Predicting the risk of developing oropharyngeal cancer for Canadians: current evidence and models

Hamed Ghanati

Faculty of Dental Medicine and Oral Health Sciences

McGill University

Montreal, Québec, Canada

June 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

© Hamed Ghanati 2022

Dedication

This work is dedicated to my beloved spouse, my supportive parents, and my much-loved younger brother

Acknowledgments

Words are certainly not enough to thank my first supervisor, Dr. Belinda Nicolau, for providing me with a golden opportunity to study at one of the high-ranked universities worldwide and learn valuable skills essential for my future career. I definitely consider my acquaintance with her a milestone in my life. She was outstanding academic support during my studies at McGill, without whom this endeavor would have never been possible. She always boosted me with generous encouragement and advice. Being substantial sympathetic support during my tough first days of immigration in Canada, she has been a sister I never had. THANK YOU, Belinda, for everything!

I am deeply indebted to my second supervisor, Dr. Sreenath Madathil, for his tremendous support during my studies at McGill. I have never met someone as patient, helpful, and kindhearted as he. The resources he referred me to, the academic and life lessons he taught me, and the constructive advice he provided me will undoubtedly pave my way toward a successful future. I was also extremely fortunate to have him beside me through each hurdle in conducting this thesis project. Despite his super-hectic schedule, he always kindly answered all my questions when I approached his office. Without his invaluable help, advice, and constructive feedback on statistical techniques, this project would have never been initiated and completed. I can never thank you enough, Sreenath!

I would like to express my most profound sense of gratitude to my committee members, Drs. Paul Allison and Anita Koushik, for expertly guiding me through the whole process of my thesis. Their invaluable comments during the committee meetings helped me address this thesis project's issues.

I would also like to thank Ms. Crystal Noronha and Ms. Despoina Moirakidou, the Administrative and Student Affairs Coordinators in the McGill Faculty of Dental Medicine and Oral Health Sciences, for their kind help and timely reminders of the deadlines, financial opportunities, and essential steps to be taken to fulfill the requirements of my Masters.

I am so grateful to my parents, Mrs. Masoumeh and Mr. Ebrahim, and my younger brother, Mohipilot, for their affection, love, and positive energy that kept me active during my stay in Canada.

Reading countless pages of acknowledgments from countless resources, I have never imagined how difficult it would be to find proper words to appreciate the unparalleled love, help, support, understanding, and grief delivered by my beloved spouse, the light of my life, Aylar. Since the time we met each other, she has always been my best friend who endured hardships to pave my way toward my goals. I will never ever forget the tough days we experienced during the first days of our landing in Canada, when she tremendously lent substantial emotional support while she, herself, was enduring extreme levels of stress. I must admit that she is the toughest person I have ever met. As time goes on, I am becoming more and more grateful to God for sending me one of his Angles. Thank you, lovely angle, Aylar!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

HNC: Head and Neck Cancer

HPV: Human PapillomaVirus

OPC: OroPharyngeal Cancer

UADT: Upper AeroDigestive Tract

SEP: Socioeconomic Position

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

PROBAST: Prediction model Risk Of Bias ASsessment Tool

Somers' D: Sumers' Delta

C-Statistic: Concordance statistic

ROC: Receiver Operating Characteristic

AUC: Area Under the ROC Curve

O/E: Observed over Expected values' rate

H-L test: Hosmer-Lemeshow test

ROB: Risk Of Bias

CAA: Concerns About Applicability

PPV: Positive Predictive Value

NPV: Negative Predictive Value

GOF: Goodness-Of-Fit

Abstract

Background: Every year, more than half a million people are diagnosed with Head and Neck Cancers (HNCs). Among different cancers, HNC has a high mortality and morbidity rate. While the etiology of HNC has been known for many years, there has been a rise in the incidence of a subset of these cancers, mainly oropharyngeal cancer, in high income countries including Canada over the past decades. A considerable part of this rise has been attributed to the human papilloma virus (HPV). Therefore, preventive interventions such as vaccination against HPV infection are expected to reduce the number of new oropharyngeal cancer cases. To have efficient prevention, the interventions need to be targeted at high-risk individuals. Risk prediction models can improve the efficiency of these preventive programs by estimating the individualized risk of developing HNC and identifying the high-risk population. Different risk prediction models have been developed worldwide; however, little is known about these models and their applicability in the Canadian context.

Objectives: This thesis aims to: 1) review the literature on the HNC risk prediction models and 2) validate a risk prediction model on a sample of the Canadian population.

Methods: First, we reviewed the published articles on HNC risk prediction modeling. We included the full-text of peer-reviewed publications that reported at least one model for predicting the risk of developing HNC. We only considered the models that can be used in the primary clinical settings, thus, excluded the ones with genetic markers. Based on the TRIPOD checklist, we extracted the data on the study participants, analytical methods, and models' characteristics and performance. The quality of studies and the models' risk of bias and applicability were appraised using PROBAST. This review identified a model that was potentially applicable to the Canadian context. The model was developed to predict the one-year risk of developing oropharyngeal cancer in the US population. In the second step of this thesis project, we validated the predictions of this model on the dataset derived from the Canadian site of the HeNCe Life study, a case-control investigation on the etiology of HNC through a life-course framework in Canada. Based on the model's development study, we derived a dataset from HeNCe Life comprising 214 cases of oropharyngeal cancers and 433 controls, frequency matched to the cases by sex and 5-year age categories. We replicated the model and tested its predictions on the derived dataset. We

evaluated the model's overall prediction performance by measuring Somers' D, Brier scores, and R^2 (Nagelkerke). The discrimination ability was tested using C-Statistics and discrimination indices. The model's calibration was assessed by evaluating the calibration slope and intercept values.

Results: The first step of this thesis identified nine peer-reviewed HNC risk prediction modeling studies that overall reported 16 models. Six of these studies were conducted in Asia, and only three were published from Western countries, but none from used Canadian data. Most of the models were developed by multivariable logistic regression analysis. All included studies had a high risk of bias, and two of them had high concerns about applicability of the models. Although we did not identify any article reporting a development or validation of a model for the Canadian population, the review found an oropharyngeal cancer risk prediction model, developed in a sample of the US population, that is reproducible and potentially applicable in the Canadian context. This model was developed using multivariable logistic regression analysis. Its predictors comprised age, sex, race, pack-years of smoking, previous year's alcohol consumption, number of lifetime sexual partners, oral HPV infection status, and two-way interaction between sex, pack-years of smoking, and oral HPV infection status. During the overall performance assessment, the model presented a Somers' D, Brier score, and R^2 of 0.49, 0.29, and 0.25, respectively. The model's discrimination index was 0.19, and the C-Statistics was 0.75 (0.69-0.79). The model overestimated the predictions by 4.64 points, which was the calibration intercept. The calibration slope also was 0.57. In summary, although the model showed a moderately high level of discrimination, it had poor calibration performance.

Conclusion: Limited numbers of HNC risk prediction modeling studies provide sufficient information to judge the models' quality and applicability. However, the review identified one model that may still be used in the Canadian context. However, since the model presented acceptable discrimination but a poor calibration, it needs to be recalibrated and updated on the Canadian context before implementation in practice. Future studies are needed to understand this model's applicability in the Canadian clinical settings.

Résumé

Contexte: Chaque année, plus d'un demi-million de personnes reçoivent un diagnostic de cancer de la tête et du cou (CTC). Parmi les différents cancers, les CTC ont un taux élevé de mortalité et de morbidité. Bien que l'étiologie du CTC soit connue depuis de nombreuses années, l'incidence d'un sous-ensemble de ces cancers, principalement le cancer de l'oropharynx, a augmenté dans les pays à revenu élevé, dont le Canada, au cours des dernières décennies. Une part considérable de cette augmentation a été attribuée au virus du papillome humain (VPH). Par conséquent, les interventions préventives telles que la vaccination contre l'infection par le VPH devraient réduire le nombre de nouveaux cas de cancer de l'oropharynx. Pour avoir une prévention efficace, ces programmes doivent cibler les personnes à haut risque. Les modèles de prédiction des risques peuvent améliorer l'efficacité de ces programmes de prévention en estimant le risque individualisé de développer un CTC et en identifiant la population à haut risque. Différents modèles de prévision des risques ont été développés dans le monde; cependant, on en connait peu sur ces modèles et leur applicabilité dans le contexte canadien.

Objectifs: Cette thèse vise à : 1) faire une revue de la littérature sur les modèles de prédiction du risque de CTC et 2) valider un modèle de prédiction du risque sur un échantillon de la population canadienne.

Méthodes: Tout d'abord, nous avons examiné les articles publiés sur la modélisation de la prévision des risques du CTC. Nous avons inclus le texte intégral des publications évaluées par des pairs qui ont rapporté au moins un modèle pour prédire le risque de développer un CTC. Nous n'avons considéré que les modèles qui peuvent être utilisés dans les contextes cliniques primaires, donc exclu ceux avec des marqueurs génétiques. Sur la base de la liste de contrôle TRIPOD, nous avons extrait les données sur les participants à l'étude, les méthodes d'analyse, les caractéristiques et les performances des modèles. La qualité des études, le risque de biais et l'applicabilité des modèles ont été évalués à l'aide de PROBAST. À partir de cet examen, nous avons identifié un modèle potentiellement applicable au contexte canadien. Le modèle a été développé pour prédire le risque sur un an de développer un cancer de l'oropharynx dans la population américaine. Dans la deuxième étape de ce projet de thèse, nous avons validé les prédictions de ce modèle sur l'ensemble de données dérivé de la partie canadienne de l'étude

HeNCe Life, une enquête cas-témoins sur les étiologies du CTC à travers un cadre de parcours de vie au Canada, en Inde, et au Brésil. Sur la base de l'étude de développement du modèle, nous avons dérivé un ensemble de données de HeNCe Life comprenant 214 cas de cancers de l'oropharynx et 433 témoins, fréquence appariée aux cas par sexe et catégories d'âge de 5 ans. L'âge des participants à l'ensemble de données dérivé variait de 30 à 79 ans. Nous avons reproduit le modèle et testé ses prédictions sur l'ensemble de données dérivées. Nous avons évalué les performances de prédiction globales du modèle en mesurant le D de Somers, les scores Brier et R2 (Nagelkerke). La capacité de discrimination a été testée à l'aide de C-Statistics et d'indices de discrimination. L'étalonnage du modèle a été évalué en évaluant la pente d'étalonnage et les valeurs d'interception.

Résultats: La première étape de cette thèse a identifié neuf études de modélisation de prédiction des risques de CTC évaluées par des pairs qui ont rapporté 16 modèles. Six de ces études ont été menées en Asie, et seulement trois études ont été publiées dans les pays occidentaux, mais aucune du Canada ou utilisant des données canadiennes. La plupart des modèles ont été développés par analyse de régression logistique multivariée. Toutes les études incluses présentaient un risque élevé de biais, et deux d'entre elles avaient de fortes inquiétudes quant à l'applicabilité des modèles. L'examen des études incluses a révélé la nécessité d'élaborer ou de valider un modèle de prédiction des risques pour la population canadienne, car aucun article n'a été publié au Canada. De plus, nous avons identifié un modèle de prédiction du risque de cancer de l'oropharynx développé dans un échantillon de la population américaine qui est reproductible et potentiellement applicable dans le contexte canadien. Le modèle a été développé à l'aide d'une analyse de régression logistique multivariée. Ses prédicteurs comprenaient l'âge, le sexe, la race, les paquets-années de tabagisme, la consommation d'alcool de l'année précédente, le nombre de partenaires sexuels au cours de la vie, le statut d'infection orale au VPH et l'interaction bidirectionnelle entre le sexe, les paquets-années de tabagisme et le statut d'infection orale au VPH. Lors de l'évaluation de la performance globale, le modèle a présenté un D de Somers, un score Brier et un R2 de 0,49, 0,29 et 0,25, respectivement. L'indice de discrimination du modèle était de 0,19 et la statistique C était de 0,75 (0,69-0,79). Le modèle a surestimé les prévisions de 4,64 points, ce qui correspondait à l'interception de l'étalonnage. La pente d'étalonnage était

également de 0,57. En résumé, bien que le modèle ait montré un niveau de discrimination modérément élevé, ses performances d'étalonnage étaient médiocres.

Conclusion: Un nombre limité d'études de modélisation de la prévision des risques de CTC fournit suffisamment d'informations pour juger de la qualité et de l'applicabilité des modèles. Cependant, le modèle identifié peut toujours être utilisé dans le contexte canadien pour aider à des interventions préventives personnalisées. Considérant que le modèle présentait une discrimination acceptable mais un mauvais calibrage, il doit être recalibré et mis à jour sur le contexte canadien avant d'être mis en pratique. Des études futures sont nécessaires pour comprendre l'applicabilité de ce modèle dans les milieux cliniques canadiens.

Preface

This project is a manuscript-based thesis written according to the updated standards established by McGill Graduate and Postdoctoral Studies for fulfilling the requirements of a Master's degree in the Dental Sciences-Thesis program. The two manuscripts follow the primary goal of in this thesis by providing the Canadian context with an HNC risk prediction model. The first manuscript reviews the literature on the HNC risk prediction modeling aiming to understand the current status in this field and investigate the existing models' strength, risk of bias, and applicability in the new settings. The second manuscript investigates the generalizability of a model identified in the first manuscript in the Canadian context. Manuscripts are logically coherent with this thesis work and share a unified theme. Based on the standards of McGill University, each manuscript comprised a separate set of appendices and reference lists.

The first chapter of this thesis includes an introduction to the topic, which is then followed by reviewing the literature providing the current knowledge in the field. Supported by the second chapter, the rationale of this thesis project is then provided in the third chapter. The fourth chapter includes detailed objectives, while the fifth chapter provides the methods followed in conducting this thesis project. *Manuscripts I and II* are the next two chapters that stand alone regarding tables, figures, appendices, and reference lists. The eighth chapter comprehensively discusses this thesis work, followed by the ninth chapter, where the overall conclusions are provided.

The two manuscripts comprise multiple authors whose contributions to each manuscript are provided in the next section.

Contribution of Authors

Manuscript I

What is the current status of prognostic risk prediction modeling for head and neck cancer? A scoping review

Hamed Ghanati, Master's Candidate, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Conceived objective of the investigation, screened and selected studies, extracted data, assessed the quality of studies, statistical analyses, performed statistical analysis, and wrote the manuscript.

Mohammad Al-Tamimi, Master's Candidate, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: participated in the conceptual development, contributed to writing the manuscript.

Ziad Al-Asmar, Master's Candidate, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Participated in the screening and selection process.

Martin Morris, Schulich Library of Physical Sciences, Life Sciences and Engineering, McGill University, 809 rue Sherbrooke O, Montréal, QC, H3A 0C1, Canada: Conducted the search process, contributed to writing the manuscript.

Sreenath Madathil, Assistant professor, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Supervised all steps, contributed to the conceptual development, selection process, quality assessment, statistical analyses, interpretation, and writing the manuscript.

Belinda Nicolau., Professor, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Supervised all steps, contributed to the conceptual development, selection process, quality assessment, interpretation, and writing the manuscript (Corresponding author).

Manuscript II

Validating an oropharyngeal cancer risk prediction model on a Canadian population

Hamed Ghanati, Master's Candidate, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Conceived objective of the investigation, prepared data, performed statistical analysis, interpreted the results, and wrote the manuscript.

Belinda Nicolau, Professor, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Supervised all steps, contributed to the conceptual development, provided the databases, interpretation, and writing the manuscript.

Paul Allison, Professor, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: participated in the conceptual development, contributed to writing the manuscript.

Anita Koushik, Associate professor, Centre de recherche du CHUM and Department of Social and Preventive Medicine, Université de Montréal, QC, H2X 0A9, Canada: participated in the conceptual development, contributed to writing the manuscript.

Sreenath Madathil, Assistant professor, Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada: Supervised all steps, performed statistical analysis, contributed to the conceptual development, study design, interpretation of results, and writing the manuscript (Corresponding author).

1 Introduction

More than half a million people are diagnosed with cancers of the lips, oral cavity, oropharynx, hypopharynx, and larynx, collectively known as Head and Neck Cancers (HNCs)[a] each year around the world[1]. Despite its relatively low incidence, HNC have high survival rates compared to more common cancers such as breast and prostate cancers[2,3]. Last year in Canada, for example, an estimated 2,100 deaths occurred from 7,400 HNC cases[4]. Due to their location on the body, psychological impacts, and the side effects of treatments, HNCs are among the cancers with high morbidity and suicide rates[4-6].

More than 50 years of research have identified HNC risk factors including, but not limited to, age, sex, race, tobacco smoking, alcohol consumption, human papillomavirus (HPV) infection, dietary habits, body mass index (BMI), and socioeconomic position[4,7]. A significant proportion of global HNC occurrence is associated with tobacco and alcohol consumption. Nonetheless, there has been a rise in HPV-related HNC incidence, especially in high income countries countries[8]. Considering that most of the HNC risk factors are preventable (e.g., smoking), the preventive interventions are expected to reduce the number of new cases of these cancers.

Risk prediction models have recently become more popular in medical decision-making[9]. These models are mathematical equations that can estimate the probability of currently having a disease (diagnostic risk prediction models) or developing a disease in the future (prognostic risk prediction models) for an individual[9]. The predictions from these models can be used to identify the high-risk population and fulfill an effective preventive intervention[9,10].

Canada has witnessed a rise in the incidence of HPV-related HNC over the past few decades[11]. While Canadian federal government has promoted different programs to prevent HPV related diseases[12,13], HNC prognostic risk prediction models may be used to identify populations at high risk of developing HNC; thus making these programs more cost effective.

[a] In this thesis, the term "head and neck cancer" excludes the cancers in the area other than lips, oral cavity, oropharynx, hypopharynx, and larynx due to the difference in etiology.

Although different risk prediction models have been developed for HNC worldwide[14-17], Canada has not benefitted from any of them because these models need to be developed according to the population in which they will be used. In other words, they are context-specific. Therefore, there is a need to develop risk prediction models for HNC in Canada. However, before developing a new model, the potential applicability of the existing models should be evaluated to avoid redundancy[9].

To achieve this, we need a comprehensive review of the current models to identify those that can be applied to the Canadian context. In addition to helping identify potential models for use in Canada, such a review may also shed light on where we are standing in this field and reveal the existing models' quality and generalizability. To the best of our knowledge, there has been no such comprehensive knowledge synthesis on the current status of risk prediction models for HNC.

This thesis project addresses these gaps in the literature by first conducting a scoping review in HNC risk prediction modeling, describing the reported models' strengths and limitations. Subsequently, we use data from a Canadian population to investigate the performance of a model potentially applicable in the Canadian context.

2 Literature review

This chapter comprises an overview of HNC epidemiology and its risk prediction modeling, supporting the objectives of this thesis project.

2.1 HNC definition

HNC is defined as any type of malignancy that develops in or around the upper aerodigestive tract[18]. The anatomical locations routinely included in this definition are lips, oral cavity, oropharynx, pharynx, hypopharynx, and larynx. The general definition of HNC excludes the cancers of nasopalatine, sinuses, brain, and esophagus because of their different etiology. Recent publications also reported differences in the etiology of the cancers of lips and hypopharynx[19]. Therefore, we define HNC in this thesis as the cancers originating in the oral cavity, oropharynx, pharynx, and larynx.

About 90% of HNC begin with the malignant genetic change in the squamous cells that lines the head and neck mucosal surfaces (e.g., oral, buccal, or tongue mucosa)[8,20]. These cancers can also begin in the salivary glands, sinuses, or nerves in the head and neck area, but these types of malignancies are much less common than squamous cell carcinoma (SCC)[18,21]. The SCCs are considered as oral cavity SCCs if they are located in the buccal mucosa, the floor of the mouth, anterior tongue, alveolar ridge, retromolar trigone, and hard palate. SCCs of the oropharynx are defined as the tumors of tonsils, the base of the tongue, the posterior pharyngeal wall, and the soft palate. The larynx SCCs are considered as tumors in the supraglottis, glottis, and subglottis. Further, the SCCs stage I or II are small primary tumors without cervical lymph-node involvement, while stages III or IV are the large tumors or the ones with lymph-node involvement[18].

2.2 HNC descriptive epidemiology

The global burden of cancer is expected to rise from 18 million cases in 2018 to 29 million cases by 2040, primarily because of the global population growth and aging[22]. Approximately 10 million new cancer cases are diagnosed annually across the globe, of which 900,000 are the HNC cases[23].

Typically diagnosed in older patients in association with heavy use of tobacco and alcohol, HNC is slowly declining globally, partly because of decreased use of tobacco[24]. Conversely, cases of HPV-associated oropharyngeal cancer, induced primarily by HPV type 16[25], are increasing predominantly among the younger population in North America and northern Europe[26]. For example, Johnson-Obaseki et al.[11] reported that there has been an increase in incidence of HPV-related oropharyngeal cancer and a decrease in the incidence of oral cavity and other subsites of HNC among Canadians from 1992 to 2006. These authors also reported the decrease in age range of the patients with HPV-associated oropharyngeal cancer during these years. As another example of a north American country, the fraction of HNC cases diagnosed as HPV-positive oropharyngeal cancers in the United States rose from 16.3% during 1984–1989 to more than 72.7% during 2000–2004[27] due to the changes in sexual norms and behaviors that occurred in the late 1960's and early 1970s[28].

Approximately 400,000 deaths occur annually due to HNC worldwide[1]. In 2021, an estimated 2,100 deaths occurred among 7,400 Canadians diagnosed with HNC[4]. In the United States, 3% of all diagnosed malignancies are HNC (66,000 cases), of which 15,000 deaths occur annually[29]. A study published in 2015 by Gatta et al.[30] estimated that HNC cases comprised 4% of the overall European cancer burden in 2012. Notwithstanding the geographical difference in incidence, HNC is more prevalent among males than females, with a male/female ratio of 2:1 to 4:1[31]. The HNC incidence rate in males of the Indian subcontinent, Central and Eastern Europe, France, Spain, Italy, and Brazil can reach 20 per 100,000[32]. The prevalence of HNC subsites is not similar around the world. While oral and lip cancers are most prevalent in the Melanesia, oropharyngeal and laryngeal cancers are respectively more prevalent in Western Europe and Caribbean countries[33].

HNC is also more prevalent among African Americans than in the white population in the United States. Most of this burden is related to laryngeal cancer, which has a 50% higher incidence among African Americans. A study in 2009 reported a significantly higher mortality rate of African Americans with cancers of the oropharynx compared to a similar population with cancers in other HNC subsites[34]. Interestingly, HPV infection is less prevalent among these populations, and this

may explain the higher mortality as the HPV-related HNC has a relatively lower mortality rate compared to other types of HNC[35,36].

2.3 HNC Risk factors

HNC is recognized to be related to a set of specific risk factors. Some of these risk factors are demographical characteristics such as age, sex, and race, and some are behavioral factors or the ones related to a person's lifestyle such as smoking, alcohol, HPV infection, diet, BMI, level of education, and socioeconomic position. The following section elaborates on the current evidence on these risk factors.

2.3.1 HNC behavioral risk factors

2.3.1.1 Tobacco consumption

Tobacco consumption in any form (cigarettes, cigars, pipes, or hookah) is considered one of the factors that directly relate to the increased risk of developing HNC[37]. The causal relationship between cigarette smoking and the risk of HNC has been reported in a monograph published by the International Agency for Research on Cancer (IARC) in 1986[38]. This relationship also is shown in multiple prospective studies[39]. It is reported that the risk of developing HNC in heavy smokers is 5 to 25 times higher compared to non-smokers. The tobacco type, intensity, and duration are related to the increased risk of developing HNC[40]. Tobacco-HNC relationship is reported to be dose-dependent[40]. Exposure to secondhand tobacco smoke is also positively associated with an increased risk of developing HNC later in life[41].

Different types of tobacco consumption exist worldwide, but cigarette smoking is the most common form. It is estimated that over 1.1 billion people around the world smoke cigarette[42]. The prevalence of cigarette smoking varies in different parts of the world[43]. While 35% of men and 22% of women smoke cigarettes in high income countries, these rates are 50% and 9% in the middle and low income countries, respectively[44].

Other types of tobacco consumption are also reported to be positively associated with an increased risk of developing HNC. For example, results from a pooled analysis conducted in 2013 using data from the International Head and Neck Cancer Epidemiology (INHANCE) consortium reported a positive independent association between cigar and pipe consumption and an increased risk of developing HNC [37].

Tobacco smoking also differently affects the subsites of HNC. A meta-analysis revealed that the relative risk (RR) of developing laryngeal cancer in current smokers compared to non-smokers is 6.98 (with a 95% Confidence interval (CI) of 3.14-15.52)[39]. Pharyngeal cancer also presented a similar association (RR= 6.76; 95% CI: 2.86–15.98), while the relative risk of oral cancer was 3.34 (CI: 2.86-15.98)[39].

2.3.1.2 Alcohol consumption

The causal relationship between alcohol consumption and increased risk of developing HNC is reported by the IARC working group on evaluating carcinogenic risks to humans in 2010[45]. A recent study exploring 26 case-controls using data from the INHANCE consortium revealed that alcohol consumption is associated with cancers of the oral cavity, oropharynx, and larynx with an odds ratio of 7.95 (95% CI: 4.60–13.00), 12.86 (95% CI: 7.20–23.70), and 6.6 (95% CI: 4.90–9.00), respectively[46]. In a pooled analysis conducted in 2009 by Hashibe M. *et al.* [47], an effect measure modification was reported between alcohol and tobacco consumption, where the joint effects of tobacco and alcohol consumption increased the risk of developing HNC. According to this study, ever alcohol consumption alone had a weaker association with HNC (OR[b]= 1.06, CI: 0.88–1.28) compared to ever tobacco smoking alone (OR=2.37, 95% CI: 1.66–3.39), the joint effects of ever alcohol and ever tobacco consumption had the strongest association with HNC (OR=5.73, 95% CI: 3.62–9.06).

[b] OR = Odds Ratio adjusted for age, sex, education, race/ethnicity, and study center

2.3.1.3 HPV infection

By the time of writing this thesis, 229 types of HPV have been identified[48]. About 19 types of HPV are identified as high-risk HPVs for HNC. These are HPVs 16, 18, 26, 31, 33, 35, 39, 45, 51–53, 56, 58, 59, 66, 68, 70, 73, and 82[8]. Even though all these high-risk HPVs have been detected in oral, oropharyngeal, and laryngeal samples, 0 to 80% of them are cleared in healthy individuals within the 6 to 20 months of infection. According to a recent systematic review of literature on oral HPV samples of cancer-free subjects, HPV-16 has the highest clearance time (7 to 22 months), resulting in a longer persistence in the oral cavity[49]. An approximate 85% of the HPV-related HNC cases are positive for HPV-16 and HPV-18[8]. The strongest association is reported between these genotypes and oropharyngeal cancer compared to other sub-sites of HNC[50].

The prevalence of oral HPV infection is different among various age groups, with a peak at 25-30 years and 55-60 years of age[51,52]. HPV infection incidence is also increasing, specifically in the high-income countries. A strong sex and birth cohort-specific trend is reported for the western countries, while younger men are the most affected cohort[27,51,53,54].

The most prevalent way of HPV infection transmission is through sexual contacts[55]. A recent study of sexual behaviors in young women revealed that greater numbers of lifetime and past-year sexual partners, deep kissing, oral-genital, and coital sexual exposures are similarly associated with oral HPV infection.

2.3.2 Other risk factors

Apart from tobacco, alcohol, and HPV infection, other factors are known to be associated with HNC. One of these risk factors is indicators of socioeconomic position (SEP); those with low SEP have an increased risk of developing HNC[56-58]. Moreover, diet poor in fresh fruits and vegetables[59,60], sedentary lifestyle[61], poor oral health status (periodontal health)[62,63], poor-fitting dentures[64-66], and diseases of the oral mucosa (e.g., pre-malignant conditions)[67] have been associated with the increased risk of developing HNC.

2.4 Risk prediction models

2.4.1 Definition

Risk prediction models are mathematical equations that estimate the likelihood of a condition (dependent variable) based on a set of independent factors (predictors)[68]. These models can be developed using statistical techniques such as regression analysis. For example, if the outcome of interest is a binary or dichotomous variable, which is the case in predicting the risk of developing HNC, the logistic regression analysis is used to estimate the predicted probability of the outcome. Formula 2-1 represents a logistic regression risk prediction model:

Equation 2-1: $Logit\big(P(Y=1)\big) = \beta_0 + \sum_{i=1}^{n} \beta_n x_n$

Where P is the probability that the outcome happens (probability of Y = 1), β_0 is the intercept value, β_n is the value of coefficients of the independent variable n, and x_n represents the value of the independent variable n. The left side of the formula 2-1 can be unraveled as the formula below:

Equation 2-2: $Logit\big(P(Y=1)\big) = log(\frac{P(Y=1)}{1-P(Y=1)})$

In the medical field, these models are used to predict the likelihood of the disease or medical condition that currently exists (diagnostic risk prediction models) or to estimate the probability of the development of an outcome or medical condition occurring in the future (prognostic risk prediction models)[9,69] given individuals' demographics and behavioral profile (e.g., age, sex, tobacco smoking, and alcohol consumption), disease characteristics, and diagnostic test results. The predictions from these models are used for different purposes, such as predicting the risk of development or recurrence of a disease or medical condition, assessing the risk in recruiting participants for trials or helping the public health policies where health policy-makers need to predict the diagnosis or prognosis of disease to make decisions on screening, therapeutic, or preventive measurements[9]. Using the probabilistic risk estimates derived from the risk such

models, the health care providers or even the individuals themselves can decide on further disease management.

2.4.2 Prognostic and diagnostic risk prediction models

As discussed above, risk prediction models can be categorized into two main groups: i) diagnostic risk prediction models, and ii) prognostic risk prediction models.

A diagnostic risk prediction model predicts an individual's likelihood of having the disease at present. As an example in the HNC field, a diagnostic risk prediction model for predicting the malignancy of a suspicious oral lesion may use the patient's characteristics (e.g., age, sex, and race), patient's behavioral factors (e.g., smoking, alcohol, diet), and the pathological test results (e.g., size, color, and texture of the lesion) as predictors[67].

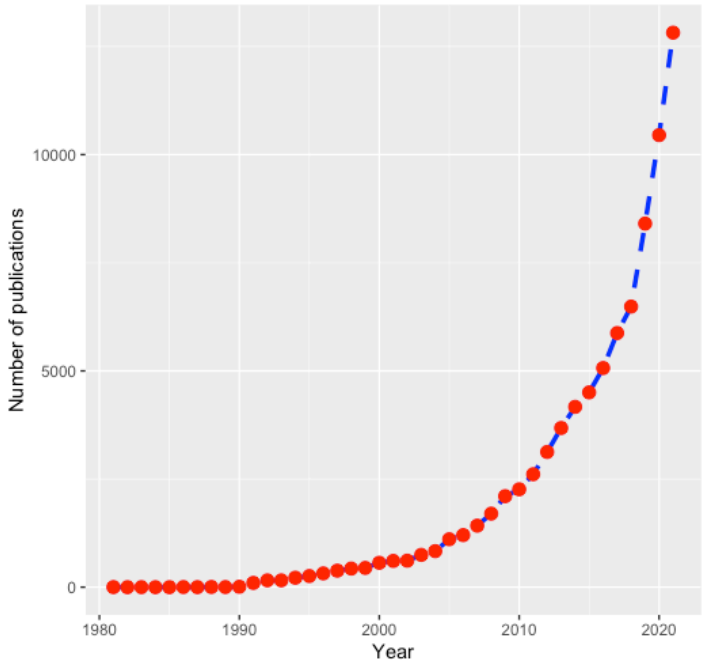
On the other hand, a prognostic risk prediction model estimates an individual's probability of developing an outcome or medical condition in the future [9,69]. The predictors of these models are usually, but not limited to, the individual's sociodemographic and behavioral characteristics related to the outcome of interest. An example of prognostic HNC risk prediction is the model developed by Lee YA *et al.* that uses an individual's age, sex, race, level of education, smoking, and alcohol consumption to predict the risk of developing HNC in the next 20 years[14]. Depending on the outcome of interest, the prognostic risk prediction models may also comprise the predictors related to the treatment of the outcome. For example, a prognostic risk prediction model can predict the risk of major postoperative adverse events after HNC surgery based on the patient's age, sex, smoking habit, and the factors related to the surgical treatment such as operative time and wound classification[70]. For simplicity, the word "risk prediction model" will be used as the synonym for "prognostic risk prediction model", in the rest of this thesis.

2.4.3 A brief historical overview of risk prediction modeling

To better situate the reader in the subject, this section provides a historical overview of risk prediction modeling. The first paper indexed in the Web of Science with the keywords of "risk"

and “prediction” and “model” was published by Kager *et al*. in 1981[71]. The paper was about a clinical model for predicting the changes in the colon microflora after using a new antibiotic in colorectal cancer prophylaxis. Since then, the publications on risk prediction modeling have exponentially increased. The number of indexed scientific publications on the Web of Science with the same keywords published in 2018 has doubled and reached 12,820 publications in 2021. Figure 2-1 demonstrates this sharp increase.

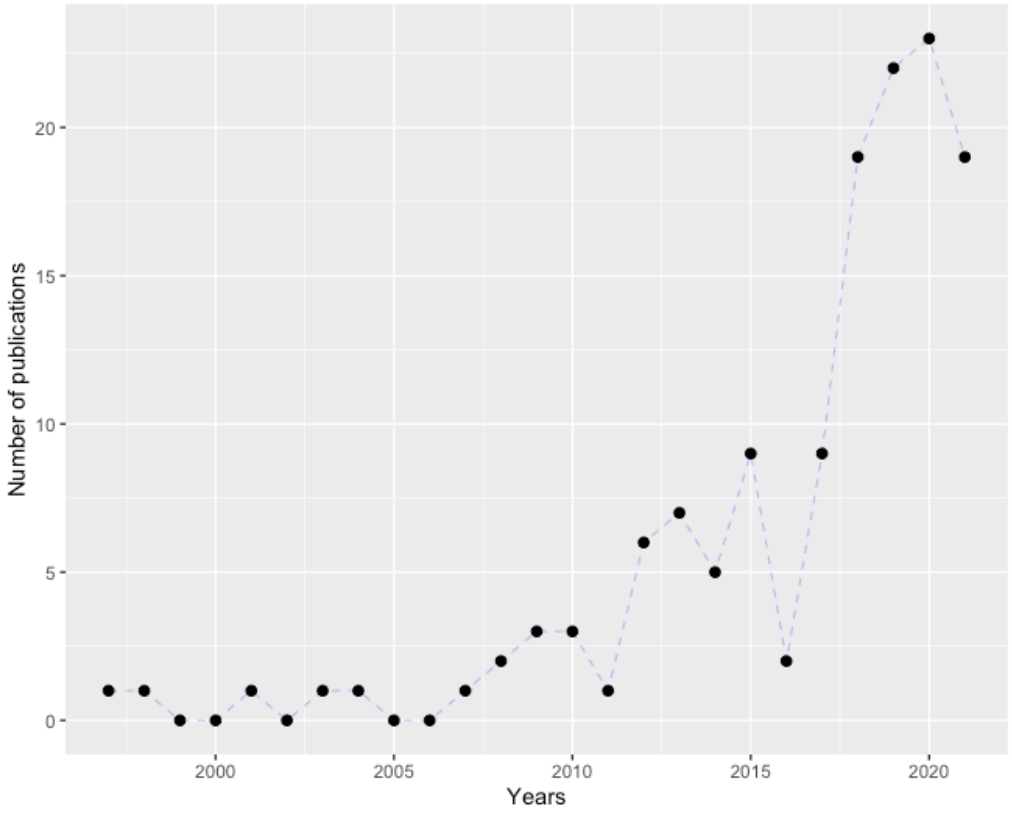
Figure 2-1: Number of publications indexed in Web of Science with the words “risk” and “prediction” and “model”.



Almost half of these articles are related to the health sciences, of which 781 are the share of oral health sciences, including the HNC risk prediction modeling. The first published risk prediction modeling study in oral health sciences was conducted by Abernathy *et al.* in 1987[72]. The study reported the development of a model for predicting the risk of dental caries in 1st and 5th–grade children in the National Preventive Dentistry Demonstration Program[73]. A search on the MEDLINE database through Ovid for the keywords of “head and neck cancer” and “risk” and “prediction” and “model” results in 143 publications. One of the earliest risk prediction modeling studies in the HNC field was published by Cole *et al.* in 1987[74], which reported the development of a model

for predicting the risk of wound sepsis in HNC oncologic surgery. One of the oldest report of an HNC prognostic risk prediction model is published by Baatenburg de Jong *et al.* in 2001[75]. The report used a cox proportional hazard regression model to predict the survival of patients newly diagnosed with head and neck squamous cell carcinoma. After that, there was a period of stagnation in HNC risk prediction modeling until 2015. Since then, a considerable rise in the subject has been observed, with a more noticeable increase in the past five years (Figure 2-2).

Figure 2-2: Number of publications on MEDLINE through Ovid with the keywords of "risk" and "prediction" and "model" and "head and neck cancer".



A review of the literature reveals that the majority of the HNC risk prediction models are developed to predict the survival[76] or post-surgery complications[77-79]; there are a limited number of models developed for predicting the risk of developing HNC in the future (HNC prognostic risk prediction models).

2.4.4 Evidence of using HNC risk prediction models

A quick literature review reveals that different risk prediction models have been developed to predict the prognostic outcomes related to HNC. Some models have considered HNC as a single outcome[80,81], but others considered one or more subsites of HNC as separate outcomes of interest[17,76]. Overall, nine peer-reviewed articles have reported the development of models predicting the prognostic risk of developing HNC or its subsites[14,15,17,82-87]. These models have served as an assistant tool in different settings. For example, the model developed by Cheung *et al.*[82] identified a high risk population in a risk-based selection for a screening program in Kerala, India. As an example of implementing the HNC risk prediction models in public health, the models introduced by Tikka *et al*.[16,88] served as a complementary tool in refining the HNC patients' referral guidelines in the UK. Tota *et al*.[17] also explored the geographical transportability of an HNC risk prediction model in the US. However, to our knowledge, there is no comprehensive reviews in HNC risk prediction models. Importantly, little is known about the quality and applicability of these models and other models in different settings. The first part of this thesis project aims to fill this knowledge gap by conducting a systematic scoping review.

2.4.5 Modeling strategies

Certain approach should be employed to ensure a good model development and validation. Steyerberg and Vergouwe (2014)[89] suggested a seven-step strategy that should be adopted by every prediction modeling study that uses logistic regression technique. In this section, I will elaborate these steps in the context of the thesis and provide more details on the challenges in developing or validating a good HNC risk prediction model.

2.4.5.1 Defining prediction problem and inspecting data

Understanding the prediction problem and the context in which the study is being conducted are essential to identify an appropriate data source and conduct a good risk prediction modeling study. Consider the development of a HNC risk prediction model study as an example. Preferably, the data are derived from a longitudinal study (e.g., cohort) that is specifically designed and conducted for the modeling study[9,90]. However, HNCs are diseases with long latency period (i.e.,

HNCs take a long period to present as an overt pathology) and relatively low incidence compared to other medical conditions, thus, to design a HNC prediction modeling longitudinal study will require long follow ups and large sample size, which is not cost effective[9,10]. To overcome this challenge, researchers routinely use data derived from a retrospective study (e.g., case-control) or cancer registries to develop HNC risk prediction modeling[9,90]. It is important to mention that when data are collected for a purpose other than the modeling study, the model's predictions are in danger of a high risk of bias. Moons et al.[90] provided detailed information on how a model might be in a high risk of bias when the data source is different from the modeling study.

The quality of dataset also is important to ensure a good modeling study. One of the important factors determining the dataset's quality is the number of missing values. Most statistical software, by default, excludes the subjects with missing values, thus, addressing this problem is one of the essential steps before conducting any modeling study. One way to address this issue is to use imputation techniques, which replace a missing value with some substitute value to retain most of the data information. Altman and Bland (2007)[91] have emphasized on the advantages of imputing missing values over the available case analysis (excluding the participants with missing data).

2.4.5.2 Coding predictors

In addition to having a robust dataset, the predictors from it needs to be coded appropriately to ensure unbiased estimations from the model. Preferable, numerical continuous variables (e.g., age, tobacco smoking, or ethanol consumption) should be kept in their original form. However, these variables are often measured in categories for ease of data collection in many studies. Furthermore, categorizing the numerical continuous variables has raised heated debates for a long time[92-95]. The important point is that categorizing a continuous variable involves choosing arbitrary cut points (e.g., five-years age groups). These cut points should be determined based on the literature on the predictor's association with the outcome. It needs to be highlighted that dichotomizing a continuous variable may result in information loss, and a model that includes dichotomized continuous predictors will have poor predictive ability[93,96,97].

Categorical variables also need specific considerations before modeling. A single category with infrequent participant can be collapsed with others to enhance the model's fit and predictive ability[89]. For example, certain types of HPV are associated to the increased risk of developing HNC; however, the frequency of HPV 16 and 18 is much higher in the histological samples of HNC than other types of high-risk HPVs[2]. Therefore, other high-risk HPVs can be combined into the "other high-risk HPV" category to improve the model's fit as well as clinical interpretability.

2.4.5.3 Selecting predictors and specifying the model

It is suggested to select the predictors for a risk prediction model based on the literature review and previous knowledge on the association between the outcome and candidate predictors[9,90,98,99]. There are also statistical techniques that can help in selecting the predictors[99]. The simplest statistical method is to conduct a univariate analysis by fitting a model with a single predictor and decide about keeping the variable based on its univariate correlation with the outcome[99]. This method has been criticized because it may result in wrong predictor selection by dropping the potentially effective predictors[90,99]. Indeed, this method ignores the fact that some potential variables recognized as insignificant predictors during the univariate analysis may still significantly contribute to variation in the outcome when they are combined with other predictors. To address this shortcoming, backward and forward predictor selection techniques have been proposed[99,100].

Backward selection begins with a full model that includes all candidate variables and removes one variable at a time until all the significantly correlated predictors are remaining. The advantage of this method is that it makes it available to assess the joint predictive ability of different variables[100]. However, in this method once a variable is deleted from the model it will not be re-entered again. Thus, it is not possible to check the significance of the dropped variable in the final model that contains different sets of predictors[99].

Forward selection, as its name indicates, starts with a single variable model and continues adding the predictors until no variable that are added to the model can make any significant contribution

to the outcome variable[100]. This method has the advantage of low susceptibility to the collinearity between two or more predictors in a model. However, in forward selection, the existing variables in the model can be rendered non-significant by adding a new variable in the next step, though the existing variables cannot be discarded[99].

To keep the advantages of backward and forward predictor selection methods while addressing their shortcomings, the stepwise predictor selection method has been suggested[99]. In this method, the predictor selection process goes in both directions and, at each step, once a new variable is added to or deleted from the model, all other included variables are checked, and the non-significant variables are eliminated based on their contribution to the outcome. Stepwise method is the most commonly used technique in selecting the predictors for the risk prediction models in medicine[89]. However, like all the statistical techniques, this method also needs to be used with caution[9,89]. In fact, the predictor selection based on the statistical techniques might be instable when the sample size is small[89]. A low number of events results in extreme regression coefficients during the selection process, thus a distorted estimation of the performance of the selected model[100]. Moreover, although statistical methods help in selecting the significant predictors based on the available dataset, they may not always produce a clinically useful risk prediction model because the best fitted model might contain a predictor that is hard to be measured in a clinical setting[9]. For example, studies have reported that aldehyde dehydrogenase 2 (ALDH2) gene polymorphism makes the individual susceptible to the carcinogenic effects of alcohol[101-103]. Koyanagi et al.[15] reported that including the variable for ALDH2 polymorphism increases the model's predictive performance. Statistical methods in predictor selection may include this variable in a model. However, it is hard or costly to measure the polymorphism of ALDH2 in the routine clinical practice, the applicability of the model for the primary clinical settings is questionable[15].

Indeed, investigators should use a combination of statistical methods, clinical knowledge, and evidence from previous studies to ensure that an appropriate set of predictors, which are clinically applicable, are included in the models[9,89,90].

2.4.5.4 Estimating coefficients and optimism check

Following the predictor selection, the weight (coefficient) of each variable will be estimated. For regression based risk prediction modeling, often a maximum likelihood estimation process is used[89]. The important point here is to make the model as generalizable as possible. Particularly, optimism of the model should be checked to ensure that its predictions are not overfitted to the dataset used to develop the model. Once the value of optimism is calculated, the model's performance metrics (e.g., calibration) can be corrected to make the model generalizable. Optimism check can be done by internal validation of the model[9]. Section 2.4.5.6 provides more details on different methods of internal validation.

2.4.5.5 Risk prediction model performance measurement

A risk prediction models' predictive performance requires careful evaluation before implementing the model in practice. Steyerberge et al.[104] proposed a framework for assessing the prediction models' performance. This framework mainly focuses on assessing the overall performance, discrimination ability, and calibration, of a prediction model. I will elaborate on this framework in this section.

2.4.5.5.1 Overall performance

The overall performance of a risk prediction model can be assessed by measuring the Somers' Delta (Somers' D) statistics, Brier score, and Nagelkerke R^2 index.

Somers' D statistics, developed in 1962 named after his originator Robert H. Somers[105], evaluates the rank correlation between the predicted probability and the actual value of the outcome. This metric can take values between -1 and 1, with -1 indicating that all pairs of prediction-outcome disagree and 1 indicating that all pairs of prediction-outcome agree. Somers' D (D_{yx}) and can be calculated by the formula below:

Equation 2-3: $D_{yx} = P(y = 1|x = 1) - P(Y = 1|x = 0)$

where y is the value of predicted probability of outcome variable derived from the model and x is the value of actual outcome[106].

The Brier score, first proposed by Glenn W. Brier in 1950[107], is the square of the differences between the predicted values and the actual values of the outcome. The Brier score shows how accurate the model predictions are. The formula for calculating the Brier score is as follows:

Equation 2-4: $Brier\ Score = \frac{1}{N} \sum_{i=1}^{N} (Y - P)^2$

where N is the number of observations (participants in the dataset), Y is the value of the observed outcome, and P is the predicted value of the outcome variable. As inferred from the model, the lower the Brier score, the better the overall predictive performance of a model. The best performed risk prediction model that has no error in the predictions receives a Brier score of 0. If a model is non-informative, that is, it assigns the same probability to all the predictions, the model receives a Brier score of 0.25. The Brier score is usually scaled and reported as a percentage. To scale a Brier score, the maximum possible Brier score of the model is used following the below formula:

Equation 2-5: $Scaled\ Brier\ Score = 1 - \frac{Brier\ Score}{Maximum\ Brier\ Score}$

Where the maximum Brier Score is achieved by:

Equation 2-6: $Maximum\ Brier\ Score = mean(P) \times \big(1 - mean(P)\big)$

Where $mean(P)$ is the average of the predicted probability value for the outcome variable. As mentioned above, the scaled Brier score ranges from 0% for the non-informative model to 100% for the perfect model[9,104].

Nagelkerke R^2 is also one of the useful metrics in overall performance measurement. Measuring the R^2 (Explained variation) is the common method of performance assessment for continuous outcomes. Nagelkerke R^2 is similar to Pearson's R^2, but estimated for a generalized linear models such as logistic regression. This metric can be calculated using the formula below:

Equation 2-7: $\mathrm{R}_{\mathrm{N}}^{2} = (\mathrm{Y} \times \log(\mathrm{P})) + ((\mathrm{Y} - 1) \times \log(1 - \mathrm{P}))$

where Y is the value of the actual outcome and P is the predicted probability derived from the model. Based on the formula 2-7, the less the value of R_N^2, the better the overall performance of the model[9,104].

2.4.5.5.2 Discrimination ability

Discriminative ability of a model is defined as the capability of a model in discriminating between participants with and without the outcome. Discrimination capability of a model can be evaluated by measuring the Concordance statistic (C-Statistic) and discrimination slope (index).

The C-statistic of the risk prediction model is defined as the probability that a randomly selected participant, with outcome of interest, will receive a higher predicted probability compared to a randomly selected participant without the outcome[9]. In fact, C-statistic indicates how good the model classifies the participants into two groups. With a binary outcome, the value of C-statistic equals to the area under the receiver operating characteristic (ROC) curve (AUC). ROC curve can be derived from plotting the sensitivity of the model (true positive rate) over 1–specificity (false positive rate)[108].

Discrimination slope (index) is another metric that can measure how well the model separates the subjects with and without the outcome[109]. This metric also can be obtained by calculating the absolute difference in average predictions for those with and without the outcome[104]. Discrimination slope can also be visualized by plotting a histogram or boxplot of the predictions

for those who have and have not the outcome[104]. The more the value of discrimination slope, the better the model can distinguish between those with and without the disease.

2.4.5.5.3 Calibration

Calibration of a risk prediction model is the agreement between the predicted and observed values of the outcome variable[110]. A well-calibrated model correctly predicts a decile of probability 100P % of the time with P confidence. Considering a perfect calibrated model, for instance, if the model predicts 10% risk of developing HNC for a number of participants, 10% of these participants should be HNC cases.

Calibration of a risk prediction model can also be visually evaluated by plotting the predictions and observed values in a two-dimensional graph. If the outcome variable is a continuous, the calibration plot is a scatter plot. If the outcome is binary, such as predicting the risk of developing HNC, the calibration plot's Y-axis will be the observed proportion of participants with the outcome of interest at a specific predicted risk level. For ease of interpretation, predicted probabilities are grouped based on percentiles (e.g., deciles) while generating the calibration plot. The x-axis of such calibration plot contains the groups of predicted probabilities. To refine the calibration plot, the loess algorithm can be used to smooth the calibration plot[106]. A diagonal line in the calibration plot represents the ideal calibration. That is for a perfectly calibrated model, the predicted probabilities(x-axis) are equal to observed proportions (y-axis). Evaluating the performance of a model through the calibration plot, entails visualization and comparison of a smoothed calibration line of the model to the ideal calibration line[9]. For example, if the smoothed calibration line is above the ideal line, the model is underestimating the probabilities of the outcome at that point. Whereas, if the smoothed calibration line is below the ideal line the model is overestimating the probabilities of the outcome at that point.

Apart from the calibration plot, the deciles of predictions (groups of participants with the same value of predicted risk) should also be compared to the related deciles of observed values to understand the model's goodness-of-fit. This comparison is done by Hosmer Lemeshow

goodness-of-fit test. This test groups the predictions and compares the proportion of data with outcome of interest in that decile[9].

2.4.5.6 Validation of the risk prediction models

The predictions of a risk prediction model must be valid for new settings. Generally speaking, the dataset that the model is derived from (origin dataset) is only a means to learn for the future. In simple words, the model is implemented for the individuals whose outcome status is unknown at the time of implementing the model. Therefore, the model's predictions should be validated before implementation. There are two types of model validation: Internal validation and external validation.

Internal validation is generally conducted during the model development process, where the model's optimism and performance are tested using the origin dataset or a sample from origin dataset. The external validation is always done by checking the model's performance on a dataset with different population of interest or the same population but different time. External validation helps to understand the model's generalizability.

2.4.5.6.1 Internal validation

Internal validation is defined as validating the predictions of model using the origin dataset. Split sample validation, K fold cross-validation, and bootstrapping are three of the most commonly used internal validation techniques in risk prediction modeling.

For split sample validation, the origin dataset is randomly split into two parts. One part (training dataset) is used to train the model, and the other part (validation dataset) remains for testing the model's predictive performance. Testing the model in this technique sometimes is mistakenly interpreted as external validation as the validation dataset literally comprises different participants. However, the training and validation datasets are derived from a same study with the same research question, thus this method should be considered as internal validation[9]. This

method is usually criticized because partitioning the datasets may drastically reduce the sample size that can be used for leaning the model[9,10].

As an alternative to the splitting technique, the K-fold cross-validation has been introduced to avoid sample size reduction. This is a resampling strategy that has a single parameter known as K which refers to the number of smaller sets that a dataset is divided into. In other words, a given data sample is divided into K folds in which one of them is kept for testing the model and the remaining are used to training the model. This process repeats K times, thus K models are developed. The overall model performance is reported based on the average estimated performances of the K models. As a more elegant method of cross-validation, each trained model is tested on the whole dataset, and then, the average of models' performance is reported.
The K fold cross-validation has the advantage of using a larger amount of sample size for model training than split-sample validation. Nevertheless, it is necessary to repeat the process several times (e.g., 10 times), which can be to obtain stable performance measurements[9].

There is a modification of cross-validation similar to the Jackknife technique[111,112], in which the cross-validation repeats for every participant. In simple words, if N is the number of participants in the dataset, the model is trained on N–1 participants and tested on the remaining 1 participant. This process repeats for N times. Therefore, the number of developed models in this technique equals to the number of participants. It is obvious that the Jackknife technique is inefficient when the dataset is large. Also, this technique might underestimate the models' variability. In other words, the difference between the training and validation datasets is only one participant, and typically, a more or less the same set of predictor coefficients are used for every model. This issue also arises when a higher number of folds is implemented, and the size of validation dataset is relatively small[9].

To overcome the underestimation of variability in K-fold cross validation, bootstrap validation has been introduced. In bootstrapping, different datasets are created by resampling from the origin dataset. Each bootstrapped dataset contains different replacements of the origin dataset's

participants. During the bootstrap validation, the model's performance is evaluated on every bootstrapped dataset. The mean difference in the model's performance on the origin dataset and bootstrap datasets reflects the amount of model optimism. Bootstrapping is sensitive to the sample size, thus, it is recommended to repeat the bootstrap validation at least 500 times to reach to the stable results[9].

2.4.5.6.2 External validation

To understand a model's generalizability and applicability in a new setting, its predictions should be validated on a different dataset (external dataset). This validation, which is usually called "external validation", is indeed testing the developed model's predictive performance on a different dataset. Depending on the external dataset's population of interest, the external validation study can answer the questions regarding the model's temporal, geographical, methodological, or spectrum transportability[9].

To assess a model's temporal transportability, a dataset is obtained from the same population of interest at a different time period (often more recent than origin dataset period). Temporal external validation ensures that the model's predictions are still valid for the recent same settings. For example, two nested case-control studies can be conducted using the data from a prospective cohort study. One of these nested case-controls recruits participants' information regarding a specific year, while the other uses the same information regarding some years later. The first case-control is used to develop the model and the second provides the dataset for temporal external validation[9].

Geographical transportability of a model can be assessed using the data from a population from other places (e.g., other countries, or geographical areas). The process that Tota *et al.*[17] followed to validate their risk prediction model for oropharyngeal cancer is an example of validating the geographical transportability of a model. They developed the model using a case-control study conducted at the Ohio State University Comprehensive Cancer Center[17] and validated the model on a dataset derived from a case-control conducted at the Johns Hopkins University[9,113,114].

Methodological transportability reveals that if the model's predictions are valid when different data collection procedures are used. Considering a HNC risk prediction model with HPV laboratory test result as a predictor, an external validation study answering the model's methodological transportability may use the "number of sex partners" instead of the HPV laboratory test result and check the model's performance[9].

Spectrum transportability is also checked to ensure that the model can work on a different care setting. For example, if the model derived from the data regarding the secondary care setting predicts well on the data related to a primary care setting, the model has a good spectrum transportability[9].

Specific considerations are needed to ensure a good external validation process. Choosing an appropriate model plays a significant role in the external validation success[9].

The ideal candidate model for external validation is the one that has been internally validated, and the details of its different characteristics have been fully published[9,90]. Such a model could be easily replicated, and its predictions are comparable to the development study's results.

The candidate model's population of interest should also be similar to that of the external validation study because the considerable difference between two populations may result in the significant variation in the predictors' effect on the outcome variable. Imagine a model developed to predict the risk of developing oropharyngeal cancer in India, where HPV infection incidence is relatively low compared to high income countries. The model gives a relatively low weight to the HPV variable, thus its predictions may not be applicable to the Canadian context, where HPV-related oropharyngeal cancer is more prevalent than India[115,116]. Therefore, if we want to externally validate this model on a Canadian population, its estimations would be incorrect.

It is also recommended that the external validation process starts with replicating the model without changing its predictor set or the values of coefficients and intercept. The exact replication of the model makes it possible to compare the model's performance during the external validation with that of the development study. Such a comparison reveals the model's strengths, limitations, and the areas for improving the model's performance in the new setting[9].

In case the model's performance is suboptimal in the external validation, model updating can be considered. Different updating methods have been proposed. Ewout W. Steyerberg has provided details on the recommended model updating[9]. Briefly, the model's update starts with recalibrating the model and adjusting its predictions for the baseline risk of the target population (Calibration-in-the-large). Next step will be updating and recalibrating the values of intercept (baseline risk) and slope (coefficients of variables). Updating can further be proceeded to the model revision, in which the recalibration of slope is accompanied with the selective re-estimating the coefficient values (refitting the model and selective updating the coefficients) on the external validation study's dataset. As the extensive updating method, model extension could be conducted by selectively adding the predictors to the recalibrated and revised model. The important point about model updating is that all of the updating methods result in a change in the original model. Therefore, the updated model is not same with the original model. In other words, updating the models is considered as a small-scaled model development, which should follow the standards of model development methods such as internal validation after updating the model[9].

2.4.5.7 Presenting the model

Appropriate presentation of a risk prediction model helps in its applicability in practice. The output of these models is a number as a predicted probability of developing the outcome based on the information about the specific predictors. Researchers are generally interested in the models' characteristics and specifications, therefore, presenting the model as the regression equation is one of the essential steps[9]. Clinicians also are also interested in the outcome of the models. A wide variety of paper-based tools are available to help in the model's applicability in

routine clinical practice, including score charts and nomograms[106]. Recently, there has been an interest in presenting the risk prediction models as the mobile applications or web-based tools[89]. It is estimated that risk prediction models will be integrated into clinical decision aids and electronic patient records in the future, helping clinicians, researchers, and public health administrators in different settings[89,104].

2.4.6 Applicability in clinical settings, public health, and epidemiology

Risk prediction models are helpful tools in evidence-based medicine. Their predictions may support the decisions made in the clinics, public health, and epidemiology. Table 1 provides an example of how risk prediction models can be used to answer the questions in these three main domains.

Table 2-1: Examples of implementing the risk prediction models in different settings

Domain	**Example question**	**Decisions based on model predictions**
Clinic	"Will the patient develop HNC in future?"	Motivation counselling
Public health	"How to increase the HPV vaccination's efficiency?"	Prioritizing the high-risk population for HPV vaccination
Epidemiology	"How many HNC should be expected in the control arm of a new HPV vaccination's trial?"	Determine the minimum sample size for the trial

In clinical setting, model's predictions provide patients and the clinicians with an overview of the risk of developing the outcome. This risk overview will be a useful complementary tool in decision making about management of the disease, counselling processes, and patient classifications among. For example, a HNC risk prediction model can be used to decide about requesting further diagnostic tests or warn a high-risk individual about risky behaviors such as heavy smoking or

alcohol consumption. This model can also act as an individualized encouraging tool and a person can track the risk changes during the time after high-risk behaviour amendment. The model's predictions also help clinicians in classifying the patients based on the risk of developing HNC. Such a risk-based classification is helpful in communication among physicians[9].

In public health, the risk predictive models could help in recognizing high-risk individuals for targeted preventive interventions. It has been reported that the individuals with the highest risk of developing a disease will benefit the most from a preventive measurement[117]. Therefore, the efficacy of public health preventive measurements such as intensive screening for oropharyngeal cancer or chemoprophylaxis of suspicious oral lesions is expected to improve by targeting these measurements to high-risk individuals. Consider HPV vaccination as another example of a public health preventive strategy. The public health policy makers can better plan for effective prevention by prioritizing the HPV vaccination candidates based on the model's predicted risk[9].

The risk prediction models can also act as helpful tools in different epidemiological research. In a randomized controlled trial (RCT), the models' prediction can assist in recruiting the participants in different arms. Imagine an RCT project that aims to evaluate the effect of an HPV preventive mouthwash. The participants in different arms should have a balanced baseline prognostic risk of developing HNC to avoid biased results. The estimations of a HNC risk prediction model can help in creating this balance between different arms[9].

It is worth highlighting that the risk prediction models' results are supportive material in decision making and must not be considered as the final determinant. The models' predictions contain a degree of error, thus, there is no utopian model that can perfectly predict every individual's risk. However, the predictions are useful in directing the decisions and supporting the evidence.

2.5 Summary of literature review

This chapter presented a literature review on HNC epidemiology, definition and strategies for developing and validating a good risk prediction model with specific emphasis on HNC. Also, an

update on recommended metrics for performance assessment and validation methods was provided.

Based on what was discussed here, HNC risk prediction modeling studies need to follow standard methods to have their models applicable in practice. Although different papers have reported development and validation of risk prediction models for HNC[14,15,17,82,83,85-87,118], little is known about these models' quality, risk of bias, and applicability in new settings. Moreover, we did not find any study using HNC risk prediction models in a Canadian population. It seems logical that the first step to choose a model for the Canadian population is to review the existing models and investigate the possibilities of using these models in the Canadian context. Next chapter will provide more details by shedding light on why we aimed to review the scope of HNC risk prediction modeling and externally validate a potential model in the Canadian context.

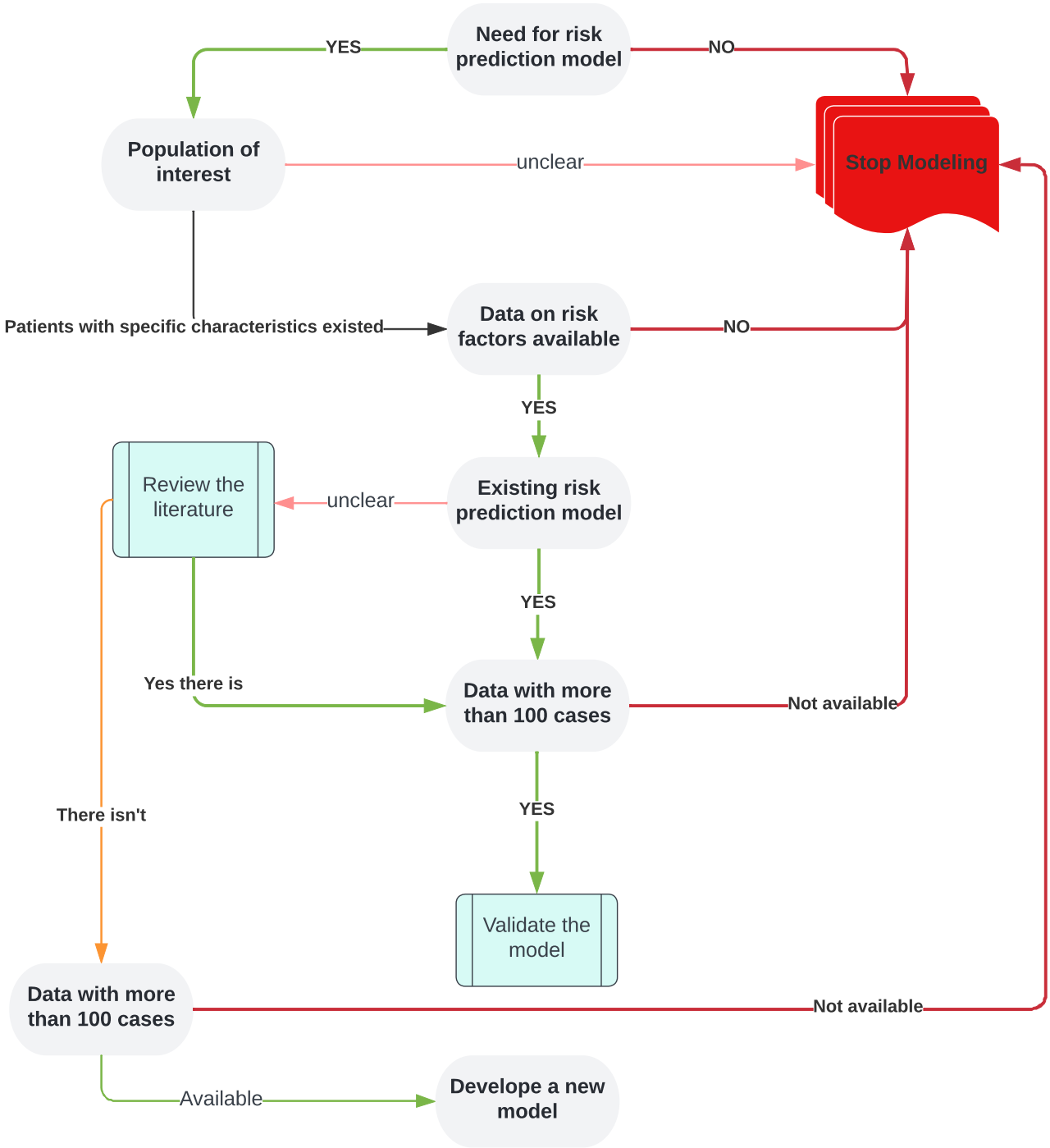
3 Rationale

It is estimated that 7,500 Canadians will be diagnosed with HNC in 2022, of which 2,100 deaths will occur in this year[4]. After several years of stable or decline of HNC incidence, mainly due to anti-tobacco policies and regulations, there is recent sharp rise in HNC. This increase has been mainly in the oropharyngeal subsite, which is primarily driven by HPV infection [119-121]. Indeed, there has been a substantial increase in oropharyngeal cancer in high income countries, with the HPV infection being responsible for most of its new cases in North America and Western Europe[122-124].

Effective preventive interventions should prioritize and target high-risk populations[12]. In this way, prognostic risk prediction models for HNC may play a significant role in assessing the risk and recognizing Canadians at high-risk of these diseases. Although several HNC prognostic risk prediction models have been proposed in the literature, none of them has used data from Canadian population. Moreover, knowledge is lacking on these models' quality, risk of bias, and applicability in new settings. Considering the strategies for cancer control in Canada[125], sound HNC as well as oropharyngeal cancer risk prediction models are necessary to identify high-risk Canadians.

Ewout W. Steyerberg, in the book "Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating," suggests model developers examine the existing risk prediction models' performance and check their applicability before developing new risk prediction models[9]. This approach will prevent redundancy and help in refining the existing developed models. Flowchart 3-1 displays an adaption of the diagram published by Maarten van Smeden, which helps to understand the best practice of developing a risk prediction model[126].

Figure 3-1: The guide for risk prediction model developers, adapted from the flowchart provided by Maarten van Smeden.



This flowchart provides specific questions to be answered before conducting a modeling study. These questions could be framed in modeling HNC in Canada:

1- Is there a need for a risk prediction model for HNC, especially for oropharyngeal subsite, in Canada?
2- What is the target Canadian population of interest?
3- Are the necessary data about HNC and its subsites in the target population of interest available?
4- Can any existing model for HNC or its subsites be validated or updated in the Canadian context?
5- Is there any Canadian dataset containing a large sample size?
6- Are the model predictors identified by solid research on the risk factors?

The first question's answer is a definite "Yes" based on what is discussed in the previous paragraphs. Answering the second question, the model is needed for all Canadian at high risk of developing HNC, especially HPV-related oropharyngeal cancer. Considering the third question, different studies have explored HNC and its risk factors in Canada[11,63,115,116,127,128], providing a rich data bank. Among these studies, the HeNCe Life study specifically investigated the risk factors of HNC and captured valuable solid data on the lifetime exposure to the different risk factors of HNC, containing 460 cases (including 219 cases of oropharyngeal cancer) and 458 controls[63,115,116,129]. Therefore, its dataset can be used to develop or validate the model. To answer the fourth question, understanding the current status in HNC risk prediction modeling and identifying the models potentially generalizable to the Canadian context is necessary. Therefore, before planning to develop any new model for Canada, we need to review the literature in HNC risk prediction modeling and explore the potentially generalizable models' performance and applicability in the Canadian context. This thesis project was designed to address the above gaps in knowledge.

4 Aims and Objectives

The main goal of this thesis project is to identify and validate HNC risk prediction models that are potentially applicable in the Canadian context. To achieve this goal, this thesis project aims to:

1- Conduct a scoping review to systematically map the literature on HNC prognostic risk prediction modeling, assess the existing models' performance, risk of bias, and applicability in practice. Such a review will help in identifying the models potentially generalizable to the Canadian context.

2- Externally validate an identified oropharyngeal cancer risk prediction model, developed and validated on a sample of the USA population, using data from a Canadian case-control study in HNC.

Manuscript I entitled "What is the current status of prognostic risk prediction modeling for head and neck cancer? - A scoping review" addresses the first aim of this thesis by systematically reviewing the published papers on the HNC risk prediction modeling. The specific objectives of this manuscript are to identify: 1) types of study designs and data sources used to develop HNC risk prediction models, 2) types of statistical and machine learning models developed to predict the individualized HNC risk, and 3) modeling strategies implemented to develop and validate the HNC risk prediction models.

The results of the scoping review (*Manuscript I*) provide valuable information on the current models' key features, performance, risk of bias, replicability, and generalizability. Importantly, it identifies a model for HNC specifically for oropharyngeal cancer that can potentially be validated in the Canadian context, which is an essential information to fulfill the second objective of my thesis. The models has been developed and validated by Tota et al. (2019) on a sample of US population[17].

Manuscript II titled "Validating an oropharyngeal cancer risk prediction model on a Canadian population" aims to externally validate this model using data derived from a hospital-based case-control study of HNC in a sample of Canadian population.

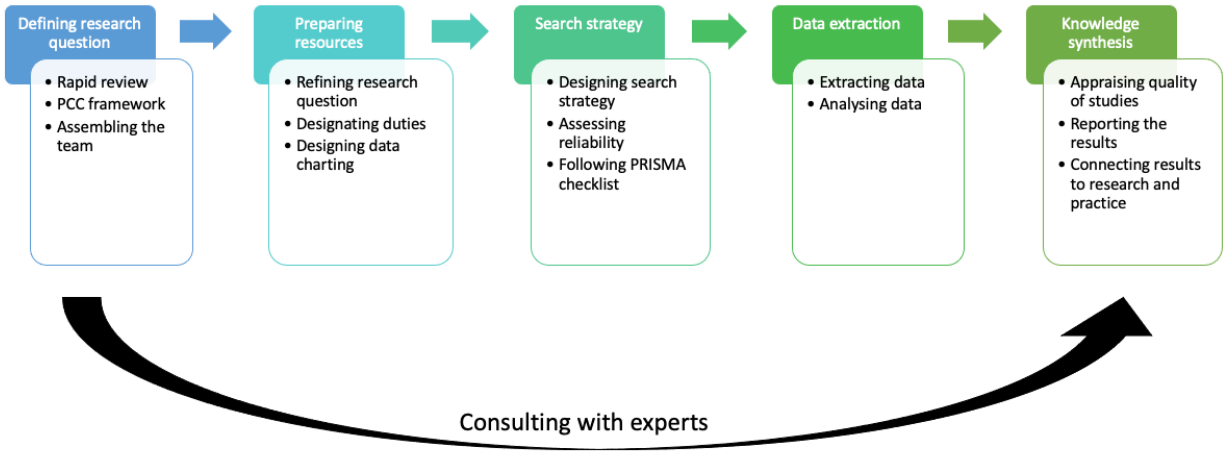
5 Methods

To facilitate the reader's understanding, I describe the methods used in this project according to each study's objectives.

5.1 Objective 1: Scoping review of the literature on HNC risk prediction modeling

To address the first objective of my thesis, I conducted a team-based scoping review on HNC risk prediction modeling (*Manuscript I*). The review project systematically maps the existing developed or validated HNC risk prediction models and evaluates their strengths and limitations. I used the Arksey-Omaley's methodology[130] that provides a five-stage framework for conducting a team-based scoping review. Figure 5-1 demonstrates a graphical representation of this methodology published by Westphaln et al. (2021)[131].

Figure 5-1: Graphical representation of Arksey-Omaley's methodology for scoping reviews adapted from Westphaln et al. (2021)



Manuscript I provides in detail the methods and approaches followed at each stage of the scoping review. Below, I give an overview of this study describing the sections that were less highlighted in this manuscript.

At the first stage, a rapid literature review was undertaken to recognize the important aspects of the research question and determine the framework that could be used to answer the review question, "What is the current status in prognostic HNC risk prediction modeling?"

At the second stage, a librarian developed a systematic search strategy to identify peer-reviewed published articles on the HNC risk prediction modeling[132]. The search strategy can be found in the chapter six of this thesis. We searched several databases including Medline (Ovid), for Embase (via Ovid)[133], CAB Abstracts[134], Scopus,[135] and Web of Science[136] from inception to June 2021 without language restriction. We also scanned the reference lists of included articles to determine if any other useful publication had been missed by the search.

During the third stage, two blinded reviewers selected the relevant articles using the Rayyan web app[137], which facilitates the blinded shortlisting process. Once the articles were selected, the agreement between the reviewers was checked and any conflict between them were resolved by integrating the expert consultation. Subsequently, the interrater reliability was calculated based on the abstract selections of the reviewers using the Cohen's kappa coefficient (κ)[138], which uses the formula below:

Equation 5-1: $\kappa = \frac{\Pr(OA)-\Pr(CA)}{1-\Pr(CA)}$

where $\Pr(OA)$ is the observed agreement and $\Pr(CA)$ represents the chance agreement. $\Pr(OA)$ and $\Pr(CA)$ are calculated based on the number of papers the raters (reviewers) included and excluded:

		Reviewer A	
		Included papers	Excluded papers
Reviewer B	Included papers	a	b
	Excluded papers	c	d

Equation 5-2: $\Pr(OA) = \frac{a+d}{N}$

&

Equation 5-3: $\Pr(CA) = \left(\frac{a+c}{N} \times \frac{a+b}{N}\right) + \left(\frac{b+d}{N} \times \frac{c+d}{N}\right)$

,

Equation 5-4: $N = a + b + c + d$

The value of this coefficient is interpreted as no agreement $\kappa \leq 0$, none to slight agreement $0.01 \leq \kappa \leq 0.20$, fair agreement $0.21 \leq \kappa \leq 0.40$, moderate agreement $0.41 \leq \kappa \leq 0.60$, substantial agreement $0.61 \leq \kappa \leq 0.80$, and almost perfect agreement $0.81 \leq \kappa \leq 1$[139]. The agreement in this review was 0.75, indicating substantial interrater agreement.

At stage 4, detailed data of the included studies were extracted and analyzed based on the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist. The details of data extraction process and tables are provided in the *Manuscript I* and appendix I of this thesis. The quality of the included studies and their models was appraised using the PROBAST (Prediction model Risk Of Bias ASsessment Tool), which facilitates the assessment of models' risk-of-bias and applicability in the new settings.

At the final stage, the extracted data and results were synthesized and reported according to the PRISMA checklist, and the replicable models that can be validated on the Canadian context were identified.

5.2 Objective 2: Externally validating an identified oropharyngeal cancer risk prediction model

The scoping review identified one model[17] potentially applicable in the Canadian context. In the *Manuscript II*, I tested the performance and applicability of this models through the external validation on the data from a sample of Canadian population.

Details about the model validation study is provided in the *Manuscript II* chapter of this thesis, however, the additional information on the models' preparation are presented here.

5.2.1 Models' development study

The model has been developed by Tota et al. (2019)[17] using data derived from a population representative case-control study created by weighted oversampling of data from 241 new cases of oropharyngeal cancer at the Ohio State University Comprehensive Cancer Center from 2011 to 2015 and 9327 noninstitutionalized controls aged between 30 to 69 years from the National Health and Nutrition Examination Survey (NHANES) from 2009 to 2014[17,52,140,141]. The model has been developed using multivariable logistic regression to predict a single outcome defined as histopathological diagnosed malignancy in the base of the tongue, soft palate, palatine tonsils, and posterior pharyngeal wall.

The predictors comprise age, sex, race, pack-years of smoking, last year's alcohol consumption, number of lifetime sex partners, HPV infection status, the interaction of sex and HPV, and the interaction of smoking and HPV.

Sex, HPV infection, and alcohol (≤ 14 & > 14 drinks per week) were considered as binary variables while age (years) and smoking (pack-years of smoking) were kept as continuous variables. Race (black, white, other) and lifetime number of sex partners (0-1, 2-5, 6-10, >10) were considered as categorical variables. HPV infection was categorized into two levels: 1) HPV negative, 2) Positive for HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59.

During the validation study, the estimates of this model was validated on the data derived from HeNCe Life study. Details about this study have been published elsewhere[63,115,116,129,142]. Nonetheless, I provide below a summary of this study and detail the methods I followed to prepare the dataset for the validation study.

5.2.2 HeNCe Life study

HeNCe Life study is an international hospital-based case-control study conducted in three countries of Brazil (from 2003 to 2005), Canada (from 2005 to 2013), and India (from 2008 to 2012) investigating the etiology of HNC through a life course approach. The Canadian part of this study recruited 458 HNC cases and 460 controls from the four major referral health care centers, with in-house facilities for the histological diagnosis of HNC, in Montreal, Quebec. These four health centers were Montreal Jewish General Hospital, Montreal General Hospital, Notre-Dame Hospital, and Royal Victoria Hospital. HNC cases were consecutive newly diagnosed patients with the squamous cell carcinoma (SCC) of the oral cavity, oropharynx, hypopharynx, and larynx. A total of 214 cases of histopathological confirmed oropharyngeal SCCs, stages I to IV were included in this study. These cancers comprise SCCs at the base of tongue, soft palate, palatine tonsils, oropharynx, and uvula related to the ICD-10 codes of C01, C02.4, C05.01, C05.2, C09, C10, C12 and, C14[143]. A blinded expert histopathologist confirmed the cancers based on the standard definition for the SCC issued by the National Cancer Institute[144]. Those participants who were undergoing cancer treatment were excluded from the study to eliminate the effects of treatments with the biological markers examined (e.g., HPV infection). Furthermore, prevalent cases of HNC were excluded in order to avoid survivor bias, temporal ambiguity, and reverse causality[145,146]. Non cancer controls frequency matched to cases according to sex and 5-year age categories were randomly selected within the maximum of three months of recruiting the case participants from outpatient clinics (e.g., Orthopedics, Nephrology, Gynecology, Ophthalmology, Dentistry, Ear, nose, and throat (ENT), Neurology, Dermatology) at the same hospitals as the cases were recruited from. The HeNCe Life study collected an array of life course exposures using a questionnaire and the life grid technique, which is suggested to reduce the recall bias[147]. Data collected included participants' demographic characteristics, life course indicators of socioeconomic position, and behavioral factors (e.g., lifetime tobacco smoking and alcohol consumption). This study also collected biological oral transepithelial samples from the participants for genetic and HPV analysis.

5.2.3 Variable preparation

5.2.3.1 Tobacco smoking

HeNCe Life collected detailed information on the number, type, and intensity of the tobacco consumption of all participants who reported using tobacco for at least one year during their lifetime. The questionnaire (Appendix III) was structured in such a way that can capture periods of change in the habits throughout the individual's life course. Data collected included duration (using information on the age of initiation and cessation) and consumption (how many cigarettes per day or per week, or per month). The brand used and the type of cigarette (filtered or non-filtered) were also recorded. Based on the different commercial types of tobacco consumption, we standardized the smoking measurement by converting it to the standard pack of cigarette smoking, using below formula[148]:

Equation 5-5: $1\ pack\ of\ standardized\ cigarettes = 20\ commercial\ cigarettes\ (filtered\ or\ non-filtered) = 4\ hand-rolled\ cigarettes\ = 4\ cigars\ =\ 5\ pipes$

Using this information, a variable representing the lifetime intensity of tobacco was created. This cumulative exposure variable, called 'pack years', was calculated using the formula below:

Equation 5-6: $Pack-Years\ of\ smoking = \sum_1^i Smoking\ frequency_i \times Smoking\ duration_i$

where i is the number of same frequency periods during a participant 's smoking history.

Translating this formula, it means 1 pack-year is equivalent to smoking 1 pack per day for 1 year, or 2 packs per day for half a year. Pack-years of smoking was coded as a continuous variable.

5.2.3.2 Alcohol consumption

Similar to the tobacco consumption, HeNCe Life collected valuable lifetime exposure to the ethanol. Therefore, we could calculate the one-year exposure to the ethanol by using the ethanol concentration of every alcoholic beverage consumed during the last year of the recruitment. The

concentration of ethanol was derived for every alcoholic beverage using the standardized glass of alcoholic drink[149]. Accordingly, beer was considered as containing 5% ethanol, hard liquor 50%, and toddy, aperitif, wine, and other types each 10% of ethanol. The total amount of daily ethanol consumption in milliliters (ml) during the one year ending to the recruitment in HeNCe Canada was calculated. The milliliters of ethanol exposure were summed up and converted to the number of drinks by dividing the results into 17.05 based on the standard amount of alcohol per drink in Canada[150]. Laprise et al. (2019)[115], provides a full description of how this variable was created. The alcohol exposure variable was categorized by No drink, less or equal to 14 drinks per week, and more than 14 drinks per week based on the models' development study[17].

5.2.3.3 Other variables

HeNCe Life study provided a lifetime information on the participants' number of sex partners, making it possible to prepare the variable according to the development study. Similar to the models' development study, the number of sex partners variable was categorized by cut points of 1, 5, and 10 partners. HPV infection status also was categorized into three levels of HPV negative, HPV-16 positive, and positive for other high-risk HPVs infection.

5.2.4 Missing data

Considering the solid and enriched dataset the HeNCe Life study provided, there was no missing value regarding age, sex, race, smoking and alcohol habits, and number of sex partners variables. However, since some participants of HeNCe Canada did not consent to provide the saliva samples, 13.2 % of the data related to the HPV status was missing. We did not impute these missing values in *Manuscript II* to have an overview on the available case analysis.

5.2.5 Sample size estimation

We calculated the minimum sample size needed for an acceptable external validation of a risk prediction model with a binary outcome based on the closed-form sample size calculations as below[151]:

Equation 5-7: $N = \frac{(1-\varphi)}{\varphi(SE(\ln(\frac{O}{E}))^2}$

Where N was the sample size, φ represented the proportion of the cases in the filtered HeNCe Canada dataset, O/E represented the estimated ratio of the total number of observed cases by the total number of expected (predicted by model) cases, and SE was the standard error of the $\ln(\frac{O}{E})$. Ideally, O/E is 1. We calculated φ as 0.478 and aimed for a 95% confidence interval width of 0.2 for O/E to have a good calibration-in-the-large.

5.2.6 Model performance assessment

Models' discrimination power was assessed by calculating C-Statistics and brier. We also assessed the models' calibration by calculating the calibration-in-the-large (the mean absolute difference in observed and predicted probabilities or O/E), calibration slope, and visual assessment of the calibration plot. *Manuscript II* provides more details on models' performance assessment.

Preface to Manuscript I

According to the literature review chapter of this thesis, risk prediction models play significant roles in decision-making in different settings. For example, an HNC risk prediction model may assist a clinician in providing motivation counseling for a high-risk individual or a researcher in calculating the minimum sample size for trials of new HPV preventive measures. Different risk prediction models have been developed for HNC and its subsites worldwide. Nonetheless, little is known about these models' type, quality, risk of bias, and clinical applicability. The *Manuscript I* will fill this knowledge gap by reviewing the papers publishing these models' characteristics. Based on McGill University's guidelines for a manuscript-based thesis, the next chapter stands alone as *Manuscript I* with an independent reference list and appendices.

6 Manuscript I

What is the current status of prognostic risk prediction modeling for head and neck cancer? A scoping review

Hamed Ghanati[1], Mohammad Al-Tamimi [1], Ziad Al-Asmar [1], Martin Morris, [2] Sreenath Madathil [1], Belinda Nicolau [1*]

[1] Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada

[2]Schulich Library of Physical Sciences, Life Sciences and Engineering, McGill University, 809 rue Sherbrooke O, Montréal, QC, H3A 0C1, Canada

*Corresponding author:

2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada

Key words:

Prognosis; Models, Statistical; Head and Neck Neoplasms; Review [Publication Type]

Word count:

6421

Abstract

Background: Annually, over 700,000 individuals are being diagnosed with head and neck cancer (HNC) worldwide. Identifying high-risk individuals is an important factor in preventive measures and managing HNC. Prognostic risk prediction models may estimate the individualized probability of developing HNC in the future, providing valuable information to be used in managing the disease and leading personalized preventive interventions. Different HNC risk prediction models have been developed worldwide; however, the quality and clinical applicability of these models have not yet been synthesized. We aimed to review the literature on HNC prognostic risk prediction modeling and assess their quality while providing suggestions for future studies.

Method: Following the PRISMA-ScR, we conducted a systematic scoping search on MedLine (Ovid) and included the full-text-available peer-reviewed published papers on developing or validating at least one prognostic HNC risk prediction model. Data were extracted according to the TRIPOD checklist. The studies' quality was appraised using the PROBAST tool.

Results: Nine papers were included, which collectively reported 16 models for HNC and its subsites. Most of them were conducted in Asian countries. While oral cancer was the most frequent outcome, sex, age, smoking, alcohol, and education were the most frequent predictors. Most models were developed using multivariable logistic regression analysis. Although all included studies had a high risk of bias, mainly in the analysis domain, only two studies had high concerns in applicability.

Conclusion: Currently published HNC prognostic risk prediction modeling studies provide insufficient information about the model development and validation, making it difficult to judge their quality and applicability. We suggest that future investigations follow the standards in reporting the prediction modeling studies. This study also recognized a need for an HNC risk prediction model developed and validated for Canada, Australia, Brazil, and many European, Latin American, and African countries.

Introduction

Every year, more than 700,000 new cases of cancers of the lips and oral cavity, oropharynx, hypopharynx, and larynx are diagnosed around the world[1]. Approximately, 400,000 deaths from these cancers, also collectively known as Head and Neck Cancers (HNC), occur annually. In Canada, an estimated 2,100 out of 7,400 HNC patients died from HNC in 2021[2]. HNC have a 5-year survival rate of around 50%[3], which has remained stable for many decades[4]. These cancers have one of the highest morbidity rates of all cancers because of where they are located on the body, their treatments' direct or indirect impacts on physical functioning (e.g., the impact of surgical resection of mandible on speech and swallowing), and their possible psychosocial impacts (e.g., depression, anxiety).[5] Past studies have consistently shown that tobacco smoking and alcohol consumption and human papilloma virus (HPV) are the main risk factors of these diseases[6,7]. Despite this knowledge, the incidence of HNC has remained relatively stable[8-10]. Importantly, the incidence of a subset of these cancers related to HPV is increasing in several countries[11,12], specifically in the developed countries[13] including Canada[8,14]. There is, therefore, a need to devise prevention strategies to reduce HNC incidence.[15]

The use of risk prediction models in medical decision-making has become increasingly popular[16-18]. These models estimate the probability of having a disease (diagnostic prediction model) or future occurrence of a disease (prognostic prediction model) in a person based on a set of the individual's sociodemographic and behavioral characteristics[16,19]. The predictions of these models may assist health care professionals and individuals themselves to evaluate the risk of developing a disease (in this case HNC) based on the risk factor profile[20]. This may then lead to personalized prevention intervention strategies[21]. Prediction models are also useful in identifying high-risk individuals for screening programs[22] or clinical trials for new prevention measures[23]. Successfully identifying individuals at high-risk of lung cancer for screening programs is a good example of how these models may be applied[22,24,25]. We can, therefore, expect that HNC prognostic risk prediction models may help to identify high-risk individuals for screening programs[26] and/or clinical trials (e.g., trials to prevent oral HPV infection[27]).

Although several prognostic HNC risk prediction models have been developed[26,28-35], limited evidence exists on the quality of them, putting the applicability of these models in clinical setting, public health, or clinical trials into question. Investigating the strength and limitations of the current models proposed in the literature can help to identify the best developed and performed models. More importantly, such an investigation will provide valuable information for developing future prediction models. To the best of our knowledge, no publication exists that synthesis the knowledge on prognostic HNC risk prediction modeling. Given this, we aimed to systematically map the literature on HNC prognostic risk prediction modeling and investigate the existing models' performance, risk of bias, and applicability in practice.

Method

We followed the five-stage methodology proposed by Arksey and O'Malley for scoping reviews[36,37]. After a quick review of literature on prognostic HNC risk prediction modeling and consulting with the experts in HNC epidemiology, our team reached a consensus on the research question of: "What is the current status in prognostic HNC risk[c] prediction modeling?"
This question could be refined as three detailed questions:

- What types of study designs and data sources are used to develop prognostic HNC risk prediction models?
- What types of statistical and machine learning models are used to predict the individual risk of developing HNC?
- What modeling strategies are used to develop and validate prognostic HNC risk prediction models?

We used the Population, Concepts, and Context (PCC) framework to define our research question[37]. According to this framework, our population of interest was any type of prognostic model developed to predict the individual risk of developing HNC. Our concept was the model

[c] Risk: individual risk of developing head and neck cancer based on demographic and behavioral factors (e.g., age, sex, alcohol or tobacco consumption).

development and/or validation strategy and the model performance metrics. The context comprised all studies reporting the development or validation of at least one prognostic model to predict an individual's risk of developing HNC.

Information source and literature search

A medical librarian trained in literature searching for knowledge syntheses (MM) created a systematic scoping search strategy[38] for Medline (Ovid), which is provided in the Appendix. The strategy comprised a combination of Medical Subject Headings, title/abstract key words, truncations, adjacency operators and Boolean operators, and included the concepts of head and neck cancers, epidemiology, and computer modelling. The strategy was subsequently translated for Embase (via Ovid)[39], CAB Abstracts[40], Scopus,[41] and Web of Science[42]. All databases were searched from inception to 18 June 2021, and the combined library was deduplicated in Endnote 20[43]. In addition, we scanned the reference lists of included articles to determine if any other useful publications had been missed by the search.

Study screening and selection process

Following the deduplication on EndNote 20 Desktop Version[43], we uploaded a copy of the EndNote library on the Rayyan web app for systematic reviews[44]. Two blinded reviewers (HG & ZA) read the title and abstracts and shortlisted the papers based on the inclusion and exclusion criteria. The inclusion criteria include the full-text-available peer-reviewed published papers on developing or validating at least one prognostic model to predict the risk of developing HNC. Initially, we included and read the review papers to understand the current status, get more knowledge on the context, and check their references to see if any paper is missed from our main search. Subsequently we excluded the review papers and any other article that did not meet inclusion criteria. We also excluded the articles related to the models that contain genetic predictors (e.g., DNA methylated genes as predictor) because we aimed to investigate the models that are easy to be implemented in the clinical settings.

The conflicts between the two reviewers were resolved during a discussion with two experts (SM & BN). We calculated the Cohen's Kappa coefficient[45] to assess the inter-reviewer agreement.

Data extraction process

Following the shortlisting process, we developed a data extraction table according to the TRIPOD criteria[46]. The extracted information comprised study objective(s), source of data, type of study used to derive the model, number of participants, outcome, and predictors used in the final model, sample size, amount and management of missing data, model type (statistical/machine learning), modelling strategy, internal and external validation techniques used, performance metrics, and models' limitations, interpretations, and implications.

Methodological quality appraisal

Although to appraise methodological quality or risk of bias is not a necessary step in a scoping review, it helps to identify gaps in the literature related to low quality of research. To assess the quality of included studies, we used the PROBAST tool[47]. This tool is designed to appraise the quality of modeling studies in two ways: Assessment of risk of bias and assessment of concerns regarding applicability

It comprises 20 signaling questions across 4 domains: participants, predictors, outcome (s), and data analysis. Signaling questions are answered as yes, probably yes, no, probably no, or no information.

Using PROBAST, a structured way and detail guidelines to identify potential risk of bias (ROB) in prediction modeling studies, we judged each domain according to the answer for each question. "Yes" and "No" responses indicate low and high ROB, respectively. If the information is not sufficient to confidently judge, the ratings "probably yes" and "probably no" are included. While Yes and No answers are intended to have similar implications to responses "probably yes," and "probably no", respectively, they have a subtle distinction in which yes/no is something we know and "probably yes"/"probably no" is something that is likely to be the case. In our review, any

question answered with "No" or "probably no," were reported the overall assessment of that section as "High ROB." If at least one question was answered as "No information provided," the overall assessment for that domain was judged as "Unclear." Assessing concerns regarding applicability, we used information in three domains about participants, predictors, and outcome(s) of the studies. We did not include data analysis as this domain relates to how the data analysis was carried out and limitations with the data both of which are not related applicability assessment. If any of these items failed to meet our research question's requirements, we marked it as a high concern of applicability. For full details on this toll please refer to the published resources for PROBAST[47,48].

We followed the updated methodological guidance of the PRISMA-ScR[49] to conduct and report our investigations.

Results

Literature search

Our search strategy gathered 1554 articles, of which 192 were duplicates (Figure 6-1). 15 papers met the inclusion criteria, of which five were review articles. None of these reviews comprised additional information or references; thus, we excluded all of them. One of the papers was withdrawn from the publication, and we could not find its full text. After a discussion with a librarian at McGill University, we excluded it. Therefore, nine articles were retained for this review[26,28-35]. The Cohen's Kappa coefficient for the shortlisting process between the two reviewers was 75.34%, which shows a good inter-reviewer agreement.

General characteristics of the models

Table 1 displays the general characteristics of the included studies. Four papers were published in the last three years[26,29,31,33]. Among the remaining five studies, four were conducted between 2016 to 2019[28,30,32,34] and one was published in 2010[35]. Data used for modeling was derived from a population-based cohort[29], case control designs [28,30-35], and randomized controlled screening trial[26]. Regarding geographical diversity, most studies (60.0%) used data from Asian countries

(India[26,28,30], Malaysia[35], Japan[34], and Taiwan[29]). The remaining three articles were from the USA [31,32] and UK[33,]. The sample size for the model development ranged from 255 to 1,836,888. The number of cases also ranged between 84 to 117,697.

Overall, 15 models were developed from the nine included studies. The number of developed models is more than the number of included articles because some of them developed more than one model[31,34,35]. Oral cancer was the outcome of six models[26,29-31,35]. Three models were developed for HNC[31,33,34], two models were specifically developed for oropharynx[31,32], two models for upper aerodigestive tract cancer (the combination of HNC and esophageal cancer)[28,34], one for hypopharynx[31], and one for larynx[31]. One of the articles reported development of a separate model for esophagus cancers[34], but we did not consider it in our model performance assessments because this outcome was not part of our inclusion criteria. The analysis for model development in the reviewed papers comprised Fuzzy regression and Fuzzy Neural Network (one study)[35], Cox Proportional Hazard regression (one study)[26], and Multivariable Logistic Regression (seven studies)[28-34].

Analytical characteristics of the models

Table 2 summarizes the model development and assessment techniques in each study. Seven studies[26,28,30-34] (77.8%) reported missing values and two studies[29,35] (22.2%) did not provide this information. Only three studies[26,32,34] (33.3%) used imputation techniques to manage missing values. One study[30] excluded the participants with missing values from the analysis, and one study[31] resolved the data inconsistency by communication with the source dataset investigators. Three studies[32-34] (33.3%) externally validated the developed model. All studies reported Area Under the Receiver Operating Characteristic Curve (AUC) score to report the model discrimination performance.

Table 3 summarizes the model type, outcome, and the discrimination performance score reported for each developed and tested model. AUC of the models during internal validation ranged between 0.69 and 0.96, and during external validation ranged between 0.73 and 0.91.

The best-performed model is a multiple logistic regression developed by Gupta B. et al.[28] in 2017, with an overall AUC=95.8 (95% CI of [93.6–97.4]), positive predictive value = 74.8% and negative predictive value = 96.6% after internally validating by bootstrapping with 1000 replications. For calibration, two studies[30,34] (22.2%) reported Hosmer-Limeshow goodness-of-fit, two[26,32] (22.2%) reported calibration score only (observed/expected ratio), and two[31,33] (22.2%) demonstrated calibration plots additional to the calibration score reporting. Only one study[34] reported calibration plot and reported calibration score besides Hosmer-Limeshow goodness-of-fit test results. Three articles[28,29,35] (33.3%) did not report calibration measurements. For the internal validation, splitting, cross validation, and bootstrapping were used in three[31,32,35], one[26], and two[28,30] studies, respectively. Three studies[29,33,34] (33.3%) did not provide information about the internal validation.

Table 4 displays the predictors of the models. These predictors comprised age, sex, race, ethnicity, smoking, tobacco chewing, Mishri consumption, alcohol, HPV infection, ALDH2 genotype, education, BMI, socioeconomic factors, diet, medical history, family history of HNC, marital status, lesion or swelling, sexual behavior, and other behavioral factors such as rinsing mouth with water after eating or smoking. The most frequently used predictors were sex (91%), age (88.9%), tobacco smoking (77.8%), alcohol consumption (66.7%), tobacco chewing (44.4%), and education (44.4%). Only one study[32] considered HPV as a predictor. One[28] of the studies used lifetime consumption of alcohol and tobacco smoking.

The results of quality appraisal are summarized in tables 3-5. Overall, the studies had a high ROB, which was mainly driven by the "analysis" domain. Three[28,33,35] studies had a high ROB in the "participants" domain. All studies except one[29] had a high ROB in "predictors" domain. One study[35] did not provide enough information to judge ROB in the "predictors" domain. While the majority of studies had low ROB in the "outcome" domain, two studies[29,35] did not report sufficient information to judge this domain. Figure 6-3 displays the assessment of Concerns about Applicability (CAA). Two[26,29] studies' models had high overall CAA. One[29] of them had high CAA in the "Participants" domain, and the other one[26] had high CAA in the two domains of

"Predictors" and "Outcome". Figures 6-2 and 6-3 explicitly demonstrate the quality appraisal results using the template provided by the PPROBAST[48].

Discussion

This study reviewed nine papers that reported at least one prognostic risk prediction model for HNC. Similar to other health outcomes, prognostic risk prediction modeling for HNC is a relatively new topic with most of the papers published within the past six years[26,28-34]. All studies had a high ROB in the overall assessment, which was predominantly driven by analytical issues including improper managing or lack of reports on missing values and calibration of the models. According to the PROBAST, 7 [28,30-35] out of eleven studies had low concerns related their applicability. However, only 3 studies[32-34] externally validated their models, thereby assessing their clinical applicability.

Geographical diversity

The studies included in the review were from several countries representing no specific geographical pattern. Two studies using data from the US population[31,32] developed five models to predict overall HNC risk and according to its subsites. Considering a recent sharp rise in HPV-related HNC[50], especially the oropharyngeal cancer, incidence in the US, these models can help in preventing oropharyngeal cancer in this country.

Koyanagi et al.[34] developed 3 models for predicting the risk of cancer of the oropharynx, and esophagus and HNC overall in a Japanese population. Similarly, three studies[26,28,30] used data from India to develop the models, two[26,30] of which were for oral cancer and one for HNC [28]. While these studies are useful to predict HNC in their specific population, they cannot be used in the whole country; India population is highly diverse[51], and thus, the baseline risk should be assessed and adjusted before implementing the models on a different population in that country.

Even though HNC is prevalent in European and Latin American countries[52], we identified only one[33] study from these regions, which was from the UK. We also have no prediction model for

the Canadian and Australia populations, despite 7,400 and 5,104 new cases of HNC were diagnosed in 2021[2 53] in these countries, respectively.

Participants and source of data

The source of data is an essential factor in risk prediction modeling. The optimum dataset comes from longitudinal investigations specifically designed and conducted for the modeling study[48]. However, these studies are expensive to run and the routine practice is, therefore, to use data from the existing cohorts or registries or conduct a case-control study to capture more number of events[54].

However, using secondary data poses major challenges including data consistency. Often variables are not measured and recorded consistently for all participants, which is essential in modeling studies[48]. Therefore, investigators must perform data quality check before modeling. Also, the modelling studies using data from case-control designs require adjustment of outcome frequency in the source population to avoid the risk of biased estimations[48]. These two challenges need to be considered in a modeling study to avoid risk of biased estimations. Three[28,33,35] studies failed to address these challenges among the seven[28,30-35] studies that used pre-existing data or case-control studies. We, thus, classified them as high ROB in the "Participants" domain.

Predictors

Most of studies in this review included age, sex, tobacco and alcohol consumption, HPV infection, socioeconomic position, and dietary habits as the predictors of their models. Additionally, some models used area-specific risk factor such as Mishri[28], bidi smoking[28] or betel chewing[29,30] consumption. While it is important to include area specific behavioral predictors, their inclusion affect the model's applicability. For example, the model including Mishri consumption[28] is not applicable to other populations who do not consume Mishri.

HPV infection, a major risk factor for a subset of HNC especially in Western countries[55] must be included in HNC prognostic models. However, to detect HPV infection requires laboratory tests

that are not always available. One way to overcome this challenge is to estimate HPV infection using behavioral factors such as sex behavior[32]. While this is not a precise estimation of HPV infection, it can be used as a predictor because prognostic models are complementary tools in the primary care settings.

Another concern in the risk prediction modeling studies relates to the assessment of predictors. Ideally, the assessment of predictors should be blinded to the outcome status of a participant[48]. In other words, having prior knowledge on a person's HNC status results in biased estimation measurement of predictors. This issue is even more critical in modeling studies using data from case control studies because the data collectors are aware of the outcome status of patients. Seven[28,33,35] studies derived data from case-control studies, and thus, were assigned high ROB in the "predictors" domain.

We recognized a high ROB and high CAA for the model developed by Cheung et al.[26] as one of its predictors was not replicable. The authors used a cluster-randomized controlled screening trial to create the dataset and included the "Screening arm" as a predictor in the final model. Obviously, the screening arm is irreplicable, and therefore, the model is not applicable to other settings.

Outcome

The definition of the outcome is an important factor in a modeling study. Suboptimal methods to ascertain the outcome may lead to misclassification, resulting in biased performance measurement and high ROB in this domain. The diagnosis of HNC, the outcome measure in our review, requires histopathological tests performed by an expert. In our review, two[29,35] studies lacked detailed information on the assessment of the outcome and thus we ranked them as a high ROB in the "Outcome" domain.

Analysis

As previously stated, the majority of the studies included in this review had issues in their analytical strategies. The following guidelines have been suggested to ensure a low ROB in the "analysis" domain[16,48]:

First, the dataset used for modeling must have enough participants to ensure enough number of events per variable (EPV). One study[35] in our review with a sample size of 84 case participants used the Fuzzy Neural Network technique to develop the model. We assessed high ROB in the "Analysis" domain of this study because machine learning-based models require at least 200 EPV to avoid overfitting[16]. In addition to leading to an overfitted model, results from studies with small sample sizes are not applicable in clinical settings as they lack enough generalizability.

Second, coding continuous variables as dichotomized or categorized variables in the risk prediction model causes information loss[16,48]. However, sometimes categorization is done to improve the model's clinical interpretability[16]. In this case, the categorization must be done based on the predefined and widely accepted cut points to avoid a high risk of bias in the model's predictions[48]. "Age" variable is the most common continuous variable in medicine that is sometimes categorized into five groups of children, youth, adult, old, or into five-year age groups (0-5, 6-10, 11-15, etc.)[56]. Only three studies[26,32,33] considered "Age" as a continuous variable. Others, categorized[29,31,34] or dichotomized[35] it. Two studies[28,30] did not include the "Age" variable in their models.

Managing the missing values is another essential factor in the prognostic risk prediction modeling. Improper missing data management results in biased model performance and biased estimation. Only four studies[26,31,32,34] appropriately managed the missing data (multiple imputations). Others excluded[30] participants or did not provide information in this regard[28,29,33,35], thus we assessed them as high ROB.

Moreover, the best approach for predictor selection is using nonstatistical methods. In this approach, the predictors are selected based on their importance according to the literature and clinical applicability. Some statistical methods such as univariable analysis or using Bayesian Information Criterion and Akaike Information Criterion in a multivariable analysis also help selecting predictors. However, if a study uses analytical methods to select the predictors, the developed model should be assessed and tested for optimism by internal validation because the analytical methods for predictor selection will make the model overfitted to the source dataset endangering the model's applicability and clinical usefulness—question 4.5 of PROBAST specifically addresses this issue in the prediction modeling studies. Cheung et al.,[26] used Akaike information Criterion to find the linear relationship between the outcome and possible predictors. However, the authors conducted five-fold cross validation to check for the model's optimism. Therefore, we classified this model as low ROB. One study[33] used univariable analysis to select the model predictors. No information about the model optimism check was provided. Therefore, we recognized a high ROB in the analysis domain for this study.

The predictions of a prognostic risk prediction model should represent the actual risk of developing the outcome in the target population. The problem arises when a case-control design is used to increase EPV, where the model's estimations are not real due to the difference in the fraction of case and controls between the study participants and the target population of interest. One way to address this is to weight the control participants by the inverse of their sampling fraction (adjustment for sampling fraction). Otherwise, the risk estimations will not represent the absolute outcome probabilities. Three[31,32,34] out of the seven[28,30-35] case-control studies reported adjustment for sampling fraction. The studies failed to consider this adjustment were classified as high ROB in our review.

Modeling studies must also use appropriate measurements to assess the performance of a predictive model. Accurate predictions discriminate between individuals with and without the outcome of interest. Concordance (C)-Statistics, a common performance measure used to indicate the discriminative ability of models, evaluates the relationship between true (sensitivity)

and false (specificity) positive rate. In a case of a binary outcome, C-Statistics is the same to the area under the receiver operating characteristic (UAC-ROC) curve, which plots true positive rate against false positive rate at a variety of thresholds for the probability of an outcome. However, this is not solely enough to judge about a model's performance. We still need to understand how well the model is distinguishing between presence or absence of outcome. This can only be achieved by measuring the distance between the predicted outcome and actual outcome. The distances are related to the concept of 'goodness-of-fit' of a model, with smaller distances between predicted and observed outcomes determining the best models. The approaches used to quantify how far the predictions are from the actual outcome, includes measures such as explained variation (e.g., R^2) and the Brier score. We can further estimate model performance using the Hosmer-Lemeshow "goodness-of-fit" test.

Finally, there are other performance measurements that can be used when implementing the model (e.g., external validation), such as checking the positive predictive value, negative predictive value, and accuracy. In our review, all studies reported AUC, but none reported brier score or any other measure of 'goodness-of-fit'. Also, none of the studies[32-34] that externally validated the model reported positive and negative predictive values as well as accuracy.

Moreover, calibration needs to be reported in predicting model studies to understand the rate of the observed outcome over the expected outcome, that is, the agreement between predicted probability and observed risk. The calibration can be reported as a score and a plot, but the latter is crucial for understanding the model's performance when the sample size is relatively small. Furthermore, calibration needs to be appropriate to the type of outcome. If the outcome is predicting the time to develop the HNC, the calibration must account for the censoring time. Only Cheung et al.,[26] used the time to event outcome and appropriately considered the censoring time in their calibration. Four articles[28-30,35] did not or inappropriately reported the calibration measurements.

Furthermore, the developed models are prone to overfit to the original dataset and overestimate the risk. Therefore, the fitted models must be internally validated on their source dataset to avoid optimism. The routine internal validation methods are splitting, cross-validation, and bootstrapping. Splitting is a basic method in which the source dataset is randomly split into two parts. While one part is used for model derivation (fitting and developing the model), the other part is kept for testing the model. This technique has been recognized as an inefficient method as it reduces the sample size used for model derivation[16]. The less sample size for the model derivation, the more imbalanced outcome or predictors distribution, the less reliability of model performance assessment. When a dataset is randomly split into two parts, the distribution of predictors and outcome will affect the model performance. For example, if the split derivation dataset comprised more participants with outcome (cases) than the remaining (test) dataset, the outcome distribution will be different. Thus, the model will overperform during the development process compared to the testing process.

Three studies[31] [32,35] used splitting for internal validation. Even though the sample size of one study[31] seems sufficient to have a good number of EPV, the splitting technique seem not to be the best choice. Cross-validation is an alternative technique in which a dataset is divided into K folds (usually five to ten folds, depending on the number of events per variable). One-fold is kept for testing the model and the remaining folds are used for deriving the model. This process repeats for K times, and the mean performance of the model in all testing folds is considered the model's performance during the internal validation.

Three papers[29,33,34] did not provide information on the internal validation method. One[29] of these used a sufficiently large dataset with 117,697 cases and 1,719,191 controls. These numbers are big enough to have sufficient EPV. According to our quality assessment, this study had a low ROB in terms of internal validation as the derived model is not prone to overfitting due to a good number of EPV, though reporting internal validation would have added to this study's value.

Another issue with the reviewed studies was the lack of information on different components of the final models. The exact model's characteristics and coefficients of each variable included in the final model should be fully reported. This would allow to replicate and to implement or test the model's performance in different settings. Only one study[32] reported all necessary components of the model whereas others did not report[26] or incompletely reported[28-31,33-35] their models' components.

It is also worth noting that the primary goal of a prognostic risk prediction model study is to develop a tool to be used in new settings. To achieve this goal, the developed models should be validated on the different populations. The external validation helps assess the developed model's performance in the new settings. Ignoring this process results in a model with biased estimation and the cannot be used in a new clinical settings because its performance is not clear.

To implement a model in a new setting and in a new population, most of the time, we need to revise the model which can be done by updating the baseline risk (intercept parameter), the coefficients (odds ratio), or both. However, the predictor set of the model must remain intact. Otherwise, the validated model is not the same as the original model.

It is of note that the splitting techniques (which is an internal validation technique) should not be misinterpreted with external validation[16]. When the source study is the same as the testing dataset, the process is called internal validation, not external validation. Indeed, an external validation could be done on the datasets that derived from the same population, but in different times (time that data is driven)[16]. While three studies[32-34] followed standard methods to externally validate their models, one study[31] used splitting technique to perform this procedure. Accordingly, we still cannot consider this as an external validation, thought the authors called this technique "validating the model".

Current status and future pathways

Different types of statistical analyses have been used to develop models for predicting the individual risk of developing HNC. Most studies[26,28-34] (91%) used statistical techniques to develop a model. Standard multivariable logistic regression was the most frequent technique used to develop the model. A standard multivariable logistic regression model is efficient in training and easy to interpret. However, it has a major limitation: the assumption of linearity between exposure and outcome. We rarely have such a linear relationship between dependent and independent variables in the real world. Also, standard multivariable logistic regression models cannot estimate the effects of exposure time on the outcome. Cox proportional hazard models are alternatives for regression models. When the "time to event" variable is achievable from the dataset, Cox proportional hazard models can predict the time of disease incidence. Cheung et al.,[26] used a Cox proportional hazard ratio model to predict the 7-year incidence of oral cancer.

Recently, the onset of big data registry in medicine has made it possible to use Artificial Intelligence (AI) for accurate risk prediction modeling. AI is widely being implemented in predicting different medical conditions (e.g., breast cancer, prostate cancer, diabetes); however, only Rosma et al.,[35] used AI techniques to develop a prognostic model to predict the risk of oral cancers. This study developed two models, one with Fuzzy Neural Network modeling method and another with Fuzzy regression analysis. The authors compared the performance of these two models with the clinicians' prediction and concluded that the AI-based models have slightly better performance, though the difference was not significant.

An important point was that Rosma et al.'s study did not use the big registrars to enhance their models' credibility and generalizability. The sample size of an AI-based model should be big enough (>200 EPV) to have unbiased predictions. Rosma et al.'s study had relatively small sample size (84 cases and 171 controls). Recent use of AI in HNC prediction modeling mainly focuses on the diagnostic models[57] or the ones predicting the recurrence[58] of or survival[59,60] from HNC, not its development. Therefore, future works in predicting the risk of developing HNC should use big

data registries and benefit from new techniques in AI such as Bayesian neural network, deep learning, or decision tree to develop generalizable risk prediction models for clinical settings.

Our overall assessment revealed a high ROB in the analysis part of all included models. All points mentioned above are embedded in the TRIPOD and PROBAST tools[46-48], and following them is suggested to ensure a good modeling. The TRIPOD's article[46] was published in 2015. Considering that all the included studies were published after 2015, we expected less deficiency in reporting, specifically in the model performance measurements. Also, the PROBAST tool[47,48] was published in 2019. This fact highlights an urgent need for communicating essential reporting checklists and valuable tools in prognostic HNC risk prediction modeling. It is also suggested that future studies use participants as representative of the target population as possible, provide sufficient information on the predictors and outcome measurements, and ensure appropriate data analysis and missing data management. The future models' performance needs to be assessed following the standards of the modeling studies. The models also need to be externally validated on a different population to ensure good generalizability and clinical applicability.

Strengths and limitations

The current study assessed all the published papers that reported at least one prognostic risk prediction model for HNC and its subsites. This is the first study that systematically reviews these papers, assesses the developed models' performance, and identifies current status while providing insights for future studies. We followed a specific quality assessment tool (PROBAST) to recognize the ROB and concerns about the applicability of the included papers. We also organized this article using a standard reporting checklist for modeling studies (TRIPOD).

Our study is limited in defining a specific outcome in the research question. We decided to consider all types of cancers in the upper aerodigestive tract because our initial search revealed limited numbers of modeling papers published for each HNC subsite. This study also did not report the distribution of predictors according to the HNC subsite. Although predictors differ among different subsites of HNC, there was a high ROB in reporting HNC predictors. Only one

included article[31] reported separate sets of predictors for each HNC subsite. Therefore, we could not fully report different sets of predictors related to the specific HNC subsites in most of the models. We also did not report the final models' characteristics, such as the coefficients, because there was limited information on some of the models' coefficients.

We only assessed the prognostic risk predictions to limit the number of the included studies and have a detailed assessment of the current status in the HNC risk prediction. There is still a need for a literature review on diagnostic risk prediction models for HNC.

Reproducibility and implementation of the HNC prognostic risk prediction models

A fully reported and shared model helps other researchers and clinicians recalibrate the model in the new setting and benefit from it. Therefore, the reproducibility of a modeling study needs to be highlighted in future studies. Only one study[32] fully reported the final model's characteristics. We recognized an urgent need for reproducible prognostic HNC risk prediction models for every country (population of interest). Future modeling studies should fully report the modeling process and the final developed model's characteristics. Sharing the data analysis codes also helps build blocks on the previously developed models.

The ultimate goal of prognostic risk prediction modeling is to develop a tool that recognizes high-risk people in primary care settings so that clinicians can warn these people about their high-risk behaviors. Prognostic models can also be used as an encouraging tool. A high-risk individual who will change the high-risk behaviors, such as decreasing the number of cigarettes or decreasing alcohol consumption, may use these models to track the risk changes during different time periods.

It is worth noting that the practitioners should not solely rely on the risk prediction models in decision-making. The prognostic risk prediction models are complementary tools that help practitioners, but the models' estimations should not be considered as the only factor considered in the decision making.

Conclusion

Many prognostic modeling studies fail to provide sufficient information to judge their models' performance. HNC prognostic risk prediction still needs a well-developed and well-performed model to help clinicians in critical dilemmas. Risk prediction models are complementary tools, and their estimates should not be considered the only means for clinical decision-making.

Prognostic risk prediction models are generalizable and applicable only to the source population. Therefore, a model derived from the data related to one specific area in a country (e.g., province or state) is not applicable for the whole population of that country. As a result, there is always a need for a well-developed and updated model for each geographical area and population of interest.

Contributions

Our team comprised of Drs. Sreenath Madathil (SM) and Belinda Nicolau (BN) as the main supervisors, Mohammad Al-Tamimi (MA) and Ziad Al-Asmar (ZA), and Martin Morris (MM) as the co-investigators. Hamed Ghanati (HG) is the chief investigator of this study. SAM, BFN, and HG developed the idea. MM has contributed to the search strategy. HG and ZAA shortlisted the studies. All authors contributed to drafting the manuscript.

Conflict of interest

Authors declare that they have no conflict of interest.

Tables and figures

Table 6-1: Table 1 General Characteristics of included studies

				Sample size			
First author (year)	**Analysis type**	**Study Design**	**Study Setting**	**Cases**	**Total**	**Outcome**	**Region/Country of source data**
Cheung (2021)	Cox regression	Cluster-randomized controlled screening trial	Community	395	191,870	Oral cancer incidence	Trivandrum, India
Gupta (2017)	Multivariable logistic regression	Case-control	Hospital	240	480	Cancers of lip, oral cavity, oropharynx, hypopharynx, upper third of esophagus (UADT)	Pune, Maharashtra, India
Hung (2020)	Multivariable logistic regression	Population based cohort	Community	117,697	1,719,191	Oral cancer incidence	Taiwan
Krishna Rao (2016)	Multivariable logistic regression	Case-control	Hospital	180	452	Oral cancer	Karnataka, India
Amy Lee (2020)	Multivariable logistic regression	Case-control from registry	Community	7,299	10,301	An invasive tumor of oral cavity, oropharynx, hypopharynx, or larynx	The USA
Tota (2019)	Multivariable logistic regression	Case-control from registry	Hospital and community	241	9,568	Oropharynx cancers	The USA
McCarthy (2020)	Multivariable logistic regression	Nested case-control	Community	389	502,177	Head and neck cancer excluding laryngeal cancer	The UK
Koyanagi (2016)	Conditional logistic regression	Case-control	Hospital	1,284	3,198	Cancers of upper aerodigestive tract, head and neck, and esophageal	Development & Validation: Nagoya, Japan
Rosma (2010)	Fuzzy neural network & Fuzzy regression	Case-control	Not provided	84	171	Oral cancer	Malaysia

Table 6-2: Model development characteristics of each study.

First author (year)	Study type	Missing data	How missing data was managed	Discrimination measurement	Calibration measurement	Adjustment for outcome frequency (population baseline risk)	Internal validation method
Cheung (2021)	Development only	Yes	Imputation	C-statistics	Observed/expected ratio (Calibration score)	Yes	Five-fold cross validation
Gupta (2017)	Development only	Yes (2% of the participants)	Not provided	C-statistics	Not provided	Not provided	Bootstrapping (1000 replications)
Hung (2020)	Development only	Not provided	Not provided	C-statistics	Not provided	Yes (Cohort study)	Not provided
Krishna Rao (2016)	Development only	Yes (<10%)	Excluded from analysis	C-statistics	Hosmer–Lemeshow gof test	Yes	Bootstrapping (200 replications)
Amy Lee (2020)	Development only	Yes	Inconsistencies resolved by discussion	C-statistics	Calibration score and plot	Yes	Splitting (70% training, 30% testing)
Tota (2019)	Development and validation	Yes	Imputation	C-statistics	Observed/expected ratio (Calibration score)	Yes	Splitting
McCarthy (2020)	Development and validation	Yes (<1% [except exercise 5.9%])	Not provided	C-statistics	Calibration score and plot	Not provided	Nothing done
Koyanagi (2016)	Development and validation	Yes	Imputation (coded as dummy variables)	C-statistics	Hosmer–Lemeshow gof test & Calibration plot	Yes	Not provided
Rosma (2010)	Development only	Not provided	Not provided	C-statistics	Not provided	Not provided	Splitting

Table 6-3: Model type, outcome, predictors, and performance of models.

Model	Model Type	Outcome	Predictors	Performance metrics
Cheung (2021)	Cox regression	Oral cancer incidence	Time, age, sex, education, BMI, tobacco chewing, smoking, alcohol, interaction between chewing duration and smoking intensity, and study group or arm	AUC overall[d]: 0.84 (0.77–0.90) AUC ever T&A[e]: 0.75 (0.67–0.83) O/E overall[f]: 1.08 (0.81–1.44) O/E ever T&A[g]: 1.07 (0.77–1.43)
Gupta (2017)	Multivariable logistic regression	Cancers of lip, oral cavity, oropharynx, hypopharynx, upper third of esophagus	Socio-demographic profile, chewing tobacco, smoking tobacco, Mishri, alcohol Diet, housing, secondhand smoke, medical history, family history, BMI	AUC = 95.8 (93.60–97.40) PPV: 74.80% NPV: 96.60%
Hung (2020)	Multivariable logistic regression	Oral cancer incidence	Age, sex, smoking, betel nut chewing, educational level, marital status, Rural/urban residency, diabetes, other cancers, comorbidity severity, other catastrophic illnesses, salary	AUC = 0.7306 PPV: 63.90% NPV: 71.10%
Krishna Rao (2016)	Multivariable logistic regression	Oral cancer	Sex, chewing tobacco chewing, alcohol, BMI, diet, rinsing mouth with water after smoking/eating, socioeconomic status, parent's education, paternal alcohol drinking habits, family history of upper aerodigestive tract cancers	AUC = 0.869 PPV: 77.30% NPV: 83.00%
Amy Lee (2020)	Multivariable logistic regression	An invasive tumor of oral cavity, oropharynx, hypopharynx, or larynx	First set: age, sex, education, race/ethnicity, alcohol, cigarette smoking, and/or family history of HNC[h].	AUC was 0.70 or higher, except for oropharyngeal cancer in men (0.643).

[d] AUC related to the internal validation on overall population

[e] AUC related to the internal validation on ever tobacco and/or alcohol users

[f] Observed/Expected ratio

[g] Ever tobacco and/or alcohol users

[h] In the second set family history of HNC was added to the models, except for oropharyngeal cancer in both sexes and laryngeal cancer in men, where a clear risk was not observed in the data.

Tota (2019)	Multivariable logistic regression	Oropharynx cancers	age, sex, race, smoking, alcohol, lifetime number of sexual partners, HPV infection	Internal: AUC: 0.94 (0.92-0.97) O/E: 1.05 (0.67-1.44) External: AUC: 0.87 (0.84-0.90) O/E: 0.91 (0.57-1.25)
McCarthy (2020)	Multivariable logistic regression	Head and neck cancer	Age, sex, smoking, alcohol, diet and exercise, BMI, medical history, socio-demographics, >=6 lifetime sexual partners	AUC: 0.69 (0.66-0.71) Calibration slope (external): 0.83
Koyanagi (2016)	Conditional logistic regression	Cancers of upper aerodigestive tract, head and neck, and esophageal	Genetic model: age, sex, ALDH2 genotype Environmental model: age, sex, cumulative smoking & alcohol Inclusive model: age, sex, cumulative smoking, combination of ALDH2 genotype&alcohol	AUC: Genetic UADT: 0.65 (0.62–0.68), external validation: 0.65 (0.62–0.68) Genetic HNC (internal): 0.59 (0.55–0.62) (external): 0.54 (0.49–0.58)
Rosma (2010)	Fuzzy neural network & Fuzzy regression	Oral cancer	age gender cigarette smoking alcohol tobacco chewing[i]	AUC: Fuzzy neural network: 0.804 Fuzzy regression: 0.799 Clinicians' predictions: 0.631

[i] 12 different input variable sets based on these five variables. The best model's predictors are age, gender, cigarette smoking, alcohol, tobacco chewing, and the mixed variable of "age, alcohol, tobacco chewing, and cigarette smoking"

Table 6-4: Predictors used in the final model.

	Cheung (2021)	Gupta (2017)	Hung (2020)	Krishna Rao (2016)	Amy Lee (2020)	Tota (2019)	McCarthy (2020)	Koyanagi (2016)	Rosma (2010)
Age	●		●		●	●	●	●	●
Sex	●		●	●	●	●	●	●	●
Race/Ethnicity					●	●			
Tobacco smoking	●	●		●	●	●	●		●
Tobacco Chewing	●	●		●					●
Alcohol	●	●			●	●	●		●
HPV infection						●			
Education	●		●		●		●		
Paternal education				●					
Occupation							●		
Income			●				●		
Housing		●					●		
Urban/Rural living			●						
BMI	●	●		●					
Mishri		●							
Bidi smoking		●							
Betel chewing			●	●					
Very spicy food		●		●					
Diet with fresh fruits and vegetables				●			●		
Exercise							●		
Second-hand smoking		●							
Family history				●	●				
Medical history							●		
Other diseases			●						
Marital status			●						
ALDH2 genotype								●	
Screened or non-screened	●								
Lifecourse data	No	Yes	No	No	No	No	No	No	No
Urban/rural			●						
Rinsing mouth with water after eating and smoking				●					
Sexual partners						●			

Figure 6-1: Flow diagram of selection process

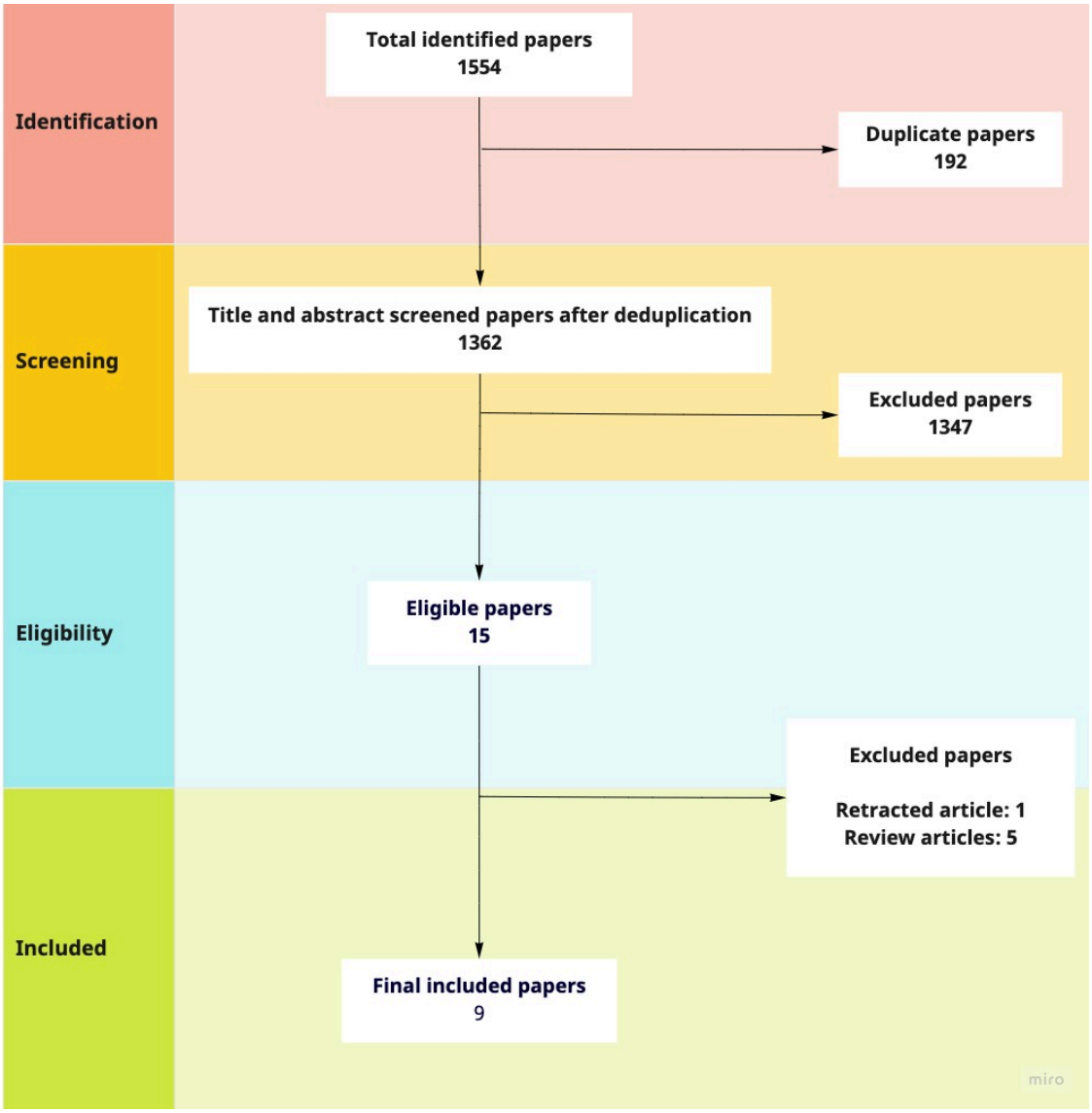


Figure 6-2: Assessment of risk of bias

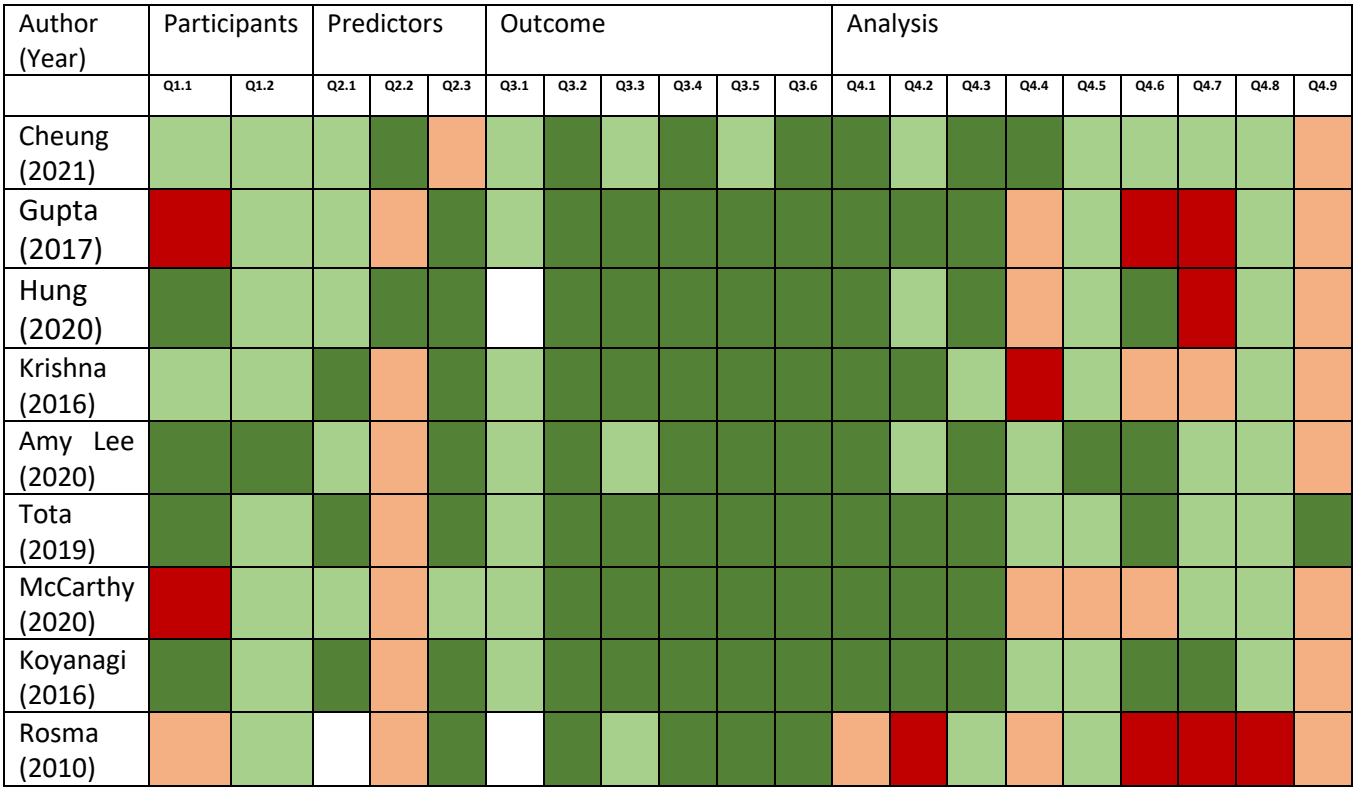


Figure 6-3: Assessment of concerns about applicability

	Participants	Predictors	Outcome
Cheung (2021)			
Gupta (2017)			
Hung (2020)			
Krishna (2016)			
Amy Lee (2020)			
Tota (2019)			
McCarthy (2020)			
Koyanagi (2016)			
Rosma (2010)			

Table 6-5: PROBAST results

	ROB				Applicability			Overall	
Author (published year)	**Participants**	**Predictors**	**Outcome**	**Analysis**	**Participants**	**Predictors**	**Outcome**	**ROB**	**Applicability**
Cheung (2021)	+	-	+	-	+	-	-	-	-
Amy Lee (2020)	+	-	+	-	+	+	+	-	+
Hung (2020)	+	+	?	-	-	+	-	-	-
McCarthy (2020)	-	-	+	-	+	+	+	-	+
Rosma (2010)	-	-	?	-	+	+	+	-	+
Tota (2019)	+	-	+	+	+	+	+	-	+
Gupta (2017)	-	-	+	-	+	+	+	-	+
Koyanagi (2016)	+	-	+	-	+	+	+	-	+
Krishna (2016)	+	-	+	-	+	+	+	-	+

Table 6-6: Search strategy on Medline (Ovid)

1. "Head and Neck Neoplasms"/ or exp Facial Neoplasms/ or exp Mouth Neoplasms/ or exp Otorhinolaryngologic Neoplasms/ or exp Tracheal Neoplasms/
2. ((cancer* or tumo?r* or neoplas* or metaplas* or carcinoma* or metastasi* or squamous cell carinoma? or SCC or HNSCC or malignan*) adj5 (head or neck or uadt or "upper aero-digestive" or "upper aerodigestive" or face or facial or oral* or intra-oral* or intraoral* or mouth or buccal or gingiv* or gum* or lip? or labial* or palat* or lingual* or mandib* or maxill* or jaw? or tongue* or glossal* or otor?inolaryngolog* or throat or ear? or auricle* or auricular or larynx* or laryngeal* or nose* or nasal* or paranasal* or sinus or hypopharynx or hypopharyngeal* or nasopharynx or nasopharyngeal* or oropharynx or oropharyngeal or tonsil* or trachea* or cheek* or pharynx or pharyngeal or retromolar or alveolar or tonsil* or sinonasal or sinus* or vestib* or piriform or post-cricoid or glottic or subglottic or superglottic or transglottic or "unknown primary" or trigone or maxillofacial*)).tw,kf.
3. (exp Head/ or exp Neck/) and exp Neoplasms/
4. (head and neck surgery).mp.
5. 1 or 2 or 3 or 4
6. exp Epidemiology/
7. exp Epidemiologic Methods/
8. exp Epidemiologic Studies/
9. exp Models, Statistical/
10. exp Statistics as Topic/
11. exp mathematical concepts/
12. ((computer* or likelihood or predicti* or prognosti* or linear or log-linear or binomial or polynomial or probabili* or statistical or two-parameter or logistic or logit or hazard? or cox) adj1 (model? or tool? or regression? or function? or test? or estimate? or ratio?)).tw,kf.
13. (statistics or statistical or nomogram? or partin table? or algorithm* or stochastic or machine learning or probability or pattern analysis or training set or deep learning).tw,kf.
14. (ep or sn).fs.
15. computing methodologies/ or exp computer simulation/ or exp mathematical computing/
16. (computer* adj1 (simulation? or mathematic* or statistic*)).tw,kf.
17. ((probability adj3 approach) or in silico?).tw,kf.
18. exp Risk/ or exp Risk Assessment/
19. or/6-18
20. 5 and 19
21. 9 or 10 or 11 or 12 or 13 or 15 or 16 or 17
22. 5 and 21
23. 1 or 2
24. 21 and 23
25. 9 or 11 or 12 or 13 or 15 or 16 or 17
26. 23 and 25
27. risk?.tw,kf.
28. 20 and 27
29. 26 and 28

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249.
2. Lee S. Cancer Statistics. Canadian Cancer Society. https://cancer.ca/en/research/cancer-statistics. Published 2022. Accessed 3/20/2022.
3. Tiwana MS, Wu J, Hay J, Wong F, Cheung W, Olson RA. 25 year survival outcomes for squamous cell carcinomas of the head and neck: population-based outcomes from a Canadian province. *Oral Oncol.* 2014;50(7):651-656.
4. 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.
5. Kligerman MP, Sethi RKV, Kozin ED, Gray ST, Shrime MG. Morbidity and mortality among patients with head and neck cancer in the emergency department: A national perspective. *Head Neck.* 2019;41(4):1007-1015.
6. 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.
7. Toporcov TN, Znaor A, Zhang Z-F, et al. Risk factors for head and neck cancer in young adults: a pooled analysis in the INHANCE consortium. *International journal of epidemiology.* 2015;44(1):169-185.
8. Johnson-Obaseki S, McDonald JT, Corsten M, Rourke R. Head and neck cancer in Canada: trends 1992 to 2007. *Otolaryngology--Head and Neck Surgery.* 2012;147(1):74-78.
9. 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. *International journal of cancer.* 2005;114(5):806-816.
10. Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. Paper presented at: Mayo Clinic Proceedings2008.
11. Argirion I, Zarins KR, Defever K, et al. Temporal changes in head and neck cancer incidence in Thailand suggest changing oropharyngeal epidemiology in the region. *Journal of Global Oncology.* 2019;5:1-11.

12. Joseph AW, D'Souza G. Epidemiology of human papillomavirus-related head and neck cancer. *Otolaryngol Clin North Am.* 2012;45(4):739-764.
13. Curado MP, Hashibe M. Recent changes in the epidemiology of head and neck cancer. *Current opinion in oncology.* 2009;21(3):194-200.
14. Habbous S, Chu KP, Lau H, et al. Human papillomavirus in oropharyngeal cancer in Canada: analysis of 5 comprehensive cancer centres using multiple imputation. *Cmaj.* 2017;189(32):E1030-e1040.
15. Hashim D, Genden E, Posner M, Hashibe M, Boffetta P. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. *Ann Oncol.* 2019;30(5):744-756.
16. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. In: Second edition. ed. Cham, Switzerland: Springer; 2019: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2204948

https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5836937

https://doi.org/10.1007/978-3-030-16399-0

https://link.springer.com/book/10.1007%2F978-3-030-16399-0

http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9783030163990.

17. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis.* 2019;11(Suppl 4):S574-s584.
18. Chen L. Overview of clinical prediction models. *Annals of Translational Medicine.* 2020;8(4).
19. Hendriksen JM, Geersing G-J, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *Journal of Thrombosis and Haemostasis.* 2013;11:129-141.
20. Domchek SM, Eisen A, Calzone K, Stopfer J, Blackwood A, Weber BL. Application of breast cancer risk prediction models in clinical practice. *Journal of Clinical Oncology.* 2003;21(4):593-601.
21. Silveira A, Monteiro E, Sequeira T. Head and Neck Cancer: Improving Patient-Reported Outcome Measures for Clinical Practice. *Curr Treat Options Oncol.* 2018;19(11):59.

22. Tammemaegi MC. Application of risk prediction models to lung cancer screening: a review. *Journal of thoracic imaging.* 2015;30(2):88-100.
23. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics.* 2016;17(7):392-406.
24. Guo L-W, Lyu Z-Y, Meng Q-C, et al. A risk prediction model for selecting high-risk population for computed tomography lung cancer screening in China. *Lung Cancer.* 2022;163:27-34.
25. Tammemägi MC, Ruparel M, Tremblay A, et al. USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): interim analysis of a prospective cohort study. *Lancet Oncol.* 2022;23(1):138-148.
26. Cheung LC, Ramadas K, Muwonge R, et al. Risk-Based Selection of Individuals for Oral Cancer Screening. *J Clin Oncol.* 2021;39(6):663-674.
27. Diana G, Corica C. Human Papilloma Virus vaccine and prevention of head and neck cancer, what is the current evidence? *Oral Oncol.* 2021;115:105168.
28. Gupta B, Kumar N, Johnson NW. A risk factor-based model for upper aerodigestive tract cancers in India: predicting and validating the receiver operating characteristic curve. *J Oral Pathol Med.* 2017;46(6):465-471.
29. Hung LC, Kung PT, Lung CH, et al. Assessment of the Risk of Oral Cancer Incidence in A High-Risk Population and Establishment of A Predictive Model for Oral Cancer Incidence Using A Population-Based Cohort in Taiwan. *Int J Environ Res Public Health.* 2020;17(2).
30. 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.
31. 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.
32. Tota JE, Gillison ML, Katki HA, et al. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer.* 2019;125(24):4407-4416.

33. 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.
34. 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.
35. Rosma MD. The Use of Artificial Intelligence to Identify People at Risk of Oral Cancer : Empirical Evidence in Malaysian University. 2010.
36. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology.* 2005;8(1):19-32.
37. Westphaln KK, Regoeczi W, Masotya M, et al. From Arksey and O'Malley and Beyond: Customizations to enhance a team-based, mixed approach to scoping review methodology. *MethodsX.* 2021;8:101375.
38. Morris M, Boruff JT, Gore GC. Scoping reviews: establishing the role of the librarian. *J Med Libr Assoc.* 2016;104(4):346-354.
39. Embase. https://www.wolterskluwer.com/en/solutions/ovid/embase-903. Accessed 3/20/2022.
40. CAB Abstracts. CABI.org Web site. https://www.cabi.org/publishing-products/cab-abstracts/. Accessed 3/20/2022.
41. Scopus preview. https://www.scopus.com/. Accessed 3/20/2022.
42. Document search - Web of Science Core Collection. https://www.webofscience.com/wos/woscc/basic-search. Accessed 3/20/2022.
43. Gotschall T. EndNote 20 desktop version. *J Med Libr Assoc.* 2021;109(3):520-522.
44. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev.* 2016;5(1):210.
45. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* 1960;20(1):37-46.

46. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *European Urology.* 2015;67(6):1142-1151.
47. Wolff RF, Moons 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.
48. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med.* 2019;170(1):W1-w33.
49. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth.* 2020;18(10):2119-2126.
50. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol.* 2011;29(32):4294-4301.
51. Xing J, Watkins WS, Hu Y, et al. Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol.* 2010;11(11):R113.
52. 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.
53. Australia C. Head and neck cancer in Australia statistics. Cancer Australia 2022. https://www.canceraustralia.gov.au/cancer-types/head-and-neck-cancer/statistics. Published 2022. Accessed 3/20/2022.
54. Lewallen S, Courtright P. Epidemiology in practice: case-control studies. *Community Eye Health.* 1998;11(28):57-58.
55. Sabatini ME, Chiocca S. Human papillomavirus as a driver of head and neck cancers. *Br J Cancer.* 2020;122(3):306-314.
56. Chen H, Cohen P, Chen S. Biased odds ratios from dichotomization of age. *Statistics in medicine.* 2007;26(18):3487-3497.

57. Kazmierski M, Welch M, Kim S, et al. A Machine Learning Challenge for Prognostic Modelling in Head and Neck Cancer Using Multi-modal Data. *arXiv preprint arXiv:210111935.* 2021.

58. Peng Z, Wang Y, Wang Y, et al. Application of radiomics and machine learning in head and neck cancers. *International Journal of Biological Sciences.* 2021;17(2):475.

59. Huynh B-N, Ren J, Groendahl AR, Tomic O, Korreman SS, Futsaether CM. Comparing deep learning and conventional machine learning for outcome prediction of head and neck cancer in PET/CT. Paper presented at: 3D Head and Neck Tumor Segmentation in PET/CT Challenge2021.

60. Salmanpour MR, Hosseinzadeh M, Modiri E, et al. Advanced survival prediction in head and neck cancer using hybrid machine learning systems and radiomics features. Paper presented at: Medical Imaging 2022: Biomedical Applications in Molecular, Structural, and Functional Imaging2022.

Preface to Manuscript II

Based on what is discussed in chapters 3 and 4 of this thesis, the first step in providing a population of interest with a risk prediction model is to investigate the applicability of the existing models. The *Manuscript I* confirmed that no model was developed or validated in the Canadian context. Nevertheless, it identified two studies reporting the models developed and validated on the US population. These models are potentially generalizable to the Canadian context because they include the main risk factors of HNC in Canada; however, only one of them is replicable as its investigators reported full details of the model characteristics (e.g., values of intercept and coefficients). This model is for predicting the one-year risk of developing oropharyngeal cancer. *Manuscript II* uses data from a sample of people living in Montreal, Quebec, to assess this model's applicability in Canada. The next chapter provides the full text of this manuscript, and, similar to the previous chapter, comprises independent appendices and a reference list, according to the requirements specified by the McGill Graduate and Postdoctoral Studies.

7 Manuscript II

Validating an oropharyngeal cancer risk prediction model on a Canadian population

Hamed Ghanati[1], Belinda Nicolau [1], Paul Allison[1], Anita Koushik[2], Sreenath Madathil [1 *]

[1] Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada

[2] Centre de recherche du CHUM and Department of Social and Preventive Medicine, Université de Montréal, Montréal, QC, H2X 0A9, Canada

*Corresponding author:

Dr. Sreenath Madathil

Faculty of Dental Medicine and Oral Health Sciences, McGill University, 2001 Avenue McGill College, Montréal, QC, H3A 1G1, Canada

Keywords:

Head and Neck Neoplasms; Prognosis; Models, Statistical; Probability.

Word count:

3,385

Abstract

Background: Recently, there has been an increase in the Human papillomavirus (HPV) related oropharyngeal cancer (OPC) incidence rate in North America. Predicting who is at high risk of developing OPC may lead to more targeted preventive interventions. However, there is a lack of OPC risk prediction models that are validated for the Canadian population. We aim to externally validate an OPC risk prediction model developed elsewhere using a Canadian dataset and discuss its potential clinical applicability.

Method: We selected a model developed on a case-control study from the US. Based on this model's development study subject's characteristics, the validation data were derived from the HeNCe Life study, a hospital-based case–control investigating head and neck cancer etiology. The validation dataset comprised 214 cases of oropharyngeal squamous cell carcinoma and 433 controls frequency matched to the cases by age and sex, recruited from the four referral hospitals in Montreal, Quebec. Participants aged between 30 to 79 years. The predictors comprised age, sex, race, lifetime pack-years of smoking, last year's alcohol, number of lifetime sex partners, HPV infection, the interaction between the number of sex partners and HPV, and interaction of smoking and HPV as predictors. Different performance metrics and plots assessed the model's overall performance, discriminative ability, and calibration.

Results: The model presented Somers' D, scaled Brier score, Nagelkerke's R^2, AUC, and calibration slope values of 0.49, 68.84%, 0.24, 0.74 (95% CI:0.69–0.79), and 0.57 (95% CI:0.45–0.68), respectively. Calibration-in-the-large of 4.93 (95% CI:4.71–5.16) points indicated systematic mis-calibration of the predictions over the validation dataset.

Conclusion: The model's overall predictive performance and discrimination ability was above average. However, its poor calibration demonstrates the need for recalibrating it on the HeNCe life study's dataset. Further studies are needed to evaluate this model's performance after updating it on the Canadian context.

Introduction

With less than 50% survival rate, the head and neck cancers (HNC) are categorized among the cancers with high mortality rates[1]. It is estimated that 2,100 out of 7,500 Canadians with HNC will die in 2022[2]. Despite global efforts in controlling HNC[3-5], the incidence of one of its subsites, oropharyngeal cancer (OPC), has substantially increased over the past few decades, predominantly in high income countries such as Canada[2,6]. A considerable part of this increase is attributed to human papillomaviruses (HPV) infection.[7,8] Indeed, recent studies have reported a strong sex and birth cohort-specific trend in HPV infection in Western countries, with younger men being the most affected group[9,10]. By the time of writing this paper, 229 types of HPVs have been identified[11]; however, only HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 are known to be associated with the increased risk of HPV-related OPC[12,13]. Among these, HPV-16 is the most prevalent type detected in the OPC tumor samples[8,13]. In addition to the HPV infection, numerous studies have reported an association between smoking and alcohol consumption and an increased risk of developing OPC[14-18].

Detecting OPC in the early stages is crucial to improve its survival rate[19-22]. However, due to the lack of symptoms during initial stages and limited visibility to the oropharynx area diagnosis of OPC during the early stage is challenging[23-25]. Preventive interventions may reduce HPV-related OPC and improve the overall OPC mortality rate[26]. Assessing the risk of developing OPC and detecting high-risk individuals will assist in conducting personalized preventive interventions and leading public health policies toward controlling OPC.

We recently reviewed the HNC risk prediction modeling[j] studies and found no model developed or validated for the Canadian population. We also identified an OPC risk prediction model, developed and validated by Tota et al. (2019) on a sample of the US population[12], potentially generalizable to the Canadian context. However, should a model be implemented on a different population, its estimates need to be validated before using it in practice[27,28]. Therefore, this study

[j] To avoid the complexity in this paper, we use the term “risk prediction model” as of “prognostic risk prediction model”.

aimed to externally validate the identified model on a sample of the Canadian population and assess its performance.

Methods

Development study

Details about the identified model are published elsewhere[12]. Briefly, Tota et al. reported a multivariable logistic regression model developed using data from a case-control study conducted by weighted oversampling of data from 241 incidence cases of OPC at the Ohio State University Comprehensive Cancer Center from 2011 to 2015 and 9327 US population-representative controls collected from the participants of the National Health and Nutrition Examination Survey (NHANES) from 2009 to 2014[29]. The model was internally validated using the split-sample technique and externally validated on 116 OPC cases from Johns Hopkins University series[30,31] and 3,237 controls from NHANES 2013-2014[29].

The original study reported strong discrimination ability during both internal and external validation with the AUC scores of 0.94 (95% CI:0.92–0.97) and 0.87 (95% CI:0.84–0.90), respectively. The model's calibration was also good during both the internal and external validation process, with calibration slope values (the overall observed/expected ratio) of 1.01 (95% CI:0.70–1.32) and 1.08 (95% CI:0.77-1.39), respectively. The model's predictors were age, sex, race, pack-years of smoking, last year's alcohol consumption, number of lifetime sex partners, HPV infection status, the interaction of sex and HPV, and the interaction of smoking and HPV.

Source of validation dataset

To validate these models, we used a dataset from the HeNCe Life study, a hospital-based case-control investigating the etiology of HNC in Canada[7,8,32,33]. The HeNCe Life study was conducted at four major HNC referral hospitals in Montreal, Quebec, Canada from 2005 to 2013.

Study participants

The participants of HeNCe Life aged between 25 to 93 years. However, we filtered the dataset for the participants aged between 30 to 79 years to reflect the age range applicable to Tota et al.'s study[12]. The filtered dataset comprised 214 consecutive incidence cases diagnosed with the histologically confirmed Squamous Cell Carcinoma (SCC) in the base of the tongue, soft palate, palatine tonsils, oropharynx, and uvula related to the ICD-10 codes of C01, C02.4, C05.01, C05.2, C09, C10, C12 and, C14. The cancer ascertainments were done by the expert histopathologists at the referred hospitals using the standard definition for the SCC issued by the National Cancer Institute[34].

Controls comprised 433 individuals who were recruited within the maximum of three months of the related case subject's enrollment. Controls frequency-matched with cases by sex and age (5-year categories) randomly selected from the list of diseases not related to smoking, alcohol in several outpatient clinics of the same referral hospitals. To be part of HeNCe Life study, participants had to: (i) be born in Canada, (ii) speak English or French, (iii) be 18 years of age or older, (iv) have no history of cancer, immunosuppressive condition, or mental disorders, and (iv) live within 50km area of the referred hospitals. Table 7-1 displays the characteristics of the participants.

Predictors Information

HeNCe Life employed semi-structured one-on-one interviews using a questionnaire with the life grid technique to collect life course data on sociodemographic and behavioral characteristics such as age, sex, race, lifetime number of sex partners, and life course exposure to tobacco smoking and alcohol consumption. Data on the HPV infection was collected using HPV DNA detection from the oral cell samples. Data collection was supervised by an expert oral epidemiologist to have a similar predictor assessment for all the participants and avoid biased. Details about the HeNCe Life study data collection have been published elsewhere[7,8,32,33].

Data analysis

We calculated the minimum sample size following the method published by Riley et al. (2021)[35] considering the ideal Observed/Expected (O/E) ratio of 1 and the outcome event proportion of 0.478 in the HeNCe Life dataset. We aimed for a 95% confidence interval of 0.2 for O/E to ensure a good calibration-in-the large in the external validation.

All variables were coded based on the development study[12]. Accordingly, age, smoking, and the number of lifetime sex partners were considered as continuous variables. To create the pack-years of smoking variable, we calculated the cumulative number of cigarette standard packs based on the estimated tobacco content of different smoking types. Accordingly, one pack of standardized cigarettes equals 20 commercial filtered or unfiltered cigarettes, five pipes, four hand-rolled cigarettes, or four cigars[36]. We then multiplied the daily smoking packs by number of smoking years to obtain the pack-years variable. Based on the development study, this predictor was coded as a continuous variable presenting logarithm of smoking pack-years.

We may replicate the model as below multivariable logistic regression equation:

$$
\begin{aligned}
Logit(Probability\ of\ developing\ OPC) \\
&= -12.40 + \left(0.09 \times x_{Age}\right) - \left(1.09 \times x_{Gender_{Female}}\right) \\
&- \left(0.25 \times x_{Race_{Black}}\right) - \left(1.30 \times x_{Race_{Other}}\right) + \left(0.16 \times x_{Smoking}\right) \\
&+ \left(0.60 \times x_{Alcohol}\right) + \left(0.16 \times x_{\ 2\ to\ 5\ sex\ partners}\right) \\
&+ \left(0.32 \times x_{\ 6\ to\ 10\ sex\ partners}\right) + \left(0.48 \times x_{\ >10\ sex\ partners}\right) \\
&+ \left(3.80 \times x_{\ HPV\ positive}\right) + \left(1.38 \times x_{female} \times x_{HPV\ positive}\right) \\
&- \left(0.33 \times x_{smoking} \times x_{HPV\ positive}\right)
\end{aligned}
$$

Following Tota et al.'s cateorizaion, we considered the ethanol concentration of every alcoholic beverage and summed up the milliliters of ethanol exposure within the last year ending to the recruitment. We then converted the milliliters to "number of drinks" by dividing it by 17.05 based on the standard amount of ethanol per drink in Canada[37]. The alcohol exposure variable was categorized into "less or equal to 14 drinks per week" and "more than 14 drinks per week" levels

based on the models' development study. Race and HPV status were also coded and categorized into the "White", "Black", and "Other", and "HPV negative" and "HPV positive[k]", respectively.

Model performance assessment

We followed the framework proposed by Steyerberge et al.[38] for assessing the model's performance. We calculated the Sumers' Delta (Sumers' D) score, scaled Brier score, and Nagelkerke's R^2 to evaluate the model's overall performance and assessed the model's discrimination ability by measuring its C-Statistics and discrimination index. Calibration of the model was evaluated using the calibration plot and the values of calibration-in-the-large, calibration slope, and Hosmer-Lemeshow goodness-of-fit (H-L test). We obtained the smoothed calibration plot by *loess* algorithm[39] through the val.prob function in "Regression Modeling Strategies" statistical package version 6.3.0[40]. We investigated the model's predictions' clinical usefulness by plotting clinical decision curves based on different cut-points. All analyses were conducted on R, version 4.2.0[41]. We followed the TRIPOD checklist for Prediction Model Development or Validation[42] to conduct and report this study.

Results

Based on the sample size calculation, we needed a minimum overall sample size of 448 participants, with a minimum of 214 cases (events), to ensure a good calibration-in-the-large and acceptable external validation. Table 7-1 illustrates the characteristics of the validation dataset. The majority of participants aged between 53 and 67 years. While most of the participants were male (71.60%), Caucasian White race was dominated both cases and controls. Almost half of the cases (49.1 % of the cases) had more than 20 lifetime pack-years of smoking. The majority of both cases and controls had drunk less than or equal to 14 alcoholic drinks per week during the last year of recruitment in the HeNCe Life study. The mean number of sex partners the prevalence of HPV infection among cases were higher than that of the controls.

[k] HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59.

Only the HPV variable had missing values (7.00%) in the validation dataset. Missing data was more prevalent among cases (13.60%) than in controls (3.70%). Table 7-2 provides detailed information on the performance assessment.

The model's overall performance was above medium level (Somers' D score = 0.49). The scaled brier score of 68.84% also supported this finding. Although C-statistics was above 0.7 (Figure 7-1), the discrimination slope indicated that the model is not well discriminating between those with and without the outcome (Figure 7-2).

Figure 7-3 displays the calibration plot for the model over the validation data. Even though the H-L test chi-square was significant (p=0.00) with nine degrees of freedom[27], the calibration-in-the-large value of 4.93 (95% CI:4.71-5.16) indicated that the predictions were systematically too high (overestimated). Calibration slope (CS) also was 0.57 (95% CI:0.45–0.68). Considering the difference of this value with the perfect calibration (CS=1), the model is mis-calibrated on the validation dataset.

Discussion

This study externally validated a risk prediction model developed by Tota et al. (2019)[12]. One reason for choosing this model was the similarity between the OPC epidemiology in the US and Canada. The incidence of oropharyngeal SCCs, the most dominant form of OPC[43], is consistently increasing in these countries, especially among young men[20,44]. Also, a systematic review published in 2016 identified evidence supporting a similar prevalence of HPV infection in oropharyngeal SCCs diagnosed in these countries[45]. Therefore, it was expected that the predictions of a model developed for the US population could potentially be generalizable to the Canadian context. In a recent scoping review, we identified studies reporting the development of OPC risk prediction models for a sample of the population in the United States[12,46]. This review identified the Tota et al.'s model[12] easily replicable. Therefore, we proceeded to investigate its potential applicability in the Canadian context using the data from the HeNCe Life study.

The model presented a suboptimal calibration. Although the H-L test result was significant (P=0.00), the calibration slope was far lower than 1, indicating a poor calibration. Also, the calibration-in-the-large value showed that model predictions are generally mis-calibrated. Such a miss-calibration could be due to the difference between the baseline risk of OPC (i.e., the risk of developing OPC among people with the lowest levels of risk factor exposure[27]) in the populations of interest with the temporal and geographical heterogeneity[47]. Some case studies reported that the factors outside of the model, such as disease prevalence or baseline risk, are attributable to the mean difference in the predicted probability derived from a model[48,49]. Therefore, it is expected that adjusting the predictions of the current model for the baseline risk of developing OPC among residents of Montreal, Quebec, and for the amount of mis-calibration will improve the model's predictive performance. The calibration-in-the-large value derived from this study can be used in updating the model in future studies. One reason that this study did not proceed to the updating the models relates to challenges in calculating the baseline risk in the HeNCe life population. Canadian Cancer Society publishes the data on different cancers; however, the incidence of OPC is usually reported combined with HNC incidence[2], making the calculations of baseline risk difficult.

There are specific limitations related to our study. The model was not recalibrated before the performance assessment. Our goal was to evaluate the model's net prediction performance before updating it to better understand the strength and limitations that needed to be considered in future studies. Having the net performance in hand also makes it possible to compare this model's performance before and after the updating. Moreover, we did not quantify the clinical usefulness, that is, the net true predictions for cases gained by using the model compared to no model at different thresholds. Although this test could shed light on the model's clinical applicability, the results' interpretation would not be accurate as the model was not calibrated on the validation dataset. Future steps would be to recalibrate the model and assess clinical usefulness based on the decision curves at different thresholds. It needs to be highlighted that assessing an OPC risk prediction model's clinical usefulness based on the risk thresholds is

challenging because there is insufficient evidence supporting a specific risk threshold for decision-making on OPC clinical management[27,38].

There are also challenging issues using the HeNCe Life dataset.

First, the HeNCe Life study used a strict inclusion criterion that could have eliminated control participants who were smokers and heavy alcohol drinkers. Although the distribution of these risk factors among the control group was similar to the Quebec population[50,51], we cannot guarantee that the sample is representative of the people living in Montreal. Therefore, the representativeness of the model's predictions may be questionable. This seems to be a common problem for models developed using secondary data, in which the studies were designed to address different research questions[42]. While the best practice is to conduct a study specifically designed for modeling purposes[27,42], conducting an independent study for a model's validation for rare diseases such as OPC is not cost-effective.

Second, different studies have reported that incorporating the participants' cancer status into the predictor assessments can put the predictor's measurements at a high risk of bias[42,52-54]. Due to the retrospective design of HeNCe Life, predictors' assessments were inevitably made with knowledge of the participants' outcome status, putting the current model's predictions at a high risk of bias.

Third, the data source for developing or validating a risk prediction model should ideally come from a prospective longitudinal cohort study. In a case-control study (except nested case-control), the predictor measurements are at a high risk of recall bias[55,56]. As mentioned above, conducting a cohort study to assess OPC risk factors is time-consuming and inefficient. Therefore, developing or validating a model using a case-control study might put the model's predictions at a high risk of bias due to the recall biases in the original study's predictor assessments. The HeNCe Life study used the life grid technique that is suggested to minimize the recall bias in a

retrospective observational study[7,57,58]. Therefore, our model predictions may be at low risk of participants' recall bias.

Finally, the HeNCe Life study comprised a small fraction of the Canadian population who resided in Quebec province. Therefore, results may not be generalizable to the whole Canada. Further studies are needed to understand the validated model's performance on Canadians residing in other provinces.

Some advantages can also be highlighted concerning this validation study. To our knowledge, this is the first study to tested an OPC risk prediction model in a Canadian population. Despite a rising trend in HPV-related OPC in North American countries, Canada's health care system has not benefited from an OPC risk prediction model. Therefore, this study's results may help researchers develop valuable models for different settings in Canada. The HeNCe Life study had sufficient number of cases that allowed us to avoid the biased model predictions due to small sample size. Different studies have suggested that an external validation study should comprise at least 100 cases to obtain unbiased model predictions[59-61]. Furthermore, the HeNCe Life included 647 participants with 214 cases, which was sufficient for unbiased predictions based on our sample size calculations (Minimum 448 participants and 214 cases).

Moreover, evaluating various performance metrics regarding the model's overall performance, discrimination ability, and calibration enabled us to better understand the model's strength, limitations, and implications in practice. Depicting the ROC curve and Calibration plot also helped understand the models' discriminative performance and mis-calibration.

Suggestions for future studies

This study studied the model's performance without altering or updating its characteristics. Although the model's calibration was suboptimal, recalibrating may improve this metric[38]. Updating a model is a hierarchical process that can start with adjusting the calibration-in-the-

large and continue updating different parts of the model, e.g., updating the coefficient values or adding new predictor(s)[27].

It is always preferable to use updated data to develop or validate a risk prediction model. The recruitment for the HeNCe Life study finished in 2013[7,8]. Considering the recent constant change in the OPC epidemiology in North America[45], testing the current model on an updated dataset can be beneficial.

As the final note, parsimonious methods such as validating the existed models are often preferable to the new model development[27]. Therefore, it is suggested that future studies start with simply updating this model by methods such as adjustment for the baseline risk and calibration-in-the-large before proceeding with the complex modeling strategies.

Conclusion

The validated model presented suboptimal calibration, although its overall performance and discrimination ability were above average. Nevertheless, the model may still be used in Canada's clinical settings after updating. Further studies are needed to understand this model's performance after updating and recalibrating on the Canadian context.

Conflict of interest

None of the authors had conflict of interest regarding this project.

Contributions

HG and SM conceived objective of the investigation, prepared data, and performed statistical analysis. BN and SM supervised all steps of the project. PA and AK participated in the conceptual development. All authors contributed to the conceptual development and writing the manuscript.

Tables and figures:

Table 7-1: Demographic and behavioral characteristics of the validation dataset

Predictors	Control (N=433)	Case (N=214)	Total (N=647)
Age			
Mean [SD]	60.2 [9.88]	59.2 [8.91]	59.8 [9.57]
Median [Q25, Q75]	61.0 [54.0, 67.0]	58.0 [53.0, 66.0]	59.0 [54.0, 67.0]
Min, Max	30.0, 79.0	32.0, 79.0	30.0, 79.0
Sex			
Female	129 (29.8%)	55 (25.7%)	184 (28.4%)
Male	304 (70.2%)	159 (74.3%)	463 (71.6%)
Race			
White	428 (98.8%)	213 (99.5%)	641 (99.1%)
Black	2 (0.5%)	1 (0.5%)	3 (0.5%)
Other	3 (0.7%)	0 (0%)	3 (0.5%)
Smoking (Packs-years)			
Never	114 (26.3%)	49 (22.9%)	163 (25.2%)
<= 20	130 (30.0%)	60 (28.0%)	190 (29.4%)
> 20	189 (43.6%)	105 (49.1%)	294 (45.4%)
Alcohol (Drinks per week)			
=<14 drinks	375 (86.6%)	189 (88.3%)	564 (87.2%)
>14 drinks	58 (13.4%)	25 (11.7%)	83 (12.8%)
nSexPart_cat			
0-1	32 (7.4%)	11 (5.1%)	43 (6.6%)
2-5	184 (42.5%)	56 (26.2%)	240 (37.1%)
6-10	88 (20.3%)	64 (29.9%)	152 (23.5%)
>10	129 (29.8%)	83 (38.8%)	212 (32.8%)
HPV status			
Negative	385 (88.9%)	80 (37.4%)	465 (71.9%)
Positive	32 (7.4%)	105 (49.1%)	137 (21.2%)
Missing	16 (3.7%)	29 (13.6%)	45 (7.0%)

Table 7-2: Model's predictive performance on the validation dataset

	Performance Metric	Score
Overall performance	Somers' D	0.49
	$Brier_{Scaled}$	68.84%
	Nagelkerke's R^2	0.24
Discrimination	C –Statistics	0.74 (0.69 – 0.79)
	Discrimination slope	0.18
Calibration	Calibration-in-the-large	4.93 (4.71–5.16)
	Calibration slope	0.57 (0.45–0.68)
	H-L test	Chi-square 29275.71, p= 0.00

Figure 7-1: Receiver operating characteristics (ROC) curve for the model in the validation dataset with 647 participants. The Area Under the ROC Curve (AUC) is 0.74 with a 95% confidence interval between 0.69 and 0.79.

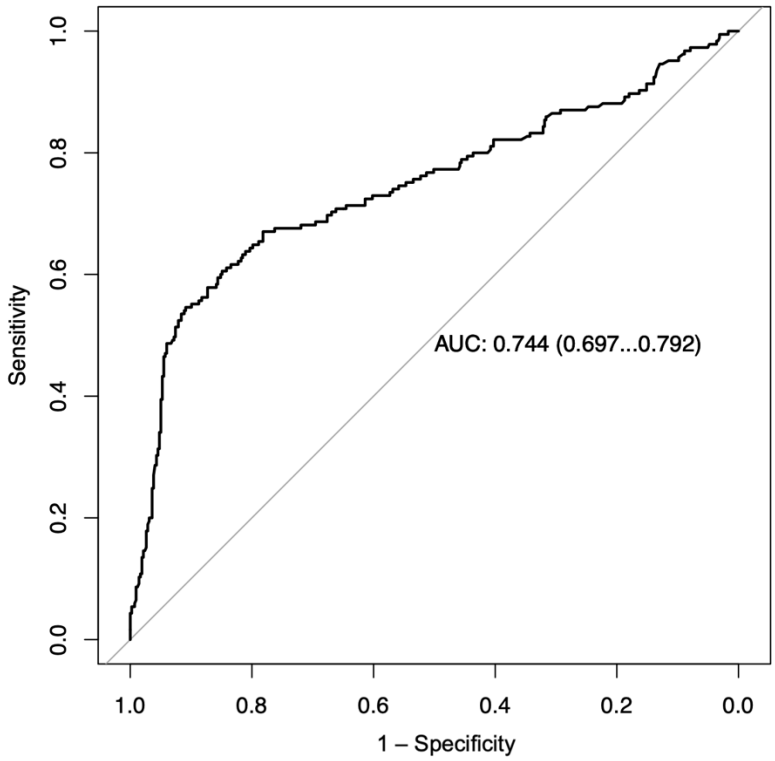


Figure 7-2: Box plot related to the model's predicted probabilities among cases and controls. The discrimination slope presents the difference between the mean predicted probability for the two groups.

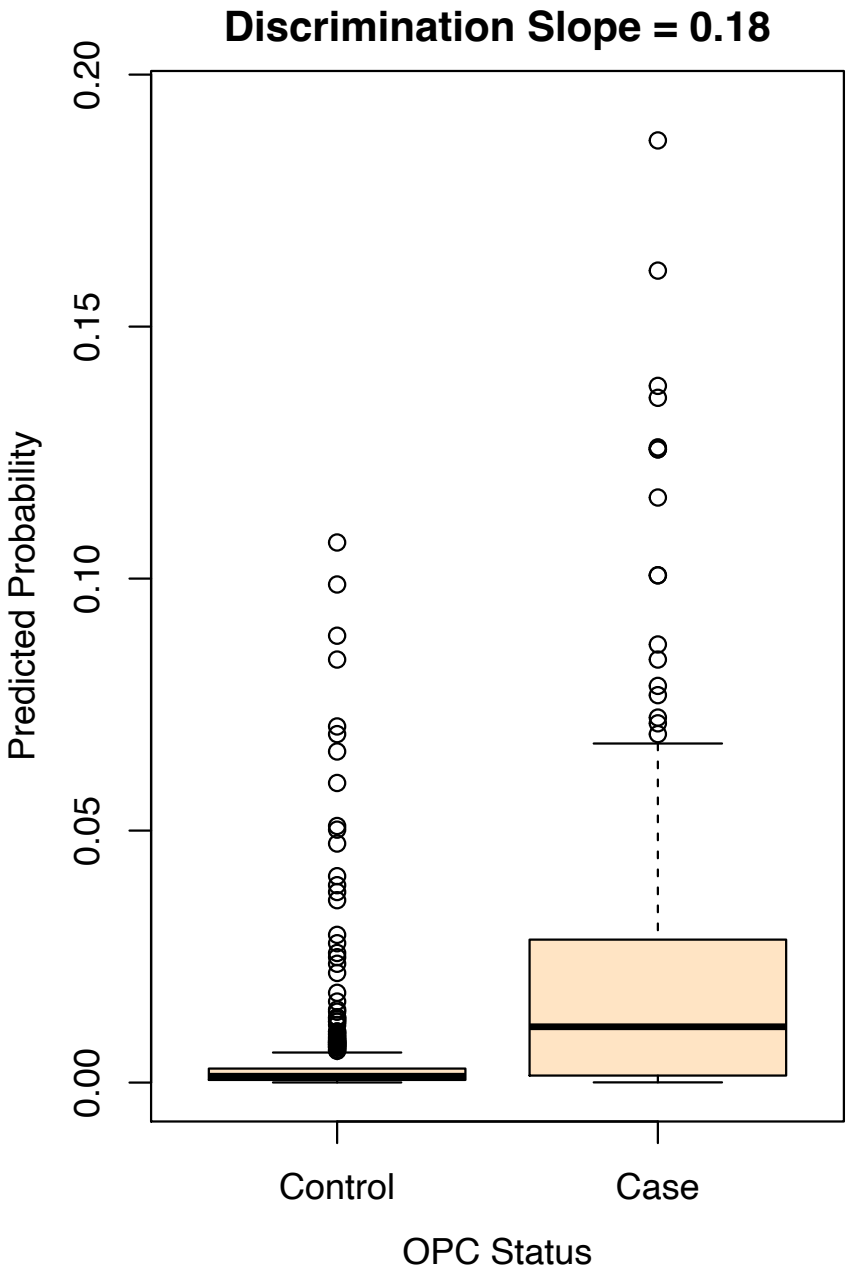
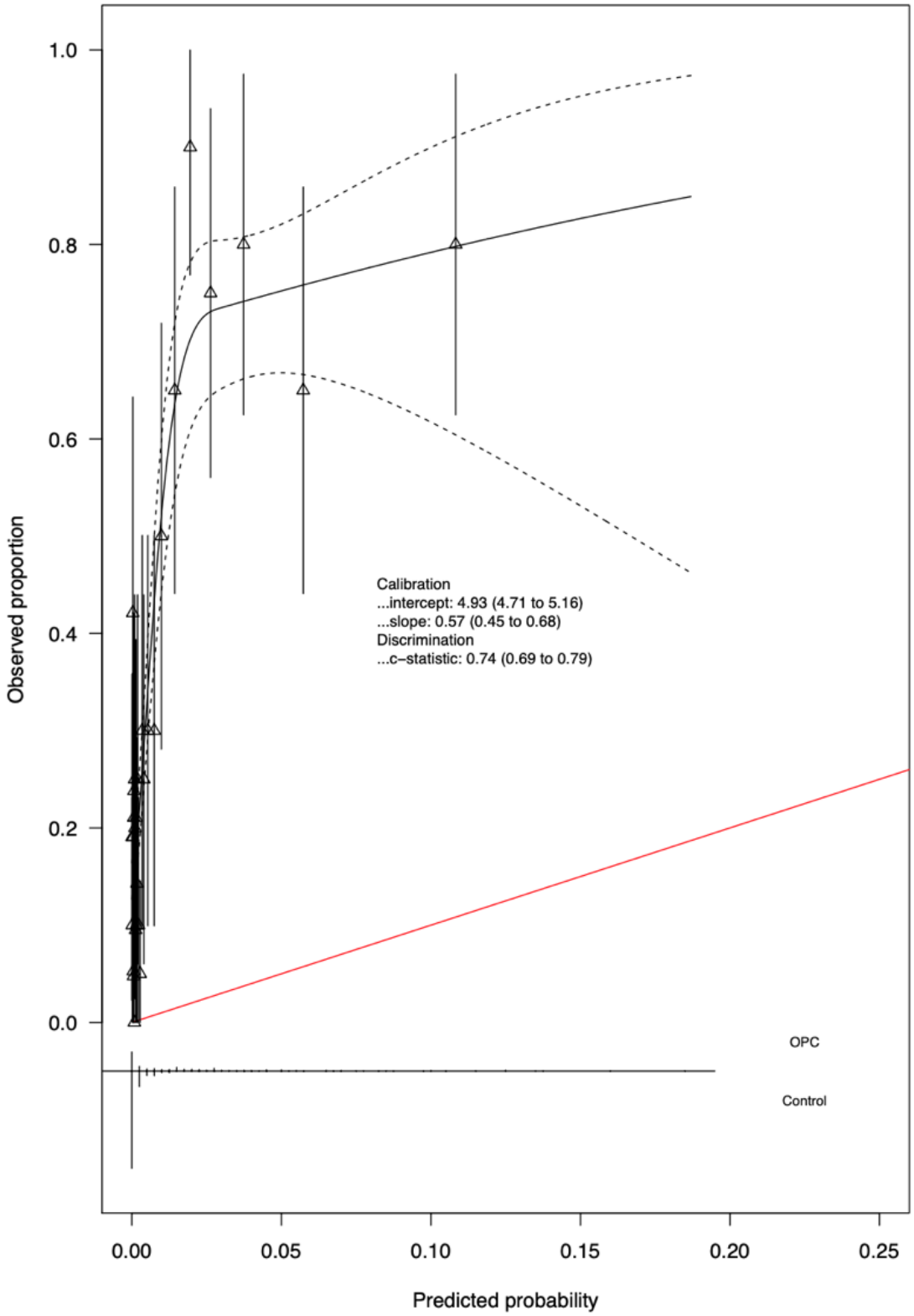


Figure 7-3: Calibration plot for the model over the validation data.



References:

1. Marcu LG, Yeoh E. A review of risk factors and genetic alterations in head and neck carcinogenesis and implications for current and future approaches to treatment. *Journal of cancer research and clinical oncology.* 2009;135(10):1303-1314.
2. Lee S. Cancer Statistics. Canadian Cancer Society. https://cancer.ca/en/research/cancer-statistics. Published 2022. Accessed 3/20/2022.
3. Fagan JJ, Noronha V, Graboyes EM. Making the best of limited resources: improving outcomes in head and neck cancer. *American Society of Clinical Oncology Educational Book.* 2021;41:279-289.
4. Lo Nigro C, Denaro N, Merlotti A, Merlano M. Head and neck cancer: improving outcomes with a multidisciplinary approach. *Cancer Manag Res.* 2017;9:363-371.
5. Control of oral cancer in developing countries. A WHO meeting. *Bull World Health Organ.* 1984;62(6):817-830.
6. Chaturvedi AK, Anderson WF, Lortet-Tieulent J, et al. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *Journal of clinical oncology.* 2013;31(36):4550.
7. Laprise C, Madathil SA, Schlecht NF, et al. Increased risk of oropharyngeal cancers mediated by oral human papillomavirus infection: Results from a Canadian study. *Head Neck.* 2019;41(3):678-685.
8. Laprise C, Madathil SA, Schlecht NF, et al. Human papillomavirus genotypes and risk of head and neck cancers: Results from the HeNCe Life case-control study. *Oral Oncol.* 2017;69:56-61.
9. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol.* 2011;29(32):4294-4301.
10. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol.* 2015;33(29):3235-3242.

11. Karolinska International human papillomavirus (HPV) reference center. https://www.hpvcenter.se/human_reference_clones/. Accessed 3/20/2022.
12. Tota JE, Gillison ML, Katki HA, et al. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer.* 2019;125(24):4407-4416.
13. Tumban E. A Current Update on Human Papillomavirus-Associated Head and Neck Cancers. *Viruses.* 2019;11(10).
14. Anderson KS, Dahlstrom KR, Cheng JN, et al. HPV16 antibodies as risk factors for oropharyngeal cancer and their association with tumor HPV and smoking status. *Oral oncology.* 2015;51(7):662-667.
15. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral oncology.* 2009;45(4-5):309-316.
16. Gormley M, Dudding T, Sanderson E, et al. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. *Nat Commun.* 2020;11(1):6071.
17. Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer research.* 1988;48(11):3282-3287.
18. Gillison ML. Current topics in the epidemiology of oral cavity and oropharyngeal cancers. *Head Neck.* 2007;29(8):779-792.
19. Survival Rates for Oral Cavity and Oropharyngeal Cancer. https://www.cancer.org/cancer/oral-cavity-and-oropharyngeal-cancer/detection-diagnosis-staging/survival-rates.html. Accessed 3/20/2022.
20. Damgacioglu H, Sonawane K, Zhu Y, et al. Oropharyngeal cancer incidence and mortality trends in all 50 states in the US, 2001-2017. *JAMA Otolaryngology–Head & Neck Surgery.* 2022;148(2):155-165.
21. Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol.* 2022;19(5):306-327.

22. Goy J, Hall SF, Feldman-Stewart D, Groome PA. Diagnostic delay and disease stage in head and neck cancer: a systematic review. *The Laryngoscope.* 2009;119(5):889-898.
23. Kowalski L, Franco EL, Torloni H, et al. Lateness of diagnosis of oral and oropharyngeal carcinoma: factors related to the tumour, the patient and health professionals. *European Journal of Cancer Part B: Oral Oncology.* 1994;30(3):167-173.
24. Teppo H, Alho OP. Relative importance of diagnostic delays in different head and neck cancers. *Clinical Otolaryngology.* 2008;33(4):325-330.
25. Carvalho AL, Pintos J, Schlecht NF, et al. Predictive factors for diagnosis of advanced-stage squamous cell carcinoma of the head and neck. *Archives of Otolaryngology–Head & Neck Surgery.* 2002;128(3):313-318.
26. Timbang MR, Sim MW, Bewley AF, Farwell DG, Mantravadi A, Moore MG. HPV-related oropharyngeal cancer: a review on burden of the disease and opportunities for prevention and early detection. *Human vaccines & immunotherapeutics.* 2019.
27. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. In: Second edition. ed. Cham, Switzerland: Springer; 2019: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2204948

https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5836937

https://doi.org/10.1007/978-3-030-16399-0

https://link.springer.com/book/10.1007%2F978-3-030-16399-0

http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9783030163990.

28. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. *Eur J Cardiothorac Surg.* 2018;54(2):203-208.
29. Control CfD, Prevention. NHANES questionnaires, datasets, and related documentation. *National Health and Nutrition Examination Survey.* 2018.
30. Gillison ML, D'Souza G, Westra W, et al. Distinct risk factor profiles for human papillomavirus type 16–positive and human papillomavirus type 16–negative head and neck cancers. *Journal of the National Cancer Institute.* 2008;100(6):407-420.

31. D'Souza G, Kreimer AR, Viscidi R, et al. Case–control study of human papillomavirus and oropharyngeal cancer. *New England Journal of Medicine.* 2007;356(19):1944-1956.
32. Madathil S, Rousseau MC, Durán D, Alli BY, Joseph L, Nicolau B. Life Course Tobacco Smoking and Risk of HPV-Negative Squamous Cell Carcinomas of Oral Cavity in Two Countries. *Front Oral Health.* 2022;3:844230.
33. Laprise C, Madathil SA, Allison P, et al. No role for human papillomavirus infection in oral cancers in a region in southern India. *Int J Cancer.* 2016;138(4):912-917.
34. Definition of squamous cell carcinoma - NCI Dictionary of Cancer Terms - NCI. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/squamous-cell-carcinoma. Published 2011. Updated 2011/02/02/07:00. Accessed 3/20/2022.
35. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med.* 2021;40(19):4230-4251.
36. Schlecht NF, Franco EL, Pintos J, Kowalski LP. Effect of smoking cessation and tobacco type on the risk of cancers of the upper aero-digestive tract in Brazil. *Epidemiology.* 1999;10(4):412-418.
37. Canada H. Low-risk alcohol drinking guidelines. https://www.canada.ca/en/health-canada/services/substance-use/alcohol/low-risk-alcohol-drinking-guidelines.html. Published 2021. Updated 2021/07/05/. Accessed 3/29/2022.
38. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-138.
39. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med.* 2014;33(3):517-535.
40. *rms: Regression Modeling Strategies* [computer program]. Version 6.3-02022.
41. *R: A Language and Environment for Statistical Computing* [computer program]. R Foundation for Statistical Computing, Vienna, Austria. 2020.
42. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.

43. Chow LQ. Head and neck cancer. *New England Journal of Medicine.* 2020;382(1):60-72.
44. Ghazawi FM, Lu J, Savin E, et al. Epidemiology and patient distribution of oral cavity and oropharyngeal SCC in Canada. *Journal of cutaneous medicine and surgery.* 2020;24(4):340-349.
45. Stein AP, Saha S, Kraninger JL, et al. Prevalence of Human Papillomavirus in Oropharyngeal Cancer: A Systematic Review. *Cancer J.* 2015;21(3):138-146.
46. 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.
47. Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects. *Diagn Progn Res.* 2017;1:12.
48. Chan SF, Deeks JJ, Macaskill P, Irwig L. Three methods to construct predictive models using logistic regression and likelihood ratios to facilitate adjustment for pretest probability give similar results. *J Clin Epidemiol.* 2008;61(1):52-63.
49. Morise AP, Diamond GA, Detrano R, Bobbio M, Gunel E. The effect of disease-prevalence adjustments on the accuracy of a logistic prediction model. *Med Decis Making.* 1996;16(2):133-142.
50. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol.* 1992;135(9):1029-1041.
51. Farsi N. *Epidemiology of Human Papilloma virus related head and neck cancers*, McGill University (Canada); 2015.
52. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *Bmj.* 2009;338:b375.
53. Moons KG, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol.* 2002;55(7):633-636.
54. Whiting PF, Rutjes AW, Westwood ME, Mallett S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;66(10):1093-1104.
55. Sedgwick P. What is recall bias? *BMJ.* 2012;344.

56. Raphael K. Recall bias: a proposal for assessment and control. *International journal of epidemiology.* 1987;16(2):167-170.
57. Arekunnath Madathil S. *Life course behavioural risk factors of head and neck cancers: A Bayesian approach*, McGill University; 2018.
58. Berney L, Blane D. The lifegrid method of collecting retrospective information from people at older ages. *Research, Policy and Planning.* 2003;21(2):13-22.
59. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of clinical epidemiology.* 2016;74:167-176.
60. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Statistics in medicine.* 2016;35(2):214-226.
61. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of clinical epidemiology.* 2005;58(5):475-483.

8 Discussion

The following section provides an overview of the objectives, rationale, and results of this thesis, followed by the discussion on the strengths and limitations, as well as future research directions, and ends with the implications of the results.

8.1 Overview of rationale and results

Approximately 60% to 70% of HNC patients present at the advanced stages of this disease[29,152]; when medical treatment is less efficient, surgical treatments may cause considerable damage to speech and swallowing-related organs, and patients will more likely experience poor after-treatment quality of life[153-155]. Despite the advances in HNC treatments[156,157], its five-year survival has not considerably increased over the past decades[2,152,157-159]. Several factors contribute to the low survival rate and poor prognosis of HNC, including but not limited to the patients' risk-associated behaviors such as tobacco smoking, alcohol consumption[160,161],and HPV infection[18,37,46,47,119,160,162,163].

Risk assessment and preventive motivation counseling efforts for controlling the high-risk behaviors are expected to improve HNC incidence, cost burden, and mortality[152]. In both aspects, risk prediction models play a significant role in helping personal preventive interventions by identifying high-risk individuals in the clinical, public health, or epidemiological research setting.

Even though numerous HNC risk prediction models have been developed worldwide, the literature lacks on studies synthesizing this information and discussing the existing models' strengths, limitations, risk of bias, and applicability concerns. The first part of this thesis project was designed and conducted to address this knowledge gap. A clear perspective of the current status in HNC risk prediction modeling could help researchers identify the existing models and their characteristics for use in future model developments. The results of the *Manuscript I* revealed that most modeling studies have specific limitations regarding the data used to develop or validate the model and the comprehensive reporting of the modeling stages, making it hard

to decide on the model's quality. The quality appraisal based on PROBAST[90,98] also indicated a high risk of bias in the analysis domain of most developed models. Even though guidelines for developing and reporting a good risk prediction model have already been published[9,10,89,104], there seems to be a need to emphasize these guidelines and the necessity of the transparent reporting of the HNC risk prediction modeling studies based on the standards such as the TRIPOD checklist[164].

Considering the recent rise in HPV-related HNC incidence in Canada[116,119,165], it is of utmost importance to advocate HNC preventive interventions in this country Therefore, a risk prediction model that contains HNC main risk factors such as smoking, alcohol, number of sex partners, and HPV status is needed in Canada. The *manuscript I* revealed that no model had been developed or validated using data from Canada. Therefore, the *Manuscript II* followed the recommended methods[9,10,89,104] for modeling (Figure 3-1) and investigated the possibility of generalizing one of the identified models[17] that comprised the main risk factors for a subsite of HNC, oropharyngeal cancer. We validated the model without altering its characteristics (e.g., updating the values of intercept and coefficients or the set of predictors). This method is recommended[9] because it helps understand the model's net generalizability in a new setting and identify the parts of the model that need improvement in future studies. The validated model described in *Manuscript II* presented an above-average overall performance and discrimination ability on the validation dataset, indicating that it is potentially a good model that could be generalizable to the validation dataset's population of interest (Montreal residents). This was expected because, firstly, the model's development study reported its excellent discrimination ability during the internal and external validation of the US population (AUC in internal validation: 0.94 [0.92–0.97] and external validation 0.87 [0.84–0.90]). Secondly, the epidemiology of HPV-related HNC in Canada and the US is similar[4,29]. However, the model still needs improvement in performance before implementation in practice. The model's mis-calibration (calibration-in-the-large: 4.93 [4.71–5.16] & calibration slope: 0.57 [0.45–0.68]) also indicated that there is still a need for updating this model before use in the Canadian context.

8.2 Strengths and limitations

Despite the title of "scoping review," which conveys a comprehensive review of peer-reviewed and pre-print publications, the scoping review part of this thesis project used a restrict inclusion criteria, resulting in the exclusion of pre-print HNC modeling studies. However, our primary goal was to extract detailed information on the existing HNC prognostic risk prediction models and assess their risk of bias and clinical applicability. The detailed information could help us understand where we stand in the HNC risk prediction modeling field and what we should do to develop or validate a good risk prediction model for the Canadian population. Including pre-print papers may have captured poorly developed or reported models into the scoping review, making it hard to appraise the quality of all developed models comprehensively.

The scoping review also excluded the diagnostic risk prediction models. Apart from the fact that discussing diagnostic risk prediction models was out of our scoping review context, their inclusion could have made it hard to design and organize the report of all models' characteristics (e.g., table of predictors). Therefore, we organized the search strategy based on capturing only the peer-reviewed published HNC risk prediction models.

The *Manuscript I* identified some studies that have developed a separate model for each subsite of HNC. This is preferable[14] as different subsites of HNC have different weights of risk factors. For example, HPV infection is considered the main risk factor for oropharyngeal cancer[115,116], while smoking and alcohol consumptions have more weight in developing oral cancer [142,166]. Therefore, a single model for all subsites of HNC might underestimate this variation and provide biased predictions. Considering the modest improvement in the 5-year survival of HNC in Canada, it is of utmost importance to have a good risk prediction model for every subsite of the HNC in Canada. The validated model in the *Manuscript II* covers only one subsite of HNC (oropharyngeal cancer) in Canada. Therefore, there is still a need for a separate model for other subsites of HNC developed and validated in the Canadian context. Similar to the method we implemented in *Manuscript II*, future studies may validate the models potentially applicable in the Canadian context. Our scoping review (*Manuscript I*) identified such models. Based on its results, Lee et al.

[14] have developed separate models for each subsite of HNC using data from the US. Therefore, these models are expected to be potentially generalizable to the Canadian population. This thesis project did not validate Lee et al.'s models because the development study[14] published insufficient information regarding the models' detailed characteristics (e.g., values of intercepts and coefficients), making it difficult to replicate them. Future studies may communicate with the researchers who developed these models and validate them on the data from Canada before developing new models for the Canadian context (Figure 3-1).

The key element in the risk prediction model development or validation is having representative data on the risk factors of a disease in a population of interest[9,89,90]. These data preferably should come from a prospective longitudinal study with prespecified and consistent inclusion and exclusion criteria[90]. The data also should contain comprehensive information on the behavioral factors. In the case of oropharyngeal cancer in Canada, for example, the data should preferably contain lifetime exposure to tobacco smoking, alcohol consumption, and HPV infection. Although some studies have investigated the HNC and its risk factors in different provinces in Canada[120,167,168], their data rarely contain comprehensive information on the lifetime exposure to the HNC risk factors. This thesis project benefited from a solid dataset derived from the HeNCe Life study, which captured the lifetime exposures to the risk factors of HNC through the consistent and supervised method with specific inclusion and exclusion criteria. However, using this dataset for external validation of a risk prediction model have a specific limitation; that is, the subjects were not the random sample of the Canadian population nor the population living in the Quebec province. While several procedures were in place to assure the internal validity of HeNCe Life study (e.g., control selected from several outpatient clinics with diseases unrelated to smoking and alcohol, cases and controls had to live 50km of the hospital area), the external validity of a hospital-based case control design is questionable. This is a common limitation of all modeling study using data from a case-control study that is designed and conducted to answer a different research question[9,90]. As an alternative option, we could design and conduct a prospective longitudinal study. However, considering that the HNC is a relatively rare disease, a prospective study needed to last a long time to capture enough samples for risk prediction

modeling; thus, it would be inefficient. Using data from the Canadian Cancer Registry (CCR)[169] could have been an alternative option. Such data would have the advantage of containing HNC cases on the scale of the Canadian population level, making the model's predictions more representative than when using HeNCe Canada's data. However, using data from CCR would put our external validation project in considerable challenges. The CCR's data are not collected based on a consistent method needed for a good modeling study[90]. It also does not capture data on lifetime exposure to the risk factors of HNC, making it hard to replicate the model. Considering this limitation, there should be an emphasis on capturing comprehensive information on the life course exposure to the HNC risk factors during the data collection for CCR.

8.3 Future research directions

This thesis project provided an overview of what is there in the literature on HNC risk prediction modeling and the possibility of generalizing one of the existing models in the Canadian context.

Based on the results of *Manuscript I*, most existing modeling studies have not fully reported the modeling stages and the model's characteristics, making it challenging to reproduce and generalize the existing models. Therefore, this thesis project emphasizes the importance of transparent reporting of the modeling stages, especially the analysis part.

The second part of this thesis project (*Manuscript II*) also presented the possibility of implementing a potential model for oropharyngeal cancer in the Canadian context. The validated model displayed an above-average overall performance and discrimination ability but poor calibration, displaying a need for updating this model before application in practice. Future studies are needed to improve this model and assess its performance after updating. Ewout W Steyerberg[9] suggested a step-by-step procedure for updating and improving a prediction model's performance. Based on his suggestions, a model's performance would considerably improve after adjusting for calibration-in-the-large. Therefore, the first step in future updating of the validated model in this thesis project could be to recalibrate it for calibration-in-the-large derived from *Manuscript II*. Should the model's performance improvement be insufficient to rely on its

predictions, updating the intercept value (adjusting for the baseline risk) will be the next step. It is worth highlighting that updating this model by adjusting for the baseline risk of oropharyngeal cancer among Montreal residents poses a particular challenge because the incidence of these cancers is not being reported separately in the Quebec province[4]. Canadian Cancer Society only releases the overall incidence of HNC (combination of all subsites), making it difficult to calculate the baseline risk of each subsite of HNC in different parts of Canada. It is suggested that the Canadian cancer society releases detailed epidemiological data on HNC and its subsites to help future researchers easily access to the data and develop collaborations in risk prediction modeling.

It is of note that the model validated in the second manuscript predicts the risk of one of the subsites of HNC. The Canadian context still needs a risk prediction model for other HNC subsites. Future studies are suggested to benefit from the methods followed in the *Manuscript II* and validate the potential models in the Canadian context.

8.4 Implications of results

The perspective of the current status in HNC risk prediction modeling provided in this thesis project may help researchers and public health providers in different settings. For example, future studies' investigators could refer to *Manuscript I* and use its results to develop or validate efficient models. Our scoping review also identified no model developed or validated for most countries worldwide. Future studies may build up on *Manuscript I* results and develop or validate proper risk prediction models in different countries. Regarding the public health implications, the global health policymakers could benefit from the results of our scoping review and advocate and support the development or validation of good risk prediction models for HNC in different parts of the world.

The validated model in the *Manuscript II* also can be implemented in different settings. In epidemiological research, this model may help in recruiting the participants for the clinical trials of the new HPV infection preventive interventions. Also, risk prediction model developers could

use this model in determining the minimum sample size needed for prospective or retrospective studies related to oropharyngeal cancer in Canada. With a constant rise in the HPV-related HNC incidence in Canada[120], Canadian public health policymakers can use this model to identify high-risk populations for targeted HPV preventive interventions to reduce the incidence of HPV-related HNC in Canada. In the clinical setting in Canada, this model could be a helpful tool for the clinicians to assess the risk of developing HNC in the future year of an individual, helping in the personalized preventive interventions such as motivation counseling.

Suffice it to say all abovementioned implications could only be possible after the appropriate updating of this model in the Canadian context.

8.5 Knowledge translation

As part of knowledge translation activities, we aim to present the results of both manuscripts using different platforms.

The scoping review project will be published as a journal article helping researchers in future HNC risk prediction modeling. Presenting it at different conferences also may help in publishing our achievements. We have already presented a part of this research at the 2022 AADOCR/CADR Annual Meeting & Exhibition[170] and at several local conferences including McGill dentistry research day and Réseau de Recherche en Santé Buccodentaire et Osseuse (RSBO)[171] conferences.

We also aim to publish the second manuscript's result in a journal. The validated model could also be presented on different platforms. The regression formula related to the model will help researchers in updating this model or using it in different epidemiological investigations related to OPC. The validated model could also play as a base for paper-based risk assessment tools such as score charts or nomograms, helping identify individuals with a high risk of developing OPC. The model also could be embedded in the web-based or mobile app, helping clinicians in critical

decision-making. The predictions of this model could also be inserted into the electronic patient records, assisting future epidemiological research, or supporting clinical decision-making.

9 Conclusion

- Most of the existing HNC risk prediction models have non-ignorable levels risk of bias, especially in the analysis part, making it hard to replicate the models.

- There is a need to follow reporting standards and model development guidelines in HNC risk prediction modeling to ensure transparent modeling.

- The oropharyngeal cancer risk prediction model developed and validated by Tota et al.[17] needs recalibration and improvements before implementing in practice in Canada.

- Future studies are needed to investigate this model's performance after updating in the Canadian context.

10 References

1. Global Cancer Observatory. https://gco.iarc.fr/. Accessed 7/7/2022.
2. Tiwana MS, Wu J, Hay J, Wong F, Cheung W, Olson RA. 25 year survival outcomes for squamous cell carcinomas of the head and neck: population-based outcomes from a Canadian province. *Oral Oncol.* 2014;50(7):651-656.
3. Marcu LG, Yeoh E. A review of risk factors and genetic alterations in head and neck carcinogenesis and implications for current and future approaches to treatment. *J Cancer Res Clin Oncol.* 2009;135(10):1303-1314.
4. Lee S. Cancer Statistics. Canadian Cancer Society. https://cancer.ca/en/research/cancer-statistics. Published 2022. Accessed 3/20/2022.
5. Osazuwa-Peters N, Barnes JM, Okafor SI, et al. Incidence and risk of suicide among patients with head and neck cancer in rural, urban, and metropolitan areas. *JAMA Otolaryngology–Head & Neck Surgery.* 2021;147(12):1045-1052.
6. Osazuwa-Peters N, Arnold LD, Loux TM, Varvares MA, Schootman M. Factors associated with increased risk of suicide among survivors of head and neck cancer: A population-based analysis. *Oral Oncology.* 2018;81:29-34.
7. 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 journal of oncology.* 2018;9(3):80.
8. Tumban E. A Current Update on Human Papillomavirus-Associated Head and Neck Cancers. *Viruses.* 2019;11(10).
9. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. In: Second edition. ed. Cham, Switzerland: Springer; 2019: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2204948

https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5836937

https://doi.org/10.1007/978-3-030-16399-0

https://link.springer.com/book/10.1007%2F978-3-030-16399-0

http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9783030163990

https://link.springer.com/book/10.1007/978-3-030-16399-0.

10. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLoS Medicine.* 2012;9(5). doi:10.1371/journal.pmed.1001221.
11. Johnson-Obaseki S, McDonald JT, Corsten M, Rourke R. Head and neck cancer in Canada: trends 1992 to 2007. *Otolaryngology--Head and Neck Surgery.* 2012;147(1):74-78.
12. Goyette A, Yen GP, Racovitan V, Bhangu P, Kothari S, Franco EL. Evolution of Public Health Human Papillomavirus Immunization Programs in Canada. *Current Oncology.* 2021;28(1):991-1007.
13. Canada Leads the World in Educating and Raising Awareness of HPV. https://fmwc.ca/canada-leads-the-world-in-educating-and-raising-awareness-of-hpv/. Accessed 7/4/2022.
14. 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.
15. 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. *European Journal of Cancer Prevention.* 2017;26(1):38.
16. Tikka T, Pracy P, Paleri V. Refining the head and neck cancer referral guidelines: a two centre analysis of 4715 referrals. *British Journal of Oral and Maxillofacial Surgery.* 2016;54(2):141-150.
17. Tota JE, Gillison ML, Katki HA, et al. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer.* 2019;125(24):4407-4416.
18. Chow LQ. Head and neck cancer. *New England Journal of Medicine.* 2020;382(1):60-72.
19. Döbróssy L. Epidemiology of head and neck cancer: Magnitude of the problem. *Cancer and Metastasis Reviews.* 2005;24(1):9-17.
20. Head and Neck Cancers - NCI. https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet. Published 2021. Updated 2021/06/09/08:00. Accessed 5/29/2022.

21. Son E, Panwar A, Mosher CH, Lydiatt D. Cancers of the Major Salivary Gland. *J Oncol Pract.* 2018;14(2):99-108.
22. The Burden of Cancer. *The Cancer Atlas.*
23. Stenson KM, Brockstein BE, Ross ME. Epidemiology and risk factors for head and neck cancer. *UpToDate.* 2014.
24. Jethwa AR, Khariwala SS. Tobacco-related carcinogenesis in head and neck cancer. *Cancer Metastasis Rev.* 2017;36(3):411-423.
25. Anderson KS, Dahlstrom KR, Cheng JN, et al. HPV16 antibodies as risk factors for oropharyngeal cancer and their association with tumor HPV and smoking status. *Oral Oncol.* 2015;51(7):662-667.
26. Majchrzak E, Szybiak B, Wegner A, et al. Oral cavity and oropharyngeal squamous cell carcinoma in young adults: a review of the literature. *Radiol Oncol.* 2014;48(1):1-10.
27. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol.* 2011;29(32):4294-4301.
28. Ryser MD, Rositch A, Gravitt PE. Modeling of US Human Papillomavirus (HPV) Seroprevalence by Age and Sexual Behavior Indicates an Increasing Trend of HPV Infection Following the Sexual Revolution. *J Infect Dis.* 2017;216(5):604-611.
29. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7-33.
30. Gatta G, Botta L, Sánchez MJ, Anderson LA, Pierannunzio D, Licitra L. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. *Eur J Cancer.* 2015;51(15):2130-2143.
31. Lambert R, Sauvaget C, de Camargo Cancela M, Sankaranarayanan R. Epidemiology of cancer from the oral cavity and oropharynx. *Eur J Gastroenterol Hepatol.* 2011;23(8):633-641.
32. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer.* 2013;132(5):1133-1145.

33. Gupta B, Johnson NW, Kumar N. Global epidemiology of head and neck cancers: a continuing challenge. *Oncology.* 2016;91(1):13-23.
34. DeSantis C, Naishadham D, Jemal A. Cancer statistics for African Americans, 2013. *CA Cancer J Clin.* 2013;63(3):151-166.
35. Settle K, Posner MR, Schumaker LM, et al. Racial survival disparity in head and neck cancer results from low prevalence of human papillomavirus infection in black oropharyngeal cancer patients. *Cancer Prev Res (Phila).* 2009;2(9):776-781.
36. Jiron J, Sethi S, Ali-Fehmi R, et al. Racial disparities in Human Papillomavirus (HPV) associated head and neck cancer. *Am J Otolaryngol.* 2014;35(2):147-153.
37. Wyss A, Hashibe M, Chuang SC, et al. Cigarette, cigar, and pipe smoking and the risk of head and neck cancers: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Am J Epidemiol.* 2013;178(5):679-690.
38. Tobacco smoking. *IARC Monogr Eval Carcinog Risk Chem Hum.* 1986;38:35-394.
39. Gandini S, Botteri E, Iodice S, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer.* 2008;122(1):155-164.
40. Humans IWGotEoCRt. Tobacco smoking. In: *Personal Habits and Indoor Combustions.* International Agency for Research on Cancer; 2012.
41. Troy JD, Grandis JR, Youk AO, Diergaarde B, Romkes M, Weissfeld JL. Childhood passive smoke exposure is associated with adult head and neck cancer. *Cancer Epidemiol.* 2013;37(4):417-423.
42. Reitsma MB, Kendrick PJ, Ababneh E, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *The Lancet.* 2021;397(10292):2337-2360.
43. Smoking Rates by Country 2022. https://worldpopulationreview.com/country-rankings/smoking-rates-by-country. Accessed 6/16/2022.
44. Petti S. Lifestyle risk factors for oral cancer. *Oral Oncol.* 2009;45(4-5):340-350.
45. Humans IWGotEoCRt. Alcohol consumption and ethyl carbamate. *IARC monographs on the evaluation of carcinogenic risks to humans.* 2010;96:3-1383.

46. Di Credico G, Polesel J, Dal Maso L, et al. Alcohol drinking and head and neck cancer risk: The joint effect of intensity and duration. *British journal of cancer.* 2020;123(9):1456-1463.

47. Hashibe M, Brennan P, Chuang 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. Karolinska International human papillomavirus (HPV) reference center. https://www.hpvcenter.se/human_reference_clones/. Accessed 3/20/2022.

49. Wood ZC, Bain CJ, Smith DD, Whiteman DC, Antonsson A. Oral human papillomavirus infection incidence and clearance: a systematic review of the literature. *J Gen Virol.* 2017;98(4):519-526.

50. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer.* 2017;141(4):664-670.

51. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol.* 2015;33(29):3235-3242.

52. Chaturvedi AK, Graubard BI, Broutian T, et al. NHANES 2009-2012 Findings: Association of Sexual Behaviors with Higher Prevalence of Oral Oncogenic Human Papillomavirus Infections in U.S. Men. *Cancer Res.* 2015;75(12):2468-2477.

53. Ndiaye C, Mena M, Alemany L, et al. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. *Lancet Oncol.* 2014;15(12):1319-1331.

54. Combes JD, Franceschi S. Role of human papillomavirus in non-oropharyngeal head and neck cancers. *Oral Oncol.* 2014;50(5):370-379.

55. Cook RL, Thompson EL, Kelso NE, et al. Sexual behaviors and other risk factors for oral human papillomavirus infections in young women. *Sex Transm Dis.* 2014;41(8):486-492.

56. Adeyemi BF, Olusanya AA, Lawoyin JO. Oral squamous cell carcinoma, socioeconomic status and history of exposure to alcohol and tobacco. *J Natl Med Assoc.* 2011;103(6):498-502.
57. Johnson S, McDonald JT, Corsten M, Rourke R. Socio-economic status and head and neck cancer incidence in Canada: a case-control study. *Oral Oncol.* 2010;46(3):200-203.
58. Purakkal AST. *Association between Life course Socioeconomic position and Oral cancer among a sample of Indian subjects.* McGill University (Canada); 2012.
59. Freedman ND, Park Y, Subar AF, et al. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. *Int J Cancer.* 2008;122(10):2330-2336.
60. Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. Dietary factors and oral and pharyngeal cancer risk. *Oral Oncol.* 2009;45(6):461-467.
61. Nocini R, Sanchis-Gomar F, Lippi G. Physical activity and laryngeal cancer. *Annals of translational medicine.* 2019;7(23).
62. Tezal M, Sullivan MA, Reid ME, et al. Chronic periodontitis and the risk of tongue cancer. *Arch Otolaryngol Head Neck Surg.* 2007;133(5):450-454.
63. Laprise C, Shahul HP, Madathil SA, et al. Periodontal diseases and risk of oral cancer in Southern India: Results from the HeNCe Life study. *Int J Cancer.* 2016;139(7):1512-1519.
64. Balaram P, Sridhar H, Rajkumar T, et al. Oral cancer in southern India: the influence of smoking, drinking, paan-chewing and oral hygiene. *Int J Cancer.* 2002;98(3):440-445.
65. Franco EL, Kowalski LP, Oliveira BV, et al. Risk factors for oral cancer in Brazil: a case-control study. *Int J Cancer.* 1989;43(6):992-1000.
66. Guha N, Boffetta P, Wünsch Filho V, et al. Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies. *Am J Epidemiol.* 2007;166(10):1159-1173.
67. Alhazmi A, Alhazmi Y, Makrami A, et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med.* 2021;50(5):444-450.
68. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. *Eur J Cardiothorac Surg.* 2018;54(2):203-208.

69. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost.* 2013;11 Suppl 1:129-141.
70. Mascarella MA, Muthukrishnan N, Maleki F, et al. Above and Beyond Age: Prediction of Major Postoperative Adverse Events in Head and Neck Surgery. *Ann Otol Rhinol Laryngol.* 2021:34894211041222.
71. Kager L, Ljungdahl I, Malmborg AS, Nord CE, Pieper R, Dahlgren P. Antibiotic prophylaxis with cefoxitin in colorectal surgery: effect on the colon microflora and septic complications--a clinical model for prediction of the benefit and risks in using a new antibiotic in prophylaxis. *Ann Surg.* 1981;193(3):277-282.
72. Abernathy JR, Graves RC, Bohannan HM, Stamm JW, Greenberg BG, Disney JA. Development and application of a prediction model for dental caries. *Community Dent Oral Epidemiol.* 1987;15(1):24-28.
73. Bohannan HM, Disney JA, Graves RC, Anderson PE, Klein SP. Operation of the National Preventive Dentistry Demonstration Program. *J Public Health Dent.* 1985;45(2):75-82.
74. Cole RR, Robbins KT, Cohen JI, Wolf PF. A predictive model for wound sepsis in oncologic surgery of the head and neck. *Otolaryngol Head Neck Surg.* 1987;96(2):165-171.
75. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S. Prediction of survival in patients with head and neck cancer. *Head Neck.* 2001;23(9):718-724.
76. Russo D, Mariani P, Caponio VCA, et al. Development and Validation of Prognostic Models for Oral Squamous Cell Carcinoma: A Systematic Review and Appraisal of the Literature. *Cancers (Basel).* 2021;13(22).
77. Mascarella MA, Muthukrishnan N, Maleki F, et al. Above and Beyond Age: Prediction of Major Postoperative Adverse Events in Head and Neck Surgery. *Ann Otol Rhinol Laryngol.* 2022;131(7):697-703.
78. Weed HG, Lutman CV, Young DC, Schuller DE. Preoperative identification of patients at risk for delirium after major head and neck cancer surgery. *The Laryngoscope.* 1995;105(10):1066-1068.

79. Kuo P-J, Wu S-C, Chien P-C, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. *Oncotarget.* 2018;9(17):13768.

80. Hoesseini A, van Leeuwen N, Offerman MPJ, et al. Predicting survival in head and neck cancer: External validation and update of the prognostic model OncologIQ in 2189 patients. *Head Neck.* 2021;43(8):2445-2456.

81. Huang Z-D, Yao Y-Y, Chen T-Y, Zhao Y-F, Zhang C, Niu Y-M. Construction of Prognostic Risk Prediction Model of Oral Squamous Cell Carcinoma Based on Nine Survival-Associated Metabolic Genes. *Frontiers in physiology.* 2021;12:609770.

82. Cheung LC, Ramadas K, Muwonge R, et al. Risk-Based Selection of Individuals for Oral Cancer Screening. *Journal of Clinical Oncology.* 2021;39(6).

83. Gupta B, Kumar N, Johnson NW. A risk factor-based model for upper aerodigestive tract cancers in India: predicting and validating the receiver operating characteristic curve. *Journal of Oral Pathology & Medicine.* 2017;46(6):465-471.

84. Hung L-C, Kung P-T, Lung C-H, et al. Assessment of the risk of oral cancer incidence in a high-risk population and establishment of a predictive model for oral cancer incidence using a population-based cohort in Taiwan. *International journal of environmental research and public health.* 2020;17(2):665.

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

86. 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. *International Journal of Oncology.* 2020;57(5):1192-1202.

87. Rosmai MD, Sameemii AK, Basir A, Mazlipahiv IS, Norzaidi M. The use of artificial intelligence to identify people at risk of oral cancer: empirical evidence in Malaysian University. *International Journal of Scientific Research in Education.* 2010;3(1):10-20.

88. Tikka T, Paleri V, MacKenzie K. External validation of a cancer risk prediction model for suspected head and neck cancer referrals. *Clinical Otolaryngology.* 2018;43(2):714-717.

89. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal.* 2014;35(29):1925-1931.
90. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med.* 2019;170(1):W1-w33.
91. Altman DG, Bland JM. Missing data. *Bmj.* 2007;334(7590):424.
92. Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. *American Journal of Neuroradiology.* 2011;32(3):437-440.
93. Altman DG, Royston P. The cost of dichotomising continuous variables. *Bmj.* 2006;332(7549):1080.
94. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *International journal of epidemiology.* 1999;28(5):964-974.
95. Baneshi MR, Talei A. Dichotomisation of continuous data: review of methods, advantages, and disadvantages. 2011.
96. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst.* 1994;86(11):829-835.
97. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127-141.
98. Wolff RF, Moons KGM, Riley RD, et al. PROBAST : A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine.* 2019;170(1).
99. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Family medicine and community health.* 2020;8(1).
100. Ratner B. Variable selection methods in regression: Ignorable problem, outing notable solution. *Journal of Targeting, Measurement and Analysis for Marketing.* 2010;18(1):65-75.
101. Boffetta P, Hashibe M. Alcohol and cancer. *The lancet oncology.* 2006;7(2):149-156.

102. Yoshida A, Huang IY, Ikawa M. Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proc Natl Acad Sci U S A.* 1984;81(1):258-261.

103. Li Y, Zhang D, Jin W, et al. Mitochondrial aldehyde dehydrogenase-2 (ALDH2) Glu504Lys polymorphism contributes to the variation in efficacy of sublingual nitroglycerin. *J Clin Invest.* 2006;116(2):506-511.

104. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-138.

105. Somers RH. A new asymmetric measure of association for ordinal variables. *American sociological review.* 1962:799-811.

106. Harrell FE. Regression modeling strategies : with applications to linear models, logistic and ordinal regression, and survival analysis. In: Second edition. ed. Cham: Springer; 2015: http://books.scholarsportal.info/viewdoc.html?id=/ebooks/ebooks3/springer/2017-08-17/9/9783319194257

https://doi.org/10.1007/978-3-319-19425-7

http://link.springer.com/openurl?genre=book&isbn=978-3-319-19424-0

http://link.springer.com/10.1007/978-3-319-19425-7

https://link.springer.com/openurl?genre=book&isbn=978-3-319-19424-0

https://0-link-springer-com.pugwash.lib.warwick.ac.uk/book/10.1007/978-3-319-19425-7

http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9783319194257

http://www.vlebooks.com/vleweb/product/openreader?id=Exeter&isbn=9783319194257

http://public.eblib.com/choice/PublicFullRecord.aspx?p=6315870.

107. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly weather review.* 1950;78(1):1-3.

108. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565-574.

109. Yates JF. External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance.* 1982;30(1):132-156.

110. Hilden J, Habbema JD, Bjerregaard B. The measurement of performance in probabilistic diagnosis. II. Trustworthiness of the exact values of the diagnostic probabilities. *Methods Inf Med.* 1978;17(4):227-237.
111. Efron B, Tibshirani R. *An introduction to the bootstrap.* New York: Chapman & Hall; 1994.
112. McIntosh A. The Jackknife estimation method. *arXiv preprint arXiv:160600497.* 2016.
113. Gillison ML, D'Souza G, Westra W, et al. Distinct risk factor profiles for human papillomavirus type 16–positive and human papillomavirus type 16–negative head and neck cancers. *Journal of the National Cancer Institute.* 2008;100(6):407-420.
114. D'Souza G, Kreimer AR, Viscidi R, et al. Case–control study of human papillomavirus and oropharyngeal cancer. *New England Journal of Medicine.* 2007;356(19):1944-1956.
115. Laprise C, Madathil SA, Schlecht NF, et al. Increased risk of oropharyngeal cancers mediated by oral human papillomavirus infection: Results from a Canadian study. *Head Neck.* 2019;41(3):678-685.
116. Laprise C, Madathil SA, Schlecht NF, et al. Human papillomavirus genotypes and risk of head and neck cancers: Results from the HeNCe Life case-control study. *Oral oncology.* 2017;69:56-61.
117. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *Jama.* 2007;298(10):1209-1212.
118. Hung LC, Kung PT, Lung CH, et al. Assessment of the Risk of Oral Cancer Incidence in A High-Risk Population and Establishment of A Predictive Model for Oral Cancer Incidence Using A Population-Based Cohort in Taiwan. *Int J Environ Res Public Health.* 2020;17(2).
119. Corsten M, Rourke R, Johnson SE, McDonald T. HPV and Head and Neck Cancer in Canada: Trends 1992 to 2008. *Otolaryngology—Head and Neck Surgery.* 2011;145(2_suppl):P66-P67.
120. Habbous S, Chu KP, Lau H, et al. Human papillomavirus in oropharyngeal cancer in Canada: analysis of 5 comprehensive cancer centres using multiple imputation. *Cmaj.* 2017;189(32):E1030-e1040.

121. Timbang MR, Sim MW, Bewley AF, Farwell DG, Mantravadi A, Moore MG. HPV-related oropharyngeal cancer: a review on burden of the disease and opportunities for prevention and early detection. *Human vaccines & immunotherapeutics.* 2019.
122. Mehanna H, Franklin N, Compton N, et al. Geographic variation in human papillomavirus–related oropharyngeal cancer: Data from 4 multinational randomized trials. *Head & neck.* 2016;38(S1):E1863-E1869.
123. Taberna M, Mena M, Pavón M, Alemany L, Gillison M, Mesía R. Human papillomavirus-related oropharyngeal cancer. *Annals of Oncology.* 2017;28(10):2386-2398.
124. Pytynia KB, Dahlstrom KR, Sturgis EM. Epidemiology of HPV-associated oropharyngeal cancer. *Oral oncology.* 2014;50(5):380-386.
125. Luciani S, Berman NJ. Canadian strategy for cancer control. *Chronic Diseases in Canada.* 2000;21(1):23-25.
126. Maarten van S. Should a risk prediction model be developed? In. *Twitter*2018.
127. Forte T, Niu J, Lockwood GA, Bryant HE. Incidence trends in head and neck cancers and human papillomavirus (HPV)-associated oropharyngeal cancer in Canada, 1992–2009. *Cancer Causes & Control.* 2012;23(8):1343-1348.
128. Hwang E, Johnson-Obaseki S, McDonald JT, Connell C, Corsten M. Incidence of head and neck cancer and socioeconomic status in Canada from 1992 to 2007. *Oral oncology.* 2013;49(11):1072-1076.
129. Arekunnath Madathil S. Life course behavioural risk factors of head and neck cancers: A Bayesian approach. 2018.
130. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International journal of social research methodology.* 2005;8(1):19-32.
131. Westphaln KK, Regoeczi W, Masotya M, et al. From Arksey and O'Malley and Beyond: Customizations to enhance a team-based, mixed approach to scoping review methodology. *MethodsX.* 2021;8:101375.
132. Morris M, Boruff JT, Gore GC. Scoping reviews: establishing the role of the librarian. *J Med Libr Assoc.* 2016;104(4):346-354.

133. Embase. https://www.wolterskluwer.com/en/solutions/ovid/embase-903. Accessed 3/20/2022.
134. CAB Abstracts. CABI.org Web site. https://www.cabi.org/publishing-products/cab-abstracts/. Accessed 3/20/2022.
135. Scopus preview. https://www.scopus.com/. Accessed 3/20/2022.
136. Document search - Web of Science Core Collection. https://www.webofscience.com/wos/woscc/basic-search. Accessed 3/20/2022.
137. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev.* 2016;5(1):210.
138. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* 1960;20(1):37-46.
139. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276-282.
140. Gillison ML, Broutian T, Pickard RK, et al. Prevalence of oral HPV infection in the United States, 2009-2010. *Jama.* 2012;307(7):693-703.
141. Control CfD, Prevention. NHANES questionnaires, datasets, and related documentation. *National Health and Nutrition Examination Survey.* 2018.
142. Madathil SA. *Life course behavioural risk factors of head and neck cancers: A Bayesian approach*, McGill University (Canada); 2018.
143. ICD-10 Version:2010. https://icd.who.int/browse10/2010/en#/C00-C14. Accessed 3/20/2022.
144. Definition of squamous cell carcinoma - NCI Dictionary of Cancer Terms - NCI. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/squamous-cell-carcinoma. Published 2011. Updated 2011/02/02/07:00. Accessed 3/20/2022.
145. Rothman KJ, Greenland S, Lash TL. Case-Control Studies: Overview. *Wiley StatsRef: Statistics Reference Online.* 2014.
146. Begg CB, Gray RJ. Methodology for case-control studies with prevalent cases. *Biometrika.* 1987;74(1):191-195.

147. Blane DB. Collecting retrospective data: development of a reliable method and a pilot study of its use. *Social science & medicine.* 1996;42(5):751-757.
148. Schlecht NF, Franco EL, Pintos J, Kowalski LP. Effect of smoking cessation and tobacco type on the risk of cancers of the upper aero-digestive tract in Brazil. *Epidemiology.* 1999:412-418.
149. Gajalakshmi V, Hung RJ, Mathew A, Varghese C, Brennan P, Boffetta P. Tobacco smoking and chewing, alcohol drinking and lung cancer risk among men in southern India. *Int J Cancer.* 2003;107(3):441-447.
150. Canada H. Low-risk alcohol drinking guidelines. https://www.canada.ca/en/health-canada/services/substance-use/alcohol/low-risk-alcohol-drinking-guidelines.html. Published 2021. Updated 2021/07/05/. Accessed.
151. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med.* 2021;40(19):4230-4251.
152. Hashim D, Genden E, Posner M, Hashibe M, Boffetta P. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. *Ann Oncol.* 2019;30(5):744-756.
153. van der Schroeff MP, de Jong RJB. Staging and prognosis in head and neck cancer. *Oral oncology.* 2009;45(4-5):356-360.
154. Hammerlid E, Silander E, Hörnestam L, Sullivan M. Health-related quality of life three years after diagnosis of head and neck cancer--a longitudinal study. *Head Neck.* 2001;23(2):113-125.
155. Hammerlid E, Bjordal K, Ahlner-Elmqvist M, et al. A prospective study of quality of life in head and neck cancer patients. Part I: at diagnosis. *The laryngoscope.* 2001;111(4):669-680.
156. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *The Lancet.* 2008;371(9625):1695-1709.
157. Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. Paper presented at: Mayo Clinic Proceedings2008.

158. 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. *International journal of cancer.* 2005;114(5):806-816.

159. Yeole BB, Sankaranarayanan R, Sunny L, Swaminathan R, Parkin D. Survival from head and neck cancer in Mumbai (Bombay), India. *Cancer: Interdisciplinary International Journal of the American Cancer Society.* 2000;89(2):437-444.

160. Cadoni G, Giraldi L, Petrelli L, et al. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. *Acta Otorhinolaryngol Ital.* 2017;37(6):458-466.

161. Sivasithamparam J, Visk CA, Cohen EE, King AC. Modifiable risk behaviors in patients with head and neck cancer. *Cancer.* 2013;119(13):2419-2426.

162. Gritz ER, Carr CR, Rapkin DA, Chang C, Beumer J, Ward PH. A smoking cessation intervention for head and neck cancer patients: trial design, patient accrual, and characteristics. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 1991;1(1):67-73.

163. Aupérin A. Epidemiology of head and neck cancers: an update. *Curr Opin Oncol.* 2020;32(3):178-186.

164. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *European Urology.* 2015;67(6):1142-1151.

165. Menezes FdS, Fernandes GA, Antunes JLF, Villa LL, Toporcov TN. Global incidence trends in head and neck cancer for HPV-related and -unrelated subsites: A systematic review of population-based studies. *Oral Oncology.* 2021;115:105177.

166. Madathil S, Rousseau MC, Durán D, Alli BY, Joseph L, Nicolau B. Life Course Tobacco Smoking and Risk of HPV-Negative Squamous Cell Carcinomas of Oral Cavity in Two Countries. *Front Oral Health.* 2022;3:844230.

167. Habbous S, Chu KP, Qiu X, et al. The changing incidence of human papillomavirus-associated oropharyngeal cancer using multiple imputation from 2000 to 2010 at a Comprehensive Cancer Centre. *Cancer epidemiology.* 2013;37(6):820-829.

168. Ghazawi FM, Lu J, Savin E, et al. Epidemiology and patient distribution of oral cavity and oropharyngeal SCC in Canada. *Journal of cutaneous medicine and surgery.* 2020;24(4):340-349.

169. Canada S. Canadian Cancer Registry (CCR). In: Statistics Canada Ottawa, ON; 2014.

170. American Association for Dental, Oral, and Craniofacial Research. https://aaam2022.secure-platform.com/a/. Accessed 3/20/2022.

171. RSBO - Réseau de recherche en santé buccodentaire et osseuse. RSBO Web site. https://www.rsbo.ca/. Accessed 3/20/2022.