Population Characteristics

Between July 2001 and December 2007, 135 consecutive patients had major oncologic and reconstructive surgery because of a primary squamous cell carcinoma of the oropharynx or oral cavity. The study population consisted of 80 men and 55 women. The age of the patients ranged from 24 to 83 years, with a median age of 59 years. Approximately 92% of the tumors were located in the oral cavity. Of all tumors, 89.6% were in stage III or stage IV. Surgery consisted of a combined mandibular approach, followed by tumor resection, neck dissection, tracheotomy, and reconstruction with a free vascularised tissue flap. The most frequently used donor site for reconstruction was the radial forearm (54.1%), followed by the fibula (28.1%). Most patients (88.9%) received postoperative radiotherapy. The average duration of the surgical procedure was 11.3 hours (anesthesiologic preparation included) with a minimum of 6.0 hours and a maximum of 15.4 hours. Further demographic data and tumor characteristics are shown in Table 1.

Table 1. Baseline characteristics of the study population (n = 135).

Characteristics	No. (%)	Characteristics	No. (%)
Year of diagnosis		M-stage	
2001	8 [5.9]	M0	135 [100.0%]
2002	22 [16.3]	M1	0 [0.0%]
2003	27 [20.0]	Sex	
2004	17 [12.6]	Male	80 [59.3%]
2005	23 [17.0]	Female	55 [40.7%]
2006	19 [14.1]	Age at diagnosis	
2007	19 [14.1]	< 50 years	24 [17.8%]
Tumor location		50-59 years	48 [35.6%]
Oral cavity	124 [91.9]	60-69 years	41 [30.4%]
Oropharynx	11 [8.1]	≥ 70 years	22 [16.3%]
T-classification		Treatment modalities	
T1	2 [1.5]	Surgery only	15 [11.1%]
T2	35 [25.9]	Surgery followed by RTx	117 [86.7%]
T3	43 [31.9]	Surgery followed by CRTx	3 [2.2%]
T4	55 [40.7]	Donor sites	
N-classification		Free radial fore arm	73 [54.1%]
N0	53 [39.3%]	Anterolateral thigh	21 [15.6%]
N1	27 [20.0%]	Fibula	38 [28.1%]
N2	53 [39.3%]	Intoxications	
N3	2 [1.5%]	Tobbaco	105 [78.8]
		Alcohol	71 [58.5]

Abbreviations: RTx: radiotherapy; CRTx: chemoradiotherapy

Major Cardiovascular Complication Occurrence

Twenty-two patients (16.3%) developed 23 major cardiovascular complications. A total of 4 myocardial infarctions were confirmed, 18 patients had heart failure, and 1 patient had ECG-confirmed ventricular fibrillation.

Lee Index Scores and Major Cardiovascular Complication Prediction

The distribution of Lee Index Scores is shown in Table 2. Eighty-five of 135 patients (43.0%) had a Lee Index Score of II or higher. From the 22 patients that developed a major cardiovascular complication, 16 patients (72.7%) had a Lee Index Score of II or higher (Table 3).

In univariate analyses, the LCRI was associated with an increased risk for major cardiovascular complication development (chi-square = 22.5; $p < .01$). An age >70 years showed a significant association as well (chi-square = 21.9; $p < .01$).

In the multivariate regression analysis, both the LCRI and age >70 years remained significant variables. Relative risks are shown in Table 4. There is a clear distinction between the impact of each Lee Index Score. A patient with a Lee Index Score II has a 1.7-fold higher risk for major cardiovascular complication development than that of a patient with Lee Index score I. Our data indicate that a Lee Index Score above II is associated with an 11- to almost 12-fold higher risk.

In model performance analysis, the AUC for the Lee index alone was 0.73, indicating a good predictive value. Addition of the variable age >70 years to the model showed an improvement of the AUC to 0.84. This indicates the additional prognostic value of the variable age.

Table 2. Distribution of LCRI scores, ASA scores and ACE-27 grades in the study population [N = 135]

LCRI	No. [%]	ASA	No. [%]	ACE-27	No. [%]
Lee I	77 [57.0]	ASA 1	19 [14.1]	Grade 0	41 [30.4]
Lee II	39 [28.9]	ASA 2	71 [52.6]	Grade 1	47 [34.8]
Lee III	15 [11.1]	ASA 3	19 [14.1]	Grade 2	36 [26.7]
Lee IV	4 [3.0]	ASA 4	1 [0.7]	Grade 3	11 [58.1]
		No Data	25 [18.5]		

Abbreviations: LCRI, Lee Cardiac Risk Index; ASA, American Society of Anesthesiologists; ACE-27, Adult Comorbidity Evaluation.

ACE-27 Severity Grades and Major Cardiovascular Complication Prediction

The distribution of overall ACE-27 severity grades is shown in Table 2. This distribution is based on a total of 153 identified preoperative comorbid conditions (Table 5). From the 22 patients who developed a major cardiovascular complication, 14 patients (63.4%) had an ACE-27 grade >1. In univariate analyses the ACE-27 was associated with an increased risk for major cardiovascular complications (chi-square = 18.6; $p < .01$).

In the multivariate regression analysis both the ACE-27 and age >70 years remained significant variables. Relative risks are shown in Table 4. There is a clear distinction between the impact of each ACE-27 severity grade. A patient with grade 1 comorbidity has a 5.7-fold higher risk for major cardiovascular complication development than that of a patient without comorbidity. Both moderate and severe decompensations are associated with at least a 9.5-fold higher risk.

In model performance analysis, the AUC for the ACE-27 alone was 0.75, indicating a good predictive value. Addition of the variable age >70 years to the model showed an improvement of the AUC to 0.84 and is similar to the LCRI model.

ASA Classification and Major Cardiovascular Complication Prediction

The distribution of ASA scores is shown in Table 2. From 135 patients, 110 (81.5%) designated ASA classifications were available for statistical analysis. In the remaining 25 patients, for unknown reasons, no ASA score was reported.

In univariate analysis, the ASA classification did not show a significant relationship between major perioperative cardiovascular complications (chi-square = 3.1, $p = .38$).

Table 3. Major cardiovascular complications and Lee index scores [n = 22]

Major cardiovascular complications	No. [%]	Lee index score	No. [%]
Acute myocardial infarction	4 [3.0]	I: 0 risk variables	1 [25.0]
		II: 1 risk variable	2 [50.0]
		III: 2 risk variables	1 [25.0]
		IV: > 2 risk variables	NA
Heart failure	18 [13.3]	I: 0 risk variables	5 [27.8]
Ventricular fibrillation / cardiac arrest	1 [0.7]	III: 2 risk variables	1 [100.0]
Complete heart block	DNO	NA	NA

Abbreviations: DNO, did not occur; NA, not applicable.

Influence of Preoperative Cardiovascular Medication

In our study population, 24 patients used prescribed beta blockers before surgery. From these 24 patients, 4 patients developed a major cardiovascular complication (16.7%). Twenty-eight patients used prescribed statins before surgery. From these 28 patients, 10 patients developed a major cardiovascular complication (35.7%).

Univariate analysis confirmed the cardioprotective effect of beta-blocker use, whereas statin use showed no significant protective relationship (chi-square = 0.00; p = .99; chi-square = 8.5; p < .01, respectively).

Table 4. Multivariate relationship between variables and major cardiovascular complications

LCRI and age above 70 model	No. [%]	p value	Adjusted Hazard Ratio
Lee index score I	77 [57.0]	<.01	1.0 [RC]
Lee index score II	40 [29.6]		1.7
Lee index score III	14 [10.4]		11.5
Lee index score IV	4 [3.0]		11.8
Age ≤ 70 years	113 [83.7]	<.01	1.0 [RC]
Age > 70 years	22 [16.3]		8.5
ACE-27 and age above 70 model	N [%]	p value	Adjusted Hazard Ratio
ACE-27 Grade 0	41 [30.4]	<.01	1.0 [RC]
ACE-27 Grade 1	47 [34.8]		5.7
ACE-27 Grade 2	36 [26.7]		9.5
ACE-27 Grade 3	11 [8.1]		36.0
Age ≤ 70 years	113 [83.7]	<.01	1.0 [RC]
Age > 70 years	22 [16.3]		7.6

Abbreviations; RC; reference category

Prolongation of Hospitalization and Short-Term Mortality

The mean duration of hospitalization for patients without any perioperative complication (n = 87) was 18 days (range, 10–44 days). Patients with a perioperative major cardiovascular complication (n = 22) had a mean hospitalization of 25 days (range, 13–110 days), which means an average hospitalization prolongation of 7 days. Nine patients had a prolonged stay on the intensive care unit, 1 patient was admitted to a cardiac care unit, and the remaining 11 patients were transferred to the ENT ward where the consulting cardiologist prescribed diuretics and cardiac medication and performed cardiologic follow-up.

It is important to mention that during admission, none of the patients died because of a major cardiovascular complication. Follow-up data showed a short-term mortality rate of 11.1%. The causes of death in these 15 patients were tumor recurrence or metastasis in 8 patients, septic shock in 4 patients, and unknown or multifactorial in 3 patients.

Table 5. Presence and severity of comorbidity conditions in the study population [N = 135]

	Grade 1: mild	Grade 2: moderate	Grade 3: severe
Overall ACE27 score	47 [34.8%]	36 [26.7%]	11 [58.1%]
Specific ACE27 categories			
Cardiovascular system	42 [27.4%]	17 [11.1%]	3 [2.0%]
Respiratory system	10 [6.5%]	5 [3.3%]	0 [0.0%]
Gastro-intestinal system	7 [4.6%]	8 [5.2%]	0 [0.0%]
Renal system	1 [0.7%]	0 [0.0%]	0 [0.0%]
Endocrine system	22 [14.4%]	4 [2.6%]	1 [0.7%]
Neurological system	10 [6.5%]	2 [1.3%]	0 [0.0%]
Psychiatric	3 [2.0%]	1 [0.7%]	0 [0.0%]
Rheumatological system	0 [0.0%]	0 [0.0%]	0 [0.0%]
Immunological system	ND	ND	ND
Prior malignancy	3 [2.0%]	1 [0.7%]	0 [0.0%]
Substance abuse	4 [2.6%]	7 [4.6%]	1 [0.7%]
Body weight	0 [0.0%]	1 [0.7%]	0 [0.0%]

Abbreviations: ND: no data on Immunological disease (AIDS / HIV) due to privacy reasons

Discussion

Patients undergoing non-cardiac surgery can be at high risk of life-threatening cardiac complications [8]. In general, the risk for perioperative complications depends on the condition of the patient before surgery, any comorbidities, and the invasiveness and duration of the surgical procedure. Specifically, cardiac complications can be expected in patients with documented or hidden coronary artery disease, heart failure, or aortic valve disease, and who undergo procedures that are associated with prolonged hemodynamic or cardiac stress. Because of the strong causal relationship between these comorbid ailments and alcohol and tobacco abuse (2 main risk factors for HNSCC as well), major cardiovascular complications can be expected in the patient with HNSCC who undergoes extensive surgery. In this study, this was illustrated by a major cardiovascular complication rate of 16.3%. This percentage is slightly higher than derived incidences from literature, ranging from 7.0% to 13.0%. This could be explained by the heterogeneity of these study populations (less stages III and IV tumor and inclusion of larynx carcinoma) and a different definition of major cardiovascular complications than that formulated by Lee [4,9,10]. The prevention of life-threatening perioperative complications is of medical and economical importance. Preoperative identification of significant risk factors should therefore be a priority. For this purpose, several instruments are available to us. The ASA classification is a widely used anesthesiologic instrument and the ACE-27 has proven to be a valid instrument, with strong association toward complication development and overall survival. Despite the good predictive value of the ACE-27, its specific association with major cardiovascular complications was not investigated in previous studies.

In this study, the predictive value of the LCRI was investigated and compared with the predictive value of the ASA classification and the ACE-27. Both the Lee Index and ACE-27 showed a strong and comparable association between major cardiovascular complication developments, whereas the ASA classification did not. The latter likely reflects the fact that the majority of our patients had medical comorbidities, making them ASA class II or class III. This skewed the data, making it difficult to establish significance. Another explanation can be that different anesthesia providers tend to assign different scores to the same patient, making it a more subjective instrument [11-12]. Furthermore, the term ‘‘systemic’’ can cause confusion. For example, myocardial infarction is a local disease, so a patient with a recent or old myocardial infarction in the absence of any other systemic disease does not truly fit in any category of the ASA classification.

Comparable predictive performances of the ACE-27 and Lee Index were not expected. Although 5 of the 6 Lee Index variables are comorbid conditions that can more or less be coded in the ACE-27 as well, the presence of other comorbid conditions without a relationship toward major cardiovascular complications has contributed to higher overall ACE-27 grades. For example, in 38 patients with a Lee Index Score I (0 risk variables), the ACE-27 was grade 1 or higher. The most frequently observed comorbid conditions in these 38 patients were: hypertension grade 1, chronic obstructive pulmonary disease (COPD) grades 1 and 2, and hepatic disease grade 2. These comorbidity conditions are correlated with alcohol and tobacco abuse just as major cardiovascular complications.

This study was based on historical data without a predefined study protocol for registration of important prognostic variables. Despite the extensive sources available to us, it is therefore possible that some complications were missed or that some comorbid conditions were unrecognized. Another limitation is the absence of routine repeated serum troponin-level testing in our patients. Silent cardiovascular complication occurrence (complications without clinical symptoms) was not identified as such.

Although none of the patients died during admission, the short-term mortality rate for this study population was 11.1%. No cardiovascular causes of death could be determined in 12 patients and the cause of death remained unclear in 3 patients. Despite these findings we believe that risk factors for major cardiovascular complication development can contribute to mortality and require appropriate attention and treatment.

Several treatment strategies have been developed with the aim of safely reducing the occurrence of major cardiovascular complications. Most strategies use drugs, including statins and beta blockers that affect plaque stability or myocardial oxygen balance, or both [13-16]. Beta blockers improve myocardial oxygenation by decreasing heart rate and myocardial contractility, and promote coronary plaque stability by reducing mechanical and shear stresses [17-18]. A recent randomized controlled trial (Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echo [DECREASE IV]) demonstrated that bisoprolol treatment, begun 1 month preoperatively and titrated to heart rate, significantly reduced the incidence of perioperative cardiac death and myocardial infarction, without increasing morbidity or non-cardiac mortality [19]. The POISE (PeriOperative Ischemic Evaluation) trial supports the results of DECREASE and other trials of long-acting agents in reducing perioperative cardiac events, although with an increased incidence of stroke.

As the authors of POISE show, other randomized trials of acute initiating high-dose beta blockers immediately before surgery have also shown an increased stroke rate. However, contrary to these findings, the incidence of perioperative stroke in a low-dose bisoprolol regimen started at least 7 days before surgery in the DECREASE trials was 0.4% of 3994 patients, similar to that with placebo therapy. In our hospital, the patients' preoperative medication is routinely continued during admission (anticoagulants excepted when cardiologic consent is given). To prevent medication discontinuation during admission, patients that are unable to take oral medication receive intravenous or nasogastric tube medication until they are able to switch back. In our retrospective evaluation, no withdrawal or change in type and dose of beta-blockade and statin use was found. The use of beta blockers was associated with less major cardiovascular complications. Based on this finding and the findings in the POISE and DECREASE trials we believe that it is worthwhile to investigate the protective function of bisoprolol in patients with surgically treated head and neck cancer.

Patients that developed a major cardiovascular complication had a mean prolongation of hospitalization of 7 days. In economic terms, prolongation of hospitalization on our ENT ward results in additional uncorrected costs of at least 500 Euros each day. When we would be able to prevent the occurrence of a major cardiovascular complication, this could mean an average cost reduction of at least 3500 Euros per patient. This emphasizes the additional economic benefit of complication reduction next to the medical importance.

Conclusions

Prevention of major complication occurrence after extensive head and neck cancer surgery is of medical and economical importance. Identification of risk factors for major complication development in the preoperative timeframe can lead to a greater awareness in the treating physician and possibly allow preventive measures before surgery takes place. Our results show that the ACE-27 and the Lee Cardiac Risk Index are suitable instruments for preoperative major cardiovascular complication risk assessment. Addition of the variable age >70 years shows an improvement in predictive value of both instruments.

Because of its simplicity and brevity in use, we advise the implementation of the Lee Cardiac Risk Index into preoperative HNSCC screening protocols. We advise the exploration of low-dose, long-acting beta blockers as a preventive treatment strategy.

References

1. Piccirillo JF, Lacy PD, Basu A, Spitznagel L. Development of a new head and neck cancer-specific comorbidity index. Arch Otolaryngol Head Neck Surg 2002; 128: 1172–1179.
2. Piccirillo JF, Costas I. The impact of comorbidity on outcomes. ORL 2004; 66: 180–185.
3. Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL. Prognostic importance of comorbidity in a hospitalbased cancer registry. JAMA 2004; 291: 2441–2447.
4. Ferrier MB, Spuesens EB, Le Cessie S, Baatenburg de Jong RJ. Comorbidity as a major risk factor for mortality and complications in head and neck surgery. Arch Otolaryngol Head Neck Surg 2005; 131: 27–32.
5. Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999; 100: 1043–1049.
6. Boersma E, Kertai MD, Schouten O, et al. Perioperative cardiovascular mortality in noncardiac surgery: validation of the Lee cardiac risk index. Am J Med 2005; 118: 1134–1141.
7. Burgers JS, Simoons ML, Hoes AW, Stehouwer CDA, Stalman WAB. Richtlijn ‘‘cardiovasculair risicomanagement.’’ Ned Tijdschr Geneeskd 2007; 151: 1068–1074.
8. Boersma E, Poldermans D. ß-Blockers in non-cardiac surgery: haemodynamic data needed. Lancet 2008; 372: 1930–1932.
9. Farwell DG, Reilly DF, Weymuller EA, Greenberg DL, Staiger TO, Futaran NA. Predictors of perioperative complications in head and neck patients. Arch Otolaryngol Head Neck Surg 2002; 125: 505–511.
10. Clark JR, McCluskey SA, Hall F, et al. Predictors of morbidity following free flap reconstruction for cancer of the head and neck. Head Neck 2007; 29: 1090–1109.
11. Little JP. Consistency of ASA grading. Anaesthesia 1995; 50; 658–659.
12. Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. Anaesthesia 1995: 50; 195–199.
13. Mangano DT, Layug EL, Wallace A, Tateo I. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. N Engl J Med 1996; 335: 1713–1720.
14. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. N Engl J Med 1999; 341: 1789–1794.
15. POISE Study Group. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomized controlled trial. Lancet 2008; 371: 1839–1847.
16. Schouten O, Poldermans D, Visser L, et al. Fluvastatin and bisoprolol for the reduction of perioperative cardiac mortality and morbidity in high-risk patients undergoing non-cardiac surgery: rationale and design of the DECREASE-IV study. Am Heart J 2004; 148: 1047–1052.
17. Dawood MM, Gupta DK, Southern J, Walia A, Atkinson JB, Eagle KA. Pathology of fatal perioperative myocardial infarction: implications regarding pathophysiology and prevention. Int J Cardiol 1996; 57: 37-44.
18. Warltier DC, Pagel PS, Kersten JR. Approaches to the prevention of perioperative myocardial ischemia. Anesthesiology 2000; 92: 253–259.

19. Dunkelgrun M, Boersma E, Schouten O, et al. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing non-cardiovascular surgery: a randomized controlled trial (DECREASE-IV). Ann Surg 2009; 249: 921–926.

Random Survival Forests versus Cox Proportional Hazards Regression

Datema FR, Moya A, Krause P, Bäck T, Willmes L, Langeveld A, Baatenburg de Jong RJ, Blom HM. Novel head and neck cancer survival analysis approach: Random Survival Forests versus Cox proportional Hazards Regression. Head Neck 2012; 34: 50-8.

Chapter Seven

Abstract

Background Electronic patient files generate an enormous amount of medical data. These data can be used for research, such as prognostic modelling. Automatisation of statistical prognostication processes allows automatic updating of models when new data is gathered. The increase of power behind an automated prognostic model makes its predictive capability more reliable. Cox proportional hazard regression is most frequently used in prognostication. Automatisation of a Cox model is possible, but we expect the updating process to be time-consuming. A possible solution lies in an alternative modeling technique called random survival forests (RSFs). RSF is easily automated and is known to handle the proportionality assumption coherently and automatically. Performance of RSF has not yet been tested on a large head and neck oncological dataset. This study investigates performance of head and neck overall survival of RSF models. Performances are compared to a Cox model as the ‘‘gold standard.’’ RSF might be an interesting alternative modeling approach for automatisation when performances are similar.

Methods RSF models were created in R (Cox also in SPSS). Four RSF splitting rules were used: log-rank, conservation of events, log-rank score, and log-rank approximation. Models were based on historical data of 1371 patients with primary head-and-neck cancer, diagnosed between 1981 and 1998. Models contain 8 covariates: tumor site, T classification, N classification, M classification, age, sex, prior malignancies, and comorbidity. Model performances were determined by Harrell’s concordance error rate, in which 33% of the original data served as a validation sample.

Results RSF and Cox models delivered similar error rates. The Cox model performed slightly better (error rate, 0.2826). The log-rank splitting approach gave the best RSF performance (error rate, 0.2873). In accord with Cox and RSF models, high T classification, high N classification, and severe comorbidity are very important covariates in the model, whereas sex, mild comorbidity, and a supraglottic larynx tumor are less important. A discrepancy arose regarding the importance of M1 classification (see Discussion).

Conclusion Both approaches delivered similar error rates. The Cox model gives a clinically understandable output on covariate impact, whereas RSF becomes more of a ‘‘black box.’’ RSF complements the Cox model by giving more insight and confidence toward relative importance of model covariates. RSF can be recommended as the approach of choice in automating survival analyses.

Introduction

An important question that the newly diagnosed patient with head and neck cancer will most likely ask his or her physician is: ‘‘How are my survival chances?’’ The answer to this question is not that simple because it is based on the impact and interaction of multiple factors. Roughly, these factors can be divided into 3 main categories: tumor-specific, patient-specific, and treatment-specific. Even for the most experienced head and neck surgeon, it remains a difficult task to determine the impact and relevance of each applicable covariate on overall survival.

To aid the physician in this survival prediction dilemma, there are statistical survival analyses. The most popular and broadly used survival analysis is the Cox proportional hazards regression (Cox model). Our most up-to-date survival model is a Cox model containing 8 covariates: tumor location, tumor size (T classification), regional metastasis (N classification), and distant metastasis (M classification), sex, age at diagnosis, prior malignancies, and comorbidity severity (Adult Comorbidity Evaluation-27 [ACE-27]). This Cox model is based on the historical data of 1371 Dutch patients with primary head and neck squamous cell carcinoma and is validated internally by use of a bootstrap procedure. The predictive accuracy of the model is illustrated by a Harrell’s concordance index (C-index) of 0.73 [1].

To further improve individualized prognosis prediction, we need to enhance the performance of our Cox model. We believe that this can be achieved in multiple ways. The fist method is addition of covariates with a significant impact on overall survival. This method was recently illustrated by the addition of comorbidity as a covariate in our Cox model, which resulted in a modest but significant accuracy improvement of 3.0% [1, 2]. The second method is extension of the original database to increase power and to create a more heterogeneous study population. This is a worthwhile but time-consuming effort, especially in a single-center study setting.

Perhaps a solution lies in the introduction and implementation of electronic patient files (EPFs). When an EPF is organized to collect standardized medical data that can automatically be exported into a study database, it can greatly assist in prognostic modeling. First of all, a process of extracting patterns from data, called ‘data mining’ allows extrapolation of relevant covariates hidden in the EPF-generated database in an automated fashion. Second, by automatisation of the survival analysis itself, periodic feedback on model performance can be given after the addition of newly identified covariates and additional patient data.

Automatisation of the generally used Cox model is possible but will require substantial subjective input from the user when it is applied to a dataset in which covariates are highly interrelated. Perhaps other survival analyses that deliver accurate predictive models are more suited for automatisation. In this article, we therefore explore a relatively new survival analysis from data mining, called random survival forests (RSFs). RSF is derived from the Random Forest Modeling technique, introduced by Breiman in 2001 [3]. RSF proved to be useful in the determination of risk factors for disease-free survival of patients with breast cancer and uncovered highly complex interrelationships between covariates in a coronary artery study [4, 5]. Furthermore, RSF is a survival analysis that can easily be automated. To our knowledge, it is the first time that RSF has been applied to a large historical head and neck cancer dataset and is, therefore, considered to be a novel approach.

Study Objectives. The main purpose of this study was to test the predictive performance of RSF models and compare them with the performance of the Cox model as a ‘‘gold standard.’’ When RSF delivers a better or comparable predictive performance, it can be the survival analysis of choice for automatisation. The second purpose of this study was to gain further insights into the relative importance of each model covariate according to each modelling technique.

Materials and Methods

Cox Proportional Hazards Regression (Cox Model)

The Cox model is the most general of regression models because it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. The Cox model can be written as:

$$h(t,\mathbf{x}) = h_0(t) * HR = h_0(t) * \exp^{(X_1\beta_1 + X_2\beta_2 + .. + X_m\beta_m)}$$

Consider the model to consist of 2 parts. First, the baseline hazard $h_0(t)$ describes the hazard (risk of dying) at a specified point in time for a reference patient with all covariate values equal to 0.

Second, effect parameters ($\beta_1, \beta_2, \ldots, \beta_m$) describe how the hazard varies in response to the models' covariates ($X_1, X_2, \ldots, X_m$) and this is expressed as ($X_1\beta_1 + X_2\beta_2, \ldots, X_m\beta_m$). The model can be linearized by dividing both sides of the equation with $h_0(t)$ and then taking the natural logarithm of both sides. Then the model is written as:

$$\log h(t,\mathbf{x}) = \log h_0(t) + X_1\beta_1 + X_2\beta_2 + \ldots + X_m\beta_m$$

Because the functional form of the baseline hazard is not given but determined from the data, the Cox model is called semi parametric. A parametric form for the effect of the covariates on the resultant hazard, however, is assumed as follows:

1. A multiplicative relationship between the underlying hazard function and the log-linear function of the covariates exists. This assumption is called the proportionality assumption. In practical terms, it is assumed that any 2 subjects have hazard functions whose ratio is a constant proportion that depends on the covariates and not on the time.
2. There is a log-linear relationship between the independent covariates and the underlying hazard function.
3. The effect of the covariates is the same at all times.

In this study, the Cox model was created in SPSS for Windows (version 16.0) and in R (version 2.10.0). The 8 covariates for the Cox model were significant in prior univariate analyses. Tumor location, tumor size (T classification), regional metastasis (N classification), distant metastasis (M classification), age at diagnosis, sex, prior tumors, and comorbidity severity (ACE-27 grade) were considered categorical covariates with their reference category set as that covariate with the best prognosis. Age at diagnosis was considered a continuous covariate.

As in a multiple regression, the p value was used as a criterion to decide whether a covariate was statistically significant ($p \leq .01$) or not ($p > .01$). The Z-statistic was used to determine which covariates were most informative. The Z-value marks the ratio of each regression coefficient to its SE within the Cox model. High positive Z-values indicate informative covariates.

Random Survival Forests

RSF is an ensemble tree method for the analysis of right censored survival data. In RSF, randomization is introduced in 2 forms. First, a randomly selected bootstrap sample (approximately 67% of the original data) is used for growing the tree. This bootstrap sample can be seen as the root of the tree. Second, the root is split into 2 daughter nodes by using a splitting rule on a randomly selected covariate. The split is the best when survival difference between the daughter nodes is maximized. Eventually, as the number of tree nodes increases with every split, and dissimilar cases become separated, each node in the tree becomes homogeneous and is populated by cases with similar survival. In the RSF algorithm implemented in this paper, the tree reaches a saturation point when a terminal node (the most extreme node in a saturated tree) has at least 1 death with unique survival times. By grouping hazard estimates from terminal nodes, a cumulative hazard function for the tree can be calculated. In this study, 1000 trees were grown in a first trial to generate the ensemble cumulative hazard. Results were later compared to a replication process with 100 trees to reduce computation speed. The ensemble cumulative hazard function is produced by the average over all trees (Figure 1) [5, 6]. Version 3.6.0 of the RSF R-package provides 4 splitting rules to choose from [6]. (1) The log-rank splitting rule splits the nodes by maximization of the log-rank test statistic. The larger the value, the greater the difference between the 2 groups and the better the split is [7, 8]. (2) The conservation-of-events splitting rule splits the nodes by finding daughter nodes closest to the conservation-of-events principle. This principle states that the sum of the estimated cumulative hazard function over the observed time points must equal the total number of deaths. The value is small if the 2 groups are well separated (Naftel, Blackstone, and Turner, unpublished notes). (3) The log-rank score splitting rule, which

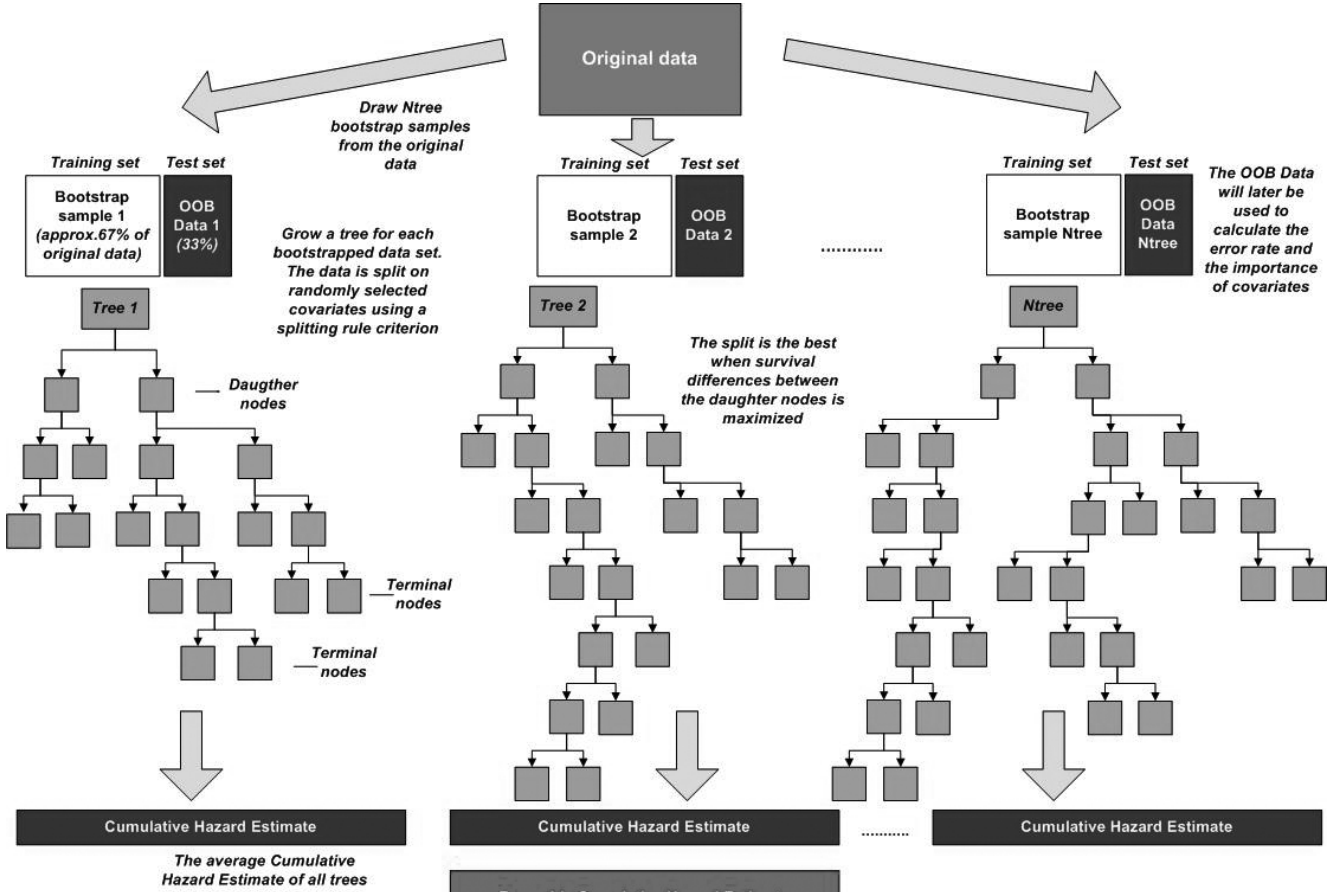
splits the nodes using a standardized log-rank statistic [9]. (4) The log-rank approximationsplitting rule. It splits the nodes by using an approximation of the log-rank test to reduce computations [8, 10].

In this study, RSF models were created with the 4 mentioned splitting rules. Analyses were performed with the 8 covariates simultaneously. The same reference categories were used as in the Cox model.

The importance of each model covariate was determined by applying the model on the data that was not used in building the model. These data (the remaining 33% of the original data) are called out-of-bag data (OOB-data). High positive importance values indicate informative covariates, whereas small positive or negative importance values indicate non-informative covariates [5, 6].

Model Performance

The quality of a predictive survival model is reflected by its prediction error. Harrell's concordance index (C-index) can be used to determine the model's quality. The C-index is related to the area under the receiver operating characteristic curve. It estimates the probability that, in a randomly selected pair of cases, the case that fails (dies) first had the worst predicted outcome. The Harrell's concordance error rate is computed as 1 minus the C-index. Error rates are between 0 and 1, with 0.5 corresponding to a procedure doing no better than random guessing. A value of 0.0 reflects perfect accuracy [11].
For the Cox model and the RSF models, the mean and SD of Harrell's concordance error rate was calculated by using 100 and 10 independent replications. Models were fit on their bootstrap data and the prediction error was estimated using the corresponding OOB-data. In RSF models, the number of trees was equal to 1000 in the first test and 100 in the second.


Original data
Draw Ntree bootstrap samples from the original data
Training set
Test set
Bootstrap sample 1 (approx.67% of original data)
OOB Data 1 (33%)
Grow a tree for each bootstrapped data set. The data is split on randomly selected covariates using a splitting rule criterion
Tree 1
Daugther nodes
Terminal nodes
Terminal nodes
Training set
Test set
Bootstrap sample 2
OOB Data 2
Tree 2
The split is the best when survival differences between the daughter nodes is maximized
Training set
Test set
Bootstrap sample Ntree
OOB Data Ntree
The OOB Data will later be used to calculate the error rate and the importance of covariates
Ntree
Cumulative Hazard Estimate
Cumulative Hazard Estimate
Cumulative Hazard Estimate
The average Cumulative Hazard Estimate of all trees

Head and Neck Cancer Data

For this study, the data of 1662 patients diagnosed with primary head and neck squamous cell carcinoma at the Leiden University Medical Center between 1981 and 1998 were available. Patients with esophageal cancer and subglottic cancer were not included because the prognosis of these patients is poor and the number of incomplete TNM classifications in this group was relatively large. Patients with carcinoma in situ were not included because the prognosis of these patients is exceptionally good. One patient was excluded because there was no follow-up data. The final study population consisted of 1371 patients with histologically proven primary squamous cell carcinoma of the lip, oral cavity, oropharynx, nasopharynx, hypopharynx, glottic larynx, or supraglottic larynx.

The TNM classification was obtained from the hospital-based oncological documentation. Oncological documentation was established in 1969 and contains patient, treatment, and follow-up data for each patient with cancer diagnosed in the Leiden University Medical Center. Trained oncological data managers store these data and safeguard an up-to-date TNM classification by staging or restaging the disease according to the most up-to-date Union Internationale Contre le Cancer manual (for this study, the fifth edition was most up-to-date at the last moment of follow-up).

Prior malignancies were coded in the presence of all preceding malignant tumors except for basal cell and squamous cell carcinoma of the skin.

Comorbidity severity was coded according to the ACE-27 manual. An overall ACE-27 grade 0 corresponds with no comorbidity, grade 1 with mild comorbidity, grade 2 with moderate comorbidity, and grade 3 with severe comorbidity [11]. We chose not to score prior malignancies as ACE-27 comorbidity because this factor is a separate covariate in the Cox model. The impact of prior malignancies on overall survival would be unjust when scored twice. In this article, we therefore use the term ‘‘adjusted’’ ACE-27.

Based on the therapeutic nil hypothesis, the type of treatment was not considered a prognostic factor for our model [12–15].

Results

Explorative Data Analyses

The mean age was 62.6 years old with an SD of 12.0 years. The majority of patients were men (1088; 79.4%). A minority of patients had a preceding malignancy outside the head and neck area (133; 9.7%). Comorbidity was found in 500 patients (36.5%) of whom 76 (5.5%) had severe comorbidity. Most tumors were located in the oral cavity (280; 20.4%) and glottic larynx (442; 32.2%). Further distribution of tumor locations and the TNM classification can be found in Table 1.

The mean follow-up time was 12.3 years and the median follow-up time was 5.3 years. During follow-up, 1048 patients (76.4%) eventually died. The remaining 323 patients (23.6%) were alive at last follow- up.

Table 1. Demographic data and tumor data of baseline study population (n = 1371).

Characteristics	No. (%)	Characteristics	No. (%)
Tumor location		M-classification	
Lip	123 (9.0%)	M0	1354 (98.8%)
Oral cavity	280 (20.4%)	M1	17 (1.2%)
Oropharynx	152 (11.1%)	Sex	
Nasopharynx	41 (3.0%)	Male	1088 (79.4%)
Hypopharynx	137 (10.0%)	Female	283 (20.6%)
Larynx-glottic	442 (32.2%)	Prior malignancies	
Larynx-supraglottic	196 (14.3%)	Yes	133 (9.7%)
T-classification		No	1238 (90.3%)
T1	516 (37.6%)	Comorbidity (adjusted) ACE27	
T2	369 (26.9%)	Grade 0: none	782 (57.0%)
T3	208 (15.2%)	Grade 1: mild	239 (17.5%)
T4	278 (20.3%)	Grade 2: moderate	185 (13.5%)
N-classification		Grade 3: severe	76 (5.5%)
N0	964 (70.3%)	No Data	89 (6.5%)
N1	145 (10.6%)		
N2	180 (13.1%)		
N3	82 (6.0%)		

Abbreviations: ACE27: Adult Comorbidity Evaluation 27

Cox Proportional Hazards Regression Analysis

Based on the given p values, the covariate sex does not impact overall survival. Furthermore, based on the used reference categories, no significant impact on overall survival was found from mild comorbidity and from tumors located in the nasopharynx, glottic larynx and supraglottic larynx (Table 2).

For the remaining significant covariates, Exp (B) can be interpreted as a multiplicative effect on the hazard of death. For example, holding all covariates constant, an additional year of age increases the monthly hazard of death by a factor of 4% (exp(B) = 1.04). Similarly, a patient with moderate comorbidity has an increased hazard of 39% (exp(B) = 1.39) compared to its reference category [RC].

Table 2. Results of the Cox Model

Predictor		B	Exp (B)	Z-value	p value	95% CI
Age		0.038	1.039	11.804	< 0.01	1.032 - 1.045
Sex	Female [RC]	--	1.000	--	--	--
	Male	0.066	1.068	0.789	0.430	0.907 - 1.258*
Tumor location	Lip [RC]	--	1.000	--	--	--
	Hypopharynx	0.628	1.874	3.417	< 0.01	1.307 - 2.686
	Oral Cavity	0.471	1.602	2.836	< 0.01	1.157 - 2.218
	Oropharynx	0.495	1.641	2.757	< 0.01	1.154 - 2.334
	Glottic larynx	0.020	1.020	0.128	0.855	0.907 - 1.258*
	Supraglottic larynx	0.266	1.305	1.555	0.125	0.933 - 1.825*
	Nasopharynx	0.211	1.235	0.801	0.425	0.737 - 2.068*
T-classification	T1 [RC]	--	1.000	--	--	--
	T2	0.289	1.335	3.204	< 0.01	1.119 - 1.592
	T3	0.437	1.549	4.022	< 0.01	1.251 - 1.916
	T4	0.708	2.029	6.820	< 0.01	1.656 - 2.487
N-classification	N0 [RC]	--	1.000	--	--	--
	N1	0.311	1.365	2.878	< 0.01	1.104 - 1.686
	N2	0.611	1.843	5.833	< 0.01	1.500 - 2.263
	N3	0.873	2.395	6.450	< 0.01	1.837 - 3.122
M-classification	M0 [RC]	--	1.000	--	--	--
	M1	1.903	6.707	6.645	< 0.01	3.826 - 11.759
Prior tumors		0.548	1.730	5.097	< 0.01	1.401 - 2.136
Comorbidity	ACE27 Grade 0 [RC]	--	1.000	--	--	--
	ACE27 Grade 1	0.057	1.059	0.648	0.524	0.891 – 1.259*
	ACE27 Grade 2	0.332	1.394	3.624	< 0.01	1.165 – 1.669
	ACE27 Grade 3	0.827	2.286	6.409	< 0.01	1.775 – 2.943

*Abbreviations: RC: reference category; Exp(B): multiplicative factor on the hazard; Z-value: ratio of regression coefficient to its standard error; 95% CI: 95% confidence interval, *: Not statistically significant.*

Random Survival Forrest Analyses

Figure 2, Figure 3, Figure 4, and Figure 5 present a graphical output of how each RSF model ranks its covariates by level of OOB-importance, based on 1000 trees. Each RSF approach shows a slightly different ranking order. However, similarities are present. Very important covariates in nearly all 4 RSF approaches were: age at diagnosis, T4 classification, N3 classification, and N2 classification. Unimportant covariates in nearly all 4 RSF approaches were: sex, mild comorbidity (grade 1), tumor location, supraglottic larynx, and nasopharynx. The remaining covariates had smaller positive importance values with somewhat different ranking within each RSF approach.

These findings were generally comparable with the results from the Cox model, with some exceptions. M1 classification is a very important and significant covariate in the Cox model, whereas in the RSF model only the log-rank splitting rule was able to confirm this finding. The glottic larynx is an insignificant covariate in the Cox model, whereas the conservation of events splitting rule ranked it as 1 of the most important covariates. The conservation of events splitting rule and log-rank score splitting rule ranked moderate comorbidity (ACE-27 grade 2) as unimportant, whereas the Cox model identified it as important.

Figure 2 Out-of-bag data (OOB-data) covariate importance values according to log-rank splitting rule

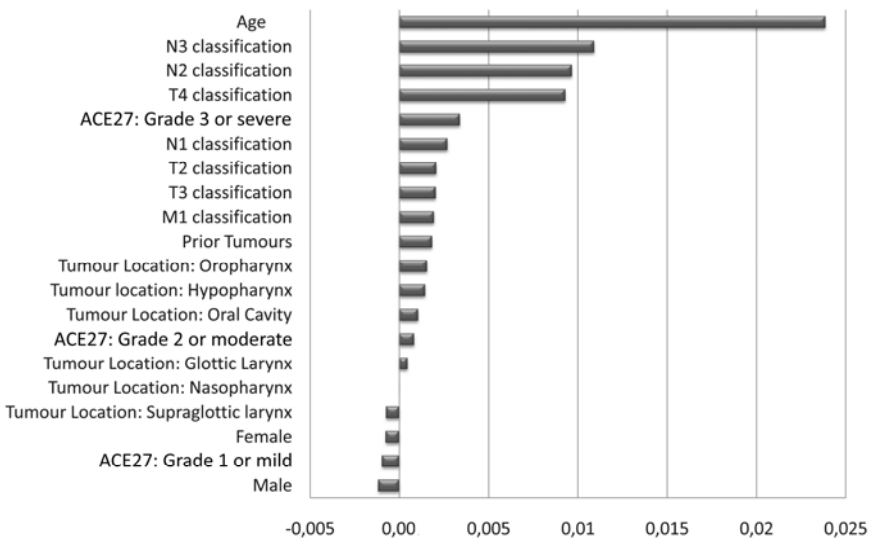


Figure 3 Out-of-bag data (OOB-data) covariate importance values according to conservation of events splitting rule

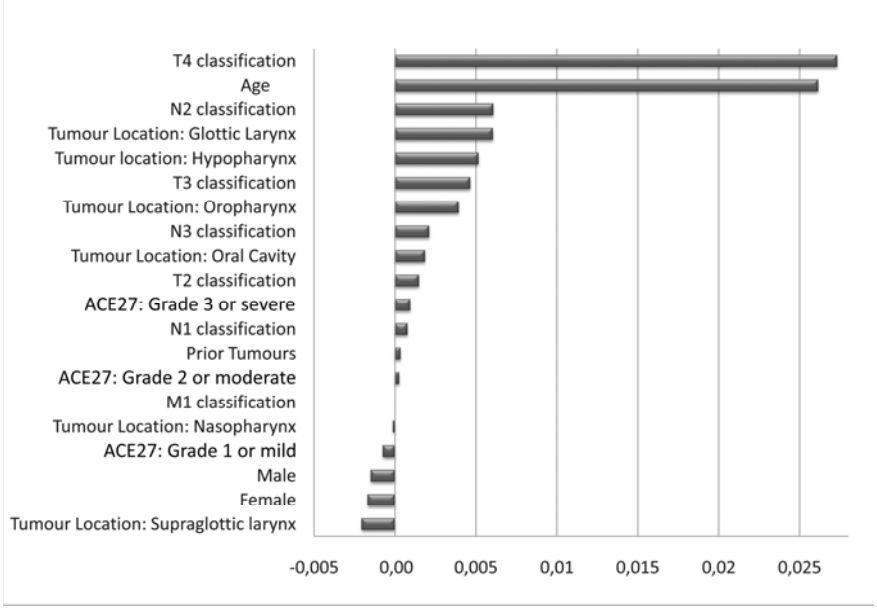


Figure 4 Out-of-bag data (OOB-data) covariate importance values according to log-rank score splitting rule.

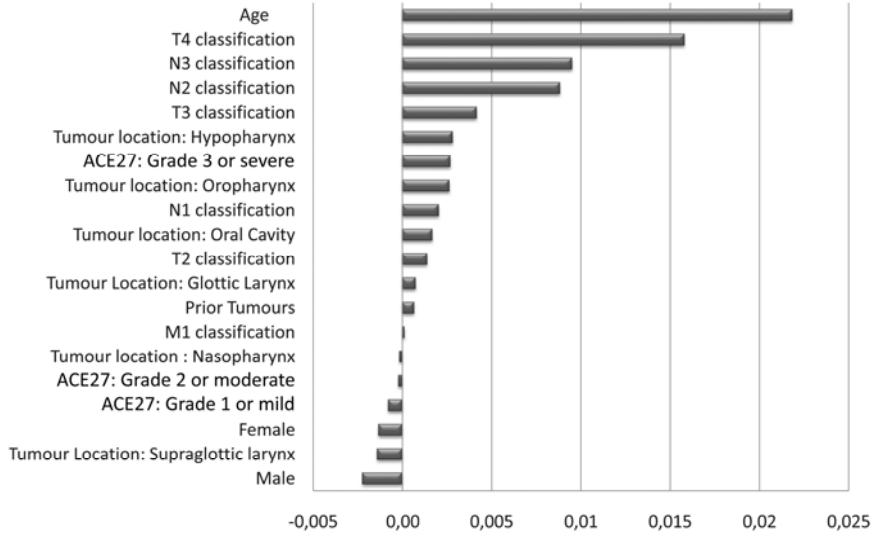
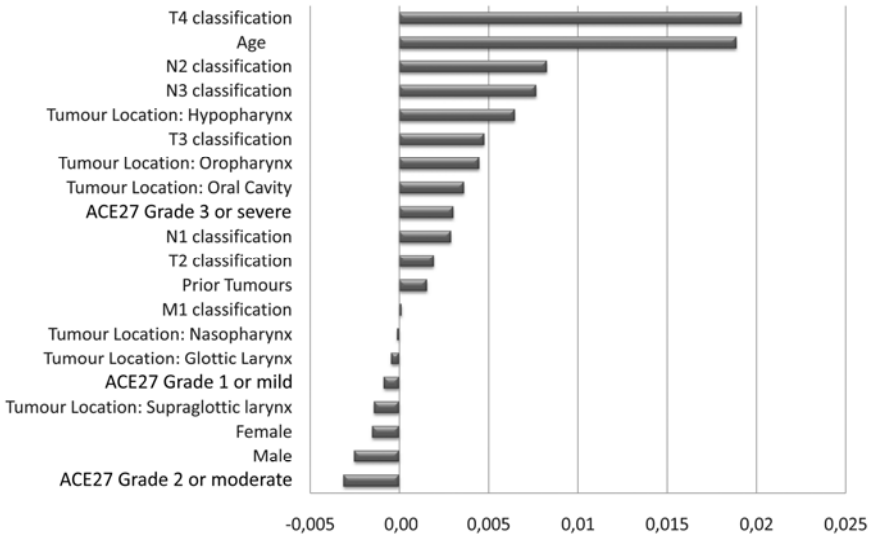


Figure 5 Out-of-bag data (OOB-data) covariate importance values according to log-rank approximation splitting rule.



Cox Model Performance versus Random Survival Forrest Model Performance

The Cox model showed a slightly better performance than all 4 RSF approaches with a Harrell's Concordance error rate (1 minus C-index) of 0.2826. The performance of each RSF approach was very similar with the best error rate (0.2873) obtained by the log-rank splitting rule with 1000 trees (Table 3). To reduce computation speed, a second RSF test with 100 trees was performed. We observed a slight increase in error rates (Table 3).

Table 3. Harrell's concordance error rate (1 minus C-index)

		Repetitions = 100		Repetitions = 10	
		Mean (ER)	SD (ER)	Mean (ER)	SD (ER)
Cox Model		0.2826	0.0120	0.2788	0.0114
		nTrees = 1000		nTrees = 100	
RSF	Log-Rank	0.2873	0.0005	0.2889	0.0014
	Conservation of Events	0.3079	0.0006	0.3099	0.0015
	Log-Rank score	0.2951	0.0007	0.2986	0.0020
	Log-Rank approx	0.2958	0.0006	0.2996	0.0013

Abbreviations: ER: error rate; nTrees: number of trees; SD: standard deviation; RSF: random survival forest

Discussion

Although at first glance RSF seems to be an unusual procedure to evaluate right-censored survival data, considerable empirical evidence has shown randomized ensembles to be highly effective. Because standard survival analyses often rely on restrictive assumption, such as proportional hazards, there is always the concern whether associations between covariates and hazards have been modeled appropriately. RSF is known to handle this problem coherently and automatically [3, 5]. Despite this advantage, RSF was not able to deliver a model with a substantially better C-index than the Cox model in this study. This could be the result of the size and content of our historical dataset. The covariates for these 1371 consecutive patients were collected with a strong assumption that they all impact overall survival based on prior univariate regression analyses. Perhaps RSF performs better as a data-mining instrument on a large dataset that does not have covariates with a predetermined impact on the model output variable. We will test this hypothesis when our EPF is able to generate and export large head and neck oncological datasets with a substantial amount of covariates. Nevertheless, the results of this study confirm what is generally found: RSF produced accurate ensemble models with performances comparable to the Cox model.

In this study, the log-rank splitting approach performed better than the other 3 RSF approaches. Users are, however, encouraged to try all 4 RSF approaches. The log-rank approximate approach has the fastest computation, followed by the conservation of events rule. The log-rank approximate approach, therefore, might be the preferred method in settings in which computational speed is essential. Comparing RSF to the Cox model, it can be said that a Cox model is more capable of extracting patterns and relationships hidden deep into medical datasets, whereas in the RSF model the ensemble tree methods become more of a ‘‘black box’’ when interpreting the model due to the sheer number of trees generated.

A predictive model not only has a certain performance, but also indicates the relative importance of its covariates. In the Cox model, we used the ratio of the regression coefficient to its SE (Z-value) to rank covariates. In RSF, the OOB-data importance values were used to rank covariates. From both methods, we learned that age at diagnosis, a high T classification, a high N classification, and severe comorbidity (ACE-27 grade 3) are very informative covariates. The covariates, sex, mild comorbidity (ACE-27 grade 1), tumor location nasopharynx, glottic larynx, and supraglottic larynx were less informative. For the remaining covariates, it can be said that they have intermediate to low importance values.

Discrepant findings between methods were present. Where M1 classification is a very informative covariate in the Cox model, it was an unimportant covariate in 3 of 4 RSF approaches. Every clinician will agree that presence of distant metastasis is correlated with a very poor prognosis. It is important to realize that when a covariate is insufficiently represented in the original data (M1 classification, n = 17; 1.2%), it is possible that some RSF approaches are unable to identify it as an important covariate, especially because the level of importance is computed on the OOB-data sample, which is approximately 33% of the original data. In this study, the log-rank approach, which delivered the lowest RSF error rate, did identify M1 classification as an important covariate. A second important finding was that despite comparable error rates, the relative importance ranking in the 4 RSF approaches was somewhat different, especially for the different tumor locations. This might make someone question the true level of importance for these model covariates.

Conclusions

The Cox model delivered the best performance of all approaches. The performance of the 4 RSF approaches were, however, almost similar. Therefore, both methods can be recommended in right-censored head and neck cancer survival analyses.
There are some considerations when using RSF. First, a Cox model clearly represents the impact of each covariate on overall survival and its relationship with other covariates. RSF does not give hazard ratios and p values and perhaps becomes more of a ‘‘black box’’ when interpreting the model due to the sheer number of trees generated. Second, the relative importance ranking of model covariates slightly differed within each RSF approach. This might make someone question the true level of importance of the respective model covariates. Third, when a covariate is insufficiently represented in the original data, RSF can miss it as an important covariate. The advantage of RSF is that it is much better suited for automatization of survival analysis than a Cox model because it requires less input from the user in data settings where covariates are highly interrelated. We will consider RSF a suited explorative ‘‘data mining’’ tool for large (future EPF generated) datasets with covariates that have an as yet undetermined effect on the model output variable. Based on the results of these analyses, RSF can assist in building Cox models with highly significant covariates with a higher level of confidence than with a single method.

Acknowledgment

We thank E. W. Steyerberg, Department of Public Health, Erasmus Medical Center, for his input and expertise on prognostic modeling.

References

1. Datema FR, Ferrier MB, van der Schroeff MP, Baatenburg de Jong RJ. Impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. Head Neck 2010; 32: 728–736.
2. 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.
3. Breiman L. Random forests. Machine Learning 2001; 45: 5–32.
4. Imran KO, Mevlut T, Fu¨sun T. The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer. Expert Syst Appl 2009; 36: 8582–8588.
5. Ishawaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat 2008; 2: 841–860.
6. Ishawaran H, KogalurUB. Random Survival Forest 3.0.1. R-package 2007 (online available at: http://cran.r-project.org).
7. Sinisi SE, van der Laan MJ. Regression trees for censored data. Biometrics 1988; 44: 35–47.
8. Leblanc M, Crowley J. Survival trees by goodness of split. J Amer Stat Assoc 1993; 88: 457–467.
9. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Comput Stat Data Anal 2003; 43: 121–137.
10. Cox DR, Oakes D. Analysis of Survival Data. Chapman and Hall; 1988.
11. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982; 247: 2543–2546.
12. Piccirillo JF, Costas I. The impact of comorbidity on outcomes. ORL J Otorhinolaryngol Relat Spec 2004; 66: 180–185.
13. Feinstein AR. Clinical biostatistics. XIV. The purposes of prognostic stratification. Clin Pharmacol Ther 1972; 13: 285–297.
14. Feinstein AR. Clinical biostatistics. XV. The purposes of prognostic stratification. I. Clin Pharmacol Ther 1972; 13: 442–457.
15. Feinstein AR. Clinical biostatistics. XVI. The process of prognostic stratification. 2. Clin Pharmacol Ther 1972; 13: 609–624.
16. Feinstein AR. Clinical biostatistics. XVII. Synchronous partition and bivariate evaluation in predictive stratification. Clin Pharmacol Ther 1972; 13: 755–768.

General discussion

Chapter Eight

In this chapter, the most important findings of research from this thesis are summarized and discussed.

Why we started this research

Since the 1950s, the TNM-classification was the instrument of choice to estimate the likely prognosis of the newly diagnosed cancer patient. The TNM-classification incorporates the size and extension of the primary tumor (T-classification), its regional lymphatic involvement (N-classification) and the presence of distant metastasis (M-classification) to stage the progression of cancer. In several decades the TNM-classification evolved into an excellent descriptive instrument. As a predictive instrument the TNM-classification showed limitations (chapter 1).

Nowadays we acknowledge that the anatomical extend of cancer alone is not the most accurate method to estimate the survival probability of the individual cancer patient. Other tumor and patient specific factors impact overall survival as well and need to be taken into account.

For a clinician it can be a difficult task to recognize all prognostically relevant factors during practice, let alone to determine the impact and interaction of these factors on prognosis. To aid the clinician there are statistical survival analyses that allow the formation of an insightful prognostic model. A prognostic model provides information about the significance, strength and independence of each covariable on prognosis and can calculate an individual overall survival estimate. We believe that prognostic models can complement medical craftsmanship in communicating prognosis to the patient.

In 2001 Baatenburg de Jong presented the first version of his head and neck squamous cell carcinoma (HNSCC) model. The model contained *seven predictors* with an independent, significant and clinically important impact on *overall survival.* These predictors were: age at diagnosis, T-, N- and M-classification, prior tumors, tumor location and sex. The model is based on the historical data of 1371 primary HNSCC patients who were diagnosed and treated in the Leiden University Medical Centre (LUMC) between 1981 and 1998. The model was internally validated.

His model formed the basis for a greater sense of awareness in both the LUMC and Erasmus Medical Centre (EMC) towards the prognostic value of patient and tumor specific factors.

Comorbidity as the eighth predictor in our HNSCC prognostic model

It is our continuous goal to improve the predictive performance of our HNSCC prognostic model. In 2005 we therefore decided to test the prognostic value of comorbidity as a possible *eighth predictor* (chapter 2). At that time, several validated instruments were available to code and quantify the severity of comorbidity in oncological patients. Based on its validity and brevity in use, we decided to use the modified Kaplan Feinstein Comorbidity Index, better known as the Adult Comorbidity Evaluation 27 (ACE27).

We found that 36.4% of our 1371 primary HNSCC patient had comorbidity. Distributed in comorbidity severity, mild comorbidity was present in 17.4%, moderate comorbidity in 13.5% and severe comorbidity in 5.5% of patients.

An interesting finding was that most comorbidity came from the cardiovascular (32.0%), gastro-intestinal (7.2%) and respiratory (6.1%) system. This can be explained by the strong causal relationship between tobacco and alcohol abuse and HNSCC and coexistent diseases from these organ systems.

The ACE27 proved a significant predictor for overall survival in univariate and multivariate regression analysis. There was a clear distinction between the hazard ratio´s of the four ACE27 severity grades. The impact on *overall survival* of a patient with severe comorbidity (ACE27 grade 3) is 2.23 times higher compared to a patient without comorbidity (ACE27 grade 0). This is comparable to the impact of a T4 tumor or a N3 neck. A hazard ratio of 1.38 was found for moderate comorbidity (ACE27 grade 2).

Mild comorbidity (ACE27 grade 1) showed little impact on overall survival with a hazard ratio of 1.04. This is probably explained by the fact that mild comorbidity is usually a historical medical event without residual damage or complications. These mild conditions, when necessary, can be treated with therapeutic or prophylactic medication and then form little mortality risk to the patient. For prognostic purposes it would be interesting to test the performance of a simplified ACE27. By excluding mild comorbidity from the ACE27, it will become an easier and faster to use instrument for daily practice. A disadvantage will be that the instrument loses its optimal descriptive capability.

The *HNSCC prognostic model* was updated with *comorbidity* as the *8th predictor*. *Internal validation* showed a good discriminative capability of the model (C-index 0.73).

Comorbidity has an independent impact on *short-term mortality* (death of all causes within six months after primary tumor diagnosis) as well. From our population, 78 patients (5.7%) died within six months. Especially moderate and severe ailments from the *cardiovascular, respiratory, endocrinological (diabetes) and gastro-intestinal system* were significant in univariate analysis. Significant results from multivariate analyses were not expected considering the low number of events in each subgroup.

The findings from this study generated a greater sense of awareness towards comorbidity in the head and neck squamous cell carcinoma patient and formed a foundation for the implementation of identification and optimization strategies. A good example is that in 2010 a specialist for internal diseases joined the head and neck oncology staff of the EMC. Her main task is to detect and optimize comorbidity before and during hospitalization in an attempt to reduce complication rates and to optimize the overall survival probability of our HNSCC patients.

The importance of external validation

The most stringent test for a prognostic model is to test its predictive performance on a cohort of primary HNSCC patients that were not part of the development data. This is called an *external validation*. An external validation is essential before implementing prediction models in clinical practice.

In 2011 we updated the prognostic model with follow-up data reaching until January 2010 and performed an external validation on 598 primary HNSCC patients from the Siteman Cancer Center/Barnes-Jewish Hospital, St. Louis, Missouri, USA (Chapter 3).

We found that the updated model achieved accurate estimates of survival in the Dutch cohort and fairly good measures of calibration and discrimination (C-index 0.69) in US patients. We believe that our prognostic model can be used for primary HNSCC patients, diagnosed and treated at medical centres in developed countries.

Severe malnutrition is a potential ninth predictor for the HNSCC prognostic model

Cancer cachexia, chewing and swallowing impairments caused by the local tumour or by side effects from oncological treatment can result in malnourishment of head and neck cancer patients. These patients are at risk for adverse events, such as body tissue catabolism and wound healing disorders. Sometimes they cannot tolerate optimal treatment. Adverse events and suboptimal treatment are two risk factors for a decreased overall survival as well. Furthermore, adverse events are associated with a decrease in quality of life, which has a proven negative impact on overall survival as well.

To expand our model with predictors that have an independent, significant and clinically relevant impact on overall survival, we saw potential in malnutrition as a ninth predictor (Chapter 5).

Between 1995 and 1998 a trained dietician joined the oncological head and neck cancer team of the LUMC. For a subset (N = 383) of our 1371 patients, data on weight loss 6 months preceding primary tumor diagnosis were available. For this study we analysed three subcategories: severe malnutrition, defined as a weight loss of > 10%, moderate malnutrition, defined as a weight loss of 5-10% and no malnutrition, defined as a weight loss of less than 5%.

We found 20 (5.2%) patients with moderate malnutrition and 28 (7.3%) patients with severe malnutrition. The majority of severely malnourished patients had a T3 or T4 tumour and 1 patient had distant metastasis.

Malnutrition proved a significant predictor for overall survival in univariate and multivariate analysis. An interesting finding from the multivariate analysis was that against our expectation, patients with moderate malnutrition did not have a significantly elevated mortality risk. However, when we look at the Kaplan Meier curves (observed overall survival), a decreased overall survival rate is seen in patients with moderate malnutrition compared to patients without malnutrition. This starts approximately 24 months after primary tumor diagnosis and continues even 10 years after diagnosis. Since we do not know the exact causes of death of our patients, we cannot properly explain this phenomenon. As expected. the mortality risk of severely malnourished patients was significant and 1.8 times higher than for patients without malnutrition. We therefore consider severe malnutrition a potential ninth predictor for a prognostic HNSCC model. Because data was only available for a subset of patients, the predictor could not be added to our existing model.

Survival visualisation

Prognostic modelling is mainly a statistical exercise. The challenge afterwards lies in the translation of statistical findings into clinical importance and relevance. To aid us in this challenge it was possible to create an on-line, freely accessible version of our model. The on-line software provides a 5-year overall survival chart, based on the applicable model covariables in the individual primary HNSCC patient. With simple clicks in a drop down-menu, model covariables can be changed, resulting in an immediate effect change of the survival chart. It is therefore possible to view the effect of each covariable on overall survival.

A limitation to the on-line software is that *age* could not be incorporated as a continuous variable. Modifications were therefore made to the database. Age was divided into six categories. In a new Cox-regression analysis, the significance and independence of all variables remained. Relative risks of the age categories are shown in chapter 5. The relative risks of the other model covariables hardly changed or remained the same. The on-line model can therefore be considered representative for the results that are presented in chapter 3 of this thesis. Despite this practical approach, we are currently making efforts to adjust the software so that it will be possible to incorporate continuous covariables.

Specific comorbidity elevates the risk for major cardiovascular complications in surgically treated head and neck cancer patients

As mentioned before, alcohol and tobacco abuse are two major risk factors for the development of HNSCC. They show a strong causal relationship with certain comorbid conditions as well. Literature reports that specific comorbid conditions from the cardiovascular, respiratory, renal and endocrinological system are risk factors for an adverse postoperative outcome after *stressful non-cardiac surgery*. We encountered substantial comorbidity from these organ systems in our cohort of 1371 HNSCC patients (chapter 2). Adverse perioperative events are therefore expected in the extensively surgically treated HNSCC population.

In 2010 we started a retrospective study towards the comorbidity of primary HNSCC patients who had extensive oncological and plastic reconstructive surgery (stressful non-cardiac surgery) as the first form of treatment in the Erasmus Medical Centre (Chapter 6). Between 2001 and 2007, 135 patients were diagnosed with a primary oral cavity or oropharynx carcinoma. They were all treated with a combined mandibular approach, tumor and (bilateral) neck dissection, tracheotomy and free flap reconstruction of the surgical defect.

We tested the performance of three risk stratification instruments that mainly depend on comorbid conditions (ASA-classification, ACE27 and Lee Cardiac Risk Index) towards major cardiovascular complication occurrence in the peri-operative time-frame. A major cardiovascular complication was defined as: cardiac death, nonfatal myocardial infarction, heart failure and cardiac arrhythmia.

16.3% of our patients developed a major cardiovascular complication. Fortunately none of these patients died during hospitalization but the average prolongation of hospitalization on the ENT ward, cardiac care unit or intensive care unit was one week. This resulted in an increase of medical costs of at least 3500 Euros per patient hereby stipulating the economical importance in addition to medical importance of complication rate reduction.

The Lee Cardiac Risk Index (LCRI) and ACE27 showed comparably good predictive performances towards major cardiovascular complication occurrence, while the ASA-classification did not. Addition of the variable “age above 70 years” resulted in an accuracy improvement of both instruments (AUC = 0.84, $p <. 01$).

Because of its simplicity and brevity in use, we advise the implementation of the LCRI into preoperative HNSCC screening protocols. The LCRI consists of six items that define an overall Lee index score of I to IV. Five items are comorbid conditions: a history of ischemic heart disease, heart failure, history of cerebrovascular disease, insulin-dependent diabetes, and kidney failure. The sixth risk factor is a high-risk type of surgery.

We found that a LCRI score of III and IV forms a severely elevated risk (relative risk 11.5 and 11.8 respectively) for major cardiovascular complication occurrence when compared to a LCRI score I. Based on these findings, we strongly advise that these patients are referred to a specialist for internal diseases for optimization strategies and observation during hospitalization.

How to proceed with prognostic research

It has been ten years since Baatenburg de Jong et. al. presented the first version of his prognostic model. To improve the performance of his model we continued research. In ten years we accomplished:

1. The identification, testing and addition of comorbidity as an eighth predictor to the model.
2. Update and external validation of the second version of the model.
3. The identification of severe malnutrition as a (potential) ninth predictor for the model.

This illustrates that improvement of prognostic model performance is a time consuming effort which requires substantial (manual) input of the researcher and model developer. Model performance can further improve by expanding the development dataset with more covariables and more patients. There is much (speed) to gain when this process is automated.

With the upcoming use of electronic patient files (EPF) we believe that this goal is reachable. When an EPF is organized to collect standardized medical data that can automatically be exported into a database, it can assist in prognostic modelling. First, a process of extracting patterns from data, called “data mining” allows extrapolation of relevant covariables hidden in the EPF-generated database. Second, by automatisation of the survival analysis itself, periodic feedback on model performance can be given after the addition of newly identified covariables and/or additional patient data.

Automatisation of Cox regression analysis is possible but will require substantial input from the user when it is applied to a dataset in which covariables are highly interrelated. We explored a relatively new modelling technique called Random Survival Forests (RSF) which is known to be easily automated.

In our study we tested the performance of four RSF models and compared them to the performance of our Cox regression model. The models delivered almost similar performance. RSF and Cox regression can therefore both be recommended for survival analyses. There are however some considerations when using RSF.

Where Cox models give a clinically understandable output with a regression coefficient, p-value and 95% confidence interval for each predictor, RSF does not provide this information and perhaps becomes more of a “black box” when interpreting the model.

RSF does give a relative importance ranking of significant predictors. It was insightful to see that in all four RSF approaches, age, severe comorbidity and high T- and N-classifications ranked the highest.

In three out of four approaches M-classification was missed as an important predictor, while M1 classification is a very important predictor in the Cox model. Every clinician will agree that presence of distant metastasis is correlated with a very poor prognosis. We learned that when a predictor is insufficiently represented in the development data (M1-classification, n = 17; 1.2%), RSF can miss it because the level of importance is computed on an out-of-bag sample, which is approximately one third of the original data.

We concluded that at this moment in time, RSF is not suited to replace Cox regression analysis. RSF however can complement Cox regression modelling by identification of highly significant covariates from the data (data mining). This will become handy when we are able to generate large “EPF datasets” with covariables that have an as yet undetermined impact on prognosis.

Lessons learned

Prognostic research is complex, interesting, fun, but most importantly clinically relevant. In the process of model enhancement and model testing several lessons were learned from which future researchers in this field can benefit:

1. The creation of a prognostic model is only possible with an extensive (historical) database that is filled with accurate and complete survival data. Most research, presented in this thesis was only possible because the LUMC invested in the ONCDOC department (since 1969). Trained oncological data managers who safeguard an adequate data registration and who monitor events during follow-up are extremely valuable for prognostic research. It is worthwhile to consider investing in a structural (prospective) data collection and data safeguarding department such as ONCDOC.

2. Electronic patient files generate an enormous amount of medical data with prognostic potential. Information in free text fields can be considered “dead data” because statistical analyses require numeric values. It is advisable to make it possible to code and quantify specific data into an EPF, so that these data can be exported into a database for statistical analysis.

3. When retrospective data is collected, it is advisable to test the intra- and inter-observer variability to test the quality of gathered data.

4. Before a prognostic model can be used in daily practice, an internal and external validation procedure is necessary.

5. Further accuracy improvement of our model is believed possible by including more covariables into the model and expansion of the development study population with more recent patients. The problem with retrospective data collection however is that there is no upfront guarantee that sufficient information will be found in the medical charts of the baseline study population to incorporate new covariables. This also applies when expanding the consecutive development dataset. The inclusion of more recent patients will possibly help to counter the 'out-of date principle' of the model. Patients who were diagnosed and treated in a period with comparable diagnostic and treatment standards as today, will have a more representative survival probability than earlier patients.

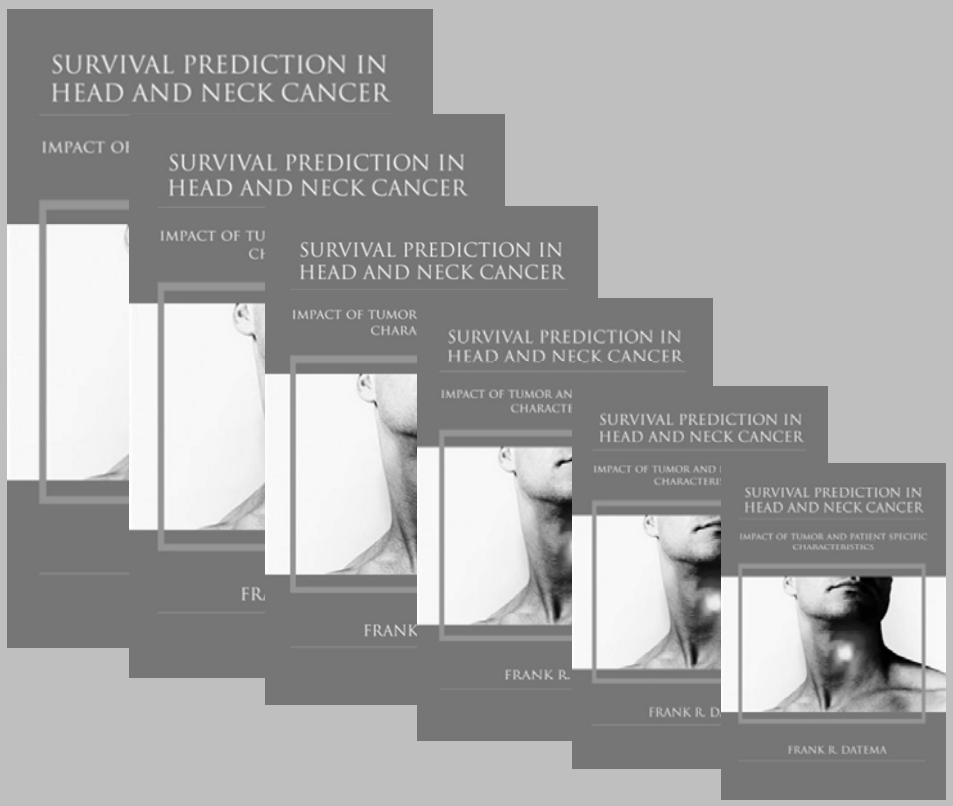
6. Currently, in the Erasmus Medical Centre, strong efforts are made to enhance the quality of prospective oncological data collection which hopefully will lead to the first *EMC prognostic HNSCC model*. For this model it is advisable that:
    a. at least the nine predictors presented in this thesis are part of a data collection form.

    b. members from the head and neck oncology department are trained to validly code all model covariables in this form.

    c. data-collection becomes an ongoing process and part of daily practice. Therefore the use of an EPF form is preferable and head and neck oncologists need to be motivated and aware of the importance of proper data collection.

    d. the collaboration with a statician with sufficient knowledge of survival analyses and model validation is at hand.

Samenvatting / Summary



Chapter Nine

HOOFDSTUK EEN: het leven van de patiënt verandert dramatisch als de diagnose kanker wordt gesteld. Onzekerheid over hoe lang het leven nog mag duren en hoe de kwaliteit van leven zal zijn na oncologische behandeling kunnen een fysieke en emotionele uitdaging vormen. Nauwkeurige informatie over het ziektebeloop en de prognose kan patiënten en hun naasten helpen om beter om te gaan met de ziekte en helpen bij de afweging tussen enerzijds de last van oncologische behandeling en anderzijds de mogelijke toename in levensverwachting en -kwaliteit. Verder kan een individueel behandelplan alleen maar het resultaat zijn van een nauwkeurige inschatting van prognose. Over- en onderschatting van prognose kan resulteren in onder- en overbehandeling.

Omdat het belang van de informatie zo groot is, zal nagenoeg elke oncologiepatiënt zijn behandelaar vragen naar prognose. Vroeger werd de TNM-classificatie gebruikt om hierover een uitspraak te doen, maar tegenwoordig onderkennen we dat alleen tumorspecifieke kenmerken onvoldoende zijn om een individuele prognose te bepalen. Andere (niet-oncologische) factoren, zoals de leeftijd en algemene gezondheid van de patiënt, moeten ook worden meegewogen. Er is veel onderzoek gedaan naar prognostische factoren, waar onder het onderzoek dat is gepresenteerd in dit proefschrift. Het zal voor de behandelaar lastig zijn om met een langer wordende lijst aan prognostische factoren, de juiste prognose te formeren. Gelukkig bestaan er statistische analyses die de significantie, sterkte en onafhankelijkheid van prognostische factoren testen op een grote groep historische patiënten met vergelijkbare tumor en patiëntgebonden karakteristieken. Een combinatie van prognostische factoren noemen we een prognostisch model. Met een prognostisch model kan (een zekere maat van onzekerheid in acht nemend) de individuele prognose van een patiënt berekend worden.

HOOFDSTUK TWEE: In 2001 presenteerden Baatenburg de Jong et. al . een prognostisch model voor primaire hoofd-hals plaveiselcel carcinoom patiënten (HHPCC) bestaande uit zeven prognostische factoren: T-, N- en M-classificatie, tumor locatie, leeftijd en geslacht van de patiënt en eerder doorgemaakte tumoren. Dit model is gefundeerd op de historische data van 1371 patienten die tussen 1981 en 1998 werden gediagnostiseerd en behandeld in het Leids Universitair Medisch Centrum. Dit model werd intern gevalideerd.

Ons blijvende doel is om de nauwkeurigheid van de prognose voorspelling door het model te verbeteren. In dit hoofdstuk beschrijven we dat na univariate en multivariate analyse bleek dat comorbiditeit (gescoord met de ACE27 handleiding) een onafhankelijke invloed heeft op de overleving van HHPCC patiënten. Comorbiditeit bleek aanwezig in 36.4% van de 1371 patiënten en betrof voornamelijk ziekten uit het cardiovasculaire, respiratoire en gastro-intestinale orgaansysteem. Dit kan worden verklaard door de relatie tussen alcohol en tabaksmisbruik (twee belangrijke risicofactoren voor het ontstaan van HHPCC) en aandoendingen uit deze orgaansystemen.

Interessant is dat milde comorbiditeit nauwlijks invloed heeft op prognose, terwijl matige en ernstige comorbiditeit een significant verhoogt mortaliteitsrisico dragen (relatief risico van 1.4 en 2.2). De invloed van ernstige comorbiditeit op overleving is daarmee vergelijkbaar met de invloed van een T4 tumor of een N3 hals. Dit geeft het belang aan van het tijdig herkennen van comorbiditeit en waar mogelijk het optimaliseren ervan. Comorbiditeit werd toegevoegd als 8^{e} voorspeller aan het model dat vervolgens intern gevalideerd werd (C-index 0.73).

HOOFDSTUK DRIE: de meest strenge test voor een prognostisch model is een externe validatie, waarbij de nauwkeurigheid van het model wordt getest op patiënten met vergelijkbare tumor en patiëntgebonden karakteristieken, maar uit een andere geografische locatie en uit een ander diagnostisch tijdperk. Een dataset van 598 primaire HHPCC patiënten, die tussen 1995 en 2000 werden gediagnostiseerd in het Siteman Cancer Centre in Amerika, kwam tot onze beschikking. Ondanks verschillen tussen de Nederlandse en Amerikaanse populatie, werden acceptabel goede calibratie en discriminatie resultaten gevonden (C-index 0.69). We hebben hiermee meer vertrouwen gekregen dat ons model ook goed presteert buiten het LUMC, maar realiseren ons dat er nog steeds ruimte is voor een verbetering van performance.

HOOFDSTUK VIER: patiënt met een ernstige ondervoeding in de periode voor kanker diagnose hebben een verhoogd risico op complicaties en verdragen soms een optimale behandeling niet. Een gecompliceerd ziektebeloop en suboptimale behandeling zijn risicofactoren voor mortaliteit. Ernstige ondervoeding werd daarom gezien als potentiele 9^{e} voorspeller voor ons model. Van een subgroep (383) van de 1371 patiënten was data beschikbaar over gewichtsverlies. Ernstige ondervoeding, gedefinieerd als meer dan 10% gewichtsverlies in de zes maanden voor tumordiagnose, werd in 5.2% aangetroffen en toonde na univariate en multivariate analyse een onafhankelijke invloed op de overleving van HPCC patienten. Het mortaliteitsrisico van deze patiënten was 1.8 keer hoger dan van vergelijkbare patiënten zonder ondervoeding.

Dit benadrukt het belang van tijdige herkenning van ernstige ondervoeding en de implementatie van interventiestrategieën. Omdat slechts van een deel van de ontwikkelpopulatie gewichtsverliesdata beschikbaar waren, kon ernstige ondervoeding helaas niet als 9^{e} variabele aan het model worden toegevoegd.

HOOFDSTUK VIJF: de meeste prognostische modellen in dit proefschrift zijn gebaseerd op Cox-regressie analyses. Tijdens een spreekuur zal het voor de behandelaar lastig blijken om met relatief ingewikkelde formules een individuele prognose te berekenen. Daarom ontwikkelden wij gebruikersvriendelijke software, waarmee de behandelaar door het aanklikken van voorspellende variabelen, direct wordt voorzien van een 1-, 2- en 5-jaars overlevingskans. Ook toont het programma een patiëntspecifieke 5-jaars overlevingscurve. We hopen dat deze software een meerwaarde heeft op het medische vakmansschap wanneer de patiënt over zijn prognose geinformeerd moet worden of wanneer er gekozen moet worden tussen geschikte behandelmodaliteiten. De software is gratis te gebruiken via www.oncologiq.nl

HOOFDSTUK ZES: Zoals werd aangetoond heeft comorbiditeit een onafhankelijke invloed op de overleving van primaire HHPCC patiënten. Bepaalde comorbiditeit verhoogt het risico op een ernstige cardiovasculaire complicatie bij patiënten die stressvolle niet-cardiale chirurgie ondergaan. Een ernstige cardiovasculaire complicatie is gedefinieerd als: acute hartdood, een niet-fataal hartinfarct, hartfalen en cardiale arytmie. Omdat deze complicaties potentieel levensbedreigend zijn is preventie zeer waardevol. Bij 16.3% van 135 primaire oropharynx en/of mondholte plaveiselcel carcinoom patiënten die uitgebreide oncologische en plastisch reconstructieve chirurgie ondergingen, trad een dergelijke complicatie op. Gelukkig overleed niemand gedurende ziekenhuisopname, maar de gemiddelde zorgkosten per patient namen wel toe met 3500 euro. Dit illustreert het economische belang van complicatie reductie, naast het medische belang.

De aangepaste Lee Cardiac Risk Index (LCRI) codeert een voorgeschiedenis met ischaemische hartziekte, hartfalen, cerebrovasculaire incidenten, insuline afhankelijke diabetes en/of nierfalen. De LCRI bleek binnen de hoofd-hals oncologie een significante voorspellende waarde te hebben voor ernstige cardiovasculaire complicaties. Een LCRI score van III of IV bleek geassocieerd met een 11 tot 12 keer hoger risico vergeleken met patiënten zonder deze specifieke comorbiditeit. De voorspellende waarde van de LCRI nam toe na het toevoegen van de variabele “leeftijd boven 70 jaar” (AUC 0.84). Wij raden aan dat deze patiënten worden verwezen naar een cardiovasculair internist voor preoperatieve screening en optimalisatie, bij voorkeur gecontinueerd tijdens en na ziekenhuisopname.

HOOFDSTUK ZEVEN: het ontwikkelen, updaten en uitbreiden van een prognostisch model dat is gefundeerd op een grote historische dataset, is een tijdrovende maar waardevolle inspanning. Wij geloven dat het model nauwkeuriger kan gaan voorspellen als er meer prognostische variabelen worden geincludeerd en/of meer patiënten aan de ontwikkeldatabase worden toegevoegd. Het probleem met retrospectief data-onderzoek is dat er van te voren geen garantie is dat er voldoende gegevens in de medische status aanwezig zijn om een extra variabele aan het model toe te kunnen voegen (hoofdstuk 4). Hetzelfde geldt voor het toevoegen van meer patiënten aan de ontwikkeldata. Dit laatste zou wel deels het `out-of-date principe´ van een model kunnen tegengaan, omdat met het includeren van patiënten die gediagnostiseerd en behandeld werden in een periode met meer gelijkwaardige diagnostische en therapeutische middelen als van vandaag, mogelijk een meer representateive overleving hebben dan patiënten uit een eerder tijdperk.

Het is onze hypothese dat er mogelijkheden voor toekomstig prognostisch onderzoek liggen bij de steeds vaker aanwezige electonische patiënten dossiers (EPD). Gedurende werktijd verzamelen we een enorme hoeveelheid aan waardevolle medische data, dat wordt opgeslagen in een EPD. Vrije tekstvelden zijn initieel zonder waarde voor onderzoek, maar als dossiergegevens gestandaardiseerd kunnen worden met een achterliggende numerieke waarde, kunnen ze worden geexporteerd naar een database waarop statistische analyses, zoals survival analyses, mogelijk zijn. Zeker wanneer data-analyse, model ontwikkeling, model validatie en updating automatisch zou kunnen, is er veel tijd te winnen. We hebben daarom een relatief nieuwe modeleringstechniek getest, genaamd Random Survival Forests (RSF). RSF staat erom bekend dat ze relatief gemakkelijk en automatisch patronen kan herkennen die verborgen liggen in grote survival databases.Er is maar weinig bekend over hoe nauwkeurig RSF modellen voorspellen in vergelijking met een gevalideerd prognostisch Cox regressie model. In onze studie gaven RSF modellen nagenoeg vergelijkbare resultaten als ons Cox regressie model. Een beperking van RSF was dat daar waar Cox regressie een klinisch begrijpelijke output genereert met een p-waarde, relatief risico en 95% betrouwbaarheidsinterval voor elke voorpeller, RSF meer een “black box” wordt. We concludeerden dat op dit moment RSF nog niet geschikt is om Cox regressie te vervangen. Wel kan RSF helpen bij het identiferen van sterk significante prognostische variabelen uit grote (EPD gegenereerde) databases met variabelen die een tot nu toe onbekende invloed hebben op prognose.

We sluiten dit proefschrift af met een aantal belangrijke lessen die geleerd zijn tijdens het ontwikkelen, updaten en valideren van de in dit proefschrift gepresenteerde modellen. Ik hoop dat deze lessen een waarde hebben voor toekomstige onderzoekers die zich bezig gaan houden met het voorspellen van een individuele prognose van de HHPC patiënt.

CHAPTER ONE: When diagnosed with cancer, the patient's life changes dramatically. Uncertainty about future life expectancy, quality of life and (side) effects of the upcoming treatment can form a physical and emotional challenge. Accurate information on what to expect from the course and likely outcome of disease, can help patients and their loved-ones to cope and prepare and to balance the burden of treatment against the possible gain in life expectancy and quality of life. Furthermore, an individualized treatment can only be the result of an accurate prognosis. Over- and underestimation of survival may result in under- and overtreatment.

In every new patient, the clinician is challenged to estimate an accurate prognosis. In the past, the TNM-classification was used for this purpose, but nowadays we acknowledge that tumor specific factors alone are not sufficient to determine an individual survival estimate. Other (non-oncological) factors, such as age or general health status need to be taken into account as well. A lot of research, including the research in this thesis, has been performed to identify and test prognostic variables in head and neck squamous cell carcinoma (HNSCC) patients. For a clinician it can be difficult to derive an accurate prognosis from an expanding list of prognostic variables. Fortunately there are statistical survival analyses that determine the significance, strength and independence of prognostic factors on overall survival from a large group of historical patients with similar tumor and patient specific characteristics. A combination of prognostic factors is called a prognostic model. A prognostic model allows (considering a certain margin of uncertainty) the calculation of an individual prognosis for the newly diagnosed HNSCC patient.

CHAPTER TWO: In 2001 Baatenburg de Jong et. al. presented a prognostic model that estimates the overall survival probability of primary HNSCC patients based on seven prognostic variables: the T-, N- and M-classification, tumor location, the age and sex of the patient and prior malignancies. The model was fitted on the historical data of 1371 primary HNSCC patients, diagnosed and treated in the Leiden University Medical Centre between 1981 and 1998. The model was internally validated.

It is our continuous goal to enhance the predictive performance of the model. In this chapter we describe that after univariate and multivariate analyses, comorbidity (coded according to the ACE27 manual) showed an independent impact on overall survival. Comorbidity was found in 36.4% of the 1371 patients and mostly came from the cardiovascular, respiratory and gastro-intestinal system. This is explained by the causal

relationship of alcohol and tobacco abuse (two major risk factors for the development of HNSCC) and diseases from these organ systems.

Interesting was that mild comorbidity hardly affected prognosis, while moderate and severe comorbidity showed a significant elevated mortality risk (relative risk 1.4 and 2.2 respectively). The impact of severe comorbidity on overall survival is comparable to the impact of a T4 tumor or N3 neck, stating the importance of early comorbidity recognition and when possible, intervention. Comorbidity was added to the model as an 8th predictor and the model was internally validated (C-index of 0.73).

CHAPER THREE: The most stringent test for a model is an external validation that tests the performance of the model on patients from a different population. This population should be similar to the development population in terms of index disease, but different in terms of geographic location and historical time period. A dataset of 598 primary HNSCC patients, diagnosed and treated at the Siteman Cancer Center, USA between 1995 and 2000 was made available to us. Despite differences between the Dutch and American populations, acceptably good calibration and discrimination results were found (C-index 0.69). We now feel more confident about the generelizability of our prognostic model but realize that there is still room for performance improvement.

CHAPTER FOUR: Patients with severe malnutrition prior to cancer diagnosis are at risk for adverse events and sometimes do not tolerate optimal treatment. Adverse events and suboptimal treatment are risk factors for a decreased overall survival. Severe malnutrition was therefore considered a potential 9th predictor for our model. From a subpopulation (383 patients) of the 1371 patients, weight loss data prior to cancer diagnosis were available for analyses. Severe malnutrition, defined as weight loss of more than 10% in the six months preceding cancer diagnosis, was encountered in 5.2% and showed an independent impact on overall survival. The mortality risk of these patients was 1.8 times higher than for patients without malnutrition. This emphasizes the importance of identifying severe malnutrition and implementation of intervention strategies. Because data was only available for a subset of the model development population, weight loss could not be added to the model as a 9th predictor.

CHAPTER FIVE: Most models that are presented in this thesis are based on Cox-regression analyses. During daily practice, it will be difficult to perform extensive statistical calculations to determine the individual prognosis of the patient. We therefore developed user-friendly software, allowing the clinician to simply select applicable predictors, resulting in a 1-, 2- and 5-year survival estimate of the patient. The software provides the corresponding 5-year survival chart as well. We hope that this software complements medical craftsmanship when communicating prognosis to the patient or when the clinician needs to choose from suitable treatment modalities. The software is freely accessible on: www.oncologiq.nl.

CHAPTER SIX: Comorbidity has an independent impact on overall survival. We performed an 'in depth' study to identify specific comorbid conditions in primary oropharynx and oral cavity carcinoma patients, receiving extensive head and neck and plastic reconstructive surgery (stressful surgery). It is known that certain comorbidity elevates the risk of major cardiovascular complications during stressful non-cardiac surgery. These complications are potentially life-threatening and therefore worth preventing. A major cardiovascular complication was defined as: cardiac death, nonfatal myocardial infarction, heart failure and cardiac arrhythmia and encountered in 16.3% of 135 patients diagnosed and treated at the Erasmus Medical Centre between 2001 and 2007. Fortunately none of these patients died during hospitalization but the average increase in medical costs was at least 3500 Euros per patient, stipulating an additional economical importance of complication rate reduction.

The modified Lee Cardiac Risk Index (LCRI) codes a medical history with ischemic heart disease, heart failure, cerebrovascular disease, insulin-dependent diabetes, and kidney failure. The LCRI proved a significant prognostic tool: for example, a LCRI score of III or IV was associated with an 11 to 12 time higher risk for major cardiovascular complication development, compared to a patient without comorbidity. The predictive performance of the LCRI increased by addition of the variable "age above 70 years" (AUC 0.84) . We strongly advise that these patients are referred to a specialist for internal diseases for optimization strategies and consultation during (and when necessary after) hospitalization.

CHAPTER SEVEN: Developing, updating and expanding a prognostic model, fitted on a large historical dataset, is a time-consuming but worthwhile effort. Further accuracy improvement of our model is believed possible by including more covariables and/or expansion of the development study population. The problem with retrospective datasets is that there is no upfront guarantee that sufficient information can be found in the medical charts to incorporate extra covariables into the model (chapter four). The same applies to addition of patients to the development population. The latter could partially counter the 'out-of date principle' of the model because inclusion of patients who were diagnosed and treated in a time period with more similar diagnostic and treatment possibilities as today, perhaps have a more representative survival probability than earlier patients.

It is our hypothesis that possibilities for future prognostic research lie with the upcoming use of electronic patient files (EPF). During practice, we gather an enormous amount of valuable medical data that is recorded in an EPF. Free text fields are initially without value for research, but when medical chart data are standardized with underlying numerical values they can be exported into databases on which statistical analyses such as survival analyses are possible. Especially when data analysis, model development, model testing and model updating can be automated there is much speed to gain. We therefore tested a relatively new survival modelling technique, called Random Survival Forests (RSF). RSF is known to easily extract patterns hidden in large survival databases in an automated fashion. Little is known about the performance of RSF models compared to Cox-regression models. In our study, RSF models delivered a similar performance as our Cox regression model. A limitation of RSF was that where Cox models give a clinically understandable output with a relative risk, p-value and 95% confidence interval for each predictor, RSF becomes more of a "black box" when interpreting the model. RSF did give insightful information on covariable importance ranking. We concluded that at the moment, RSF is not suited to replace Cox regression but that it can complement it by identification of highly significant covariates from large (EPF generated) databases with covariables that have an as yet undetermined impact on prognosis.

We end this thesis with several important lessons that were learned during the development, updating and validation of the presented prognostic HNSCC models. I hope that these lessons are of value to future researchers in the field of prognosis prediction of the individual HNSCC patient.

List of abbreviations

ACE27	Adult Comorbidity Evaluation 27
AIDS	Acquired Immune Deficiency Syndrome
ASA	American Society of Anesthesiologists
AUC	Area Under the Receiver operating characteristics Curve
BMI	Body Mass Index
C-index	Concordance Index
CIRS	Cumulative Illness Rating Scale
COPD	Chronic Obstructive Pulmonary Disease
DECREASE	Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echo
ECG	Echocardiography
EMC	Erasmus Medical Centre
EPF	Electronic Patient File
HNSCC	Head and Neck Squamous Cell Carcinoma
HIV	Human Immuno Sufficiency Virus
HPV	Human Papilloma Virus
ICED	Index of Coexistent Disease
UICC	Union Internationale Contre le Cancer
KFI	Kaplan Feinstein Comorbidity Index
LCRI	(Modified) Lee Cardiac Risk Index
LUMC	Leiden University Medical Centre
NCR	Netherlands Cancer Registry
ONCDOC	LUMC Based Cancer Registry System
POISE	Peri Operative Ischemic Evaluation
PORT	Post Operative Radio Therapy
RSF	Random Survival Forests

Dankwoord

Professor Baatenburg de Jong wil ik danken voor de mogelijkheid om in 2003 onderzoek te gaan doen voor de afdeling hoofd-hals oncologie van het LUMC. Als toenmalig geneeskundestudent kon ik absoluut niet voorzien dat ons onderzoek naar comorbiditeit een basis zou gaan vormen voor dit proefschrift, maar ook de basis voor mijn latere carrière als KNO-assistent. Ik ervaar onze samenwerking altijd als zeer prettig, constructief en uiterst leerzaam.

Doctor Blom, beste Henk. Hartelijk dank voor de mentorrol die jij hebt gevoerd gedurende mijn klinische en wetenschappelijke werkzaamheden in het HagaZiekenhuis. Jouw energie, motivatie en 'big picture' aangaande het verzamelen en verwerken van medische data naar modellen zijn zeer inspirerend geweest. Jouw duizendpoot-mentaliteit en organisatorische talent zijn jaloersmakend.

Dit proefschrift heeft niet kunnen ontstaan zonder Marciano Ferrier. Marciano, bedankt voor jouw hulp met het verwerken van data en het vertalen van onze bevindingen naar de kliniek en het papier. Bewonderenswaardig vind ik dat jij altijd bereikbaar was voor overleg en betrokken hebt willen blijven bij ons onderzoek, ook na afronding van jouw academische KNO-carrière.

Ana Moya en Peter Krausse dank ik voor hun kennis en expertise aangaande Random Survival Forest modelering. Yvonne Vergouwe dank ik hartelijk voor het delen van haar expertise aangaande interne en externe validatie van statistische modellen. Jay Piccirillo verdient niet alleen dank voor het aanleveren van de data voor externe model validatie, maar ook voor zijn ACE-27. Het bewijst wederom een zeer waardevol instrument binnen de hoofd-hals oncologie.

Collega's arts-assistenten, hartelijk dank voor een fantastische opleidingssfeer met collegialiteit, humor en het nodige drankje na het werk. Marc en Q, het was een fijne wetenschap dat wij alle drie oneindig veel oncologie data te verwerken hadden. Het meest productief was ik op momenten dat we gezamenlijk in de assistentenkamer bezig waren met ons onderzoek.

Lieve ouders, schoonouders, zusje, schoonzussen en zwagers. Dank voor jullie interesse in mijn proefschrift.

Professor dr. Steyerberg, Professor dr. van Lanschot en Professor dr. Wolvius dank ik van harte voor hun bereidheid om de wetenschappelijke waarde en leesbaarheid van dit proefschrift te toetsen.

Tot slot, lieve Jes en lieve kleine Veerle. Dank voor jullie eindeloze begrip, steun en geduld de afgelopen jaren. Het is erg fijn als je als echtgenoot en vader zonder schuldgevoel weer een hele avond in de werkkamer kan doorbrengen. Ik zal er echter geen gewoonte van maken.

Curriculum Vitae

Frank Roelf Datema werd op 4 april 1980 geboren in Breda. Hij groeide op in Woerden, waar hij in 1998 slaagde voor het examen Atheneum aan het Minkema College. In 2005 behaalde hij het arts-examen aan de Universiteit van Leiden. In 2003 maakte hij in het Leids Universitair Medisch Centrum voor het eerst kennis met het vak Keel- Neus- en Oorheelkunde. Onder supervisie van Professor dr. R. J. Baatenburg de Jong werd een onderzoek naar de invloed van comorbiditeit op de overleving van hoofd-hals oncologische patiënten verricht. Dit onderzoek vormde de basis voor dit proefschrift en mondde, na anderhalf jaar werkzaam te zijn geweest als arts-assistent KNO in het HagaZiekenhuis te Den Haag, uit in een opleidingsplaats in het Erasmus Medisch Centrum te Rotterdam. Naar verwachting zal de opleiding tot KNO-arts worden afgerond in maart 2013. Frank is naast zijn opleiding actief als bestuurslid van de Landelijke KNO assistenten vereniging en als redacteur van het Nederlands Tijdschrift voor Keel-, Neus en Oorheelkunde.

Frank is in 2007 getrouwd met Jeske van Dooren. Samen hebben zij twee dochters (Julia 2009* en Veerle 2010) en wonen in Voorburg.

List of publications

1. **Datema FR**, Blom HM. Intranasale nestvorming. *Ned. Tijdschr. KNO 2006; 12: 21-22.*
2. **Datema FR**, Holland CTQ, Blom HM. KTP/532 lasertonsillectomie en CO2 lasertonsillotomie. *Ned. Tijdschr. KNO 2007; 13: 65-70.*
3. **Datema FR**, Vemer JG, Wieringa MH, Mulder PM, Baatenburg de Jong RJ, Blom HM. A visual analog scale can assess the effect of surgical treatment in children with chronic otitis media with effusion. *Int. J. Ped. Otol. 2008; 72: 461-467.*
4. **Datema FR**, Borgstein J. A new method to solve an old problem: extraction of a sharp foreign body from the lateral basal part of the bronchial tree of a child. *Int. J. Ped. Otol. 2009; 4: 62-65.*
5. **Datema FR**, vd Schroeff MP, vd Velden LA. Orale, oropharyngeale en laryngeale lokalisatie van bulleus pemphigoïd*. Ned. Tijdschr. KNO 2009; 15: 26-29.*
6. **Datema FR**, Holland CTQ, Abedi S, Blom HM. De CO_2-lasertonsillotomie bij volwassenen onder lokale anesthesie: een beschrijving van de voorlopige resultaten van een prospectieve pilot-studie. *Ned. Tijdschr. KNO 2009; 15: 14-21.*
7. Veder LL, **Datema FR**, vd Velden LA. Stagnerende wondgenezing na hoofd-hals chirurgie: beter ten halve gekeerd dan ten hele gedwaald. *Ned. Tijdschr. KNO 2010; 16: 26 – 29.*
8. **Datema FR**, Ferrier MB, vd Schroeff MP, Baatenburg de Jong RJ. The impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. *Head Neck 2010; 32: 728-36.*
9. **Datema FR**, Poldermans D, Baatenburg de Jong RJ. Incidence and prediction of major cardiovascular complications in head and neck surgery. *Head Neck 2010; 32:1485-93.*
10. **Datema FR**, Koopman JP, Blom HM. How we do it: veel voorkomende KNO problematiek bij kinderen met het Down syndroom. *Ned. Tijdschr. KNO 2010; 16: 187 - 191.*
11. **Datema FR**, Hoeve LJ. Endoscopische sluiting van een type II posterieure larynxcleft: een case report en literatuuroverzicht – *Ned Tijdschr. KNO 2010;16: 171- 174.*
12. vd Schroeff MP, Terhaard CH, Wieringa MH, **Datema FR**, Baatenburg de Jong RJ. Cytology and histology have limited added value in prognostic models for salivary gland carcinomas. *Oral Oncol. 2010; 46: 662-6.*
13. **Datema FR.** Boekbespreking: Otolaryngology Cases: The University of Cincinnati Clinical Portfolio. *Ned Tijdschr. KNO 2010:16: 224.*
14. **Datema FR**, Moya A, Krause P, Bäck T, Willmes L, Langeveld A, Baatenburg de Jong RJ, Blom HM. Novel head and neck cancer survival analysis approach: Random Survival Forests versus Cox Proportional Hazards Regression. *Head Neck 2012; 34:50-8.*
15. **Datema FR**, Ferrier MB, Baatenburg de Jong RJ. Impact of severe malnutrition on short-term mortality and overall survival in head and neck cancer. *Oral Oncol. 2011; 47: 910-4.*

16. **Datema FR.** Redactioneel NTvKNO – *Ned. Tijschr. KNO 2011; 17: 109*
17. **Datema FR.** Boekbespreking: ENT Diseases with Head and Neck Surgery 3rd edition. – *Ned. Tijschr. KNO 2011; 17: 166*
18. **Datema FR.** Ernstige ondervoeding als potentiële 9e prognostische factor voor patiënten met hoofd-halstumoren. *Oncologie Up-to-Date 2011;2:8*
19. **Datema FR,** Ferrier MB, Vergouwe Y, Moya A, Molenaar J, Piccirillo JF, Baatenburg de Jong RJ. Update and external validation of a head and neck cancer prognostic model. *Awaiting final decision after revisions Head Neck*