Oral Oncology 74 (2017) 77–82

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

# A failure-type specific risk prediction tool for selection of head-and-neck cancer patients for experimental treatments

Katrin Håkansson[a,*], Jacob H. Rasmussen[b], Gregers B. Rasmussen[a], Jeppe Friborg[a], Thomas A. Gerds[c], Barbara Malene Fischer[d], Flemming L. Andersen[d], Søren M. Bentzen[e], Lena Specht[a], Ivan R. Vogelius[a]

[a] Department of Oncology, Section of Radiotherapy, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark
[b] Department of Otorhinolaryngology, Head & Neck Surgery and Audiology, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark
[c] Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen, Denmark
[d] Department of Clinical Physiology, Nuclear Medicine & PET, PET & Cyclotron Unit, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark
[e] Division of Biostatistics and Bioinformatics, University of Maryland Greenebaum Cancer Center and Department of Epidemiology and Public Health, University of Maryland School of Medicine, 655 W Baltimore Street, Baltimore MD21201, USA

ARTICLE INFO

*Keywords:*
Head and neck neoplasms
Decision support systems, clinical
Risk
Patient selection
Nomograms

ABSTRACT

*Objectives:* The objective of this work was to develop a tool for decision support, providing simultaneous predictions of the risk of loco-regional failure (LRF) and distant metastasis (DM) after definitive treatment for head-and-neck squamous cell carcinoma (HNSCC).

*Materials and Methods:* Retrospective data for 560 HNSCC patients were used to generate a multi-endpoint model, combining three cause-specific Cox models (LRF, DM and death with no evidence of disease (death NED)). The model was used to generate risk profiles of patients eligible for/included in a de-intensification study (RTOG 1016) and a dose escalation study (CONTRAST), respectively, to illustrate model predictions versus classic inclusion/exclusion criteria for clinical trials. The model is published as an on-line interactive tool (https://katrin.shinyapps.io/HNSCCmodel/).

*Results:* The final model included pre-selected clinical variables (tumor subsite, T stage, N stage, smoking status, age and performance status) and one additional variable (tumor volume). The treatment failure discrimination ability of the developed model was superior of that of UICC staging, 8$^{th}$ edition ($AUC_{LRF}$ = 72.7% vs 64.2%, $p < 0.001$ and $AUC_{DM}$ = 70.7% vs 58.8%, $p < 0.001$). Using the model for trial inclusion simulation, it was found that 14% of patients eligible for the de-intensification study had > 20% risk of tumor relapse. Conversely, 9 of the 15 dose escalation trial participants had LRF risks < 20%.

*Conclusion:* A multi-endpoint model was generated and published as an on-line interactive tool. Its potential in decision support was illustrated by generating risk profiles for patients eligible for/included in clinical trials for HNSCC.

## Introduction

Head-and-neck squamous cell carcinoma carcinoma (HNSCC) is a disease where the prognosis varies substantially between patients. Human papilloma virus (HPV) positive oropharyngeal carcinoma is, in relative terms, a favorable subtype with a 3-year overall survival (OS) of around 80% [1]. It has been suggested that treatment de-intensification would be possible in this group without compromising the treatment effect, and several such studies are ongoing [2]. This begs the question whether we can identify subset of HPV positive patients where

*Abbreviations:* AUC, Area under the ROC curve; Death NED, Death with no evidence of disease; DM, Distant metastasis; CI, Confidence interval; FDG, $^{18}$F-fluorodeoxyglucose; HNSCC, Head and neck squamous cell carcinoma; HPV, Human papilloma virus; IMRT, intensity-modulated radiotherapy; LRF, Loco-regional failure; $NSUV_{max}$, Maximum standard uptake value in nodes; OS, Overall survival; PSF, Point-spread-function; SUV, Standard uptake value; $TSUV_{max}$, Maximum standard uptake value in primary tumor

* Corresponding author.
*E-mail addresses:* katrin.elisabet.haakansson@regionh.dk (K. Håkansson), jacob.hoeygaard.rasmussen.01@regionh.dk (J.H. Rasmussen), gregers.bruennich.damgaard.rasmussen@regionh.dk (G.B. Rasmussen), jeppe.friborg@regionh.dk (J. Friborg), tag@biostat.ku.dk (T.A. Gerds), barbara.malene.fischer@regionh.dk (B.M. Fischer), flemming.andersen@regionh.dk (F.L. Andersen), sbentzen@som.umaryland.edu (S.M. Bentzen), lena.specht@regionh.dk (L. Specht), ivan.richter.vogelius@regionh.dk (I.R. Vogelius).

http://dx.doi.org/10.1016/j.oraloncology.2017.09.018
Received 29 June 2017; Received in revised form 8 September 2017; Accepted 17 September 2017

de-intensification may be detrimental for long term survival due to a – for the group – unfavorable disease characteristics. At the same time, other HNSCC patients have very unfavorable prognosis. A previous risk model suggests that around 25% of HNSCC patients have 2-year risk of tumor relapse exceeding 40% [3]. For high risk HNSCC patients, there are treatment intensification trials, where parts of the tumor, with a suggested higher risk of radioresistance, are treated with higher than standard radiotherapy doses, so called dose-painting [4,5]. Inclusion/exclusion of patients in such clinical trials is mainly based on risk/benefit considerations – weighting the chance of cure to the risk of toxicity. In this context, multivariate prognostic models can include complex combinations of patient characteristics, valuable in yielding a higher level of individualization than simple inclusion/exclusion criteria used in most clinical trials. Here, we present a model for risk prediction, designed to provide decision support for physicians stratifying patients into trials of systemic and/or local treatment intensification or de-intensification for HNSCC.

We consider three competing modes of first failure after treatment; loco-regional failure (LRF), distant metastasis (DM) and mortality without prior evidence of disease (death NED). They are treated as competing events, as the occurrence of one will alter or mitigate the risk of the others. Hereby, we take into account that non-cancer mortality is a significant cause of death among HNSCC patients [6]. Knowledge of a patient's risk of all of the above events is valuable information prior to counselling the patient regarding the participation in a study of e.g. local dose escalation or de-intensification. To this end, absolute risk of events occurring are more useful than hazard ratios and, consequently, several authors have constructed nomograms for HNSCC [7–10]. The current study builds on a similar concept, but taking it a step further by combining multiple endpoints into absolute risks, and by including confidence intervals for the risk estimations given. The model is presented in a user friendly internet-based display.

We assess the value of the model in clinical decision support for trial enrollment by studying two scenarios; the risk profiles of patients actually included in a phase I trial of local dose escalation at our own institution [5], and the risk profiles of patients amenable for an international de-intensification protocol (RTOG 1016) in HPV- positive patients [11].

## Materials and methods

### *Patients*

The patient cohort used for building the prognostic models consisted of 600 HNSCC patients, all treated with intensity-modulated radiotherapy (IMRT) between 2005 and 2012 at Rigshospitalet, Copenhagen. The patients were required to have a pre-treatment PET/CT, which is acquired routinely for target volume delineation at our institution. Radiotherapy was the primary treatment for all patients, supplemented by cisplatin if judged to be indicated/tolerable. Treatment details and follow-up procedures have previously been described in detail [3]. Patient and disease characteristics were acquired from retrospective chart review, supplemented by immuno-chemical p16 status [12] and data from the Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA) for tumor volume and the Mirada software (Mirada Medical Ltd, Oxford, UK) for PET imaging data. The p16+ oropharynx patients were re-staged using the 8th edition of the UICC TNM manual.

### *Imaging*

All patients were scanned for radiotherapy planning with an integrated PET/CT scanner (Biograph 40 TruePoint 40 HD-PET; Siemens Medical Solution, Malvern PA; Biograph 16 TruePoint, Siemens Medical Solutions; or Discovery LS, 4 Slice, General Electric, Milwaukee, WI). During the PET/CT scan patients were immobilized in treatment position with molded 5-point masks. $^{18}$F-fluorodeoxyglucose (FDG, 200-400 MBq depending on BMI) were intravenously administered one hour before scanning, and patients were instructed to fast for at least six hours before administration of FDG. PET/CT scan was performed from apex to thigh, with a standard duration of 3 min/bed position. The attenuation-corrected PET data were reconstructed using a 3D ordered subset expectation maximization algorithm (2D-AW-OSEM for patients scanned on the Discovery LS scanner), with or without point-spread-function (PSF) correction (introduced in 2010). Images were post-filtered with 4–5 mm Gaussian filter (Discovery LS and non-PSF corrected images) or 2 mm Gaussian filter (PSF corrected images). Standard uptake values (SUV) were corrected for body weight. Maximum SUV were extracted for the primary tumor ($TSUV_{max}$) and for nodes ($NSUV_{max}$) separately. For the analysis, SUV values were divided into quartiles, with separate cut-off values for patients scanned with and without PSF correction. Patients with N0 disease were excluded from the $NSUV_{max}$ analysis.

### *Statistics*

Time was measured from start of radiotherapy to LRF, DM, death NED or end of follow-up, whichever occurred first. The median potential follow-up time of the patients [13] was estimated with the reverse Kaplan-Meier method. The modes of failure were defined as follows:

- LRF – T and/or N site recurrence or no remission at first follow-up
- DM – M site recurrence with or without simultaneous LRF
- Death NED – Death without tumor relapse

Three cause-specific multiple Cox regression models were built for the hazard rates of LRF, DM and death NED, respectively. All three models were stratified for prescription of cisplatin and adjusted for a base set of pre-defined prediction variables:

– For the hazard rate of LRF: Smoking status, T stage, N stage and tumor subsite (p16+ oropharynx vs p16- oropharynx, larynx, oral cavity and hypopharynx).
– For the hazard rate of DM: Smoking status, T stage, N stage and p16+ oropharynx (yes/no).
– For the hazard rate of death NED: Smoking status, age, p16+ oropharynx (yes/no) and WHO performance status.

The base set of prediction variables was selected by consensus among two senior oncologists (JF and LS) and one head-and-neck surgeon in training (JHR). The combination of these three cause-specific Cox models is referred to as the basis model.

The time point for risk predictions was set 3 years after the start of radiotherapy. At this time point, the overall absolute risks of the three endpoints were estimated with the Aalen-Johansen method. For a given patient, a 3-year risk profile was obtained by combination of the three Cox regression models [14] and consisted of the 3-year absolute risks of LRF, DM and death NED, respectively. The risk profiles were displayed in triangular plots.

Four additional prediction variables – tumor volume, $TSUV_{max}$ and $NSUV_{max}$ (for LRF and DM) and albumin (for death NED) – were tested one by one for inclusion in the final multi-endpoint model by comparing the 3-year prediction performance of the basis model vs the basis model + test variable. Prediction performance was determined using the Brier score [15] as well as a time-dependent concordance index for competing risk data [16], corresponding to the area under the ROC curve (AUC), see appendix for details. When deciding if the test variables should be included, the performance of the LRF and DM predictions were higher prioritized than of the death NED predictions. The 3-year prediction performance of the final multi-endpoint model was compared to the performance of a model with UICC stage (8th edition)