Risk prediction models for head and neck cancer: A rapid review

Craig D. L. Smith BSc1,2 | Alex D. McMahon PhD1 | Alastair Ross PhD1 | David I. Conway PhD1,2
Gareth J. Inman PhD2,3

1School of Medicine, Dentistry, and Nursing, University of Glasgow, Glasgow, UK
2Institute of Cancer Sciences, University of Glasgow, Glasgow, UK
3Cancer Research UK Beatson Institute, Glasgow, UK

Correspondence
Craig D. L. Smith, Community Oral Health, Glasgow Dental Hospital and School, 378 Sauchiehall St, Glasgow G2 3JZ, UK.
Email: 2298108s@student.gla.ac.uk

Funding information
Cancer Research UK, Grant/Award Number: 315941-01

Abstract
Background: Cancer risk assessment models are used to support prevention and early detection. However, few models have been developed for head and neck cancer (HNC).

Methods: A rapid review of Embase and MEDLINE identified n = 3045 articles. Following dual screening, n = 14 studies were included. Quality appraisal using the PROBAST (risk of bias) instrument was conducted, and a narrative synthesis was performed to identify the best performing models in terms of risk factors and designs.

Results: Six of the 14 models were assessed as “high” quality. Of these, three had high predictive performance achieving area under curve values over 0.8 (0.87–0.89). The common features of these models were their inclusion of predictors carefully tailored to the target population/anatomical subsite and development with external validation.

Conclusions: Some existing models do possess the potential to identify and stratify those at risk of HNC but there is scope for improvement.

Key words
head and neck cancer, review, risk, risk assessment, risk model

1 | INTRODUCTION

Head and neck squamous cell carcinomas—generally defined as aerodigestive squamous cancers of the oral cavity, larynx, and pharynx—are a growing challenge for healthcare systems across the world: they are the eighth most common cancer, accounting for an estimated 878,348 new cases and 444,347 deaths globally in 2020.1,2 The risk profile of head and neck cancer (HNC) is also changing—with oropharyngeal cancer increasingly associated with human papillomavirus (HPV) infection,3 and inequalities in the burden of HNC associated with socioeconomic status.4

Overall HNC survival varies greatly by subsite and stage of diagnosis. Despite advancement in treatments, 5-year survival has seen no major improvements observed in recent decades.5-7 Furthermore, marginal improvements in prognosis may be undercut by the overall increased disease burden, particularly due to the changing epidemiology of HPV-associated oropharyngeal cancer.8,9 As with all cancers, prognosis is worse with advanced stage disease at presentation. Thus, a major challenge posed by HNC is its traditionally late presentation with over half of cases diagnosed at stage III or IV, when locally advanced or regional or distant metastases are present.10-14

Given the twin challenges of increasing HNC incidence and poor survival associated with late-stage detection, further attention needs to be given to primary and secondary prevention strategies—utilizing the potential of head and neck risk prediction models to identify those...
at risk and direct them to appropriate prevention and early detection/diagnosis and treatment pathways.

There has already been some success and clinical adoption of other cancer prediction models in primary care; for example, the Q-series risk prediction models or cancer Risk Assessment Tool.\textsuperscript{15,16} These models have been well evaluated, demonstrating the potential of “personalized medicine” to identify and stratify those at risk.\textsuperscript{17–21} However, they do not assess for HNC risk and there seem to be few risk prediction models for HNC developed or adopted for clinical use. Furthermore, there have been no comprehensive reviews of head and neck risk cancer prediction models or tools published. The aim of this study is to undertake such a review—via systematically searching and identifying models in the international literature, describing their characteristics and performance, quality appraising these models, and performing a narrative synthesis to compare and contrast risk prediction models for HNC.

2 | METHODS

A rapid review methodology was employed, following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.\textsuperscript{22} The review was also based on similar reviews on risk prediction models of other cancer sites/diseases.

2.1 | Search strategy

An electronic literature search of Ovid MEDLINE(R), (and In-Process, In-Data-Review & Other Non-Indexed Citations), and Embase (1947–Present, updated daily) databases was conducted using a combination of key headings and search terms associated with “head and neck cancer” and “risk/risk factor/risk assessment” and “prediction/model/tool/score” (see Supplementary Material file for the full list of search terms).

Studies were included if they satisfied all of the following criteria: (i) used a statistical model/tool to predict HNC risk including subsites and potentially malignant conditions; (ii) were published in English; (iii) considered multiple different risk factors; (iv) provided a measurement of risk; and (v) were applicable to the general population.

Given that the focus of this review was on risk prediction, studies that developed prognostic or recurrence models were excluded. Similarly, studies that only considered highly selected groups or risk variables such as highly specific genes were also excluded (as per [v] of exclusion criteria). If multiple publications of the same model were identified, the most extensive and recent report of the model was included.

The reasoning behind a statistical model/tool forming a part of the inclusion criteria was to ensure that there was a robust methodology underlying model development. Crucially, this was also to separate risk models from numerous case-control studies that considered multiple risk factors individually, often expressing these in odds ratios but not evaluating these in the form or context of a risk prediction model or tool. Reporting of a comparable measure of risk was also considered to be, at least, reporting of performance metrics (e.g., AUC) but ideally assigning a value to an individual (e.g., 5-year risk).

Articles that were ultimately sourced from search, were loaded into Endnote X9 (Clarivate Analytics) reference management software and from here were imported into Covidence (Covidence systematic review software, Veritas Health Innovation), which was used to remove duplicates and perform study screening and data extraction.

2.2 | Screening and study selection

Two reviewers (CS, DIC) independently screened search results at a title/abstract level and then at a full-text level using the eligibility criteria. In the event of a disagreement, articles were discussed and included or excluded by mutual consensus.

2.3 | Data extraction and quality assessment

Following full-text screening, data extraction was undertaken by two reviewers (CS, DIC) using a customized form containing pre-defined fields including items on: study characteristics (location, study design, cancer site/subsite, and risk factors included). The data extraction form also assessed the requirement of clinician input (based on whether reported or the nature of the data required to run the model—for instance a patient would not be able to use machine learning tools or conduct HPV serology analysis), along with items on predictive performance (discrimination, sensitivity/specificity, calibration, positive predictive value/negative predictive value [PPV/NPV] and risk threshold cut-offs) and the method of validation (if undertaken). Measures of discrimination were considered to be “acceptable” if an area under the curve (AUC) value over 0.7 was reported, and “excellent” if a value greater than 0.8 was reported.\textsuperscript{23} Measures of calibration were assessed by the expected/observed ratio or gradient of a calibration slope to the ideal value of 1 and of its intercept to the value of 0.\textsuperscript{24,25}

Two reviewers (CS, DIC) also examined the risk of bias of each model using PROBAST, a tool specifically designed to appraise clinical risk prediction models.\textsuperscript{26} Risk of bias (‘high’, “low,” or “unclear”) and applicability of the risk models was assessed using 20 questions across four domains (participants, predictors, analysis, and outcomes).

An overall quality assessment was also given to each model considering model validation, and the risk of bias, and applicability concerns assessment (from PROBAST). This was categorized as “High,” “Moderate,” or “Low.” If the model had a (i) low risk of bias, (ii) a low or unclear applicability concern, and (iii) a robust method of validation then it was considered “High” overall quality. Model performance was also considered separately by evaluating each model’s discriminative ability. If a model achieved an “excellent” AUC over 0.8 it was classed as high performing (green in table).\textsuperscript{27} Models that achieved acceptable discrimination between 0.7 and 0.8 were classed as moderate (amber
in table) and discrimination less than 0.7 was classed as poor (red in table). These classifications also reflected for confidence intervals for model AUC (where reported).

2.4 | Synthesis

The heterogeneous nature of risk prediction models makes the possibility of pooling the data between the models inappropriate. However, a narrative synthesis was conducted—focusing on the model overall quality/performance and including comparing and contrasting risk factors used in the risk prediction models—grouping them as sociodemographic factors (e.g., age, sex, socioeconomic characteristics), behaviors (smoking, alcohol), biomarkers (e.g., HPV, genetic/polygenic data), clinical information (e.g., symptoms, oral potentially malignant disorders). Models were also compared across subsites of HNC.

3 | RESULTS

Following the removal of 100 duplicates, 2945 studies were identified by the search. Of these, 2900 were excluded by title or abstract screening. Of the remaining 45 studies, 34 were excluded following full-text assessment, with reasons for exclusion noted (Figure 1). The most common reasons for exclusion were studies did not use a statistical method to develop a risk model, or did not consider multiple risk factors.
| Study ID | Study characteristics | Participants | Cancer outcome/site + sub models | Method + factors included in model | Clinician input |
|----------|-----------------------|--------------|----------------------------------|-----------------------------------|----------------|
| Amarasinghe et al. (2010) | Sri Lanka/Community-based case-control November 2006–November 2007 | 101 OPMD cases and 728 controls (men and women) | OPMD excluding lichen planus | Multivariate logistic regression was used to develop the model. Effect estimates of each factor on the risk of OPMD were derived. Gradients for each factor were given a score derived from adjusted odds ratio and ultimately used to develop a 12-point score cut off model. Predictors used included age, socioeconomic status, betel-Quid chewing, alcohol drinking, smoking. | No |
| Budhathoki et al. (Unpublished) | North America/Europe—four case-control studies, one prospective cohort—a mixture of hospital and population-based case-controls. Recruitment dates—not specified | 10,126 head and neck cancer cases and 5254 controls | Head and neck cancer (10,126), oral cavity cancer (2431), and oropharyngeal cancer (3727) Six sub models for three sites including separate models for men and women | Predictors were selected based on those that were statistically significant from univariate logistic regression via backwards stepwise selection. The final models used multivariate logistic regression to assess risk with separate models for men and women. Predictors include smoking status, drinking status, education, HPV serostatus and polygenic risk score | Yes |
| Chen et al. (2018) | China—Hospital based case-control Conducted from September 2010 to March 2017 | 978 cases and 2646 controls SEPAREATE models—1924 men and 1700 women. | Oral cancer (tongue, buccal, gingiva, floor of mouth, palate, lip, and unspecified or overlapping) (380 tongue, 135 buccal, 128 gingival, 72 floor of mouth, 69 palate, 34 lip, and 160 unspecified or overlapping) | Unconditional logistic regression was used, independently significant variables being included in the final nomogram models. Different sets of predictors were used to create separate nomograms for males and females. Men; (smoking [pack-year], alcohol drinking, tea consumption, fish, seafood, vegetables, fruits, teeth loss, regular dental visits, repetitive dental ulcer) Women; (passive smoking, cooking oil fume exposure, tea consumption, vegetables, fruits, beans, teeth loss, regular dental visits, repetitive dental ulcer, age of first intercourse) | No |
| Cheung et al. (2021) | India—Analysis of a previous Oral Cancer Randomized Control Trial conducted Screening was conducted in three waves with a 3 year interval between each indidual screening—1996–1998, 1999–2001, and 2002–2004 | 95,354 control arm 96,516 screening arm (male and female) | Oral cancer | Model was developed using a Cox regression-based risk prediction model for 7-year oral cancer incidence. Follow up time was used for a time scale and covariates were selected prior to analysis. Predictors included sex, age, education, BMI, tobacco chewing, tobacco smoking, chewing-smoking interaction, and alcohol use. | No |
| Study ID | Study characteristics | Components of model | Cancer outcome/site + sub models | Method + factors included in model | Clinician input |
|----------|-----------------------|---------------------|----------------------------------|-----------------------------------|-----------------|
| Koyanagi et al. (2017) | Japan—Hospital-based case control 2001—December 2005 365 HNC cases and 1260 controls (male and female) | Three separate models—Head and neck cancer (model of interest), upper aerodigestive tract cancer, esophageal cancer | Three models were constructed for each subsite: a genetic, an environmental and an inclusive model. The latter, categories of alcohol and ALDH2 were coded as dummy variables. Interaction terms were used to assess the combined impact of alcohol and ALDH2 interaction. The models were derived using conditional logistic regression models. Predictors included age, sex, ALDH2 genotype, cumulative smoking and alcohol consumption. | | |
| Krishna Rao et al. (2016) | India—Hospital based, unmatched case–control Conducted between July 2011 and August 2012. 180 cases and 272 controls (men and women) | Oral cavity and oropharynx cancer | Multivariable logistic regression was used to identify significant predictors. Those predictors were included in the final model. This and ROCs were used to develop the risk score and cut offs required for further referral. Predictors included—smoking, chewing tobacco, chewing quid with tobacco, alcohol, spiciness of food, fruit consumption, family history of UADT cancer, rinsing mouth with water after eating/chewing. | | No |
| Lau et al. (2018) | UK—two hospital-based case–control studies (01/07/2009–01/07/2010 [622 patients]) (1/4/2013–31/8/2013 [453 patients]) 73 cases and 932 controls (men and women) | Head and neck cancer | Regularized logistic regression was performed on 60% of the data to identify key predictors. Using information from this, the model was refined and underwent split sample (20%) and cross validation (20%). Predictors include age, sex, smoking, alcohol, and symptoms. | | Yes |
| Lee et al. (2020) | USA—Pooled analysis of 14 US-case control studies from INHANCE consortium. Conducted between 1981 and 2010 | Cancer of the oral cavity, oropharynx, hypopharynx, or larynx. (by subsite) 2388 oral cavity, 2820 oropharynx, 459 hypopharynx, and 1632 larynx (separate models for men and women considering four different sub sites—eight sub models) | Logistic Regression Models developed using 70% of the dataset. Hazard and incidence rates from the Surveillance, Epidemiology, and End Results (SEER) Program were applied to the model. Competing risk models were used to ascertain risk by individual subsite but also an overall measure of absolute HNC risk. Models considers age, sex, race/ethnicity, education, cigarette smoking duration and intensity, and/or alcohol drinking intensity; the second set of models additionally included family history of HNC, except for oropharyngeal cancer in both sexes and laryngeal cancer in men. | | No |

(Continues)
| Study ID | Country/year/study design | Participants | Cancer outcome/site + sub models | Method + factors included in model | Clinician input |
|----------|----------------------------|--------------|----------------------------------|-----------------------------------|----------------|
| Liu et al. (2017) | China—Hospital-based Cohort study March 2008 to July 2016. | 28 OLK, 41 OSCC, and 18 controls (men and women) | Risk of developing oral cancer from oral leukoplakia | Peaks RF method was used for model development. Split sample testing was used to evaluate the model in addition to 10-fold cross validation. Subsequently the training set was used to train the model which was used to test the validation set. Predictors included age, sex, site, smoking and drinking. | Yes |
| McCarthy et al. (2020) | UK—Nested case-control conducted in the UK Biobank 2006–2016 | 702 HNC cases and 423,050 controls (men and women) | Head and neck cancer excluding laryngeal cancer—no reporting of numbers by subsite | The model was developed using multivariable logistic regression, final predictors being selected upon clinical significance in literature and consultation with a patient and public involvement group. Predictors included age, sex, smoking status, townsend deprivation index, body mass index, alcohol consumption, moderate exercise and fruit and vegetable intake. | No |
| Sharma et al. (2015) | India—“retrospective chart review”—case control June 2004 to June 2009 | 1025 patients—not specified if men and women | Diagnosis of oral cavity cancer—(no reporting of numbers of cases by subsite) | Dataset underwent filter reduction to select attributes for a PNN/GRNN model—Probabilistic and General Regression Neural Network. A leave one out method was then subsequently used for internal cross-validation. Attributes included sex, socioeconomic status, clinical symptom, history of addiction, comorbid condition, gross examination, site, predisposing factor, neck nodes, and tumor size. | Yes |
| Sun et al. (2019) | China—cross-sectional study Conducted from August 2016 to May 2018 | 269 patients: OPMD (n = 192) and OSCC (n = 77) (men and women) | Risk of developing oral cancer from OPMD (leukoplakia or oral lichen planus) | Risk factors included in the final model were developed from univariate logistic regression. Multivariate logistic regression was then used to evaluate the risk factors in a model, the beta-coefficient being used to assign a risk score to each factor. Cut-offs were determined using the ROC curve considering sensitivity and specificity. Risk factors included gender, age, lesion site, local stimulus, and alcohol drinking. | Yes |
| Tikka et al. (2020) | UK—Prospective data collection, building upon previous version of model developed using case-control. January 2017 until December 2018 | 307 HNC cases and 3224 controls (men and women) | Head and Neck cancer—all primary cancers to the HaN regions (n = 247), metastatic cancers to the HaN from other regions, including lymphoma (n = 48) and cancers in neighboring regions that manifested with HaN symptoms (n = 12)* | The final multivariate logistic regression was developed using univariate logistic regression and backwards elimination of non-statistically significant variables. The final model was then used for bootstrap validation on the final model. Variables included age, gender, unintentional weight loss, smoking, alcohol, positive and negative symptoms/signs of HNC. | Yes |
factors together, or on further examination were a duplication of model already included. One conference abstract paper was excluded because the full text was not available despite attempts to contact the author.

A further three articles were identified—two of these were identified from reviewing the reference lists and the third (at the time of writing yet to be published) was identified from one of reviewer’s research collaborations (DIC). Thus, in total, 14 papers were ultimately included in this review (Figure 1). 28–41 Within these studies, three of the 14 models featured “sub-models,” using broadly similar methods but stratifying models by subsite or sex. 29,30,35 These have been reported and considered accordingly, where reported.

### 3.1 Study/model characteristics

A summary of the 14 studies including model characteristics and performance information is shown in Tables 1 and 2 respectively. Of the 14 models, three were conducted in India, one in Sri Lanka, three in China, three in the USA, three in the UK, and one in multiple centers across Northern America/Europe. The majority of the studies (11 out of 14) used a case–control design or built upon previous case–controls for model development. These were mostly of a hospital-based case–control design. One study used data collected from a randomized control trial. 31 The other two used a cross-sectional and prospective cohort study method respectively. There were 11 models which utilized a form of logistic regression analysis approach to evaluating HNC risk. Two of the three remaining models used machine learning methods, while one used a cox-regression approach to evaluate risk.

A variety of cancer outcomes were considered—one model considered the risk of OPMD, and another two included the risk of developing oral cancer from OPMD. Three further models evaluated the risk of oral cancer. Six models evaluated the overall risk of HNC, two of these stratifying risk by sex and various subsites including cancers of the oral cavity, hypopharynx, oropharynx and larynx. One model considered the risk of oral and oropharyngeal cancer. Finally, one model considered the risk of oropharyngeal cancer.

Eight of the 14 models were deemed likely to require clinician input to be used, and six could possibly be used in a self-assessment role by patients.

### 3.2 Discrimination

Discrimination, the ability of a model to discern between a positive and a negative result for disease, is a crucial performance metric of a risk model. All 14 models provided measurements of discriminatory accuracy in either their development, validation populations, or both. Ten of these models described the statistical uncertainty of their findings. Many models (n = 9) reported AUC values (and intervals where reported) greater than 0.7, achieving “acceptable” or “excellent” discrimination. One model with the highest AUC value reportedly
| Study ID | Development model performance | Validation |
|----------|-------------------------------|------------|
| **AUC (95% CI) and sensitivity/specificty (where reported)** | Calibration—E/O (95% CI), PPV, NPV (%) | Cut-offs | Method of validation | Performance |
| **Calibration**—不是一个 | **AUC of 0.87 (95% CI: 0.83–0.91), SENS 95.5%, SPEC of 75.9%** | **Not reported** | External | **AUC of 0.87 (95% CI: 0.83–0.91), SENS 95.5%, SPEC of 75.9%** |
| **E/O (95% CI)**, PPV, NPV (%) | **Cut-offs** | **Method of validation** | **Performance** |
| **SENS/SPEC: 93.7%/67.7%** | **Not reported** | **Phase 2: Suburban population of the Colombo district and in a rural population in selected PHM areas of the Bulathkohupitiya MOH area in the Kegalle district of Sri Lanka.** | **AUC of 0.87 (95% CI: 0.83–0.91), SENS 95.5%, SPEC of 75.9%** |
| **All OPMDs: 0.78 (0.75–0.81)** | **SENS/SPEC: 81.1%/67.7%** | **Exc. Lichen Planus 27.5%, 98.8% (All) PPV 50.9%, NPV 89.6%** | **AUC of 0.87 (95% CI: 0.83–0.91), SENS 95.5%, SPEC of 75.9%** |
| **Budhathoki et al. (Unpublished)** | **Not reported for development—see (internal) validation** | **SENS/SPEC—Not reported** | **Internal** | **HNC (0.72, 95% CI = 0.69–0.75 in men and 0.75, 95% CI = 0.71–0.79 in women)** |
| **Plots are provided but not quantified—good calibration** | **Not reported** | **Model was internally validated using a split sample approach; data was randomly split into a training (70%) and testing set (30%) Total N = 4601 (3030 HNC cases/1571 controls)** | **[Model + HPV serology for OPC—[AUC = 0.94, 95% CI = 0.91–0.95 in men and AUC = 0.89, 95% CI = 0.82–0.92 in women]]** |
| **Chen et al. (2018)** | **0.768 (0.723–0.813) for men** | **Plots are provided but not quantified in supplementary table. Male model had better calibration (closer to 1).** | **Internal** | **NA—see Development** |
| **0.700 (0.635–0.765) for women** | **SENS/SPEC—Not reported** | **Model was internally validated with 1000 repeat samples** | **[AUC = 0.84; (95% CI: 0.77–0.90) Calibration = 1.08; (95% CI, 0.81–1.44)]** |
| **Cheung et al. (2021)** | **Not reported for development—see (internal) validation** | **Not reported** | **Internal** | **AUC = 0.73 (0.70–0.77) Calibration = 0.97** |
| **SENS/SPEC—Not reported** | **Not reported for development—see (internal) validation** | **Not reported** | **Internally validated using 5-fold cross validation of the development population evaluated for discrimination and calibration and accounting for right censoring.** | **AUC = 0.84; (95% CI: 0.77–0.90) Calibration = 1.08; (95% CI, 0.81–1.44)** |
| **Koyanagi et al. (2017)** | **(Most extensive model) 0.72 (0.69–0.75)** | **1.00** | **External Validation was carried out with a second case control study (HERPACC-3) conducted between November 2005 and March 2013 309 head and neck cancer cases and 654 matched controls were recruited.** | **AUC = 0.73 (0.70–0.77) Calibration = 0.97** |
| Study ID                          | Development model performance | Validation |
|----------------------------------|-------------------------------|------------|
|                                  | **AUC (95% CI) and sensitivity/specificity (where reported)** | **Method of validation** | **Performance** |
|                                  | **Calibration—E/O (95% CI), PPV, NPV (%)** | **Internal** | **AUC = 0.865** |
|                                  | **Cut-offs** | **Not reported** | **SENS 0.744 (0.740–0.750), SPEC 0.851 (0.848–0.854)** |
|                                  | **Development** | **Not reported** | **PPV 0.773 (0.768–0.777), NPV 0.830 (0.827–0.833)** |
| **Study ID**                     | **SENS/SPEC** | **SENS/SPEC** | **NA** |
| **(Continued)**                  | (where reported) | (where reported) | (Continues) |
| **Krishna Rao et al. (2016)**    | 0.866 | Not reported | NA |
|                                  | SENS/SPEC: 0.746 (0.682–0.810), 0.846 (0.802–0.890) | Not reported | NA |
|                                  | AUC = 0.865 | Evaluation—E/O (95% CI), PPV, NPV (%) | NA |
|                                  | SENS 0.746 (0.682–0.810), SPEC 0.830 (0.785–0.875) | Not reported | NA |
|                                  | D—E/O (95% CI), PPV, NPV (%) | Not reported | NA |
|                                  | SENS 0.846 (0.802–0.890) | Not reported | NA |
|                                  | SPEC 0.830 (0.785–0.875) | Not reported | NA |
|                                  | PPV 0.807 (0.550–0.665) | Not reported | NA |
|                                  | NPV 0.927 (0.888–0.965) | Not reported | NA |
|                                  | **Cut-offs** | **Development** | **Internal** |
|                                  | **SENS/SPEC** | **Internal validation from original dataset with 200 bootstrap samples.** | **AUC = 0.865** |
|                                  | **Not reported** | **AUC lowest to highest = 0.643–0.820** | **SENS 0.744 (0.740–0.750), SPEC 0.851 (0.848–0.854)** |
|                                  | **Not reported** | **(A) Male oral cavity cancer (AUC = 0.752);** | **PPV 0.773 (0.768–0.777), NPV 0.830 (0.827–0.833)** |
|                                  | **Not reported** | **(B) Female oral cavity cancer (AUC = 0.718);** | **NA** |
|                                  | **Not reported** | **(C) Male oropharyngeal cancer (AUC = 0.643);** | **NA** |
|                                  | **Not reported** | **(D) Female oropharyngeal cancer (AUC = 0.745);** | **NA** |
|                                  | **Not reported** | **(E) Male hypopharyngeal cancer (AUC = 0.784);** | **NA** |
|                                  | **Not reported** | **(F) Female hypopharyngeal cancer (AUC = 0.820);** | **NA** |
|                                  | **Not reported** | **(G) Male laryngeal cancer (AUC = 0.794);** | **NA** |
|                                  | **Not reported** | **(H) Female laryngeal cancer (AUC = 0.870).** | **NA** |
| **Lau et al. (2018)**            | 0.79 | Not reported | NA |
|                                  | SENS/SPEC: Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
| **Lee et al. (2020)**            | Not reported | Provides plots for each subsite (male and female) but not quantified—mostly good calibration bar male and female hypopharyngeal cancer | Not reported |
|                                  | SENS/SPEC—Not reported | Not reported | Not reported |
|                                  | Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
| **Liu et al. (2017)**            | 1 | Not reported | NA |
|                                  | SENS/SPEC: 100.00%, 99.02% | A cut-off of 0.5 was used (50% risk) | NA |
|                                  | PPV 98.94%, NPV 100.00% | Unknown—poor reporting | NA |
|                                  | Not reported | 102 controls, 82 OLK, 93 OSCC | NA |
| **McCarthy et al. (2020)**       | 0.69 (0.66–0.71) | Plots but not quantified for development population—CIs are also included—good calibration | Not reported |
|                                  | SENS/SPEC—Not reported | Not reported | Not reported |
|                                  | Not reported | Not reported | NA |
|                                  | Not reported | Not reported | NA |
| **Sharma et al. (2015)**         | 0.9974 | Not reported | NA |
|                                  | SENS/SPEC: 98.01%, 98.68% | Not reported | NA |
|                                  | SENS 99.35%, NPV 98.01% | Internal | NA |
| **Sun et al. (2019)**            | 0.83 (0.77–0.88) | Not reported | NA |
|                                  | SENS/SPEC: 67.53%, 81.25% | Cut-off score of 3 SENS, 67.53%, SPEC 81.25% | NA |
| **(Continues)**                  | Not reported | PPV 59.09%, NPV 86.19% | Development only |
| **(Continues)**                  | (Unclear if this for same PNN/GRNN model) | (Unclear if this for same PNN/GRNN model) | (Continues) |
achieved a “perfect” model discrimination of 1.0, however, this model was constructed from a very small sample size, in addition to other key limitations and bias concerns such as a failure to report statistical uncertainty and any missing data.\(^{36}\)

### 3.3 | Accuracy

Measurements of accuracy in the form of sensitivity and specificity were described in seven of the 14 studies. These ranged from 67.53% to 100%,\(^{36,39}\) and 67.7% to 100% for each measure respectively.\(^{28,36}\) The model that reported the highest sensitivity and specificity achieved 100% in both of these metrics in their validation population.\(^{36}\)

### 3.4 | Calibration

Calibration, the degree of correspondence between the estimated probability of an outcome predicted by a model vs the outcome observed is an important measurement to consider when evaluating model performance, to minimize overfitting. Despite this, calibration is often overlooked in favor of model discrimination (AUC also known as the C-statistic).\(^{25}\) Measurements of model calibration, either in the form of an expected/observed ratio or calibration plot were described in seven out of the 14 models, however, statistical uncertainty was reported in only three of these. Calibration was evaluated in light of hierarchy definitions described by Van Calster and colleagues.\(^{42}\) Calibration was good in most, if not all, of the models where this metric was reported, with the exception of male and female hypopharyngeal cancer models in one study where calibration was sub-optimal in these particular calibration plots.\(^{35}\) Most of the models that did report calibration presented graphs or statistics that were close to the ideal calibration slope (expected/observed) value of 1, with some models slightly above this value indicating some over-prediction.\(^{24}\)

### 3.5 | PPV/NPV

PPV and NPV are defined as the proportion of patients who actually have the disease that test positive and the proportion of patients without the disease that test negative respectively. Six models reported measurements of PPV and NPV.\(^{43}\) As such only one model reported the statistical uncertainty of this.\(^{33}\) The PPV and NPV values reported ranged from 20.7% to 100% and 83% to 100%. Again, Liu and colleagues\(^{36}\) achieved the highest PPV and NPV values of 100%.

### 3.6 | Model risk cut-offs

Of the 14 models, only five reported model risk cut-offs during development. Two models used a risk probability as a cut-off.\(^{36,60}\) Three models reported cut-offs using performance metrics including AUC, sensitivity and specificity, and PPV/NPV.\(^{28,33,39}\)
3.7 Validation

Only 3 of the 14 models reported external validation—Amarasinghe et al., Koyanagi et al., and Tota et al. Three others reported robust methods of internal validation via split random sampling.  

3.8 Risk factors

Altogether, the 14 models considered over 30 various risk factors. The most common factors included were age (13 models), alcohol consumption (13 models), sex (12 models), and tobacco smoking (12 models). Notably, two models considered HPV serostatus in model development—Budhathoki et al. and Tota et al. The number of risk factors included in models ranged between 5 and 13.

3.9 PROBAST

The evaluation of each domain of the PROBAST risk of bias assessment tool is summarized in Table 3. Of the 14 models, seven were deemed to have a “high” risk of bias in at least one domain. The “analysis” section was the most common domain where a high risk of bias was identified. Common reasons for these included low numbers of the outcome of interest, a lack of external validation and limited or no internal validation, failure to report statistical uncertainty of findings and no discussion of missing data (and procedures in the event of this). Five of the 14 models were reported to have an “unclear” applicability concern whereby aspects of the model may limit its applicability but as such these were not major limitations. The “participants” section was the most common domain where applicability concerns were classified as “unclear.” Reasons for these included limited generalizability owing to the outcome considered, limiting the analysis to those of one ethnicity, use of non-primary HNC cancer sites and low-quality reporting of methods. Where models had a limitation but were otherwise fairly robust and well-developed the risk of bias was deemed as “low.” Overall, 7 of the 14 models were deemed to have an overall low risk of bias.

3.10 Overall quality assessment

The overall quality assessment of the 14 models and the components considered in this quality assessment along with model predictive performance assessment can be seen in Table 4. Of the 14 models, six were assessed as “high” quality, three as “moderate” quality, and five as “low” quality. The main components, which impacted on quality were PROBAST risk of bias, applicability concern, and validation methods.

| Study ID | Participants | Predictors | Outcome | Analysis | Overall |
|----------|--------------|------------|---------|----------|---------|
| Amarasinghe et al. (2010) | LOW | UNCLEAR | LOW | LOW | LOW | LOW | UNCLEAR |
| Budhathoki et al. (Unpublished) | LOW | UNCLEAR | LOW | LOW | LOW | LOW | UNCLEAR |
| Chen et al. (2017) | LOW | LOW | LOW | UNCLEAR | LOW | LOW | HIGH | LOW |
| Cheung et al. (2021) | LOW | LOW | LOW | LOW | LOW | LOW | HIGH | LOW |
| Koyanagi et al. (2017) | LOW | LOW | LOW | LOW | LOW | LOW | LOW | LOW |
| Krishna Rao et al. (2016) | UNCLEAR | LOW | LOW | LOW | LOW | HIGH | HIGH | LOW |
| Lau et al. (2018) | UNCLEAR | LOW | UNCLEAR | LOW | UNCLEAR | LOW | HIGH | LOW |
| Lee et al. (2020) | LOW | LOW | LOW | LOW | LOW | LOW | LOW | LOW |
| Liu et al. (2017) | HIGH | LOW | UNCLEAR | UNCLEAR | LOW | LOW | HIGH | UNCLEAR |
| McCarthy et al. (2020) | LOW | LOW | LOW | UNCLEAR | LOW | LOW | LOW | LOW |
| Sharma et al. (2015) | HIGH | UNCLEAR | HIGH | UNCLEAR | HIGH | UNCLEAR | HIGH | UNCLEAR |
| Sun et al. (2019) | HIGH | UNCLEAR | LOW | UNCLEAR | LOW | HIGH | HIGH | UNCLEAR |
| Tikka et al. (2020) (v.2) | LOW | LOW | LOW | UNCLEAR | LOW | LOW | LOW | LOW |
| Tota et al. (2019) | LOW | LOW | LOW | UNCLEAR | LOW | LOW | LOW | LOW |
In terms of performance, eight of the 14 models were high performing, with AUCs greater than 0.8, ranging from 0.83 to 1. Of the six high quality models, three had high predictive performance with good discriminative accuracy—Amarasinghe et al., Tikka et al., and Tota et al. With regards to factors included in the high-quality models, all six had some sociodemographic factors—age was evaluated in all six of them and sex in five—one model did not analyze or adjust for sex which in turn resulted in reduced applicability. Four of the models adjusted for at least one additional sociodemographic factor (education, ethnicity, or socioeconomic deprivation). Two high-quality models evaluated socioeconomic deprivation, one synthesizing educational and occupational status to define this, the other measured deprivation using an area based socioeconomic index. One model also used exercise and fruit/vegetable intake as additional factors. Two of the three models that used biomarker (genetic or HPV) data in their models were ultimately among the six assessed as high quality. One high quality model used HPV serostatus as a predictor. Another used DNA sampling to assess ALDH2 genotype. Finally, one of the six high quality models used clinical signs and symptoms as predictors to inform model design.

### 3.11 Synthesis

All six of the high-quality models were more recently developed (since 2010). Despite the heterogeneity of the models, generally, those that were assessed as high quality shared common design aspects. All of the models were developed from case–control study data with some variation in design such as hospital-, community-, synthetic-controls, or a mixture of population and hospital controls. All six high quality studies also used a form of logistic regression to derive their risk models. These included binary, multivariate or conditional logistic regression. Three of the models required clinician input to use: two of these due to HPV or genotype information, and one due to use of clinical examination information.
The critical design feature common to all of the high-quality models was robust validation methods. These included the use of external validation in another setting, a history of this in a previous developmental version of the model, or the utilization of well-conducted internal validation with a large split sample. All the high-quality models that did not use an external validation approach described this as a limitation and a next step in their model development.

In addition to the high quality models having a low risk of bias, five of the six also had a low applicability concern, while the remaining model scored “unclear” for applicability concern assessment. This was primarily due to the model evaluating the risk of OPMD only and not accounting for sex as a predictor during model development.

The high-quality models mostly had fair to good performance. One model had a sup-optimal AUC of 0.64, and two high quality models reported a fair AUC over 0.7 and three of the high quality models achieved excellent AUC values over 0.8. The better the discriminative performance, the more accurately those at risk of disease can be identified. Two of the six high quality models did not report calibration metrics. This was the main limitation of these models. Four models reported calibration, either numerically or via plots, illustrating that the models could accurately predict outcomes in line with the observed event of interest (HNC), with most of the calibration graphs or reported observed/expected values being close to the perfect prediction value of 1. Therefore, for the most part, HNC models, where reported, showed good calibration between the expected and observed risk of HNC.

Of the six high quality models, three were also high performing, achieving excellent discrimination with AUCs over 0.8. The three models predicted the risk of OPMD, HNC, and oropharyngeal cancer respectively. Two of the models were externally validated in another cohort while the other model had a history of external validation in its first version. This model has also seen some clinical use, being used to triage patients remotely during the COVID-19 pandemic. The three high-quality, high performing models all used similar sociodemographic and behavioral factors but where the three high performing models differed and ultimately excelled was in the choice of additional predictors used. These included the aforementioned use of betel quid in one model, clinical examination and symptoms in another model and the use of HPV serostatus and ethnicity in the third high quality model.

Those that were classified as moderate overall quality had at least one major methodological limitation, but generally had fair predictive performance. Studies that were classified as low quality had at least two significant limitations; some of these reported very good model discrimination, although this needs to be interpreted with caution.

4 | DISCUSSION

A range of risk prediction models for HNC were identified. These models were heterogeneous in their risk factors and outcomes, were developed with variable methodological approaches and rigor, and several models demonstrated the potential to predict and identify those at a higher risk of HNC.

The six high performing models incorporate the major HNC risk factors of tobacco smoking and alcohol consumption, in addition to the sociodemographic factors of age and sex. Additional factors are included contributing to improved performance. Four included socioeconomic factors, one: family history, one: betel quid chewing, one: HPV serology, one included a genetic marker, and one included clinical examination findings. These selected factors are consistent with the international analytical epidemiological literature which has identified tobacco smoking and alcohol drinking as the major risk factors (accounting for up to ~70% of the population attributable risk), the important role of demographics of age in cancer risk, particularly men being more predisposed to HNC. Moreover, the important role of HPV particularly in oropharyngeal cancer, and betel quid chewing in oral cavity cancer in particular populations, and the increasingly refined role of genetic factors in HNC are reflected in the models. Three of these high performing predictive models were of high quality and consistent methodological rigor—all included major behavioral and sociodemographic factors along with an additional factor. They were generally more specified models that were tailored to their target population or subsite, for example, to South Asia (the inclusion of betel quid) or to oropharyngeal cancer (the inclusion of HPV serology). Perhaps counterintuitively, some of the models that included many additional risk factors generally had lower predictive performance. This could be explained by a statistical phenomenon known as model overfitting, whereby a model becomes too tailored to a developmental dataset with unnecessary components. This violates the principle of parsimony, in turn limiting a model’s generalisability when applied to another independent dataset. The higher performing models and particularly the high performing, high-quality models generally required clinician input, reflecting the nature of the variables required.

It could be hypothesized that HPV serostatus and genetic markers could help better inform individual risk, in line with the growing popularity of “personalized medicine” in other diseases. However, this may in turn limit the practicality of a model—for example, primary care medical or dental practices where time and resources may already be limited.

Several similar reviews of risk prediction models have been undertaken for other cancer sites including colorectal and lung. The reviews of colorectal and lung cancer models both identified models with high performance (0.65–0.75, 0.76–0.96, and 0.57–0.879), while the breast cancer models generally reported poorer performance (0.56–0.63 and 0.56–0.71). The poor performance of these breast cancer models was attributed in the reviews to limited knowledge and data on risk factors for breast cancer leading to sub-optimal prediction. The performance range of the HNC models was similar to the colorectal and lung cancer reviews and might in part be attributed to the growing epidemiological research base in the field.

The methodology of this review is similar to reviews of risk models for other cancers. This review employed robust quality
and methodological assessment including the PROBAST risk of bias and applicability concerns tool, and a focus on the nature of model development and validation approach. While validation is a domain of the PROBAST tool, this was explicitly assessed separately as external validation is gold standard methodology of risk prediction model development. To our knowledge, this is the first review of HNC risk prediction models. This review has some strengths including searching multiple databases, dual article screening, as well as the comprehensive quality assessment. A detailed thematic narrative synthesis drew on the model quality and performance to identify key design characteristics. There are some limitations to this review, including not publishing a protocol. The study started as a rapid review—but ultimately became more systematic in nature—particularly in term of quality assessment methods. However, it was not feasible to register the review retrospectively, hence the review was not registered with Prospero. The PICO/research question and search/inclusion criteria were developed a priori and did not change during the review. The review was also conducted following PRISMA guidelines and was advised by a subject librarian in the field. Second, the inclusion of papers published only in English may have excluded other pre-existing models. As with most reviews, the nature and limitations of available data can influence the overall quality of evidence synthesized—the source data of this review are largely from case-control studies which do have some potential recall and selection biases. This review has also been conducted within the overarching objectives of developing a risk model for HNC and translating it to a clinical setting. Any findings of this review are intended to help inform model development.

5 | CONCLUSIONS

This review illustrates that there is a limited but growing number of HNC risk prediction models. Some of the models reviewed do have the potential to identify and stratify those at risk of HNC. Model predictor selection should include, as a minimum, well established risk factors as well as sociodemographic predictors. Additional genetic, biomarker, or clinical factors have the potential to improve predictive performance. However, care should be taken to ensure a limited number of predicting factors are chosen to avoid model overfitting. Such early identification of risk factors in the context of a HNC risk level could have important applications including using this “teachable moment” for behavior change, directing patients to preventive care pathways (e.g., for smoking cessation), or in identifying the need for tailored frequencies in recall intervals for clinical examination (e.g., with primary care dental practitioners). These models could form the basis of a personalized approach to HNC prevention. Further work can be undertaken to refine, improve and validate these models and potentially trial in the clinical setting.

ACKNOWLEDGMENTS

This work was funded and supported by Cancer Research UK as part of the TRACC programme, Beatson Institute, Glasgow (Grant number 315941-01). We would like to thank Paul Canon, University of Glasgow MVLS Librarian, for his support in helping us develop and check our review and search strategy. We would also like to thank Catherine Winchester, research Manager at the Beatson Institute, Glasgow for her role in providing feedback and proofreading this review.

CONFLICT OF INTEREST

David I. Conway is associated with two of the publications in this review: David I. Conway is a Glasgow center PI on the ARCAge multicenter study and is a co-author of one the models developed from this study (submitted)35; David I. Conway is also a member of the INHANCE Consortium from which another model was developed.35 All other authors have none to declare.

ORCID

Craig D. L. Smith https://orcid.org/0000-0001-6465-7472
Alex D. McMahon https://orcid.org/0000-0002-9425-7792
Alastair Ross https://orcid.org/0000-0003-2952-3182
Gareth J. Imman https://orcid.org/0000-0002-6264-4253
David I. Conway https://orcid.org/0000-0001-7762-4063

REFERENCES

1. GLOBOCAN – Cancer Today – Estimated number of new cases in 2020, worldwide, both sexes, all ages. 2020. https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&population=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=0&include_nmsc=0&include_nmsc_other=1
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
3. Chaturvedi AK. Epidemiology and clinical aspects of HPV in head and neck cancers. Head Neck Pathol. 2012;6(Suppl 1):S16-S24.
4. Conway DI, McMahon AD, Smith K, et al. Components of socioeconomic risk associated with head and neck cancer: a population-based case-control study in Scotland. Br J Oral Maxillofac Surg. 2010;48(1):11-17.
5. Hoffaman HT, Karnell LH, Funk GF, Robinson RA, Menck HR. The National Cancer Data Base report on cancer of the head and neck. Arch Otolaryngol Head Neck Surg. 1998;124(9):951-962.
6. Pulte D, Brenner H. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. Oncologist. 2010;15(9):994-1001.
7. Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. Int J Cancer. 2005;114(5):806-816.
8. Guo K, Xiao W, Chen X, Zhao Z, Lin Y, Chen G. Epidemiological trends of head and neck cancer: a population-based study. Biomed Res Int. 2021;2021:1738932.
9. Argirion I, Zanis KR, Defever K, et al. Temporal changes in head and neck cancer incidence in Thailand suggest changing oropharyngeal epidemiology in the region. J Glob Oncol. 2019;5:1-11.
10. Sanderson RJ, Ironside JA. Squamous cell carcinomas of the head and neck. BMJ. 2002;325(7368):822-827.
11. Northern Ireland Cancer Registry. Head and neck cancer report. 2019. https://www.qub.ac.uk/research-centres/nicr/FileStore/OfficialStats2019/Reports/Filetoupload.1188585.en.pdf
12. Lagiou P, Georgiola C, Minaki P, et al. Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. *Eur J Cancer Prev*. 2009;18(1):76-84.

13. Abrahão R, Anantharaman D, Gaboreau V, et al. The influence of smoking, age and stage at diagnosis on the survival after larynx, hypopharynx and oral cavity cancers in Europe: the ARCAGE study. *Int J Cancer*. 2018;143(1):32-44.

14. Abrahão R, Perdomo S, Pinto LFR, et al. Predictors of survival after head and neck squamous cell carcinoma in South America: the Inter-CHANGE study. *JCO Glob Oncol*. 2020;6:e486-499.

15. Cancer Research UK. Cancer decision support tools overview. 2020. [Computer program]. Julia Hippisley-Cox, Carol Coupland. 2013; https://www.qcancer.org/

16. QCancer (®) (Pancreas). 2016;14(66):1-332.

17. Jackson SE, Chester JD. Personalised cancer medicine. *Int J Cancer*. 2015;137(2):262-266.

18. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *Br J Cancer*. 2015;113(12):1645-1650.

19. Medina-Lara A, Grigore B, Lewis R, et al. Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis. *Health Technol Assess*. 2020;24(66):1-332.

20. Collins GS, Altman DG. Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer (®) (Pancreas). *Br J Gen Pract*. 2013;63(614):e636-e642.

21. Kostopoulou O, Arora K, Pálfi B. Cancer risk algorithms in primary care: can they improve risk estimates and referral decisions? *medRxiv*. 2021.

22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

23. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Third ed. Wiley; 2013.

24. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ*. 2019;365:j379.

25. Van Calster B, McMorrow DJ, van Smeden M, et al. Calibration of predictive analytics. *BMJ Med Res*. 2019;371(1):230.

26. Wolff RF, Moorens KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58.

27. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315-1316.

28. Amarasinghe HK, Johnson NW, Laloo R, Kumararaachchi M, Warnakulasuriya S. Derivation and validation of a risk-factor model for detection of oral potentially malignant disorders in populations with high prevalence. *Br J Cancer*. 2010;103(3):303-309.

29. Budhathoki S, BD, GL, et al. A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV serology and genetic markers. 2021.

30. Chen F, Lin L, Yan L, et al. Nomograms and risk scores for predicting the risk of oral cancer in different sexes: a large-scale case-control study. *J Cancer*. 2018;9(14):2543-2548.

31. Cheung LC, Ramadas K, Muxonge R, et al. Risk-based selection of individuals for oral cancer screening. *J Clin Oncol*. 2021;39(6):663-674.

32. Koyanagi YN, Ito H, Oze I, et al. Development of a prediction model and estimation of cumulative risk for upper aerodigestive tract cancer on the basis of the aldehyde dehydrogenase 2 genotype and alcohol consumption in a Japanese population. *Eur J Cancer Prev*. 2017;26(1):38-47.

33. Krishna Rao S, Mejia GC, Logan RM, et al. A screening model for oral cancer using risk scores: development and validation. *Community Dent Oral Epidemiol*. 2016;44(1):76-84.

34. Lau K, Wilkinson J, Moorthy R. A web-based prediction score for head and neck cancer referrals. *Clin Otolaryngol*. 2018;43(4):1043-1049.

35. Lee YA, Al-Temimi M, Ying J, et al. Risk prediction models for head and neck cancer in the US population from the INHANCE consortium. *Am J Epidemiol*. 2020;189(4):330-342.

36. Liu Y, Li Y, Fu Y, et al. Quantitative prediction of oral cancer risk in patients with oral leukoplakia. *Oncotarget*. 2017;8(28):46057-46064.

37. McCarthy CE, Bonnet LJ, Marcus MW, Field JK. Development and validation of a multivariable risk prediction model for head and neck cancer using the UK Biobank. *Int J Oncol*. 2020;57(5):1192-1202.

38. Sharma N, Om H. Usage of probabilistic and general regression neural network for early detection and prevention of oral cancer. *Sci World J*. 2015;2015:234191.

39. Sun Z, Liu L, Wang C, et al. Assessment of the cancerization risk for oral potentially malignant disorders by clinical risk model combined with autofluorescence and brush biopsy with DNA-image cytometry. *Eur Arch Otorhinolaryngol*. 2019;276(9):2549-2557.

40. Tikka T, Kavanagh K, Lowit A, et al. Head and neck cancer risk calculator (HaNC-RC)-V2. Adjustments and addition of symptoms and social history factors. *Clin Otolaryngol*. 2020;45(3):380-388.

41. Tota JE, Gillison ML, Kati H, et al. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer*. 2019;125(24):4407-4416.

42. Van Calster B, Nieboer D, Vergouw Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167-176.

43. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45-50.

44. Paleari V, Hardman J, Tikka T, Bradley P, Pracy P, Kerawala C. Rapid implementation of an evidence-based remote triaging system for assessment of suspected referrals and patients with head and neck cancer on follow-up after treatment during the COVID-19 pandemic: model for international collaboration. *Head Neck*. 2020;42(7):1674-1680.

45. Dhull AK, Atri R, Dhankhar R, Chauhan AK, Kaushal V. Major risk factors in head and neck cancer: a retrospective analysis of 12-year experiences. *World J Oncol*. 2018;9(3):80-84.

46. Cogliano VJ, Baan R, International Agency for Research on Cancer., et al. Preventable exposures associated with human cancers. *J Natl Cancer Inst*. 2011;103(24):1827-1839.

47. Hashibe M, Brennan P, Chiang SC, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev*. 2009;18(2):541-550.

48. Johnson DE, Burtsev B, Leemans CR, Bauman JE, Grands JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers*. 2020;6(1):92.

49. Centers for Disease Control and Prevention. Head and neck cancers. [Computer program]. 2020. [https://www.cdc.gov/cancer/headneck/index.htm](https://www.cdc.gov/cancer/headneck/index.htm)

50. Sabatini ME, Chiocca S. Human papillomavirus as a driver of head and neck cancers. *Br J Cancer*. 2020;122(3):306-314.

51. Guha N, Warnakulasuriya S, Vlaanderen J, Strålf K. Betel quid chewing and the risk of oral and oropharyngeal cancers: a meta-analysis with implications for cancer control. *Int J Cancer*. 2014;135(6):1433-1443.

52. Riaz N, Morris LG, Lee W, Chan TA. Unraveling the molecular genetics of head and neck cancer through genome-wide approaches. *Genes Dis*. 2014;1(1):75-86.

53. Beck TN, Golemis EA. Genomic insights into head and neck cancer. *Cancers Head Neck*. 2016;1(1):1.
54. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci. 2004;44(1):1-12.
55. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk prediction models for colorectal cancer: a systematic review. Cancer Prev Res (Phila). 2016;9(1):13-26.
56. Williams TG, Cubiella J, Griffin SJ, Walter FM, Usher-Smith JA. Risk prediction models for colorectal cancer in people with symptoms: a systematic review. BMC Gastroenterol. 2016;16(1):63.
57. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012;133(1):1-10.
58. Louro J, Posso M, Hilton Boon M, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. Br J Cancer. 2019;121(1):76-85.
59. Gray EP, Teare MD, Stevens J, Archer R. Risk prediction models for lung cancer: a systematic review. Clin Lung Cancer. 2016;17(2):95-106.
60. Bravi F, Lee YA, Hashibe M, et al. Lessons learned from the INHANCE consortium: an overview of recent results on head and neck cancer. Oral Dis. 2021;27(1):73-93.
61. Winn DM, Lee YC, Hashibe M, Boffetta P. The INHANCE consortium: toward a better understanding of the causes and mechanisms of head and neck cancer. Oral Dis. 2015;21(6):685-693.
62. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2020;14(1):49-58.
63. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-473.
64. Tenny S, Kerndt CC, Hoffman MR. Case control studies. StatPearls. StatPearls Publishing LLC; 2022.

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

How to cite this article: Smith CDL, McMahon AD, Ross A, Inman GJ, Conway DI. Risk prediction models for head and neck cancer: A rapid review. Laryngoscope Investigative Otolaryngology. 2022;7(6):1893-1908. doi:10.1002/lio2.982