Prognostic models for radiation-induced complications after radiotherapy in head and neck cancer patients

Takada, Toshihiko; Damen, Johanna A.A.G.; Tambas, Makbule; Spijker, René; Steenbakkers, Roel J.H.M.; Sharabiani, Marjan; Clementel, Enrico; Langendijk, Johannes A.; Moons, Karel G.M.; Schuit, Ewoud

Published in:
Cochrane Database of Systematic Reviews

DOI:
10.1002/14651858.CD014745

Publication date:
2021

Cochrane Library

Cochrane Database of Systematic Reviews

Prognostic models for radiation-induced complications after radiotherapy in head and neck cancer patients (Protocol)

Takada T, Damen JAAG, Tambas M, Spijker R, Steenbakkers RJHM, Sharabiani M, Clementel E, Langendijk JA, Moons KGM, Schuit E

Takada T, Damen JAAG, Tambas M, Spijker R, Steenbakkers RJHM, Sharabiani M, Clementel E, Langendijk JA, Moons KG, Schuit E.
Prognostic models for radiation-induced complications after radiotherapy in head and neck cancer patients (Protocol).
Cochrane Database of Systematic Reviews 2021, Issue 5. Art. No.: CD014745.
DOI: 10.1002/14651858.CD014745.

www.cochranelibrary.com

TABLE OF CONTENTS

[Prognosis Protocol]

Prognostic models for radiation-induced complications after radiotherapy in head and neck cancer patients

Toshihiko Takada[1], Johanna AAG Damen[1,2], Makbule Tambas[3], René Spijker[2], Roel JHM Steenbakkers[3], Marjan Sharabiani[4], Enrico Clementel[4], Johannes A Langendijk[3], Karel GM Moons[1,2], Ewoud Schuit[1]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [2]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [3]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [4]European Organisation for Research and Treatment of Cancer (EORTC) Headquarters, Brussels, Belgium

Contact address: Toshihiko Takada, ttakada@water.ocn.ne.jp.

Editorial group: Cochrane Haematology Group, Cochrane ENT Group.
Publication status and date: New, published in Issue 5, 2021.

Citation: Takada T, Damen JAAG, Tambas M, Spijker R, Steenbakkers RJHM, Sharabiani M, Clementel E, Langendijk JA, Moons KG, Schuit E. Prognostic models for radiation-induced complications after radiotherapy in head and neck cancer patients (Protocol). *Cochrane Database of Systematic Reviews* 2021, Issue 5. Art. No.: CD014745. DOI: 10.1002/14651858.CD014745.

ABSTRACT

Objectives

This is a protocol for a Cochrane Review (prognosis). The objectives are as follows:

Primary objective

The review question is "Which prognostic models are available to predict the risk of radiation-induced side effects after radiation exposure to patients with head and neck cancer, what is their quality, and what is their predictive performance?".

Investigation of sources of heterogeneity between studies

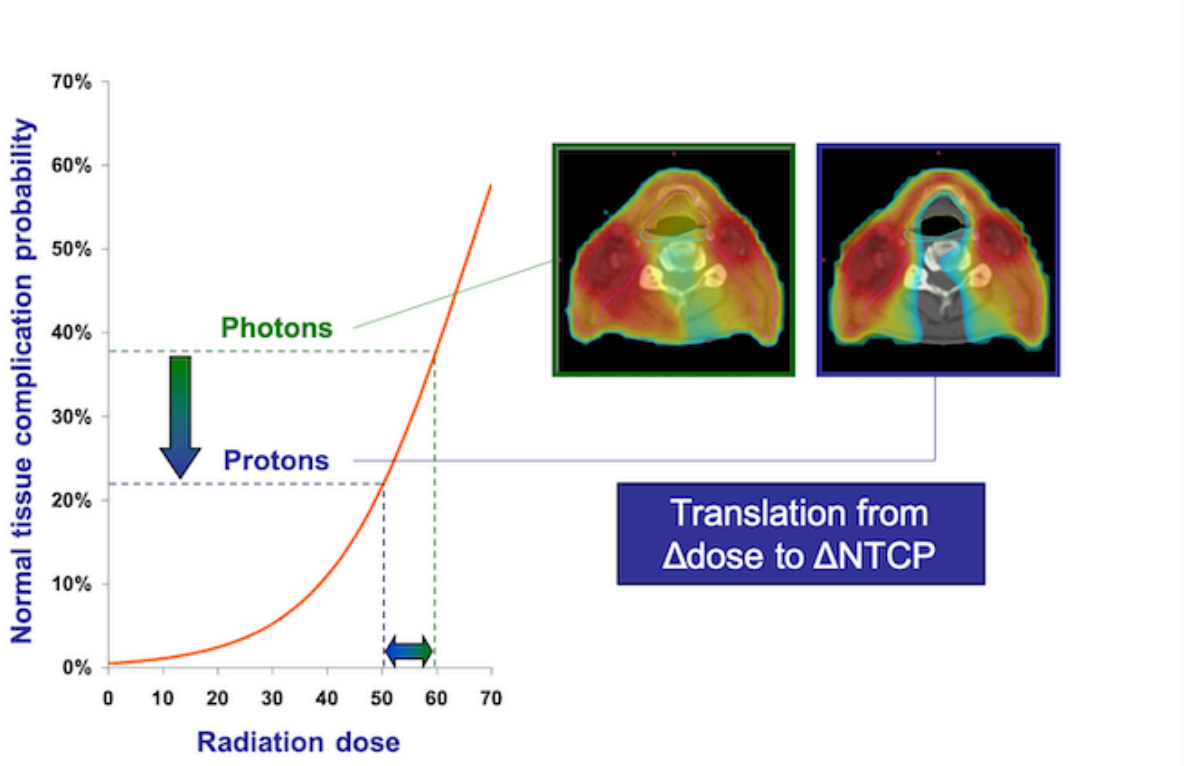
We will assess sources of heterogeneity among the prognostic models developed in the eligible studies. The potential sources are study population (e.g. site/stage of cancer, the use of other treatment [surgery and chemotherapy]), predictors, definition and incidence of the predicted outcomes, and prediction horizons. If there are multiple validation studies for the same model, the same sources of between-study heterogeneity will be investigated.

BACKGROUND

Description of the health condition and context

Head and neck cancer is the sixth most common malignant neoplasm in the world (Gupta 2016). More than 90% of head and neck cancers are squamous cell carcinomas originating from the upper airway and digestive tract, typically in the oral cavity, larynx, or pharynx (Gupta 2016). Since squamous cell carcinomas are radio-sensitive, radiotherapy (either alone or combination of surgery and/or chemotherapy) is the mainstay of the treatment for head and neck cancer (De Felice 2018). Despite its effectiveness, the challenge of radiotherapy is how to avoid the risk of radiation-induced side effects, by reducing the radiation dose to the normal tissues. The relationship between the dose distributions to normal tissues and the risk of radiation-induced side effects can be estimated by prognostic models, often referred to as so-called "normal tissue complication probability" (NTCP) models (van der Laan 2012). Head and neck cancer radiotherapy may result in a wide spectrum of radiation-induced side effects such as xerostomia (dry mouth syndrome), dysphagia (difficulty in swallowing), mucositis (inflammation of the mucous membrane), problems with taste, problems with speaking, trismus (restriction in mouth opening) and hearing loss. As these side effects have a significant impact on the more general dimensions of quality of life (Langendijk 2008), prevention of these side effects is crucial. Nowadays, more advanced radiation technologies like protons become increasingly available. Dose distributions obtained with protons have significant superiority over the currently used photons because of their excellent energy absorption profile. Proton beams can yield a uniform dose distribution across the target volume with a rapid dose fall-off distally from the target, which results in a highly accurate dose depositions in the target and much better sparing of surrounding normal tissues. Consequently, proton therapy is expected to result in less radiation-induced side effects than photon therapy (Langendijk 2013). However, because of the limited accessibility and higher costs of such a new technique, it is necessary to identify patients who are likely to benefit most from proton therapy in terms of prevention of radiation-induced side effects for efficient use of medical resources. By using prognostic models which include predictors related with radiation dose next to other none-dose related predictors, it is possible to use such models for the so-called model-based selection for proton or conventional photon therapy. In the model-based selection, first, the radiation dose distribution in the target organ is estimated for both the photon and proton therapy. Then, the risk of radiation-induced side effects is calculated for the photon and proton therapy by using prognostic models including predictors related with radiation dose distribution. Thus, individual plan comparison between the photon and proton can be made in each patient, in which the radiation dose distributions are translated to the risk of radiation-induced side effects for protons and photons (Figure 1). The difference in the risk of the radiation-induced side effects between the photon and proton therapy (ΔNTCP) can then be used to decide if a patient indeed qualifies for proton therapy or rather for conventional therapy.

Figure 1. In the model-based selection, individual plan comparison is made in each patient, in which the radiation dose distributions are translated to the risk of radiation-induced side effects for protons and photons. The difference in the risk of the radiation-induced side effects between the photon and proton therapy (ΔNTCP) can be calculated.



Description of the prognostic models under review

Various prognostic models for prediction of various types of radiation-induced side effects in patients with head and neck cancer (e.g. xerostomia, dysphagia, hearing loss, and hypothyroidism), have been developed (for example: Beetz 2012; Cheraghi 2017; Christianen 2016; Luo 2018; Wopken 2014). These models estimate the future risk of the radiation-induced side effects in head and neck regions based on multiple predictors. These predictors are mainly classified as dosimetric and non-dosimetric factors. Dosimetric factors are those related with radiation dose. Non-dosimetric factors include patient characteristics (e.g. age and gender), type/site/stage of cancer, and concurrent treatment. The purposes of these prognostic models estimating an individual's risk of these radiation-induced side effects are as follows:

i) informing patients and their families about the possible side effects caused by radiotherapy,

ii) selection of patients who may benefit most from newly developed techniques of radiotherapy,

iii) optimizing dose distributions guided by the dose parameters included in the prognostic models (i.e. model-based optimization).

Health outcomes

Radiation exposure to normal tissues surrounding head and neck cancer may cause various types of side effects (e.g. xerostomia, oral mucositis, sticky saliva, dysphagia, esophagitis, taste and hearing loss, aspiration, hoarseness, fatigue, trismus, and hypothyroidism), which may have significant impact on health-related quality of life (Bansal 2004). Depending on the type of side effects, these may occur at various time points (during treatment and years after treatment) (van der Veen 2017).

Why it is important to do this review

Prognostic models for predicting side effects of radiotherapy in patients with head and neck cancer are crucial to identify those who are at the lowest risk of developing side effects from advanced radiation technologies such as proton therapy. Since there have been several prognostic models for each type of radiation-induced side effects (Beetz 2012; Cheraghi 2017; Christianen 2016; Luo 2018; Wopken 2014), this review is important to identify all the available prognostic models predicting radiation-induced side effects, appraise their quality and applicability, and summarize the current evidence on their predictive performance obtained both from the model development as from model validation studies. Thereby, this review will be helpful to understand which model(s) and to which extent may be applied for therapeutic decision making on the preferred radiation approach, which models should not be used for this purpose and which models are promising but still require further investigation, e.g. testing their predictive accuracy in other patients or settings.

OBJECTIVES

Primary objective

The review question is "Which prognostic models are available to predict the risk of radiation-induced side effects after radiation exposure to patients with head and neck cancer, what is their quality, and what is their predictive performance?".

Investigation of sources of heterogeneity between studies

We will assess sources of heterogeneity among the prognostic models developed in the eligible studies. The potential sources are study population (e.g. site/stage of cancer, the use of other treatment [surgery and chemotherapy]), predictors, definition and incidence of the predicted outcomes, and prediction horizons. If there are multiple validation studies for the same model, the same sources of between-study heterogeneity will be investigated.

METHODS

Criteria for considering studies for this review

Types of studies

Based on the PICOTS (Population, Intervention, Comparison, Outcome, Timing) of the review question (Table 1) (Debray 2017; Moons 2014; Moons 2019), we will include studies which meet all of the following criteria:

i) study design: all retrospective and prospective cohort and nested case-control studies,

ii) data source: studies that used routine care, registry data or data from randomized trials,

iii) aim: studies that aimed to develop, evaluate (internal and/or external validation) or update prognostic models to predict radiation-induced side effects in patient with head and neck cancer who underwent radiotherapy. The timing of model usage is just before starting radiotherapy.

There will be no restriction on language.

Targeted population

The targeted population is patients with head and neck cancer undergoing definitive or postoperative radiotherapy either in combination, or not, with systemic agents (e.g. chemotherapy, cetuximab). We will include all studies, including those that include only a subgroup of this targeted population (e.g. including only patients with tongue cancer). Model development or validation studies which include this target population but combined with other target populations (such as patients with prostate cancer) are not eligible for this review. We will only include studies that included patients treated with current standard radiation technologies, such as intensity modulated radiotherapy (IMRT), volumetric modulated arc therapy (VMAT), 3D conformal radiotherapy and/or proton therapy. There will be no restriction about care setting, types of concurrent treatment, and types/stages of head and neck cancer.

Types of prognostic models

We will include all prognostic models aimed at predicting radiation-induced side effects. Regardless of the predictors included and the statistical techniques used, including logistic, ordinal, or polytomous regression, time-to-event or other machine learning modeling techniques. We will provide an overview and description of all identified models. We will also describe all studies that validated these prognostic models. If there are models that are validated multiple times, these models will be discussed in more

detail and we will attempt to perform a meta-analysis on the predictive performance measures of these models across the multiple validation studies.

Types of outcomes to be predicted

The outcomes of interest are all types of side effects caused by radiation exposure on normal organs surrounding head and neck (e.g. xerostomia, oral mucositis, sticky saliva, dysphagia, esophagitis, loss of taste, tinnitus, hearing loss, aspiration, hoarseness, fatigue, oral pain, pain in the throat, trismus, osteoradionecrosis, hypothyroidism, retinopathy, visual impairment, temporal lobe injury, carotid stenosis and second neoplasia). Since the time period needed to develop these side effects may differ and as the incidence of side effects may progress or recover over time, we will have no restriction about the timing of the prediction horizon.

Search methods for identification of studies

Electronic searches

We will conduct the search in the following databases: Ovid MEDLINE and Ovid Embase. The search strategies for each database are listed in Appendix 1 and Appendix 2, respectively. To efficiently identify prognostic model studies, we will use and modify the search filter described by Geersing and colleagues (Geersing 2012) for our purpose.

Searching other resources

We will search the following databases for ongoing trials: ClinicalTrials.gov and World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). We will check reference articles cited in the retrieved articles. In addition, clinical experts in the field will be contacted to identify any relevant prognostic models we may have missed with our search. Models published in grey literature (i.e. conference abstracts) will not be considered for inclusion.

Data collection

Selection of studies

At least two review authors (TT, MT) will independently screen the search results for eligibility on title and abstract. If large numbers of models are identified, this team will be extended to three or more review authors. Then, the same review authors will independently assess the eligibility of potentially relevant studies by reading the full-text articles. Any disagreement between the two review authors will be resolved by discussion and if needed by consulting an independent review author (ES). We will document study selection in a flow chart as recommended in the PRISMA guidelines (Moher 2009).

Data extraction and management

At least two review authors (TT, MT) will independently extract the data in accordance with the CHARMS checklist (Moons 2014). These items provide information to assess applicability to the review question and their risks of bias (see below). Any disagreement between the review authors will be resolved by discussion and if needed by consulting an independent review author (ES). We will contact authors of individual studies to obtain additional information, if necessary.

The following information will be extracted using a data extraction form based on the CHARMS checklist (Moons 2014).

- General information: author, year of publication, journal, country, language
- Source of data: study design, prospective or retrospective data collection, registry
- Participants: participant eligibility criteria and recruitment method (e.g. consecutive participants, study location, number of centers, setting, inclusion and exclusion criteria)
- Outcomes to be predicted: definition and method for measurement of outcome, time of outcome occurrence, summary of duration of follow-up, whether the assessment of the outcome was blinded from predictors included in the prognostic models.
- Candidate predictors: number and type of predictors (e.g. demographics [age and sex], patient complaint(s) at baseline, comorbidities, physical examination [baseline toxicity], laboratory testing, characteristics of cancer [stage (T and N) and location of cancer], and radiation dose to organs at risk, treatment modality, baseline complaints)
- Sample size: number of participants and number of participants with the outcome of interest
- Missing data: number of participants with any missing value, handling of missing data (e.g. complete-case analysis, imputation or other methods)
- Model development: modeling method (e.g. logistic or time-to-event modeling), assumption of applied modeling, method for selection of predictors for inclusion in multivariable modeling (e.g. clinically-relevant predictors or pre-selection based on univariable analyses), method for selection of predictors during multivariable modeling and its criteria, shrinkage methods (e.g. no shrinkage, uniform shrinkage, or penalized modeling)
- Prediction performance of the models: calibration (calibration plot, observed-expected ratio [O:E ratio], calibration-in-the-large [CITL] and slope, and Hosmer-Lemeshow test) and discrimination (c-statistic) measures with confidence intervals
- Model evaluation: method used for validation of model performance (development dataset, internal validation [e.g. random split of original data, or resampling methods], external validation [e.g. temporal, geographical, different setting, or different researchers])
- Results: all types of presentation of the prognostic models (e.g. basic, extended, or simplified) including regression coefficients, intercept and baseline survival
- · Interpretation and discussion: interpretation of presented models (confirmatory or exploratory), comparison with other studies, discussion of generalizability, strengths and limitations

Assessment of risk of bias and applicability

We will use the PROBAST-tool for assessment of risk of bias and applicability of all models reported in the included studies (Moons 2019; Wolff 2019). Risk of bias and applicability will be assessed in four domains with two to nine signaling questions (Table 2). Each domain will be judged as high, low or unclear risk of bias. Judgement of risk of bias will be facilitated by the signaling questions which can be answered “yes”, “probably yes”, “probably no”, “no”, or “no information”. For the signaling question regarding the time interval between the measurement of the predictor and outcome, an appropriate time interval will

be discussed for each outcome before we assess the risk of bias. Concerns regarding applicability will be rated similarly to risk of bias, but without signaling questions. Two review authors (TT, MT) will independently assess risk of bias and applicability. Any disagreement between the authors will be resolved by discussion or by consulting a third review author (ES).

Dealing with missing data

In case there is insufficient information in included studies, we will contact the corresponding authors to ask for additional information that is necessary for our analyses. If necessary information is not available, we will try to estimate performance measures (e.g. c-statistics, O:E ratio and their standard error using the available information (Debray 2017).

Assessment of heterogeneity

We will investigate sources of heterogeneity among the different prognostic models developed in eligible studies and between validation studies of the same model. The potential sources are study population (e.g. location/stage of cancer, the use of other treatment [surgery and chemotherapy]), predictors, definition and incidence of the predicted outcomes, and prediction horizons. Heterogeneity between validation studies of the same model will be quantified using Tau^2, I^2, and prediction intervals of the performance measures of each prognostic model. Heterogeneity between development studies will be described in a qualitative way.

Assessment of reporting deficiencies

Despite recommendation by current guidelines to report both discrimination and calibration measures (Collins 2015a), this information is often missing in prediction model publications. We will describe the reporting deficiencies in the included articles based on the TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement (Collins 2015b; Moons 2015). Additionally, we will check if there are protocol violations in studies which have their published study protocols.

Data synthesis

Data synthesis and meta-analysis approaches

For each type of radiation-induced side effects, we will provide a summary of the identified prognostic models including the following information: author, publication year, number of patients included, predictors, outcome definition, number of patients with the event, c-statistics, O:E ratios, CITL and slope. If there are multiple external validations for one or more of the identified prognostic models, and these external validation studies are sufficiently robust and comparable, we will perform meta-analyses of the model performance by aggregating c-statistics, O:E ratios, CITL and slopes. Separate meta-analyses will then be performed per model. We will use appropriate transformations (e.g. a logit transformation for c-statistics) (Snell 2018). Based on random-effects approach, we will report the pooled measure with the confidence intervals and prediction intervals. These analyses will be performed based on methods proposed by Debray (Debray 2019; Debray 2017) using the following packages of R statistics: metafor, mvmeta, metamisc, and lme4.

Subgroup analysis and investigation of heterogeneity

Among the identified models for each type of radiation-induced side effects, we plan to conduct the following subgroup analyses for comparison of the model's predictive performance – provided there are enough studies in these prespecified categories:

1. the location of cancer (oral cavity, larynx, hypopharynx, oropharynx, and nasopharynx);
2. the stage of cancer;
3. patients with/without concomitant therapy such as surgery, chemotherapy and cetuximab.

Sensitivity analysis

If there is a sufficient number of validation studies of the same model, we will perform sensitivity analyses using only studies with low risk of bias. Low risk of bias is defined as a model being rated as low risk of bias on all four PROBAST domains.

Conclusions and summary of findings

At this moment, there is no GRADE guidance available for systematic reviews of prognostic models. If this guidance becomes available during the review process, we will follow the most recent guidance. Otherwise, we will adapt the guidance for overall prognosis and prognostic factor studies (Foroutan 2020; Huguet 2013; Iorio 2015). In this case, the quality of evidence for risk prediction models will be evaluated by six factors that can decrease quality: (1) phase of investigation; (2) study limitations; (3) inconsistency; (4) indirectness; (5) imprecision; and (6) publication bias. This rating will be done for discrimination and calibration performance separately. We will justify all reasons to downgrade or upgrade the quality rating of studies in footnotes. Based on the literatures which assessed the impact of the radiation-induced side effects on patient-reported quality of life (Daugaard 2017; Langendijk 2008; van der Laan 2019), we will create “Summary of findings” tables for the following outcomes.

- Dysphagia
- Xerostomia
- Hoarseness
- Fatigue
- Nausea and vomiting
- Throat pain
- Aspiration

In each “Summary of findings” table, the following items will be presented.

- Model performance across validation studies
- Number of participants
- Number of patients with the outcome
- Certainty of the evidence (GRADE)

ACKNOWLEDGEMENTS

We would like to thank Vanessa Piechotta (Managing Editor, Cochrane Haematology) and Nina Kreuzberger (Method Editor, Cochrane Haematology) for their assistance regarding the review process. We thank Cecilia Schmalbach (Cochrane ENT), Nicole Skoetz (Cochrane Cancer) and Richard Benson (Cambridge University Hospitals NHS Trust) for their valuable comments

to improve this review protocol. We also would like to thank Ina Monsef (Information Specialist, Cochrane Haematology) for feedback on the search strategy.

REFERENCES

Additional references

Bansal 2004

Bansal M, Mohanti BK, Shah N, Chaudhry R, Bahadur S, Shukla NK. Radiation related morbidities and their impact on quality of life in head and neck cancer patients receiving radical radiotherapy. *Quality of Life Research* 2004;**13**(2):481-8.

Beetz 2012

Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, et al. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. *Radiotherapy and Oncology* 2012;**105**(1):86-93.

Cheraghi 2017

Cheraghi S, Nikoofar A, Bakhshandeh M, Khoei S, Farahani S, Abdollahi H, et al. Normal tissue complication probability modeling of radiation-induced sensorineural hearing loss after head-and-neck radiation therapy. *International Journal of Radiation Biology* 2017;**93**(12):1327-33.

Christianen 2016

Christianen ME, van der Schaaf A, van der Laan HP, Verdonck-de Leeuw IM, Doornaert P, Chouvalova O, et al. Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: clinical validation according to the model-based approach. *Radiotherapy and Oncology* 2016;**118**(2):298-303.

Collins 2015a

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Annals of Internal Medicine* 2015;**162**(10):735-6.

Collins 2015b

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Annals of Internal Medicine* 2015;**162**(1):55-63.

Daugaard 2017

Daugaard R, Kjaer T, Johansen C, Christiansen J, Andersen E, Nielsen AL, et al. Association between late effects assessed by physicians and quality of life reported by head-and-neck cancer survivors. *Acta Oncologica* 2017;**56**(2):342-7.

Debray 2017

Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2019

Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research* 2019;**28**(9):2768-86.

De Felice 2018

De Felice F, Polimeni A, Valentini V, Brugnoletti O, Cassoni A, Greco A, et al. Radiotherapy controversies and prospective in head and neck cancer: a literature-based critical review. *Neoplasia* 2018;**20**(3):227-32.

Foroutan 2020

Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology* 2020;**121**:62-70.

Geersing 2012

Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS One* 2012;**7**(2):e32844.

Gupta 2016

Gupta B, Johnson NW, Kumar N. Global epidemiology of head and neck cancers: a continuing challenge. *Oncology* 2016;**91**(1):13-23.

Huguet 2013

Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic Review* 2013;**2**:71.

Iorio 2015

Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:h870.

Langendijk 2008

Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *Journal of Clinical Oncology* 2008;**26**(22):3770-6.

Langendijk 2013

Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiotherapy and Oncology* 2013;**107**(3):267-73.

Luo 2018

Luo R, Wu VW, He B, Gao X, Xu Z, Wang D, et al. Development of a normal tissue complication probability (NTCP) model for radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. *BMC Cancer* 2018;**18**(1):575.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P . Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Medicine* 2009;**6**(7):e1000097.

Moons 2014

Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Medicine* 2014;**11**(10):e1001744.

Moons 2015

Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine* 2015;**162**(1):W1-73.

Moons 2019

Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Annals of Internal Medicine* 2019;**170**(1):W1-W33.

Snell 2018

Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-statistic and calibration measures? *Statistical Methods in Medical Research* 2018;**27**(11):3505-22.

van der Laan 2012

van der Laan HP, Christianen ME, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. *Radiotherapy and Oncology* 2012;**103**(1):76-81.

van der Laan 2019

van der Laan HP, van den Bosch L, van der Schaaf A, Steenbakkers R, Bijl H, Dieters M, et al. Quality of life based Total Cost Function (TCF) to guide treatment plan optimization for head and neck cancer. *International Journal of Radiation Oncology • Biology • Physics* 2019;**105**:S96.

van der Veen 2017

van der Veen, JNuyts, S. Can intensity-modulated-radiotherapy reduce toxicity in head and neck squamous cell carcinoma? *Cancers (Basel)* 2017;**9**(10):135.

Wolff 2019

Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**(1):51-8.

Wopken 2014

Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJ, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiotherapy and Oncology* 2014;**113**(1):95-101.

ADDITIONAL TABLES

Table 1. PICOTS of the review question based on the CHARMS checklist

Population targeted	Patients with head and neck cancer undergoing current standard radiotherapy techniques (e.g. intensity modulated radiotherapy [IMRT] or volumetric modulated arc therapy [VMAT], 3D conformal radiotherapy and/or proton therapy)
Index model(s)	All available prognostic models predicting the risk of radiation-induced side effects
Comparator model(s)	Not applicable
Outcome(s) to be predicted	All types of acute and late radiation-induced side effects in head and neck regions
Timing of making prediction and Time span of the prediction	Just before starting radiotherapy. There are no restrictions on the prediction horizon
Setting	Secondary and tertiary care

Table 2. Assessment of risk of bias and applicability

1. Participants	2. Predictors	3. Outcome	4. Analysis
Signaling questions			

Table 2. Assessment of risk of bias and applicability *(Continued)*

1.1 Were appropriate data source used, e.g. cohort, randomized controlled trial, or nested case– control study data?	2.1 Were predictors defined and assessed in a similar way for all participants?	3.1 Was the outcome determined appropriately?	4.1 Were there a reasonable number of participants with the outcome?
1.2 Were all inclusions and exclusions of participants appropriate?	2.2 Were predictor assessments made without knowledge of outcome data?	3.2 Was a prespecified or standard outcome definition used?	4.2 Were continuous and categorical predictors handled appropriately?
	2.3 Are all predictors available at the time when the model is intended to be used?	3.3 Were predictors excluded from the outcome definition?	4.3 Were all enrolled participants included in the analysis?
		3.4 Was the outcome defined and determined in a similar way for all participants?	4.4 Were participants with missing data handled appropriately?
		3.5 Was the outcome determined without knowledge of predictor information?	4.5 Was selection of predictors based on univariable analysis avoided?
		3.6 Was the time interval between predictor assessment and outcome determination appropriate?	4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted appropriately?
			4.7 Were relevant model performance measures evaluated appropriately?
			4.8 Were model overfitting, underfitting, and optimism in model performance accounted for?
			4.9 Do predictors and their assigned weights in the final model correspond to the reported multivariable analysis?
Risk of bias			
Selection of participants	Predictors of their assessment	Outcome or its determination	Analysis
Applicability			
Included participants or setting does not match the review question	Definition, assessment, or timing of predictors does not match the review question	Its definition, timing, or – determination does not match the review question	

APPENDICES

Appendix 1. Ovid MEDLINE search strategy

#	Searches
1	exp "Head and Neck Neoplasms"/
2	(((upper adj aerodigestive adj tract) or uadt) and (cancer* or carcinom* or tumor* or tumour* or neoplas* or malignan* or metasta*)).ti,ab,kf.
3	(ent adj4 (cancer* or carcinom* or tumor* or tumour* or neoplas* or malignan* or metasta*)).ti,ab,kf.
4	((head or neck or oral or tongue or lip or tonsil or nasal or oropharyn* or pharyn* or laryn* or throat or EAR or glotti* or nasopharyn* or hypopharyn* or maxillofacial or paranasal-sinus) adj4 (cancer* or carcinom* or tumor* or tumour* or neoplas* or malignan* or metasta*)).ti,ab,kf.
5	(hnscc or scchn).ti,ab,kf.
6	1 or 2 or 3 or 4 or 5
7	exp Radiotherapy/ or RADIATION ONCOLOGY/ or exp Combined Modality Therapy/ or (radiotherap$ or chemoradiotherap$ or radiation-therap$ or brachytherap$ or brachy-therap$ or irradiat$ or proton or photon or radio-therap$ or chemo-radio-therap$ or irradiat$ or radiat$ or (external* adj3 beam*) or ((multimodal$ or multi-modal$) adj3 (treat$ or therap$)) or (combi$ adj3 modalit$)).ti,ab,kf.
8	6 and 7
9	(head adj2 neck adj3 (radiation or radiotherap*)).ti,ab,kf.
10	8 or 9
11	Validat$.tw. or Predict$.ti. or Rule$.tw. or ((Predict$ or probabilit*) and (Outcome$ or Risk$ or Model$)).tw. or ((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).tw. or (Decision$.tw. and ((Model$ or Clinical$).tw. or exp Models, Statistical/)) or (Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or c-statistic or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable" or NTCP).tw.
12	(safe or safety or side-effect* or undesirable effect* or treatment emergent or tolerability or toxicity or adrs or (adverse adj2 (effect or effects or reaction or reactions or event or events or outcome or outcomes))).ti,ab.
13	(xerostomia or dry-mouth or (salivary adj2 (dysfunction or flow)) or "sticky saliva" or (salivary-duct adj2 inflammation) or dysphagia or (swallowing adj2 (dysfunction or problems)) or tube-feeding or esophagitis or "oral mucositis" or (mucosal adj2 (reactions or ulcers)) or atrophia or hypothyroidism or (thyroid adj3 dysfunction) or hearing-loss or deafness or tinnitus or ((laryngeal or mandibular) adj2 (dysfunction or necrosis)) or "voice problems" or hoarseness or chondronecrosis or osteoradionecrosis).ti,ab. or (exp Neoplasms, Radiation-Induced/ or exp Vision Disorders/ or exp Radiation Injuries/ or exp Carotid Stenosis/) or ((radiation-induced adj (cancer or neoplasm* or tumo?r)) or (visual adj impair*) or retinopath* or (temporal-lobe adj injury) or carotid stenosis).ti,ab,kf.

(Continued)

14	(ae or co).fs.
15	12 or 13 or 14
16	("expected ratio" or "observed ratio" or "E:O ratio" or "Hosmer-Lemeshow" or "H-L test" or (ROC or AUC or AUROC) or (Area adj3 (Receiver or ROC or curve)) or classif* or ((negative or positive) adj2 predictive-value*) or clinical-accuracy or (sensitivity or specificity or PPV or NPV) or (performance adj3 (classification or classifier or clinical or accuracy or validation or metrics or diagnostic)) or (calibrat* and (plot* or curve* or slope* or model or models)) or decision-curve or (discriminability or c-index or c-statistic or concordance or DCA) or ((discrimination or discriminative or discriminatory) adj3 (accuracy or ability or performance or value or model or models or power or capacity or capabilit* or efficiency))).ti,ab,kf.
17	dose-response.ti,ab,kf.
18	exp ROC Curve/ or exp Area Under Curve/ or exp "artificial intelligence"/ or "neural networks (computer)"/ or exp data mining/
19	(((machine or statistical) adj (learning or model*)) or multilayer-perceptron* or random-forest* or (bayes* adj2 network*) or (support-vector adj machine*) or nearest-neighbor* or elastic-net or naive-bayes* or ((classification or regression or estimation or decision) adj3 tree) or ridge or kernel or ensemble or bagging or bagged or boosting or boosted or fuzzy).ti,ab,kf.
20	11 or 16 or 17 or 18 or 19
21	10 and 15 and 20
22	(exp animals/ not humans/) or (case reports or review).pt.
23	21 not 22

Appendix 2. Embase search strategy

No.	Query
#12	#11 AND [embase]/lim
#11	#9 NOT #10
#10	(rat:ti OR rats:ti OR mouse:ti OR mice:ti OR swine:ti OR porcine:ti OR murine:ti OR sheep:ti OR lambs:ti OR pigs:ti OR piglets:ti OR rabbit:ti OR rabbits:ti OR cat:ti OR cats:ti OR dog:ti OR dogs:ti OR cattle:ti OR bovine:ti OR monkey:ti OR monkeys:ti OR trout:ti OR marmoset*:ti) AND 'animal experiment'/exp OR ('animal experiment'/exp NOT ('human experiment'/exp OR 'human'/exp)) OR 'case report'/exp OR 'review'/it
#9	#1 AND #2 AND #8
#8	#3 OR #4 OR #5 OR #6 OR #7
#7	(((machine OR statistical) NEAR/1 (learning OR model*)):ti,ab) OR 'multilayer perceptron*':ti,ab OR 'random forest*':ti,ab OR ((bayes* NEAR/2 network*):ti,ab) OR (('support vector' NEAR/1 machine*):ti,ab) OR 'nearest neighbor*':ti,ab OR 'elastic net':ti,ab OR 'naive bayes*':ti,ab OR (((classification OR regression OR estimation OR decision) NEAR/3 tree):ti,ab) OR ridge:ti,ab OR ker-

(Continued)

	nel:ti,ab OR ensemble:ti,ab OR bagging:ti,ab OR bagged:ti,ab OR boosting:ti,ab OR boosted:ti,ab OR fuzzy:ti,ab
#6	'dose response':ti,ab
#5	'area under the curve'/exp OR 'receiver operating characteristic'/exp OR 'machine learning'/exp OR 'machine learning' OR 'artificial intelligence'/exp
#4	'expected ratio':ti,ab OR 'observed ratio':ti,ab OR 'e:o ratio':ti,ab OR 'hosmer-lemeshow':ti,ab OR 'h-l test':ti,ab OR roc:ti,ab OR auc:ti,ab OR auroc:ti,ab OR ((area NEAR/3 (receiver OR roc OR curve)):ti,ab) OR classif*:ti,ab OR (((negative OR positive) NEAR/2 'predictive value*'):ti,ab) OR 'clinical accuracy':ti,ab OR sensitivity:ti,ab OR specificity:ti,ab OR ppv:ti,ab OR npv:ti,ab OR ((performance NEAR/3 (classification OR classifier OR clinical OR accuracy OR validation OR metrics OR diagnostic)):ti,ab) OR (calibrat*:ti,ab AND (plot*:ti,ab OR curve*:ti,ab OR slope*:ti,ab OR model:ti,ab OR models:ti,ab)) OR 'decision curve':ti,ab OR discriminability:ti,ab OR 'c index':ti,ab OR 'c statistic':ti,ab OR concordance:ti,ab OR dca:ti,ab OR (((discrimination OR discriminative OR discriminatory) NEAR/3 (accuracy OR ability OR performance OR value OR model OR models OR power OR capacity OR capabilit* OR efficiency)):ti,ab)
#3	validat*:ti,ab,kw OR predict*:ti OR rule*:ti,ab,kw OR ((predict*:ti,ab,kw OR probabilit*:ti,ab,kw) AND (outcome*:ti,ab,kw OR risk*:ti,ab,kw OR model*:ti,ab,kw)) OR ((history:ti,ab,kw OR variable*:ti,ab,kw OR criteria:ti,ab,kw OR scor*:ti,ab,kw OR characteristic*:ti,ab,kw OR finding*:ti,ab,kw OR factor*:ti,ab,kw) AND (predict*:ti,ab,kw OR model*:ti,ab,kw OR decision*:ti,ab,kw OR identif*:ti,ab,kw OR prognos*:ti,ab,kw)) OR (decision*:ti,ab,kw AND (model*:ti,ab,kw OR clinical*:ti,ab,kw OR 'statistical model'/exp OR 'statistical model')) OR (prognostic:ti,ab,kw AND (history:ti,ab,kw OR variable*:ti,ab,kw OR criteria:ti,ab,kw OR scor*:ti,ab,kw OR characteristic*:ti,ab,kw OR finding*:ti,ab,kw OR factor*:ti,ab,kw OR model*:ti,ab,kw)) OR 'stratification':ti,ab,kw OR 'discrimination':ti,ab,kw OR 'discriminate':ti,ab,kw OR 'c statistic':ti,ab,kw OR 'area under the curve':ti,ab,kw OR 'auc':ti,ab,kw OR 'calibration':ti,ab,kw OR 'indices':ti,ab,kw OR 'algorithm':ti,ab,kw OR 'multivariable':ti,ab,kw OR ntcp:ti,ab,kw
#2	safe:ti,ab OR safety:ti,ab OR 'side effect*':ti,ab OR 'undesirable effect*':ti,ab OR 'treatment emergent':ti,ab OR tolerability:ti,ab OR toxicity:ti,ab OR adrs:ti,ab OR ((adverse NEAR/2 (effect OR effects OR reaction OR reactions OR event OR events OR outcome OR outcomes)):ti,ab) OR xerostomia:ti,ab OR 'dry mouth':ti,ab OR ((salivary NEAR/2 (dysfunction OR flow)):ti,ab) OR 'sticky saliva':ti,ab OR (('salivary duct' NEAR/2 inflammation):ti,ab) OR dysphagia:ti,ab OR ((swallowing NEAR/2 (dysfunction OR problems)):ti,ab) OR 'tube feeding':ti,ab OR esophagitis:ti,ab OR 'oral mucositis':ti,ab OR ((mucosal NEAR/2 (reactions OR ulcers)):ti,ab) OR atrophia:ti,ab OR hypothyroidism:ti,ab OR ((thyroid NEAR/3 dysfunction):ti,ab) OR 'hearing loss':ti,ab OR deafness:ti,ab OR tinnitus:ti,ab OR (((laryngeal OR mandibular) NEAR/2 (dysfunction OR necrosis)):ti,ab) OR 'voice problems':ti,ab OR hoarseness:ti,ab OR chondronecrosis:ti,ab OR osteoradionecrosis:ti,ab OR 'adverse drug reaction':lnk OR 'side effect':lnk OR 'adverse device effect':lnk OR 'radiation induced neoplasm'/exp OR 'radiation injury'/exp OR 'retinopathy'/exp OR 'visual impairment'/exp OR 'carotid artery obstruction'/exp OR ((((('radiation induced' NEAR/1 (cancer OR neoplasm* OR tumor* OR tumours)):ti,ab,kw) OR ((visual NEAR/1 impair*):ti,ab,kw) OR retinopath*:ti,ab,kw OR (('temporal lobe' NEAR/1 injury):ti,ab,kw) OR carotid:ti,ab,kw) AND stenosis:ti,ab,kw)
#1	('head and neck cancer'/exp OR 'head and neck cancer' OR ((((upper NEAR/1 'aerodigestive tract'):ti,ab) OR uadt:ti,ab) AND (cancer*:ti,ab OR carcinom*:ti,ab OR tumor*:ti,ab OR tumour*:ti,ab OR neoplas*:ti,ab OR malignan*:ti,ab OR metasta*:ti,ab)) OR ((ent NEAR/4 (cancer* OR carcinom* OR tumor* OR tumour* OR neoplas* OR malignan* OR metasta*)):ti,ab) OR (((head OR neck OR oral OR tongue OR lip OR tonsil OR nasal OR oropharyn* OR pharyn* OR laryn* OR throat OR ear OR glotti* OR nasopharyn* OR hypopharyn* OR maxillofacial OR 'paranasal sinus') NEAR/4 (cancer* OR carcinom* OR tumor* OR tumour* OR neoplas* OR malignan* OR metasta*)):ti,ab) OR hnscc:ti,ab OR scchn:ti,ab) AND ('radiotherapy'/exp OR 'radiotherapy' OR 'radiation oncology'/exp OR 'radiation oncology' OR 'multimodality cancer therapy'/exp OR 'multimodality cancer therapy' OR radiotherap*:ti,ab OR radio-therap*:ti,ab OR chemoradiotherap*:ti,ab OR chemo-radio-therap*:ti,ab OR 'radiation therap*':ti,ab OR brachytherap*:ti,ab OR brachy-therap*:ti,ab OR irradiat*:ti,ab OR proton:ti,ab OR photon:ti,ab OR 'radio therap*':ti,ab OR radiat*:ti,ab OR ((extrenal*

(Continued)

	NEAR/3 beam*):ti,ab) OR (((multimodal* OR 'multi modal*') NEAR/3 (treat* OR therap*)):ti,ab) OR ((combi* NEAR/3 modal*):ti,ab))

Appendix 3. Preliminary study selection, data extraction and risk of bias forms for development and internal validation studies

Domain		Item to answer	Explanation/Classification	Reference
General information		Data extracted by	Your name	
		First author	Surname of the 1st author	
		Year of publication	yyyy	
		Sub ID	If the study reports multiple models (e.g. different outcomes, different presentations, etc), enter the items in different rows. For example, if there are two models for 30-days and 6-months mortality, label them as "1st author's surname-1" and "1st author's surname-2". Also, if the model was externally validated, enter the information about the external validation in another row.	
		Sub ID details	Describe the detail of the model to distinguish from each other. (For example of the left cell, describe as "model for predicting 30-days mortality" and "external validation", respectively.)	
Study characteristics	Study description	Purpose	- Development only - Development and internal validation - Development and external validation - Others (Specify)	CHARMS domain 1: Source of data
		Number of centers	1, 2, 3...	
		Location	- North America - Europe - Asia - Australia - Africa - South America - Central America - Combination	
Participants	Study description	Data source	- Prospective cohort - Retrospective cohort - Nested case-control - Nonnested case-control - Case-cohort - RCT - Registry data - Others (specify)	CHARMS domain 2: Participants

(Continued)

		- Unclear
	Recruitment method	Were patients recruited consecutively?
	Start date recruitment period	dd-mm-yyyy If the exact date is not reported (e.g. only year is reported), just enter available information (e.g. yyyy).
	End date recruitment period	dd-mm-yyyy If the exact date is not reported (e.g. only year is reported), just enter available information (e.g. yyyy).
	Inclusion criteria	copy and paste from the article text
	Exclusion criteria	copy and paste from the article text
	Primary site of cancer	Primary site - Oral cavity - Larynx - Hypopharynx - Oropharynx - Nasopharyx - Salivary glands - Nasal cavity - Paranasal sinus - Eye - Lymphoma - Soft tissue - Skull base - Ear - Lip - Thyroid - Others (Specify)
	Type of cancer	Pathology of cancer - SCC - Adenocarcinoma - Melanoma - Sarcoma - Lymphoma - Others (Specify) - Unclear
	TNM classification	Which version of theclassification was used? - 7th version (before 2017) - 8th version (after 2017)
	Stage	Based on TNM classification (I, II, III, IV…)
	Type of radiotherapy	- IMRT - VMAT - 3D conformal - Proton - Electron - Combined - Others (Specify) - Unclear

(Continued)

		Surgery	- No surgery - Pre-surgery radiotherapy - Post-surgery radiotherapy - Non specified - Others (Specify)	
		Type of surgery if specified	Detailed information about surgery (e.g. laryngectomy) Copy and paste from the paper	
		Chemotherapy	At baseline, - No chemotherapy - Concurrent chemotherapy - Chemotherapy before radiotherapy - Chemotherapy after radiotherapy - Case mix with different radiotherapy and chemotherapy combinations - Non-specified - Others (Specify)	
		Molecular-targeted therapy	At baseline, - No molecular-targeted therapy - Concurrent molecular-targeted therapy - Molecular-targeted therapy before radiotherapy - Molecular-targeted therapy after radiotherapy - Case mix with different radiotherapy and molecular therapy combinations - Non-specified - Others (Specify)	
		Other specific patient characteristics		
	Signaling questions	Were appropriate data sources used, e.g. cohort, RCT, or nested case-control study data?	- Yes/probably yes: If a cohort design (including RCT or proper registry data) or a nested case-control or case-cohort design (with proper adjustment of the baseline risk/hazard in the analysis) has been used. - No/probably no: If a non-nested case-control design has been used. - No information: If the method of participant sampling is unclear.	PROBAST step 3, question 1.1
		Were all inclusions and exclusions of participants appropriate?	-Yes/probably yes: If inclusion and exclusion of participants was appropriate, so participants correspond to unselected participants of interest. -No/probably no: If participants are included who would already have been identified as having the outcome and so are no longer participants at risk of developing outcome, or if specific subgroups are excluded that may have altered the performance of the prediction model for the intended target population. -No information: When there is no information on whether inappropriate inclusions or exclusions took place.	PROBAST step 3, question 1.2

(Continued)

	Risk of bias	Risk of bias introduced by participants or data sources	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgment could still be "Low risk of bias", but specific reasons should be provided why the risk of bias can be considered low. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias, except if defined at low risk of bias above. - Unclear risk of bias: If relevant information is missing for some of the signaling questions and non of the signaling questions is judged to put this domain at high risk of bias.	PROBAST step 3, risk of bias domain 1
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 1
	Applicability	Concern that included participants or the setting do not match the review question	-Low concern for applicability: Included participants and clinical setting match the review question. -High concern for applicability: Included participants and clinical setting were different from the review question. -Unclear concern for applicability: If relevant information about the participants and clinical setting are not reported.	PROBAST step 3, applicability domain 1
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Predictors	Study descriptions	List candidate predictors	A; B; C; … (Not the predictors finally included in the model, but the candidate predictors which were possibly included in the model. Separate each by ;)	CHARMS domain 4: Candidate predictors
		List predictors included in the final model	A; B; C; … (Separate each by ;)	
		Number of predictors in the final model	1, 2, 3…	
		Additional degrees of freedom of predictors	If a categorical variable has more than 2 categories, count all extra categories above 2. So for 3 categories, count 1 (3 - 2), for 5 categories count 3 (5 - 2). Do this for all categorical variables and add numbers. Add the number of interaction terms, and the number of degrees of freedom spent modeling non linearities.	
	Signaling questions	Were predictors defined and assessed in a similar way for all participants?	-Yes/probably yes: If definitions of predictors and their assessment were similar for all participants. -No/probably no: If different definitions were used for the same predictor or if predictors requiring subjective interpretation were assessed by differently experienced assessors.	PROBAST step 3, question 2.1

(Continued)

			-No information: If there is no information on how predictors were defined or assessed.	
		Were predictor assessments made without knowledge of outcome data?	-Yes/probably yes: If outcome information was stated as not used during predictor assessment or was clearly not (yet) available to those assessing predictors. -No/probably no: If it is clear that outcome information was used when assessing predictors. -No information: No information on whether predictors were assessed without knowledge of outcome information	PROBAST step 3, question 2.2
		Are all predictors available at the time the model is intended to be used?	-Yes/probably yes: All included predictors would be available at the time the model is intended to be used for prediction. -No/probably no: Predictors would not be available at the time the model is intended to be used for prediction. -No information: No information on whether predictors would be available at the time the model is intended to be used for prediction.	PROBAST step 3, question 2.3
	Risk of bias	Risk of bias introduced by predictors or their assessment	-Low risk of bias: If the answer to all signaling question is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low, e.g. use of objective predictors not requiring subjective interpretation. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put the domain at high risk of bias.	PROBAST step 3, risk of bias domain 2
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 2
	Applicability	Concern that the definition, assessment, or timing of predictors in the model do not match the review question.	-Low concern for applicability: Definition, assessment, and timing of predictors match the review question. -High concern for applicability: Definition, assessment, or timing of predictors were different from the review question. -Unclear concern for applicability: If relevant information about the predictors is not reported.	PROBAST step 3, applicability domain 2
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Outcome		Outcome	- Oral mucositis - Tube feeding dependence - Xerostomia - Disgeusia	CHARMS domain 3: Outcome

(Continued)

			- Dysphagia - Esophagitis - Hypothyoidism - Hearing loss - Tinnitus - Visual loss - Wound healing complications - Aspiration - Dermatitis - Head and neck pain - Hoarseness - Oral pain - Salivary duct inflammation - Sore throat - Weight loss - Fatigue - Osteoradionecrosis - Neurocognitive dysfunction - Cerebrovascular events - Speech problems - Jaw pain - Nausea and vomiting - Others (specify)	
		Definition	Describe the definition of the outcome (Copy and paste from the original article)	
		Time point of the outcome	Days/Months/Years	
	Signaling questions	Was the outcome determined appropriately?	-Yes/probably yes: If a method of outcome determination has been used which is considered optimal or acceptable by guidelines or previous publications on the topic. Note: this is about level of measurement error within the method of determining the outcome (see concerns for applicability about whether the definition of the outcome method is appropriate). -No/probably no: If a clearly suboptimal method has been used that causes unacceptable error in determining outcome status in participants. -No information: No information on how outcome was determined.	PROBAST step 3, question 3.1
		Was a prespecified or standard outcome definition used?	-Yes/probably yes: If the method of outcome determination is objective, or if a standard outcome definition is used, or if prespecified categories are used to group outcomes. -No/probably no: If the outcome definition was not standard and not prespecified. -No information: No information on whether the outcome definition was prespecified or standard.	PROBAST step 3, question 3.2
		Were predictors excluded from the outcome definition?	-Yes/probably Yes: If none of the predictors are included in the outcome definition. -No/probably No: If ≥1 of the predictors forms part of the outcome definition. -No information: No information on whether predictors are excluded from the outcome definition.	PROBAST step 3, question 3.3

(Continued)

		Was the outcome defined and determined in a similar way for all participants?	-Yes/probably yes: If outcomes were defined and determined in a similar way for all participants. -No/probably no: If outcomes were clearly defined and determined in a different way for some participants. -No information: No information on whether predictors are excluded from the outcome definition.	PROBAST step 3, question 3.4
		Was the outcome determined without knowledge of predictor information?	-Yes/probably yes: If predictor information was not known when determining the outcome status, or outcome status determination is clearly reported as determined without knowledge of predictor information. -No/probably no: If it is clear that predictor information was used when determining the outcome status. -No information: No information on whether outcome was determined without knowledge of predictor information.	PROBAST step 3, question 3.5
		Was the time interval between predictor assessment and outcome determination appropriate?	-Yes/probably yes: If the time interval between predictor assessment and outcome determination was appropriate to enable the correct type and representative number of relevant outcomes to be recorded, or if no information on the time interval is required to allow a representative number of the relevant outcome occur or if predictor assessment and outcome determination were from information taken within an appropriate time interval. -No/probably no: If the time interval between predictor assessment and outcome determination is too short or too long to enable the correct type and representative number of relevant outcomes to be recorded. -No information: If no information was provided on the time interval between predictor assessment and outcome determination	PROBAST step 3, question 3.6
	Risk of bias	Risk of bias introduced by the outcome or its determination	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low.If ≥1 of the answers is "No" or "Probably no", the judgement could still be low risk of bias, but specific reasons should be provided why the risk of bias can be considered low, e.g. when the outcome was determined with knowledge of predictor information but the outcome assessment did not require much interpretation by the assessor (e.g. death regardless of cause). -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information about the outcome is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.	PROBAST step 3, risk of bias domain 3
		Rationale of bias rating	'Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 3

(Continued)

	Applicability	Concern that the outcome definition, timing, or determination do not match the review question.	-Low concern for applicability: Outcome definition, timing, and method of determination defines the outcome as intended by the review question. -High concern for applicability: Choice of outcome definition, timing, and method of outcome determination defines another outcome as intended by the review question. -Unclear concern for applicability: If relevant information about the outcome, timing, and method of determination is not reported.	PROBAST step 3, applicability domain 3
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Analyses		Number of participants	Specify the number of participants included in the analysis	CHARMS domain 5: Sample size
		Number of participants with the outcome	Specify the number of participants with the outcome	
		Event per variable	Filled automatically	
		Handling of missing data	How was missing data handled in the analysis? Complete-case analysis Single imputation Multiple imputation Sensitivity analysis Variable omission Others (specify) Not reported	CHARMS domain 6: Missing data
		Number of participants with any missing predictor/outcome values	Specify number. When missing data is exclusion criteria, add 0. If not reported, write "NR".	
		Modeling method	- Linear regression - Logistic regression - Multinomial/polytomous regression - Ordinal regression - Cox proportional hazards regression - Accelerated failure time analysis (e.g. Weibull model) - Other parametric survival model - Others (specify) - Unclear	
		Method to deal with continuous variables	- Linear - Non-inear transformations - Categorized - Others (specify)	CHARMS domain 7: Model development
		Method for selection of predictors for inclusion	How were the predictors statistically selected before the modeling process? - All candidate predictors were used	

(Continued)

		in multivariable analysis	- Pre-selection based on univariable analysis - Principal component analysis - Variance inflation factor - Multicollinearity - Others (specify) - Not reported - Unclear	
		Method for selection of predictors during modeling process	How were the predictors statistically selected during the modeling process? - All predictors forced in the model - BW Stepwise selection - FW Stepwise selection - Added value - Lasso/Ridge - Not applicable - Others (specify) - Not reported - Unclear	
		Criteria for selection of predictors	E.g. P value < 0.05, AIC < 2, etc. Write "NA" if not applicable, "NR" if it is not reported.	
		Technique used for Internal validation	What kind of technique was used for internal validation? - Random split of dataset into development and testing datasets - Non-random split of dataset into Development and testing set - Internal validation by resampling of same dataset (e.g. bootstrap (specify the number of samples) - Cross-validation - Unclear method of splitting data - Not applicable - Other (specify)	
		Shrinkage	Was any shrinkage method used? - Not performed - Heuristic shrinkage (uniform) - Calibration slope assessed with bootstrapping (uniform) - Penalized maximum Likelihood estimation - Lasso/Ridge - Others (specify) - Unclear	
		Model presentation	How was the model presented? - Mathematical equation (only coefficients) - Mathematical equation (both intercept and coefficients) - Simple scoring system - Nanogram - Web calculator - Others (specify)	
	Signaling questions	Were there a reasonable number of participants with the outcome?	- Yes/probably yes: For model development studies, if the number of participants with the outcome relative to the number of candidate predictors parameters is ≥20 (EPV ≥20). For EPV between 10 and 20, the item should be rated as either probably yes	PROBAST step 3, question 4.1

(Continued)

		or probably no, depending on the outcome frequency, overall model performance, and distribution of the predictors in the model. -No/probably no: For model development studies, if the number of participants with the outcome relative to the number of candidate predictor parameters is <10 (EPV <10). -No information: For model development studies, no information on the number of candidate predictor parameters or number of participants with the outcome, such that the EPV cannot be calculated.	
	Were continuous and categorical predictors handled appropriately?	-Yes/probably yes: If continuous predictors are not converted into ≥2 categories when included in the model (i.e. dichotomized or categorized), or if continuous predictors are examined for non linearity using, for example, fractional polynomials or restricted cubic splines, or if categorical predictor groups are defined using a prespecified method. For model validation studies, if continuous predictors are included using the same definitions or transformations, and categorical variables are categorized using the same cut points, as compared with the development study. -No/probably no: If categorical predictor group definitions do not use a prespecified method. For model development studies, if continuous predictors are converted into ≥2 categories when included in the model. For model validation studies, if continuous predictors are included using different definitions or transformations, or categorical variables are categorized using different cut points, as compared with the development study. -No information: No information on whether continuous predictors are examined for non linearity and no information on how categorical predictor groups are defined. For model validation studies, no information on whether the same definitions or transformations and the same cut points are used, as compared with the development study.	PROBAST step 3, question 4.2
	Were all enrolled participants included in the analysis?	-Yes/probably yes: If all participants enrolled in the study are included in the data analysis. -No/probably no: If some or a subgroup of participants are inappropriately excluded from the analysis. -No information: No information on whether all enrolled participants are included in the analysis.	PROBAST step 3, question 4.3
	Were participants with missing data handled appropriately?	-Yes/probably yes:If there are no missing values of predictors or outcomes and the study explicitly reports that participants are not excluded on the basis of missing data, or if missing values are handled using multiple imputation. -No/probably no: If participants with missing data are omitted from the analysis, or if the method of handling missing data is clearly flawed, e.g. missing indicator method or inappropriate use of last value carried forward, or if the study had nog explicit mention of methods to handle missing data.	PROBAST step 3, question 4.4

(Continued)

		-No information: If there is insufficient information to determine if the method of handling missing data is appropriate.	
	Was selection of predictors based on univariable analysis avoided?	-Yes/probably yes: If the predictors are not selected on the basis of univariable analysis prior to multivariable modeling. -No/probably no: If the predictors are selected on the basis of univariable analysis prior to multivariable modeling. -No information: If there is no information to indicate that univariable selection is avoided.	PROBAST step 3, question 4.5
	Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted for appropriately?	-Yes/probably yes: If any complexities in the data are accounted for appropriately, or if it is clear that any potential data complexities have been identified appropriately as unimportant. -No/probably no: If complexities in the data that could affect model performance are ignored. -No information: No information is provided on whether complexities in the data are present or accounted for appropriately if present.	PROBAST step 3, question 4.6
	Were relevant model performance measures evaluated appropriately?	-Yes/probably yes: If both calibration and discrimination are evaluated appropriately (including relevant measures tailored for models predicting survival outcomes). -No/probably no: If both calibration and discrimination are not evaluated, or if only goodness-of-fit tests, such as the Hosmer-Lemeshow test, are used to evaluate calibration, or if models predicting survival outcomes performance measures accounting for censoring ae not used, or if classification measures (like sensitivity, specificity, or predictive values) were presented using predicted probability thresholds derived from the dataset at hand. -No information: Either calibration or discrimination are not reported, or no information is provided as to whether appropriate performance measures for survival outcomes are used (e.g. references to relevant literature or specific mention of methods, such as using Kaplan-Meier estimates), or no information on thresholds for estimating classification measures is given.	PROBAST step 3, question 4.7
	Were model overfitting and optimism in model performance accounted for?	-Yes/probably yes: If internal validation techniques, such as bootstrapping and cross-validation including all model development procedures, have been used to account for any optimism in model fitting, and subsequent adjustment of the model performance estimates have been applied. -No/probably no: If no internal validation has been performed, or if internal validation consists only of a single random split-sample of participants data, or if the bootstrapping or cross-validation did not include all model development procedures including any variable selection. -No information: No information is provided on whether internal validation techniques, including	PROBAST step 3, question 4.8

(Continued)

			all model development procedures, have been applied.	
		Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	-Yes/probably yes: If the predictors and regression coefficients in the final model correspond to reported results from multivariable analysis. -No/probably no: If the predictors and regression coefficients in the final model do not correspond to reported results from multivariable analysis -No information: If it is unclear whether the regression coefficients in the final model correspond to reported results from multivariable analysis.	PROBAST step 3, question 4.9
	Risk of bias	Risk of bias introduced by the analysis	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgement could still be low risk of bias, but specific reasons should be provided why the risk of bias can be considered low. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information about the analysis is missing for some of the signaling questions but none of the signaling questions is judged to put the analysis at high risk of bias.	PROBAST step 3, risk of bias domain 4
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 4
Case-mix		Age	Mean	
			SD	
			Median	
			IQR 25th percentile	
			IQR 75th percentile	
			Other	
		Sex	% of men	
		Baseline complaints about the side-effect	If the outcome of the study is disphagia, report the proportion of patients who complained disphagia at baseline.	
		Type of cancer	% of oral cavity cancer	
			% of larynx cancer	
			% of hypopharynx cancer	
			% of oropharynx cancer	
			% of nasopharynx cancer	

(Continued)

			% of others
		Stage	% of stage I
			% of stage II
			% of stage III
			% of stage IV
		Type of radiotherapy	% of IMRT
			% of VMAT
			% of 3D conformal
			% of brachytherapy
			% of Proton therapy
			Radiation target dose (Gy)
			Radiation fractionation - Conventional fractionation - Hypofractionation - Hyperfractionation - Accelerated - Accelerated hyperfractionation - Others (Specify)
		Treatment	% of concurrent chemotherapy
			% of induction chemotherapy
			% of adjuvant chemotherapy
			% of concurrent molecular-targeted therapy
Model performance	Apparent performance	Apparent performance was reported?	Yes or No
		C-statistics	
		C-statistics SE	
		C-statistics LCI	
		C-statistics UCI	
		Calibration plot	If yes, which figure?
		Other	Brier score, O:E ratio, etc… Describe as point estimate (SE) or point estimate (95% CI) as possible.

(Continued)

	Internal validation	Internal validation was performed?	Yes or No If no, move to the section of comment.
		C-statistics	
		C-statistics SE	
		C-statistics LCI	
		C-statistics UCI	
		Calibration plot	If yes, which figure?
		Other	Brier score, O:E ratio, etc… Describe as point estimate (SE) or point estimate (95% CI) as possible.
		Comments	Any other comments?

Appendix 4. Preliminary study selection, data extraction and risk of bias forms for external validation studies

Domain		Item to answer	Explanation/Classification	Reference
General information		Data extracted by	Your name	
		First author	Surname of the 1st author	
		Year of publication	yyyy	
		Sub ID	If the study reports multiple models (e.g. different outcomes, different presentations, etc), enter the items in different rows. For example, if there are two models for 30-days and 6-months mortality, label them as "1st author's surname-1" and "1st author's surname-2". Also, if the model was externally validated, enter the information about the external validation in another row.	
		Sub ID details	Describe the detail of the model to distinguish from each other. (For example of the left cell, describe as "model for predicting 30-days mortality" and "external validation", respectively.)	
Study characteristics	Study description	Validation model	Prvide the complete reference of the original study in which the validated model was developed. Authors, title, journal, year, volume, issue, pages	CHARMS domain 1: Source of data
		Indicate which version of the model was validated	- Original model with regression coefficients - Simplified model based on a risk score - Others (specify)	

(Continued)

		Number of centers	1, 2, 3...	
		Location	- North America - Europe - Asia - Australia - Africa - South America - Central America - Combination	
Participants	Study description	Data source	- Prospective cohort - Retrospective cohort - Nested case-control - Nonnested case-control - Case-cohort - RCT - Registry data - Others (specify) - Unclear	CHARMS domain 2: Participants
		Recruitment method	Were patients recruited consecutively?	
		Start date recruitment period	dd-mm-yyyy If the exact date is not reported (e.g. only year is reported), just enter available information (e.g. yyyy).	
		End date recruitment period	dd-mm-yyyy If the exact date is not reported (e.g. only year is reported), just enter available information (e.g. yyyy).	
		Inclusion criteria	copy and paste from the article text	
		Exclusion criteria	copy and paste from the article text	
		Primary site of cancer	Primary site - Oral cavity - Larynx - Hypopharynx - Oropharynx - Nasopharyx - Salivary glands - Nasal cavity - Paranasal sinus - Eye - Lymphoma - Soft tissue - Skull base - Ear - Lip - Thyroid - Others (Specify)	
		Type of cancer	Pathology of cancer - SCC - Adenocarcinoma - Melanoma	

(Continued)

			- Sarcoma - Lyphoma - Others (Specify) - Unclear	
		TNM classification	Which version of the classification was used? - 7th version (before 2017) - 8th version (after 2017)	
		Stage	Based on TNM classification (I, II, III, IV…)	
		Type of radiotherapy	- IMRT - VMAT - 3D conformal - Proton - Electron - Combined - Others (Specify) - Unclear	
		Surgery	- No surgery - Pre-surgery radiotherapy - Post-surgery radiotherapy - Non specified	
		Type of surgery if specified	Detailed information about surgery (e.g. laryngectomy) Copy and paste from the paper	
		Chemotherapy	At baseline, - No chemotherapy - Concurrent chemotherapy - Chemotherapy before radiotherapy - Chemotherapy after radiotherapy - Case mix with different radiotherapy and chemotherapy combinations - Non-specified - Others (Specify)	
		Molecular-targeted therapy	At baseline, - No molecular-targeted therapy - Concurrent molecular-targeted therapy - Molecular-targeted therapy before radiotherapy - Molecular-targeted therapy after radiotherapy - Case mix with different radiotherapy and molecular therapy combinations - Non-specified - Others (Specify)	
		Other specific patient characteristics		
	Signaling questions	Were appropriate data sources used, e.g. cohort, RCT, or nested case-control study data?	- Yes/probably yes: If a cohort design (including RCT or proper registry data) or a nested case-control or case-cohort design (with proper adjustment of the baseline risk/hazard in the analysis) has been used. - No/probably no: If a non nested case-control design has been used.	PROBAST step 3, question 1.1

(Continued)

			- No information: If the method of participant sampling is unclear.	
		Were all inclusions and exclusions of participants appropriate?	-Yes/probably yes: If inclusion and exclusion of participants was appropriate, so participants correspond to unselected participants of interest. -No/probably no: If participants are included who would already have been identified as having the outcome and so are no longer participants at risk of developing outcome, or if specific subgroups are excluded that may have altered the performance of the prediction model for the intended target population. -No information: When there is no information on whether inappropriate inclusions or exclusions took place.	PROBAST step 3, question 1.2
	Risk of bias	Risk of bias introduced by participants or data sources	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgment could still be "Low risk of bias", but specific reasons should be provided why the risk of bias can be considered low. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias, except if defined at low risk of bias above. - Unclear risk of bias: If relevant information is missing for some of the signaling questions and non of the signaling questions is judged to put this domain at high risk of bias.	PROBAST step 3, risk of bias domain 1
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 1
	Applicability	Concern that included participants or the setting do not match the review question	-Low concern for applicability: Included participants and clinical setting match the review question. -High concern for applicability: Included participants and clinical setting were different from the review question. -Unclear concern for applicability: If relevant information about the participants and clinical setting are not reported.	PROBAST step 3, applicability domain 1
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Predictors	Study descriptions	List predictors included in the final model	A; B; C; … (Separate each by ;)	CHARMS domain 4: Candidate predictors
		Additional degrees of freedom of predictors	If a categorical variable has more than 2 categories, count all extra categories above 2. So for 3 categories, count 1 (3 - 2), for 5 categories count 3 (5 - 2). Do this for all categorical variables and add numbers. Add the number of interaction terms,	

(Continued)

			and the number of degrees of freedom spent modeling non linearities.	
Predictors	Signaling questions	Were predictors defined and assessed in a similar way for all participants?	-Yes/probably yes: If definitions of predictors and their assessment were similar for all participants. -No/probably no: If different definitions were used for the same predictor or if predictors requiring subjective interpretation were assessed by differently experienced assessors. -No information: If there is no information on how predictors were defined or assessed.	PROBAST step 3, question 2.1
		Were predictor assessments made without knowledge of outcome data?	-Yes/probably yes: If outcome information was stated as not used during predictor assessment or was clearly not (yet) available to those assessing predictors. -No/probably no: If it is clear that outcome information was used when assessing predictors. -No information: No information on whether predictors were assessed without knowledge of outcome information	PROBAST step 3, question 2.2
		Are all predictors available at the time the model is intended to be used?	-Yes/probably yes: All included predictors would be available at the time the model is intended to be used for prediction. -No/probably no: Predictors would not be available at the time the model is intended to be used for prediction. -No information: No information on whether predictors would be available at the time the model is intended to be used for prediction.	PROBAST step 3, question 2.3
	Risk of bias	Risk of bias introduced by predictors or their assessment	-Low risk of bias: If the answer to all signaling question is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgment could still be "Low risk of bias" but specific reasons should be provided why the risk of bias can be considered low, e.g. use of objective predictors not requiring subjective interpretation. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information is missing for some of the signaling questions and none of the signaling questions is judged to put the domain at high risk of bias.	PROBAST step 3, risk of bias domain 2
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 2
	Applicability	Concern that the definition, assessment, or timing of predictors in the model do not match the review question.	-Low concern for applicability: Definition, assessment, and timing of predictors match the review question. -High concern for applicability: Definition, assessment, or timing of predictors were different from the review question.	PROBAST step 3, applicability domain 2

(Continued)

			-Unclear concern for applicability: If relevant information about the predictors is not reported.	
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Outcome		Outcome	- Oral mucositis - Tube feeding dependence - Xerostomia - Disgeusia - Dysphagia - Esophagitis - Hypothyoidism - Hearing loss - Tinnitus - Visual loss - Wound healing complications - Aspiration - Dermatitis - Head and neck pain - Hoarseness - Oral pain - Salivary duct inflammation - Sore throat - Weight loss - Fatigue - Osteoradionecrosis - Neurocognitive dysfunction - Cerebrovascular events - Speech problems - Jaw pain - Nausea and vomiting - Others (specify)	CHARMS domain 3: Outcome
		Definition	Describe the definition of the outcome (Copy and paste from the original article)	
		Time point of the outcome	Days/Months/Years	
	Signaling questions	Was the outcome determined appropriately?	-Yes/probably yes: If a method of outcome determination has been used which is considered optimal or acceptable by guidelines or previous publications on the topic. Note: this is about level of measurement error within the method of determining the outcome (see concerns for applicability about whether the definition of the outcome method is appropriate). -No/probably no: If a clearly suboptimal method has been used that causes unacceptable error in determining outcome status in participants. -No information: No information on how outcome was determined.	PROBAST step 3, question 3.1
		Was a prespecified or standard outcome definition used?	-Yes/probably yes: If the method of outcome determination is objective, or if a standard outcome definition is used, or if prespecified categories are used to group outcomes.	PROBAST step 3, question 3.2

(Continued)

			-No/probably no: If the outcome definition was not standard and not prespecified. -No information: No information on whether the outcome definition was prespecified or standard.	
		Were predictors excluded from the outcome definition?	-Yes/probably Yes: If none of the predictors are included in the outcome definition. -No/probably No: If ≥1 of the predictors forms part of the outcome definition. -No information: No information on whether predictors are excluded from the outcome definition.	PROBAST step 3, question 3.3
		Was the outcome defined and determined in a similar way for all participants?	-Yes/probably yes: If outcomes were defined and determined in a similar way for all participants. -No/probably no: If outcomes were clearly defined and determined in a different way for some participants. -No information: No information on whether predictors are excluded from the outcome definition.	PROBAST step 3, question 3.4
		Was the outcome determined without knowledge of predictor information?	-Yes/probably yes: If predictor information was not known when determining the outcome status, or outcome status determination is clearly reported as determined without knowledge of predictor information. -No/probably no: If it is clear that predictor information was used when determining the outcome status. -No information: No information on whether outcome was determined without knowledge of predictor information.	PROBAST step 3, question 3.5
		Was the time interval between predictor assessment and outcome determination appropriate?	-Yes/probably yes: If the time interval between predictor assessment and outcome determination was appropriate to enable the correct type and representative number of relevant outcomes to be recorded, or if no information on the time interval is required to allow a representative number of the relevant outcome occur or if predictor assessment and outcome determination were from information taken within an appropriate time interval. -No/probably no: If the time interval between predictor assessment and outcome determination is too short or too long to enable the correct type and representative number of relevant outcomes to be recorded. -No information: If no information was provided on the time interval between predictor assessment and outcome determination	PROBAST step 3, question 3.6
	Risk of bias	Risk of bias introduced by the outcome or its determination	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low.If ≥1 of the answers is "No" or "Probably no", the judgement could still be low risk of bias, but specific reasons should be provided why the risk of bias can be considered low, e.g. when the outcome was determined with knowledge of predictor information but the outcome assessment did not require much interpretation by the assessor (e.g. death regardless of cause).	PROBAST step 3, risk of bias domain 3

(Continued)

			-High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information about the outcome is missing for some of the signaling questions and none of the signaling questions is judged to put this domain at high risk of bias.	
		Rationale of bias rating	'Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 3
	Applicability	Concern that the outcome definition, timing, or determination do not match the review question.	-Low concern for applicability: Outcome definition, timing, and method of determination defines the outcome as intended by the review question. -High concern for applicability: Choice of outcome definition, timing, and method of outcome determination defines another outcome as intended by the review question. -Unclear concern for applicability: If relevant information about the outcome, timing, and method of determination is not reported.	PROBAST step 3, applicability domain 3
		Rationale of applicability rating	Provide specific reasons for classification of applicability if necessary	PROBAST step 3, applicability domain 1
Analyses		Number of participants	Specify the number of participants included in the analysis	CHARMS domain 5: Sample size
		Number of participants with the outcome	Specify the number of participants with the outcome	CHARMS domain 5: Sample size
		Handling of missing data	How was missing data handled in the analysis? Complete-case analysis Single imputation Multiple imputation Sensitivity analysis Variable omission Others (specify) Not reported	CHARMS domain 6: Missing data
		Number of participants with any missing predictor/outcome values	Specify number. When missing data is exclusion criteria, add 0. If not reported, write "NR".	
		Model updated?	Was the model updated? - Yes - No	CHARMS domain 9: Model evaluation
		The method for update	If yes, which method was used to update the model? - None - Recalibration-in-the-large - Recalibration - Complete model revision	

(Continued)

	Signaling questions	Were there a reasonable number of participants with the outcome?	- Yes/probably yes: For model validation studies, if the number of participants with the outcome is ≥100. -No/probably no: For model validation studies, if the number of participants with the outcome is <100. -No information: For model validation studies, no information on the number of participants with the outcome.	PROBAST step 3, question 4.1
		Were all enrolled participants included in the analysis?	-Yes/probably yes: If all participants enrolled in the study are included in the data analysis. -No/probably no: If some or a subgroup of participants are inappropriately excluded from the analysis. -No information: No information on whether all enrolled participants are included in the analysis.	PROBAST step 3, question 4.3
		Were participants with missing data handled appropriately?	-Yes/probably yes:If there are no missing values of predictors or outcomes and the study explicitly reports that participants are not excluded on the basis of missing data, or if missing values are handled using multiple imputation. -No/probably no: If participants with missing data are omitted from the analysis, or if the method of handling missing data is clearly flawed, e.g. missing indicator method or inappropriate use of last value carried forward, or if the study had nog explicit mention of methods to handle missing data. -No information: If there is insufficient information to determine if the method of handling missing data is appropriate.	PROBAST step 3, question 4.4
		Were relevant model performance measures evaluated appropriately?	-Yes/probably yes: If both calibration and discrimination are evaluated appropriately (including relevant measures tailored for models predicting survival outcomes). -No/probably no: If both calibration and discrimination are not evaluated, or if only goodness-of-fit tests, such as the Hosmer-Lemeshow test, are used to evaluate calibration, or if models predicting survival outcomes performance measures accounting for censoring ae not used, or if classification measures (like sensitivity, specificity, or predictive values) were presented using predicted probability thresholds derived from the dataset at hand. -No information: Either calibration or discrimination are not reported, or no information is provided as to whether appropriate performance measures for survival outcomes are used (e.g. references to relevant literature or specific mention of methods, such as using Kaplan-Meier estimates), or no information on thresholds for estimating classification measures is given.	PROBAST step 3, question 4.7
		Do predictors and their assigned weights in the final model correspond	-Yes/probably yes: If the predictors and regression coefficients in the final model correspond to reported results from multivariable analysis.	PROBAST step 3, question 4.9

(Continued)

		to the results from the reported multivariable analysis?	-No/probably no: If the predictors and regression coefficients in the final model do not correspond to reported results from multivariable analysis -No information: If it is unclear whether the regression coefficients in the final model correspond to reported results from multivariable analysis.	
	Risk of bias	Risk of bias introduced by the analysis	-Low risk of bias: If the answer to all signaling questions is "Yes" or "Probably yes", then risk of bias can be considered low. If ≥1 of the answers is "No" or "Probably no", the judgement could still be low risk of bias, but specific reasons should be provided why the risk of bias can be considered low. -High risk of bias: If the answer to any of the signaling questions is "No" or "Probably no", there is a potential for bias. -Unclear risk of bias: If relevant information about the analysis is missing for some of the signaling questions but none of the signaling questions is judged to put the analysis at high risk of bias.	PROBAST step 3, risk of bias domain 4
		Rationale of bias rating	Provide specific reasons for classification of bias if necessary	PROBAST step 3, risk of bias domain 4
Case-mix		Age	Mean	
			SD	
			Median	
			IQR 25th percentile	
			IQR 75th percentile	
			Other	
		Sex	% of men	
		Baseline complaints about the side-effect	If the outcome of the study is disphagia, report the proportion of patients who complained disphagia at baseline.	
		Type of cancer	% of oral cavity cancer	
			% of larynx cancer	
			% of hypopharynx cancer	
			% of oropharynx cancer	
			% of nasopharynx cancer	
			% of others	
		Stage	% of stage I	
			% of stage II	

(Continued)

			% of stage III
			% of stage IV
		Type of radiotherapy	% of IMRT
			% of VMAT
			% of 3D conformal
			% of brachytherapy
			% of Proton therapy
			Radiation target dose (Gy)
			Radiation fractionation - Conventional fractionation - Hypofractionation - Hyperfractionation - Accelerated - Accelerated hyperfractionation - Others (Specify)
		Treatment	% of concurrent chemotherapy
			% of induction chemotherapy
			% of adjuvant chemotherapy
			% of concurrent molecular-targeted therapy
Model performance	Model performance without update	Model performance with update was reported?	Yes or No
		C-statistics	
		C-statistics SE	
		C-statistics LCI	
		C-statistics UCI	
		Calibration plot	If yes, which figure?
		Other	
	Model performance with update	Model performance with update was reported?	Yes or No
		C-statistics	

(Continued)

	C-statistics SE	
	C-statistics LCI	
	C-statistics UCI	
	Calibration plot	If yes, which figure?
	Other	
	Comments	Any other comments?

CONTRIBUTIONS OF AUTHORS

Toshihiko Takada: protocol development, content input and medical input

Johanna AAG Damen: methodological, statistical and content input

Makbule Tambas: medical and content input

René Spijker: methodological and content input

Roel JHM Steenbakkers: medical and content input

Marjan Sharabiani: medical and content input

Enrico Clementel: medical and content input

Johannes A Langendijk: medical and content input

Karel GM Moons: methodological, statistical and content input

Ewoud Schuit: methodological, statistical and content input

DECLARATIONS OF INTEREST

Toshihiko Takada: none known.

Johanna AAG Damen: none known.

Makbule Tambas: none known.

René Spijker: none known.

Roel JHM Steenbakkers: none known.

Marjan Sharabiani: received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825162 for the current work.

Enrico Clementel: received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825162 for the current work.

Johannes A Langendijk: received research grants from the Dutch Cancer Society and EU and ravel grants from ESTRO, EHNS, BIR. The department of Radiation Oncology has several research grants with companies that are mainly focusing on improving technological developments in the clinic. The department uses models to guide treatment planning optimization in the software provided or delivered by the companies. Is a board member of the scientific advisory committee of RaySearch and IBA, which also promote using models for treatment optimization. The author will serve as a content expert and will not be involved in data extraction or assessment.

Karel GM Moons: none known.

Ewoud Schuit: received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825162 for the current work.

SOURCES OF SUPPORT

Internal sources

- Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands
- Department of Radiation Oncology, University Medical Center Groningen, Groningen University, Netherlands
- European Organisation for Research and Treatment of Cancer, Belgium

External sources

- European Union's Horizon 2020 research and innovation programme under grant agreement No 825162, Netherlands