Epidemiology and Risk Prediction Modelling of Head and Neck Cancer

Craig D L Smith, BSc (Hons)

Thesis Submitted in fulfilment of the requirements for the degree Doctor of Philosophy

School of Cancer Sciences
College of Medicine, Veterinary and Life Sciences
University of Glasgow

August 2024

Abstract

Head and neck cancer (HNC), typically defined as squamous cell carcinomas of the oral cavity, pharynx and larynx, is the 7th most common cancer type globally. In the UK there over 12,000 new cases each year, with some of the highest incidence rates in Scotland. Major risk factors of HNC include smoking and alcohol consumption, both acting with independently and synergistically on HNC risk. Other key factors include sociodemographics (age, sex and socioeconomic deprivation). Human Papillomavirus (HPV) is another major risk factor for cancers of the oropharynx. With HNC incidence continuing to increase, primarily driven by the growing burden of HPV-associated oropharynx cancer (OPC) cases, there is a need for an increased focus on further understanding epidemiological changes and developing prevention strategies.

Chapter 1 summarises the literature focusing on definitions, epidemiology and risk factors for HNC. The chapter also goes on to describe the literature on approaches to the prevention of HNC.

Chapter 2 details the overarching aims and objective of this thesis and of each of the research studies.

Chapter 3 is presented as a journal publication. In this chapter, a rapid review of the literature using a systematic search of multiple databases was conducted on existing HNC risk models. In this review, 14 studies were identified according to pre-planned search criteria. The papers were assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) framework, in addition to an independent quality assessment. Six of the 14 models were classed as high quality, and of these, three were also high performing (AUC > 0.8, 0.87–0.89). A narrative synthesis of these models was performed and found that high-quality models had tailored the selection of predictors to the target populations. The synthesis identified models should include behavioural and sociodemographic predictors, in addition to clinical variables or biomarkers, where practical. The review found that some existing models have potential to predict HNC risk, but there is scope for improvement, with many of the models lacking external validation.

Chapter 4 is presented as a journal publication. In this chapter, data from the Scottish Cancer Registry were analysed. HNC incidence trends were evaluated by calculating age-standardised rates and using Poisson regression with interaction tests to assess changes in HNC incidence over time (2001-2020). This analysis revealed that the burden of HNC is changing (RR = 1.05, 95% CI = 1.01-1.09), primarily driven by increasing OPC rates (RR = 1.78, 95% CI = 1.65-1.93), accompanied by a stabilisation of OCC rates and a decline in larynx cancer rates (RR = 0.73, 95% CI = 0.68-0.79). However, the sociodemographics of people with HNC in Scotland have remained largely unchanged. Incidence over time was consistently higher among males (RR = 2.83, 95% CI = 2.74 - 2.91) and among those aged 60-64 and 65-69 years. Large area-based socioeconomic inequalities also existed and remained wide over the study period in all subsites, with the highest rates observed among those from the most socioeconomically deprived areas when compared with those from the least (RR = 2.87, 95% CI = 2.73-3.00). The Cancer Registry analysis assessing the HNC sociodemographic profile over

time largely supported the subsequent use of a case-control study (ARCAGE) conducted in 2002-2004 for the HNC risk prediction model development and subsequent validation in the UK Biobank cohort study, which is still ongoing (2006 – present).

Chapter 5 is presented in a traditional thesis chapter format. It aimed, to investigate the associations between individual socioeconomic status (SES) and HPV-positive and HPV-negative OPC in a European multicentre case-control study (Alcohol Related Cancers And Genetic susceptibility in Europe – ARCAGE, 2002-2004). Logistic regression analysis found no conclusive evidence of an association with a higher socioeconomic status and HPV-positive OPC (University education vs primary education; OR = 3.07, 95% CI = 0.92 - 10.30). However, this analysis was constrained by limited numbers of HPV-positive OPC cases (n = 74). This work lays the foundations for future international pooled analyses of multiple case-control and cohort studies within the HEADSpAcE (Head and Neck Cancers in South America and Europe) Consortium.

Chapter 6 is presented as a journal publication, detailing the process of developing and validating a HNC risk model. Following the findings of the review and supported by the findings of the registry analysis, a clinical risk prediction model was developed with data from the ARCAGE case-control study (1926 HNC cases and 2043 controls). Using established predictors from the literature, a clinical risk prediction model was devised using predictors chosen with ease of use in primary dental care at the forefront. The model exhibited fair performance in the developmental dataset (AUC = 0.75, 95% CI = 0.74–0.77) and had acceptable, but more limited, performance in the validation dataset, the UK Biobank cohort (384,616 participants, 1177 HNC cases; AUC = 0.62, 95% CI, 0.61 - 0.64). Such a model has potential to be developed into a tool and feasibility tested in primary dental care for prompting preventive interventions (e.g. smoking cessation, alcohol counselling) and recall intervals. The model could also potentially be improved with the use of biomarkers (e.g. HPV).

Chapter 7 is presented in a traditional thesis chapter format. It aimed to describe the existing HPV data within the UK Biobank. The viability of using multiple imputation to project HPV serostatus was assessed. Analysis of available HPV data within the UK Biobank revealed some associations with HPV-serostatus, mostly sexual behaviours, but the substantial volume of missing HPV data (>98% missing) made any potential imputations untenable. The utility of exploring HPV-mediated HNCs in the UK Biobank at the time of writing, was limited by the small, random sample of HPV-serology data available.

Chapter 8 is presented as a traditional thesis discussion. It describes the key findings of this thesis, including the under-utilisation of HNC risk models, a lack of validation and feasibility testing, the unchanging sociodemographics of HNC and the performance of the ARCAGE model with recommendations for future feasibility testing and areas for improvement. This chapter also compares the findings of this thesis with the wider literature and details other advancements published over the time-period of this thesis. The strengths and limitations of each study are also detailed, followed by recommendations for practice, policy and future research.

Table of Contents

Table of Contents

List of Tables

List of Figures

Acknowledgements

If you have read this whole thing, and I have somehow pulled this off, congratulations to all of us - I have got a village of amazing people to thank. This thesis is first, among many things, dedicated to those who did not make it. In loving memory of Sandy Smail, whose kindness and passion inspired wee 15-year-old me and set me on the course I am now. A brilliant dentist, taken far too soon - I wish you could have seen the journey you started me on. I would like to think I've come a long way from injecting oranges with local anaesthetic.

A gigantic thank you to my supervisor Professor David Conway. Your ability to be so knowledgeable in so many fields and make time for people despite being so busy never fails to amaze me. Thank you for being an amazing role model, hearing me out when I had concerns and reassuring me when I needed it the most. I'll be sure not to use the words "partake" or "whilst" in any future studies. Thank you, Dr. Alex McMahon, for supervising my statistics and making them bearable with your anecdotes and stories. I never imagined being able to write code but that is a testament to yourself.

Thank you to Professor Gareth Inman for being a reliable source of expertise, and feedback and being a great asset to our team. Thank you to my "Reg"(istrar) Grant Creaney, who, much like David, spins so many plates but was always there to lend a supportive ear (or idea). Thank you to Mariel Goulart for being amazingly patient and supporting me whenever I had R queries or DAG questions. Thanks also to Donald Lyall for your help and continued support throughout my academic journey. A thank you also to the PIs of the ARCAGE (and now HEADSpAcE study) group – an amazing group of researchers and clinicians – the future is bright, and I look forward to future collaborations. Thank you to my reviewers Richard Wilson and Tom Bird for your support and enthusiasm. A particular thank you to Richard for your support in my big career change, your kindness will always be massively appreciated.

Thanks to others in Community Oral Health (COH) – Frances, Kyle, Lorenzo, Budur, Amanda, Emma, Andres, Andrea, Bill, Al, Viv, Stefan and Marie for being amazing colleagues and all the laughs along the way. Thank you also to other colleagues on the balcony: Krystyna, Lauren, Zainab, Laurie, Abdullah, Saeed, Muhanna, Suror, Hafsa and lastly, but far from least, Ejay for making the workplace a supportive and fun place to be. Many thanks to David Brunton for all your help with the clinical catchup and all your patience. You were a great teacher - even though my path ultimately lay elsewhere, I'll be always grateful for your teachings and take your lessons with me (and hopefully get to give BJJ a try sometime). Lastly, thank you, Pauline, for always being helpful, cheery and being the organisational glue that holds the COH department together.

A massive thank you to Jenny Montgomery, Cat Douglas, Claire Paterson and other colleagues in ENT and oncology. You were key to my journey and the journey that this thesis became – you gave me not only invaluable input but experience from the get-go. My first theatre case (a total laryngectomy) with Jenny was truly a butterfly moment in my journey that I will never forget. Many thanks to the QEUH OMFS department and Craig Wales for being a brilliant role model and always having ideas to point me in the right direction. A big thank

you to Kevin Ryan and Jennifer Taylor from Oral Medicine for your guidance and unmatched enthusiasm.

Thank you to many of my friends from “Intercal” – Gemma, Emma, and Iona for taking a stray dental student under your wing (during an intercalated degree in the peak of a pandemic, no less!) That one act of kindness changed my path irreversibly and helped inspire me to take my leap of faith (stats still scare me, to be honest - sorry Alex if you are reading this). Thank you also, Antony, Andrew, Paddy, Sahan, Stefan, Juliane, Rory, Denver, Wil and James for the good craic and for helping ensure that I had a social life outside of GDH. Thank you to pals in GU Muay Thai for the great training sessions and socials along the way. Thank you to my flatmates Noemi and Julie – while you both arrived in the latter half of the PhD you gave me a reason to come home after a long day.

Thank you to the patients - those that I saw in clinics, in theatre or who participated in the various studies employed in this research. At the end of the day, this work is ultimately dedicated to them and every case I saw served as an important reminder of what I was fighting for whenever I was demoralised, tired and felt like giving up.

Thank you to the TRACC (to Train and Retain Academic Cancer Clinicians) programme and Cancer Research UK for this amazing opportunity, and for selecting and funding me to undertake this thesis. I look forward to future work and collaboration in clinical academia.

To my wonderful girlfriend Julie, thank you for being there and supporting me with all your love, laughter, and silliness. Thank you for being the best conference partner, the best editor, and always being patient with me when I got home later than planned. I could not have done the final hard stretch of this thesis without you. I look forward to many more adventures and Chapters (beyond a thesis) with you. Miluji tě. 🐸

Thank you to my parents for always supporting me throughout this journey. If I become half the people you are, I know I will have done well in life. Thank you also to my sister, Catriona, for offering comedic relief and providing me with an endless stream of cat photos. Thank you for the meals, doing my washing (on the rare occasions I made it home) and providing me with many laughs over the phone during many a late night in the hospital. Especially when I was tired and stressed, your support meant the world to me.

Lastly, I want to thank me. For sticking through it – the highs and the lows and taking it all in our stride. Even though the model did not work as originally planned, you can hold your head high. You had your doubts. Your moments where you felt like you were on uncharted territory, alone and well out of your comfort zone - but you persevered. There are not many people that could be the first of your kind on a programme, do the entire application process to medical school on top of a PhD and come out with three papers and four offers. Thank you for believing in yourself – you should do that more.

Absolutely zero thanks to Eduroam.

Author's Declaration

This thesis has been submitted by alternative format. I declare that, except where explicitly stated, that this thesis is the result of my own work and has not been submitted, partly or in whole, for any other degree at the University of Glasgow or any other institution.

Research from this thesis was published to the following locations or presented at the following conferences and local research meetings.

Journals:

Laryngoscope Investigative Otolaryngology – Published Online 28th November 2022, DOI: 10.1002/lio2.982 PMCID: PMC9764804 PMID: 36544947 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9764804/
"Risk prediction models for head and neck cancer: A rapid review"

BJC Reports - (Accepted for publication 6/8/24) "Head and neck cancer incidence is rising but the sociodemographic profile is unchanging in Scotland (2001-2020): A population epidemiological study"

Head and Neck – Published Online 8th June 2024, DOI: 10.1002/hed.27834 https://onlinelibrary.wiley.com/doi/10.1002/hed.27834 "Development and external validation of a head and neck cancer risk prediction model"

International Conferences:

HEADSpAcE (IARC-WHO) Annual Meeting, Heidelberg, Germany 10-12th April 2024
Oral presentation title: *Epidemiology and Risk Prediction Modelling of for Head and Neck Cancer*

International "Early Detection of Cancer Conference" Conference (CRUK) London 10-12th October 2023
Poster presentation title: *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

HeadSPACE meeting Cartagena, Columbia (remote) – 7th October 2022
Oral presentation title: *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

International PhD Student Cancer Conference (IPSSC) Conference, Heidelberg, Germany 2022 - 8th-10th June 2022
Oral presentation title: *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

National Conferences:

British Association of Head and Neck Oncology (BAHNO) National Conference,The Business Design Centre, London – 17th May 2024

Poster Presentation title: *Development and external validation of a Head and Neck Cancer risk prediction model*

Glasgow University Surgical Society (GUSS) National Undergraduate Conference – "Beyond the Borders: Carving paths to surgical equity", Royal College of Surgeons, Glasgow – 27th April 2024
Oral presentation title: *The epidemiology and sociodemographics of HNC in Scotland (2001-2020): a population epidemiological study*

Glasgow University Medical Research Society (GUMRS) National Undergraduate Conference – "Turning Tides: Navigating novelty and Change", UofG ARC, Glasgow
Poster presentation title: *The epidemiology and sociodemographics of HNC in Scotland (2001-2020): a population epidemiological study*

TRACC TRAM BHF CRTF Clinical Academic Research Event – 18th April 2024
Oral presentation title: *Epidemiology and Risk Prediction Modelling of for Head and Neck Cancer*

Scottish Cancer Networks Collaborative Head and Neck Cancer Event 24th November 2023
Poster Presentation title: *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

Junior Trainee Group BAOMS Conference, Sheffield 11-12th November 2023
Oral presentation title: Epidemiology and Risk Prediction modelling of Head and Neck Cancer

CRUK Scotland Meeting, the Beatson institute, Glasgow – 24th August 2023
Poster Presentation title: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer

TRACC TRAM BHF CRTF Clinical Academic Research Event, Scotland Institute, Glasgow – 23rd January 2023
Oral presentation title: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer

CRUK Scotland Meeting, Edinburgh – 1st September 2022
Poster Presentation title: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer

<u>Local research Meeting:</u>

Glasgow University Clinical Academic Research Prize Evening, Wolfson Medical School, Glasgow – 23rd November 2023
Poster Presentation title: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer

Flash Presentation MVLS PGR Induction – 6th October 2022

Oral presentation title: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer

School of Medicine, Dentistry and Nursing (SOMDN), Postgraduate Research Day - 16th May 2022, 23rd May 2023, 20th May 2024
Oral presentation title (2022,2023): *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer, Epidemiology and Risk Prediction modelling of Head and Neck Cancer*
Oral presentation title 2024: *The epidemiology and sociodemographic profile of HNC in Scotland (2001-2020): a population epidemiological study*

GLAHNC Research Fellows Showcase day - 31st of August 2022, 13th March 2024
Oral presentation title (2022): *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*
Oral presentation title (2024): Epidemiology and Risk Prediction modelling of Head and Neck Cancer

SOHRC Public Health and Health Services Research Group Online Scientific Event – 20th of April 2022
Oral presentation title: *Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

Oral presentations on the project in Community Oral Health department research meetings

Prizes:

3rd Place Oral Presentation - Glasgow University Surgical Society (GUSS) National Undergraduate Conference – "Beyond the Borders: Carving paths to surgical equity" Royal College of Surgeons, Glasgow – 27th April 2024
***Oral presentation title**: The epidemiology and sociodemographics of HNC in Scotland (2001-2020): a population epidemiological study*

Best Poster (Clinical) - *Poster presentation at Glasgow University Clinical Academic Research Prize Evening, Wolfson Medical School, Glasgow – 23rd November 2023*
***Poster Presentation title**: Development and Validation of a Risk Prediction Tool for Head and Neck Cancer*

Best Undergraduate (first degree) Presentation - *Junior Trainee Group BAOMS Conference, Sheffield 11-12th November 2023*
***Oral presentation title**: Epidemiology and Risk Prediction modelling of Head and Neck Cancer*

Signature:

Printed Name: Craig D L Smith

Definitions/abbreviations

ARCAGE (Study) = Alcohol-Related Cancers And Genetic susceptibility in Europe Study
AUC = Area Under Curve (also known as C-index)
AUROC = Area Under Receiver Operating Curve
BMI = Body Mass Index
CI = Confidence Interval
E/O ratio = Expected-Observed ratio
GDP = General Dental Practioner
GRNN = General Regression Neural Network
GWAS = Genome Wide Association Studies (GWAS)
HNC = Head and Neck Cancer
HPV = Human Papilloma Virus
IARC = International Agency for Research on Cancer
OCC = Oral Cavity Cancer
OLK = Oral Leukoplakia
OPC = Oropharyngeal Cancer
OPMD = Oral Potentially Malignant Disease
OR = Odds Ratio
PICO = Patient/Population, Intervention, Comparison and Outcomes
PNN = Probabilistic Neural Network
PROBAST = Prediction model Risk Of Bias ASsessment Tool
RR = Rate Ratio
SIMD = Scottish index of Multiple Deprivation
SENS = Sensitivity
SES = Socioeconomic status
SPEC = Specificity
TRIPOD = Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
UADT = Upper Aero-Digestive Tract
WHO = World Health Organisation

1 Chapter One – Background

1.1 Introduction and Definitions

Head and neck cancer is a collective term for a series of cancers with similar aetiological patterns, primarily defined as cancers of the pharynx (International Classification of Disease (ICD) codes C09-13), larynx (ICD C32) and the oral cavity (ICD C00-06 – excluding the outer surface of the lip). (World Health Organization (WHO), 2004, Winn et al., 2015a) Pathologically, 90% of these cancers are squamous cell carcinomas, developing in the squamous cell lining epithelium that is characteristic of many anatomical structures in the head and neck region. (National Cancer Institute, 2021 , Sanderson and Ironside, 2002b) A visual representation of these primary sites can be seen in Figures 1 and 2.

The following ICD-10 definitions are used in this thesis: oral cavity (inner lip - C00.3 - C00.9, dorsal, overlapping or NOS tongue - C02, gingivae - C03, floor of mouth - C04, soft palate, uvula, palate or NOS - C05 and cheek, other and NOS mouth - C06); oropharynx (base of tongue- C01, lingual tonsil - C02.4, tonsil - C09, oropharynx - C10, pharynx - C14.0,14.2); larynx (C32); all HNC - above sites in addition to tumours of the nasopharynx (C11), piriform sinus (C12), Hypopharynx (C13) and other overlapping sites (C14.8).

Cancers of the brain or eyes are defined separately and, as such, are not included in the defintion of head and neck cancer. (Centers for Disease Control and Prevention, 2020b) It should be noted that there is some variation in how head and neck cancers are defined; some defintions also include overalpping sites of an unknown primary tumour of origin (C14), non primary sites such as the salivary glands (C07,08), paranasal sinuses, nerves, the nasal cavity or middle ear (C30). Cancers of many of these structures are rarer and differ in aetiology or anatomical features. For these reasons, they are often not included as primary subsites of head and neck cancer. Head and neck cancers are also grouped as they are generally managed clinically by multidiciplinary teams (MDT). (Homer and Winter, 2024)

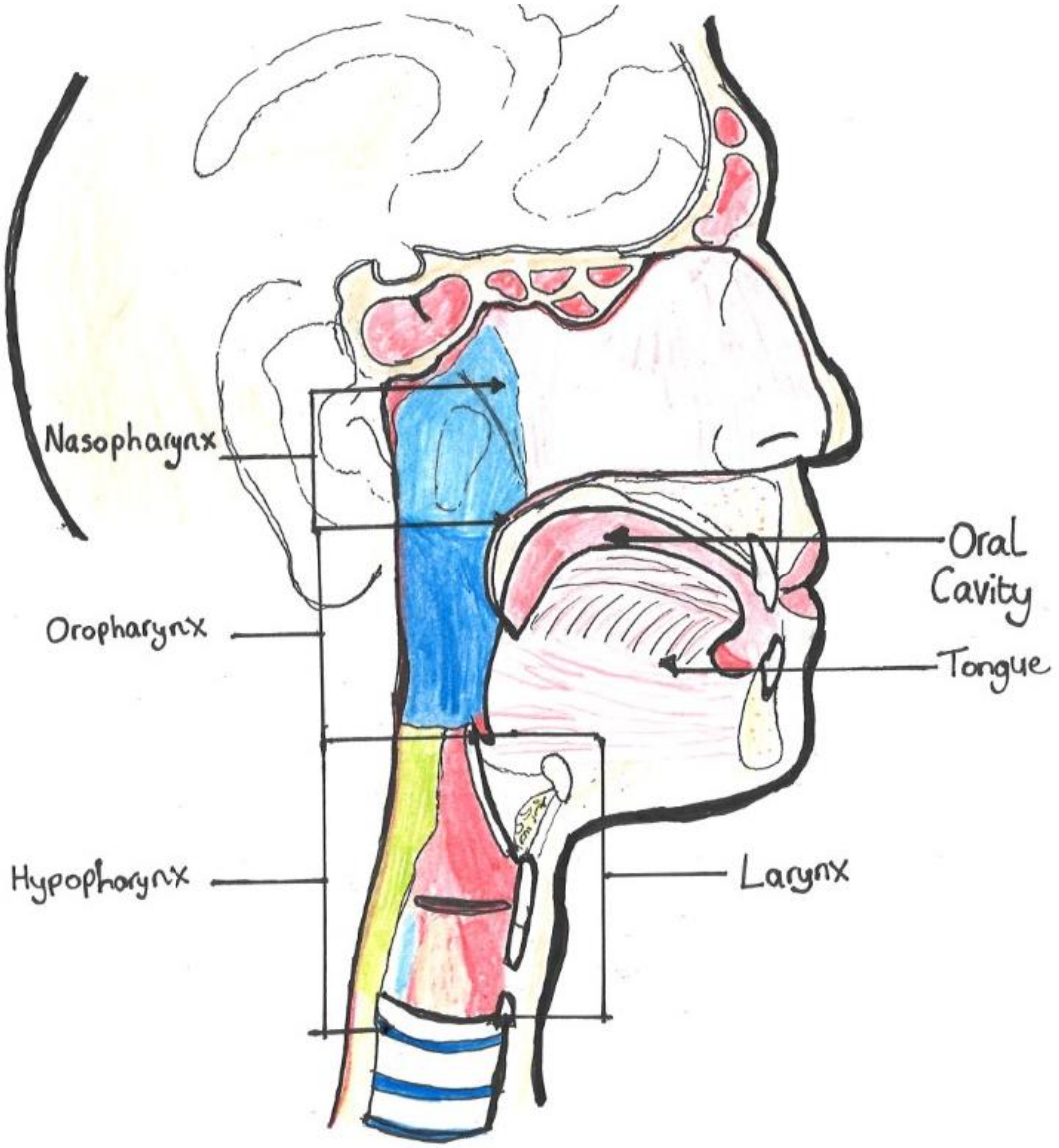


Figure 1-1: Primary Subsites of Head and Neck Cancer

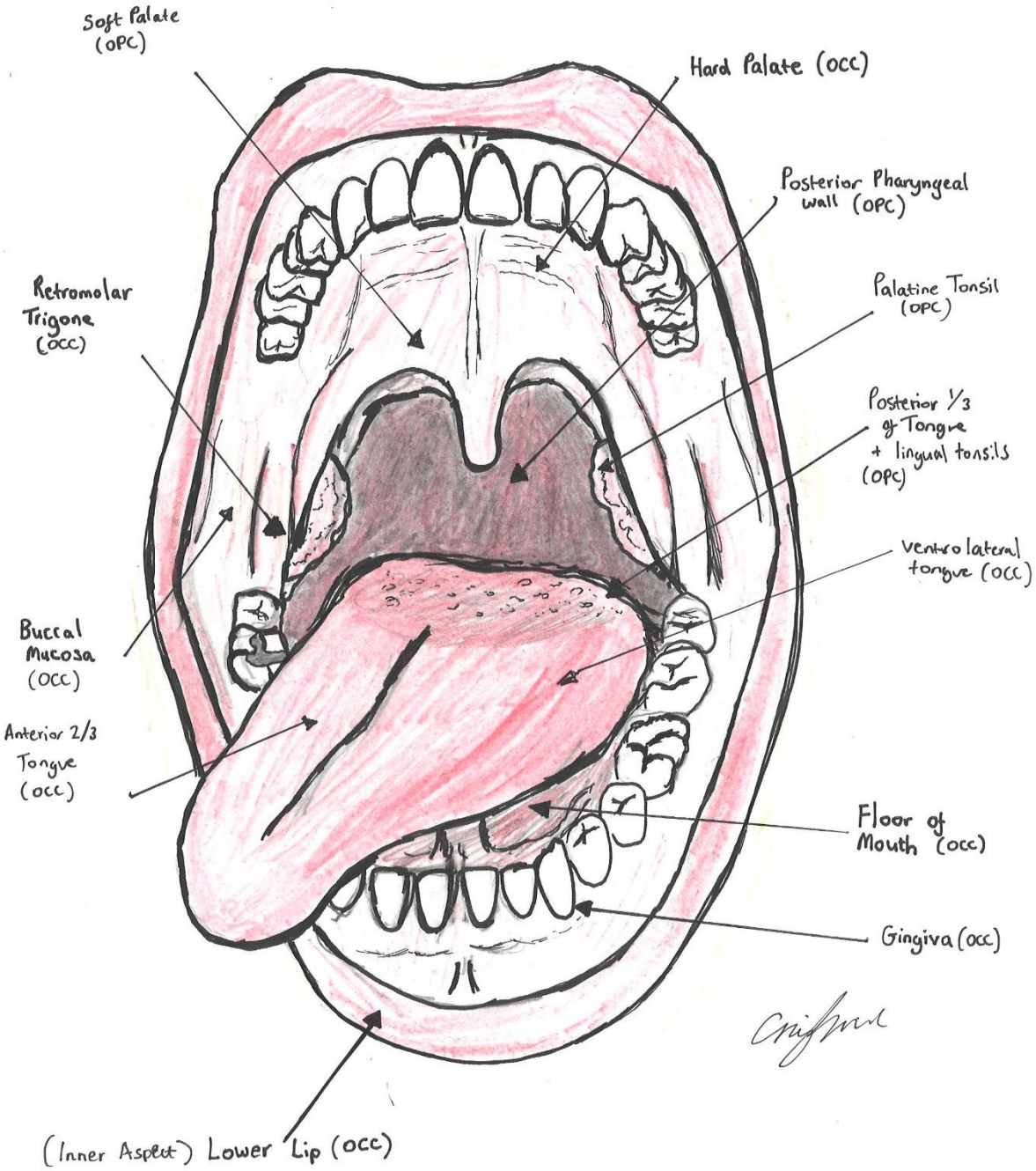


Figure 1-2: Example oral cavity cancer + oropharynx cancer Sites (Conway et al., 2018a)

In this thesis, head and neck cancer was defined as malignancies of the oral cavity, pharynx and larynx, according to defintions generally accepted and used in epidemioligcal studies, unless otherwise specified. (Conway et al., 2018a) This chapter reviews the background literaure on the presentation, descriptive epidemiology (burden, trends, survival) and analytical epidemiology (risk

factors) of head and neck cancer. In addition, the literaure on approaches to prevention will be reviewed.

1.2 Clinical Presentation of Head and Neck Cancer

Head and neck cancers may present with various signs or symptoms depending on the subsite. In some cases, people with HNC may present asymptomatically. Patients may present with symptoms such as a persistent sore throat, difficulty or pain on swallowing (dysphagia / odynophagia), stridor (a high pitched noise caused by reduced airflow and airway obstruction on inspiration), hoarseness of the voice, unintentional weight loss, referred unilateral ear pain (otalgia) or a persistent lump in the neck. (NHS Online, 2021, NHS Scotland, 2019) Further signs of head and neck cancer may include palsy of the cranial nerves, presence of a mass in the lateral aspect of the neck or the orbit, persistent red, white or mixed patches in the mouth (that cannot be gently scraped away), persistent oral ulceration, swelling or loose teeth unexplained by trauma or periodontal disease. (Mehanna et al., 2010) (NICE, 2021b) NHS Scotland guidance suggests an emergency referral should be completed in the event of severe breathing difficulties (stridor), while both Scottish and NICE guidelines advise an urgent suspicion of cancer referral (to be reviewed within two weeks) where patients present with a persistent unexplained neck lump for a period greater than three weeks, a non-healing intra-oral ulcer or lesion for greater than three weeks, or have persistent hoarseness of the voice. (NHS Scotland, 2019, NICE, 2021a)

Oral potentially malignant disorders (OPMD) are a group of pre-malignant conditions that carry a risk of progression to oral cavity cancer. (Warnakulasuriya, 2020, WHO, 2023) Common examples of OPMD include oral leukoplakia, oral erythroplakia, oral submucous fibrosis, oral lichen planus, actinic keratosis. (IARC-WHO, 2023) There is varying evidence on malignant transformation rates. A systematic review and metanalysis by Iocca et al estimated the overall malignant transformation rate of OPMDs to be 7.9%, with lichen planus carrying the lowest risk (1.4%) and proliferative verrucous leukoplakia carrying the highest risk (49.5%). (Iocca et al., 2020)

1.3 Descriptive Epidemiology of Head and Neck Cancer

Descriptive epidemiology is commonly defined as the understanding of the distribution of a disease and its determinants. (Liu, 2018) This section describes the incidence and mortality of head and neck cancer and relevant subsites at a global and national level.

1.3.1 Head and Neck Cancer Incidence

1.3.1.1 Global Disease Burden and Trends of Head and Neck Cancer

Head and neck cancers are the 7th most common cancer globally. (Johnson et al., 2020b, Sung et al., 2021) In 2022 alone, there were an estimated 892,128 novel cases of head and neck cancer. (GLOBOCAN, 2022b) The greatest overall burden of cases by continent was in Asia with an estimated 575,411 cases. When assessing age-standardised rates per 100,000, Oceania and Europe have the highest incidence rate of 11.3 and 10.7 cases per 100,000 respectively. (GLOBOCAN, 2022b)

The overall incidence of head and neck cancer is rising across the world, with an increase of 36.5% over the last decade. (Bray et al., 2018) (McDermott and Bowles, 2019) Incidence is also projected to increase by 30% by 2030. (Gormley et al., 2022b) There is some geographical variation within this changing incidence due to differing exposures and population risk profiles. (Simard et al., 2014)

1.3.1.2 Global Disease Burden and Trends by Head and Neck Cancer Subsite

In recent years, changes in the global burden of head and neck cancer by subsite have been observed. Oral cavity cancer incidence has risen, and is projected to continue to rise. (Gormley et al., 2022a) The largest burden and increase of OCC cases was recorded in Asia and the Pacific. (GLOBOCAN, 2022b) This higher burden of OCC is associated with betel quid and chewable tobacco behaviours

that are commonplace in these regions. (Warnakulasuriya et al., 2005, Shield et al., 2017)

Similarly, rising head and neck cancer rates in the USA and Europe have been largely driven by increases in oropharynx cancer, where the highest rates of OPC are recorded. (GLOBOCAN, 2022b) This has been attributed to a growing burden of Human Papillomavirus (HPV)-mediated OPC cases. (Menezes et al., 2021) The incidence rates of OPC in these regions are projected to continue to increase, primarily by HPV-related tumours, and eventually surpass OCC rates. (Purkayastha et al., 2016, Conway et al., 2018a) At a global level, larynx cancer incidence has increased but rates have declined in higher-income countries where changes and reductions in smoking behaviour patterns have occurred. (Fitzmaurice et al., 2017, Gormley et al., 2022a)

1.3.1.3 Stage of Head and Neck Cancers

Globally, the majority of HNC cases are diagnosed at advanced stages (III to IV). (Abrahão et al., 2018, Abrahão et al., 2020, Moon et al., 2022) A pre-SARS-COV2 pandemic analysis of UK cancer registries by Creaney and colleagues found that 59% of HNC cases in the UK were diagnosed at more advanced stages. (Creaney et al., 2022) This remained true across all countries in the UK. In Scotland, 65.4% of cases with stage data were diagnosed with advanced stage disease. With regards to subsite, oropharynx cancer cases constituted the largest proportion of advanced stage disease, while laryngeal cancers had the lowest proportion of advanced stage disease.

1.3.1.4 UK and Scotland Burden and Trends of Head and Neck Cancer

Head and neck cancer is the 8th most common cancer type in the UK. In 2022, there were 12,609 novel head and neck cancer cases in the UK. (GLOBOCAN, 2022a) Furthermore, the overall incidence of head and neck cancer is increasing; between 1993-1995 and 2016-2018 the UK age-standardised incidence increased by 34% from approximately 15 cases per 100,000 persons to 20 cases per 100,000 persons. (Cancer Research UK, 2020b)

Within the UK, Scotland has the highest incidence rates. There are approximately 1200 incident HNC cases per year, making it the 6th most common

cancer in the country. (Public Health Scotland, 2021) Incidence rates were highest in the west of Scotland, where some 50% of cases are diagnosed.

Across the UK, common trends in subsite incidence have been observed. Incidence has been characterised by increases in OPC across the UK. (GLOBOCAN, 2022a) This was observed in an earlier Scottish Cancer Registry analysis, with an 85% increase in OPC from 2001-2012. (Purkayastha et al., 2016) OCC rates increased in England and stabilised within Scotland, while stabilisations or declines in larynx cancer rates were also observed. (McCarthy et al., 2015, Purkayastha et al., 2016)

1.3.1.5 Incidence Trends by Sociodemographic Factors

Age, sex and area-based socioeconomic deprivation are important determinants in the risk profile of many cancers, including HNC. (Gormley et al., 2022a)

As with most cancers, head and neck cancer has a greater incidence with increasing age, most commonly developing between a patient's 5th and 7th decade of life. (Johnson et al., 2020b, Gormley et al., 2022b) In Europe, over 50% of head and neck cancer cases have been reported to be among those over the age of 60. (Mehanna et al., 2010) With the global phenomenon of aging populations, whereby the average age of populations are increasing owing to both reducing birth rates as well as reducing death rates, trends in the cancer burden are increasing across the world. (IARC-WHO, 2024)

Of the estimated 892,128 head and neck cancer cases in 2022, 679,498 cases were male and 212,630 cases were female. (GLOBOCAN, 2022b). Head and neck cancer is generally more common among males, but rates among females are also increasing. (Gormley et al., 2022b) This has been attributed to sex-specific patterns in smoking and alcohol behaviours among females. (Simard et al., 2014, Miranda-Filho and Bray, 2020) Men are two to four times more likely to develop HNC. (Johnson et al., 2020a) UK analyses reveals similar trends, males being one and a half to over four times more likely to be diagnosed with head and neck cancer. (McCarthy et al., 2015, Purkayastha et al., 2016) The greatest differences in the proportion of cases by sex were observed among cases of the larynx, while the smallest differences were among OCCs.

Area-based socioeconomic deprivation, defined as a combined measure of economic and social affluence, is an important demographic marker of HNC trends. (Baker, 2014) In descriptive epidemiology analyses, socioeconomic status is typically measured by area-based socioeconomic indices, e.g., the Scottish Index of Multiple Deprivation. (Conway et al., 2019b, Scottish Government, 2020) There are strong socioeconomic inequality patterns measured in the HNC disease burden, with the highest rates observed among those from socioeconomically deprived areas. (Purkayastha et al., 2016, Conway et al., 2021b)

1.3.2 Mortality and Survival Rates of Head and Neck Cancer

Descriptive epidemiology studies also consider the burden of cancer in terms of mortality and survival rates (typically over a one-year, five-year or 10-year period). In 2022, there were an estimated 458,486 global deaths attributed to head and neck cancer (Age Standardised Rate (ASR) = 4.6 per 100,000), with the greatest burden and mortality rate (321 212, ASR = 5.6 per 100,000) in Asia. (GLOBOCAN, 2022 -b) Mortality outcomes are poor in the UK: some 4,143 deaths were recorded from 2017-2019 (European ASR = 6.6 per 100,000). The highest UK mortality rates were observed in Scotland with 516 recorded HNC deaths from 2017-2019 (European ASR = 9.7 per 100,000). (Cancer Research UK, 2022) As with incidence, there is a strong inequality pattern in HNC mortality, with the highest mortality rates among those from the most socioeconomically deprived areas. (Brown et al., 2021)

The overall survival prognosis of head and neck cancer is poor and has seen limited improvements in recent decades. Overall, five-year survival is estimated to be over 50%, with the lowest rates observed among those with cancers of the hypopharynx. (Pulte and Brenner, 2010a, Gormley et al., 2022a) For reference, the overall five-year survival for breast and colorectal cancers are estimated to be 91% and 65%, respectively. (National Cancer Institue (NCI), 2022) (National Cancer Institue (NCI), 2024) Improvements in survival have been attributed to developments in management techniques and the changing epidemiology of HPV-positive associated head and neck cancers, however, advanced stage presentation remains a major challenge. Wide inequalities in HNC survival also exist; those from more deprived socioeconomic backgrounds have been shown to

have poorer long-term survival and quality of life outcomes. (Ingarfield et al., 2019) (Mirza et al., 2019)

Survivors of head and neck cancer can face short- and long-term complications such as infection, impaired speech, swallowing, mastication, soft tissue damage or reduced salivary flow (and be further predisposed to dental disease as a result of this). (Gomes et al., 2020) (Melo Filho et al., 2013) Survivors of advanced stage cancers may also lose their speech entirely or require extensive facial reconstruction. Such invasive treatments and their complications can have significant psychological and psychosocial ramifications. (Fundakowski, 2020) Thus, from a long-term survival perspective, the importance of prevention and early detection cannot be overstated.

1.4 Analytical Epidemiology: Risk Factors for Head and Neck Cancer

Analytical epidemiology is defined as the measurement of the association between exposures and diseases. (Thun MJ, 2003) There are various risk factors that can heighten an individual's risk of head and neck cancer. These have been rigorously investigated and appraised in the international literature by a number of systematic reviews, meta-analyses and pooled analyses. This section aims to summarise the main risk factors for HNC.

Commencing in 2004, the International Head and Neck Cancer Epidemiology Consortium (INHANCE) is a global consortium of pooled HNC studies with a cumulative total of 25,500 HNC patients and 37,100 controls. Many of the most recent and robust HNC analytical epidemiological studies have originated from this consortium. These analyses have been summarised in two overview papers. (Winn et al., 2015a, Bravi et al., 2021) Although many of the included studies are case-control in design, which are subject to their own potential recall biases and a limited representation of studies in Asia where the significant global burden of HNC occurs, the consortia serves as a large and robust resource that has informed our understanding of HNC. (Winn et al., 2015a)

The findings of the INHANCE consortium along with other published systematic reviews and meta-analyses and evidence / monographs from the International Agency for research on Cancer (IARC) of the World Health Organisation (WHO) informed the discussion of risk factors in this chapter.

1.4.1 Tobacco and Alcohol

There is a large and well-established evidence base surrounding individual behaviours and the risk of head and neck cancer. Alcohol and tobacco use (including smokeless versions) are often cited as the primary drivers of HNC, with smoking and alcohol associated with some 73-75% of head and neck cancer cases. (Hashibe et al., 2009, Anantharaman et al., 2011)

1.4.1.1 Tobacco Smoking

The WHO classifies tobacco smoking as a grade I carcinogen, attributable to many cancers including those of the head and neck. (IARC-WHO, 2004) The carcinogenic effects of smoking have been heavily studied; over 70 different carcinogens have been identified in cigarette smoke. (Jethwa and Khariwala, 2017) Smoking is one of the most significant factors for head and neck cancer, smokers being at a ten times increased risk of HNC versus non-smokers. (Jethwa and Khariwala, 2017) Smoking carries a significant and independent risk of HNC, even among never drinkers. (Hashibe et al., 2007)

Pooled analyses have revealed that there is a significant dose-response effect associated with both frequency and duration of smoking behaviours. As with lung cancer, the duration of smoking is more important than frequency of consumption (i.e. fewer cigarettes over more years carries a higher risk than greater frequency over a shorter period). (Peto, 1986, Lubin et al., 2009) These findings have remained consistent even among those with a lower frequency of smoking. (Berthiller et al., 2016) Starting smoking at a younger age is also associated with an increased risk but is also mediated by cumulative exposure. (Chang et al., 2019) The proportion of risk attributable to smoking exhibits some variations by subsite, carrying the highest risk for tumours of the larynx and a

significant, but lower, risk for oropharynx and oral cavity cancers. (Hashibe et al., 2007)

There is also evidence from analysis of the INHANCE consortium that frequent and prolonged second hand exposure to smoking may also increase individual risk of head and neck cancer among non-smokers, particularly of the pharynx and larynx. (Lee et al., 2008)

1.4.1.2 Smokeless Tobacco and Areca Nut

In low- and middle-income countries, particularly in Asia where HNC is one of the most common cancers, other forms of tobacco products present a significant risk. These can include various forms of chewable tobacco, Betel quid and Areca nut consumption. (Joshi et al., 2014) Pooled analysis of studies has shown smokeless tobacco to be a significant risk factor for HNC, particularly for tumours of the oral cavity. (Wyss et al., 2016) Whilst less common in high-income nations, an increasingly globalised world means these behaviours are becoming more common place in the USA and Europe, which is something clinicians should be conscious of when working with diverse communities.

1.4.1.3 Alcohol Consumption

Alcohol is another major risk factor for head and neck cancers. (Gormley et al., 2022a) As with smoking, the WHO classifies alcoholic drinks as a carcinogen, harmful to human health and attributable to many cancers including those of the head and neck. (Baan et al., 2007) Research has found that higher frequency alcohol consumption over a shorter period carries a greater risk than lower frequency alcohol consumption over a longer period. (Kawakita and Matsuo, 2017, Di Credico et al., 2020) High-frequency alcohol intake has been found to be an independent risk factor for HNC, even in the case of never-smokers. (Hashibe et al., 2007, Gormley et al., 2020) As with smoking, variations in attributable risk vary by subsite; alcohol consumption has been shown to carry the greatest risk for tumours of the oral cavity and oropharynx. (Lubin et al., 2011)

Research on the risks by type of alcohol is limited. Analysis of pooled studies on alcohol subtype and HNC risk concluded that the risks of HNC associated with

different types of alcohol (beer, wine and spirits) were largely comparable in risk. (Marron et al., 2012) Similarly, analysis of the INHANCE consortium showed that the risks associated with beer and spirit drinking were similar. (Purdue et al., 2009) Wine drinking was found to have a slightly weaker association with HNC, but researchers emphasised that confounding could not be excluded and that these differences were very modest. (Purdue et al., 2009)

1.4.1.4 Combined Effects of Tobacco Smoking and Alcohol Consumption

Tobacco and alcohol do not only have independent exposure risks but they also have a synergistic effect on individual risk, collectively accounting for 72% of HNC cases. (Gormley et al., 2022a) Similarly, the population-attributable risk of joint tobacco and alcohol use has been estimated to be 73-75%. These effects were the greatest among cases of the larynx (89%), reduced among cases of the oropharynx (74%) and the least pronounced among cases of oral cavity (61%). (Hashibe et al., 2009, Anantharaman et al., 2011) This can result in as much as an over 35-fold increase in risk among heavy alcohol and tobacco users. (Dal Maso et al., 2016)

1.4.1.5 Electronic Cigarettes

Electronic-cigarette usage (a.k.a. “vaping”) is a contemporary exposure. Comparatively, electronic-cigarette usage is safer than smoking and has been shown to be an effective aid in tobacco smoking cessation. (Lindson et al., 2024) Although vaping has currently been shown to have no link with an increased risk of HNC in the short term, the long-term effects, and risks of vaping for HNC and general health are unknown. (Szukalska et al., 2020) In vitro studies have suggested the presence of some carcinogens in electronic-cigarettes and subsequent genetic changes associated with exposure. (Raj et al., 2020) However, there is a lack of longitudinal studies and further research is warranted.

1.4.2 Socioeconomic Factors

There is a vast and well-described association with low socioeconomic status and poor health. (Pathirana and Jackson, 2018, Wang et al., 2024) Socioeconomic

status can be measured using area-based metrics, e.g., the Scottish Index of Multiple Deprivation (SIMD) is an area-based SES metric which uses domains such as local crime, housing, education, occupation, income and access to healthcare / services. (Scottish Government, 2020) Socioeconomic status can also be described using individual measures, such as individual income, educational and occupational data. Both are subject to their own advantages and limitations; for example, SIMD is subject to the assumption that individuals in an area are socioeconomically homogeneous, while individual metrics are generally harder to record and are subject to individual variability and biases. It is recommended, where possible, to use both types of metrics. (Conway et al., 2019b)

There is an increasingly established evidence base highlighting low socioeconomic status or position as an independent risk factor for head and neck cancer. (Gormley et al., 2022a) Education is a strong individual level measure of SES that is commonly evaluated in studies. Analysis from the INHANCE consortium revealed that those with a lower level of education were at a significantly increased risk of HNC versus their more educated peers (OR = 2.50; 95% CI = 2.02–3.09). This remained true even among those who were reported as never-smokers or drinkers (OR = 1.61; 95% CI = 1.13–2.31). (Conway et al., 2015) Similarly, relative to those with a higher income, individuals with a lower income were found to be associated with an increased risk of HNC (OR = 2.44; 95% CI = 1.62–3.67). (Conway et al., 2015)

There is some evidence to suggest that occupational exposures can increase an individual's risk of HNC. IARC have published detailed monographs on the risks of occupational exposures, with some industrial occupations and exposures particularly relevant for tumours of the larynx. (Loomis et al., 2018) Analysis from the INHANCE consortium found those in vocational jobs such as construction, cleaning or painting were found to be at a higher risk of HNC, with a dose-response relationship according to duration of employment. (Khetan et al., 2019) Additionally, occupational socioeconomic status / prestige was found to be important in another INHANCE analysis with these effects not explained by occupational exposures. (Conway et al., 2021b)

Ultimately, while behaviours (especially smoking and alcohol) are major drivers of HNC, they only partially explain the massive socioeconomic inequalities

observed in head and neck cancer. Other socioeconomic mechanisms of HNC risk could include “direct” pathways, e.g., theories of psychosocial stress and its impacts on the immune system or biological aging. (Conway et al., 2019a)

1.4.3 Ethnicity

Ethnicity is often considered as a social determinant for HNC. Generally, ethnicity is not considered to be an independent risk factor for head and neck cancer, with incidence disparities among ethnic groups often attributed to socioeconomic inequalities and behavioural (smoking and alcohol) patterns. (Daraei and Moore, 2015, Stingone et al., 2013) However, ethnic-specific risk factors have been found to be significant, for example, betel quid usage among Asian communities and variation in alcohol and smoking patterns among different ethnic groups. (Voltzke et al., 2018, Hashim et al., 2019b) (see Section 1.4.1.2).

There may be evidence to suggest that those of certain ethnic backgrounds could be more genetically predisposed, e.g., individuals with alleles of the ALDH2 (alcohol metabolism) gene more commonly found in Asian populations (see Section 1.4.9). Ethnic inequalities have also been observed in survival, even when adjusted for SES. (Russo et al., 2020)

1.4.4 Sexual Behaviours and Human Papillomavirus

There are over 100 types of Human Papillomavirus (HPV) and 15 of these are associated with carcinogenesis. The virus is associated with anogenital cancers, in addition to tumours of the oropharynx. (Nassif et al., 2022) In particular, HPV-16 is the primary type associated with OPC, which has been identified in 90% of HPV positive OPC cases. (Kobayashi et al., 2018) Reviews of primarily case-control studies have suggested that persons who engage in oral sex, have a greater number of sexual partners and a younger sexual debut are at an

increased risk of HPV infection and HPV-related OPC. (Chancellor et al., 2017, Durrant et al., 2024)

HPV has been identified as one of the main emerging drivers underlying the changing epidemiology of head and neck cancer, particularly cancers of the oropharynx. (D'Souza et al., 2007, Gormley et al., 2022a, Gillison et al., 2015) The number of HPV-associated HNC cases is increasing, with persistent infection by the virus being attributed as the most common cause of OPC. HPV infection has been attributed to roughly 30-60% of Oropharyngeal cases. (Menezes et al., 2021) HPV has not been found to have an aetiological role in other HNC subsites. (Menezes et al., 2021)The role of HPV in OPC development has been observed among both sexes at similar rates. (Sabatini and Chiocca, 2020a)

There is an increasing number of reports that the sociodemographics of people with HPV-positive OPC differ from other HNC sites. HPV-positive OPC cases have been reported to be younger than their HPV-negative counterparts and not share the behavioural characteristics of HPV-negative OPC such as a history of actively smoking and having a high alcohol intake. (Deschler et al., 2014a, Young et al., 2015) However, there is also evidence to suggest that a notable proportion (approximately 30%) of people with HPV-positive OPC also report high levels of smoking or alcohol usage. (Anantharaman et al., 2016a) The analyses of sociodemographic factors in OPC (with relation to HPV) have primarily originated from smaller clinical cohorts and clinical reports, rather than larger epidemiological studies.

1.4.5 Diet / Nutrition

Diet and nutritional factors are considered as both risk and protective factors for many cancers. (Clinton et al., 2020, Malcomson et al., 2023) Nutritional epidemiology is challenging to investigate; limitations in heterogeneity, measurement and study design mean that the role of diet in cancer and other disease is not yet fully understood. (Willett, 1987) However, various pooled case-control analyses and cohort studies have assessed the role of dietary

factors in the risk of head and neck cancer. While overshadowed by major predictors such as smoking and alcohol, INHANCE pooled analysis found that fruit and vegetable intake were the most significant dietary predictors when assessing head and neck cancer risk; relative to those with a lower fruit / vegetable intake, a higher frequency of consumption was associated with a protective effect. (Freedman et al., 2008, Chuang et al., 2012a)

Relative to those who consumed the least fruit, a higher (quartile) consumption of fruit was associated with a protective effect (4^{th} vs 1^{st} OR = 0.52, 95% CI 0.43 – 0.62). Similarly, relative to those who consumed the least vegetables, a protective effect was associated with regular consumption (4^{th} vs 1^{st} OR =0.66, 95% CI 0.49 – 0.90). (Chuang et al., 2012a) These protective effects have been observed across all age-groups. (Toporcov et al., 2015) Other notable but less significant dietary risk factors included meat and animal product consumption – in particular, red and processed meat. Dietary effects on HNC risk were also largely consistent across subsites. (Chuang et al., 2012a, Edefonti et al., 2012, De Vito et al., 2019)

1.4.6 Body Mass Index (BMI)

A higher body mass index (i.e. overweight or obese) is recognised by the WHO as a risk factor for many cancers. (Lauby-Secretan et al., 2016) In contrast, analysis of case-control studies has mostly suggested that lower BMI is a risk factor for head and neck cancer, with BMI exhibiting an inverse relationship. (Gaudet et al., 2010a, Park et al., 2011) However, studies have also noted the possibility of reverse causation underlying this. (Maasland et al., 2015b) (Park et al., 2011) Many of the analyses suggestive of low BMI as a risk factor for HNC originate from case-control studies, which are subject to potential confounding and reverse causality. Cohort studies for other cancers have suggested that, conversely, an increased BMI (i.e. being overweight or obese) is a risk factor for developing cancer. Thus, caution should be exercised when considering BMI associations and HNC risk.

1.4.7 Oral Health, Oral Hygiene, Oral Microbiome

While not considered a major risk factor, poor oral health has been suggested to be associated with a modest increase in risk of head and neck cancer. (Hashim et al., 2016a) Numerous systematic reviews and meta-analyses of case-control studies have identified an increased risk of HNC associated with poor oral hygiene, dental disease and irregular attendance at a dental practice. (Zeng et al., 2013, Gopinath et al., 2020, Bai et al., 2023) In particular, periodontal disease has been suggested to be significantly associated with head and neck cancer, even when adjusted for smoking behaviours; relative to those without periodontal disease, patients with periodontal disease were at a significantly increased risk OR = 3.17, 95% CI, 1.78 – 5.64). (Gopinath et al., 2020). Systematic review evidence has also indicated a protective role associated with regular dental attendance; compared with regular attendance at a dentist, irregular or never attendance was associated with an increased risk of all HNCs (OR = 2.24, 95% CI 1.89 to 2.65). (Gupta et al., 2019)

Currently, there is no strong review evidence to suggest that the oral microbiome is a major HNC risk factor. However, there have been suggestions of potential associations between the oral microbiome and head and neck cancer, particularly related to alcohol metabolism. (Smędra and Berent, 2023) Differences in the microbiome among HNC patients and disease-free controls have also been observed although their effects are not fully understood. (Dorobisz et al., 2023) Similarly, there may be evidence to suggest a protective effect from some commensal bacteria. (Hayes et al., 2018) Ultimately, further research in this area is needed.

1.4.8 Comorbidities and Medications

At the time of writing, there is a limited quantity of literature which evaluates the role of pre-existing comorbidities for HNC risk. However, there is some evidence to suggest comorbidities are a common and important feature among people with head and neck cancer. (Piccirillo, 2000) Analysis of the ARCAGE case-control study revealed that a history of prior infections (candidiasis, warts

and verrucae) was associated with a reduced risk of HNC. No associations with gastro-oesophageal reflux and HNC were observed. (Macfarlane et al., 2012) Other large cohort studies have observed associations between diabetes and HNC risk in Asian populations. (Tseng et al., 2014, Choi et al., 2022)

There is mixed evidence on the role of medications and HNC risk. Analysis of ARCAGE case-control study suggested a protective effect was associated with aspirin usage, however this was confined to tumours of the larynx and hypopharynx. (Macfarlane et al., 2012) A systematic review conducted by Herrán et al. found a protective effect associated with metformin and NSAID usage against HNC. (Saka Herrán et al., 2018) However, there was large heterogeneity in study quality, and many did not assess key confounders such as smoking and alcohol. Another review by Wilson et al. found no clear effect associated with NSAIDs or aspirin consumption on risk. (Wilson et al., 2011)

1.4.9 Familial and Genetic Risk Factors

There has been a growing interest in potential hereditary and genetic components to an individual's risk of head and neck cancer. Pooled analysis has shown that a family history of head and neck cancer has been shown to be a significant risk factor, with individuals being at least twice as likely to develop head and neck cancer if they have a relative with a family history of the disease, especially if the relative was a sibling. This increased significantly if those with a family history were also alcohol and tobacco users (OR = 7.21, 95% CI 5.46 – 9.54). (Negri et al., 2009)

There has been an increasing research emphasis on the role of individual genotypes in the development of HNC. (Leemans et al., 2018) Various genome wide association studies (GWAS) and case-control analyses have assessed the risk of head and neck cancer among cases and controls associated with genetic polymorphisms. GWAS have identified several loci that could be implicated in individual HNC susceptibility. Notably, protective effects against OPC were associated with variations within the Human Leukocyte Antigen (HLA) region. (Lesseur et al., 2016) (Shete et al., 2020) Polygenic risk scores, a composite

score of genetic variants, have also been used to assess genetic risks for HNC. (Lee et al., 2024)

Genes associated with tumour suppression and cell proliferation have been shown to be associated with the development of HNC. (Leemans et al., 2018) Research has also found risks that are attributable to alleles of the (ALDH2) gene, which are responsible for the metabolism of acetaldehyde in the breakdown of alcohol. (Boccia et al., 2009) These alleles are more commonly found in Asian populations and result in poorer metabolism of acetaldehyde, a known carcinogen. (Seitz and Stickel, 2010) There has also been an increasing emphasis on the epigenetics of HNC, with suggestions that DNA methylation, histone modifications, and various signalling pathways can influence individual risk of HNC. (Liu et al., 2022)

Other genetic risk factors identified by research include null or dysfunctional genotypes which encode for enzymes that perform detoxication functions. Dysfunction of the p53 gene, responsible for halting cell cycle or initiating programmed cell death, was also found to be a significant predictor of HNC. (Cadoni et al., 2012) (Hiyama et al., 2008) However, a common agreement among genomic HNC research is that genetic factors may vary by subsite and HPV serostatus, and more extensive research is required. Currently, there are no genetic biomarkers of strong risk prediction utility. (Kasradze et al., 2020)

1.5 Primary Prevention of Head and Neck Cancer

Primary prevention is defined as interventions aimed at preventing a disease in susceptible individuals or populations. (Kisling LA, 2023) There are several primary prevention strategies that can be utilised against HNC. Primary prevention against HPV-related cancers of the oropharynx can be achieved via the use of the HPV vaccine. (Kreimer, 2014) (Nassif et al., 2022) This has already been used in a preventive role against cervical and other anogenital cancers in school-age girls, but vaccination programmes have only recently been expanded to school-age boys in some nations (2019 in Scotland). (Public Health Scotland, 2023c)

Systematic review of studies has shown that oral HPV prevalence among males has reduced with the vaccine, which will have potential implications for the future OPC burden. (Tsentemeidou et al., 2021, Macilwraith et al., 2023) HPV-vaccination strategies show great promise, particularly given recent reductions in cervical cancer; no novel cervical cancer cases were recently recorded among women vaccinated at 12-13 years of age in Scotland. (Palmer et al., 2024b) However, given the longer latent period between HPV infection and development of OPC, projections suggest that it will take decades for the protective effects of vaccines against OPC to become apparent. (Zhang et al., 2021)

Other HNC primary prevention strategies pertain to the management of modifiable risk factors described in Section 1.4. At a policy level, HNC prevention can be implemented via modification of environmental factors. This can be achieved by the regulation of tobacco and alcohol, for example, smoking bans or alcohol price policies. (Wilson et al., 2012, Frazer et al., 2016, Rekve et al., 2019) As such, the WHO details alcohol and smoking policy recommendations among their “Best Buys” to reduce the burden of non-communicable disease. (WHO, 2017)

The International Agency for Research on Cancer has published evaluations and guidance on the primary prevention of HNC, with an emphasis on the benefits of cessation interventions. (Bouvard et al., 2022, Gapstur et al., 2023) Research has found that the cessation of smoking can greatly reduce an individual’s risk of head and neck cancer, a significant reduction in risk being observed in as little as 1-4 years. (Marron et al., 2009) Alcohol cessation was also found to exhibit a beneficial protective effect. However, this was only the case over a much longer period (>20 years). (Marron et al., 2009) Nevertheless, there are still major benefits in the cessation of alcohol and smoking.

At a primary and secondary care level, clinicians can support patients with a higher risk of head and neck cancer with behavioural advice, adjustment of recall intervals and signposting or referrals to support services. (Mathur et al., 2015) In primary care, this is commonly done via a “brief intervention” where behavioural advice and motivational counselling is given to patients with risk factors during a visit to a healthcare provider. (BIEN et al., 1993) As such, the

dental setting has been identified as an excellent setting for tobacco cessation interventions due to the wide population reach, regular patient contact and ability to communicate the health and aesthetic benefits of smoking cessation orally and systemically. (Tomar, 2001)

The dental setting has also been shown to be effective in the delivery of smoking cessation interventions. (Omaña-Cepeda et al., 2016) In the UK, clinical guidance advises the use of brief interventions in the prevention of HNCs. This is detailed in the “Delivering Better Oral Health Toolkit” for smoking and alcohol misuse, a clinical guideline document which has systematically appraised the literature on prevention in dental settings.(Department of Health and Social Care, 2021b, Department of Health and Social Care, 2021c) Currently, no HNC risk prediction models are used as a primary prevention tool in the dental setting to support decision making or assist with behavioural counselling. This has been identified as a gap in the literature.

1.6 Secondary Prevention of Head and Neck Cancer

Despite the global population growth and the growing incidence of all cancers, an overall reduction in cancer mortality rates has been observed over the recent decades. In addition to improved treatment outcomes and availability, this has also been attributed to screening and early detection. (Stang and Jöckel, 2018, Siegel et al., 2021, Global Burden of Disease Cancer Collaboration, 2022) Screening is a prevention strategy commonly defined as a test applied to a population that differentiates between those who may have disease and those who may not. (Wilson et al., 1968) A screening programme must strictly meet 19 criteria detailed by the United Kingdom's National Screening Committee. (UK National Screening Committee, 2003); for example, a screening programme must be cost-effective, have a proven reduction in mortality and be supported with appropriate evidence base.

In the UK, the NHS currently provides several cancer screening programmes for cancers of the cervix, breast and colon. (NHS Online, 2022). These have been successful in reducing incidence and consequent mortality rates. For example, a

systematic review and metanalysis of observational colorectal screening studies has found that screening measures reduced colorectal cancer incidence by 64% (95% CI, 50 to 74%). (Brenner et al., 2014) Similarly, assessment of breast and cervical cancer screening revealed significantly reduced mortality risks among participants of screening programmes. (Massat et al., 2016, Landy et al., 2016)

To date, no "true" screening programmes for head and neck cancer have been implemented. Head and neck, especially OCC, examination and prevention have been discussed in the literature. However, HNC examination does not always meet the specific criteria of screening for several reasons. (Speight et al., 2017) First, the evidence base for HNC screening programmes and their evaluation is limited; to date, the only trial which has evaluated the potential of screening was the Kerala randomised control trial conducted in India. (Sankaranarayanan et al., 2005) Four rounds of OCC screening using conventional intra-oral examinations were conducted over the study duration. The results from this trial found that the use of standard intra-oral examination as a screening tool did not reduce overall mortality but that the use of targeted examination in higher-risk groups could have benefits. (Subramanian et al., 2009) Similarly, other reviews have concluded that while there may be benefits from examination targeted to higher-risk groups, there is insufficient evidence to support the use of OCC screening programmes. (Brocklehurst et al., 2013, Speight et al., 2017)

The closest equivalent to HNC "screening" are opportunistic routine intra- and extra-oral exams carried out by general Dental Practitioners (with other allied dental staff) and the use of suspected lesion referral pathways. (Al-Helou, 2021) (NICE, 2015) While a sensitive examination technique, routine intra-oral exams are not specific and subject to other limitations as they do not entirely meet the criteria of an effective screening test (e.g., specificity). As such, diagnostic biopsy and histological assessment remain the gold diagnostic standard. (Walsh et al., 2021) While some biomarkers show potential, there is no clinical trial evidence to support the use of a biomarker test or test adjunct in an HNC screening programme. (Speight et al., 2017) Ultimately, intra-oral examination remains an important tool in the arsenal of secondary prevention against HNC. "Opportunistic screening" is advised with standard intra-oral examination for all patient groups as per the findings of the Kerala trial and subsequent reviews.

(Brocklehurst et al., 2013, Speight et al., 2017) Delivering Better Oral Health / NICE recall guidance indicates that in dental settings, patients at a higher risk of OCC (defined as smokers, high alcohol consumption) should have more frequent recall intervals. (Department of Health and Social Care, 2021a)

The themes of this thesis are especially relevant in the context of the recent developments in the NHS. NHS management recently published their “long term plan”; one of the long-term objectives described is a focus on early cancer detection, with the aim to have 75% of cancers diagnosed at an early stage (I or II) by 2028 (NHS UK, 2022). Given the aforementioned challenges associated with head and neck cancer, especially late presentation, an emphasis will need to be made upon primary and secondary prevention strategies.

1.7 Management / Treatment of Head and Neck Cancer

Suspected HNC is typically managed using rapid referral pathways. In the UK, NICE clinical guidelines advice an urgent suspicion of cancer (USOC) referral (aiming to provide an appointment in two weeks) for those with persisting signs and symptoms discussed in Section 1.1.2. (NICE, 2015) Referrals are usually from primary care (General Medical or Dental Practitioner), but patients can also present with emergency symptoms or be referred from other specialties. Urgent referral pathways have improved the detection rate of cancers; however, it is argued by some that the efficacy of these pathways is limited or that it requires further refinement. (Lyons et al., 2004, McKie et al., 2008, Mettias et al., 2021)

In the UK, there are standardised multidisciplinary guidelines for the management of HNC. (Homer and Winter, 2024) Following a diagnosis, a multidisciplinary team (MDT) meeting is conducted where relevant healthcare professionals meet to determine (often complex) treatment plans. Treatment can impact many of the complex systems and functions of the head and neck, such as eating and speech. Thus, treatment necessitates a coordinated effort across several specialties including surgeons (OMFS and / or ENT), oncology (radiation or medical), dietary specialists, speech and language therapy, dental restorative specialists, social workers and many more. (Taberna et al., 2020)

There are several treatment strategies that can be used for HNCs, with curative or palliative intent. Treatment guidelines vary globally; a global review identified surgery and adjuvant radiotherapy as the main treatment modalities, however, in lower-income counties with fewer resources, there is a greater emphasis on surgery with more limited use of radiotherapy treatments. A lack of clinical guidelines was identified in Oceanic and Latin American counties, where a significant burden of HNC occurs. In the UK, the primary modalities of curative treatment are surgical excision and adjuvant radiotherapy. Advanced-stage disease may involve more radical surgical approaches (e.g., total laryngectomy, glossectomy) and the use of donor sites to reconstruct lost bone or soft tissue (e.g., radial forearm, latissimus dorsi, anterolateral thigh, fibula or pectoralis major flaps). (Homer and Winter, 2024) More aggressive surgical excisions and chemo-radiotherapy can have significant quality of life implications (section 1.3.4).

Other treatment modalities include radiotherapy, chemotherapy, immunotherapy (*e.g.* cetuximab), best supportive care or combinations of these. (Homer and Winter, 2024) Where more advanced disease exists, radiotherapy and / or chemotherapy is often used adjunctly in combination with surgical excision. (Johnson et al., 2020b) (Homer and Winter, 2024) Management can depend on the stage and grade of the cancer; if caught at an early stage, some cases can be managed effectively by local surgical resection alone. Research has shown that multidisciplinary team management has a positive effect on outcomes and survival rates. (Shang et al., 2021)

1.8 Risk Prediction Models and Tools

In this thesis, a risk prediction model was defined as a statistical model that predicts the risk of an outcome of interest, combining information from several variables. (Janes et al., 2008) A risk tool was defined as a model that has been implemented in clinical practice, for example via software. (Usher-Smith et al., 2015) Algorithms and models have been used in many care settings, for example, the widely-used Glasgow Coma Score (GCS) which is used to assess a patient's

consciousness in acute care settings. (Mehta and Chinthapalli, 2019) Models and algorithms can also be used for risk assessment purposes. Risk prediction models have already seen use in prevention and screening efforts for cardiovascular disease and cancers of other subsites. Many models that assess patients' risk of breast, colorectal, lung and skin cancer have been developed. (Anothaisintawee et al., 2012, Gray et al., 2016, Louro et al., 2019, Usher-Smith et al., 2016, Williams et al., 2016, Usher-Smith et al., 2014) Many of these models have undergone extensive review processes. Some of these models have seen clinical usage, including the Q-Cancer and Cancer Risk Assessment Tools (RATs). (Clin Risk) (Cancer Research UK, 13th July 2020).

Evaluation of such risk assessment models in the primary care setting has shown that they do have predictive utility, in addition to other beneficial characteristics including prompting behavioural change. A systematic review and metanalysis of cancer risk tool trials by Walker et al. identified that health messages within tools had a positive effect on behavioural change and had the potential to improve patient knowledge and the understanding of risk. (Walker et al., 2015b) However, a common finding of reviews of cancer risk models is the necessity for further evaluation of their implementation and validation in primary care settings. (Usher-Smith et al., 2015) (Medina-Lara et al., 2020). Cross-sectional research conducted in primary care has found that such tools are often underutilised in UK general medical practices, with low uptake rates despite being less resource-intensive than other investigations (16.7%, 95% CI = 12.1 to 22.2). (Price et al., 2019) Furthermore, at the time of writing, no cancer risk tools are utilised in UK general dental practices.

It was found that there are several barriers and facilitators to the development and use of cancer risk tools in primary care. (Akanuwe et al., 2021) Common barriers included time pressures for consultations, potential to induce anxiety in patients, additional training requirements, potential for over-referral and logistical challenges of integrating novel tools into practice IT systems. Conversely, facilitators to the uptake of tools included improved tailoring of care to a patient, the potential to induce behavioural change, and aiding clinical decision making and improving referral times.

Despite the clear clinical potential of cancer risk prediction models and tools, only a limited number of head and neck cancer risk models exist. Initial literature searches identified at least four head and neck cancer risk prediction models. (Lau et al., 2018) (Lee et al., 2020) (McCarthy et al., 2020) (Tikka et al., 2020). Three of these models were developed in the UK and the remaining one was developed in the USA. The models were developed using case-control or cohort data, one of which evaluated individual risk at multiple, separate subsites (Lee et al., 2020). The remaining three calculated an overall risk of head and neck cancer. Only one risk model was found to have seen any clinical use, which was used to triage patients during the Covid-19 pandemic. (Hardman et al., 2021b)

While comprehensive reviews exist for risk models for cancers of other sites (e.g. breast and colorectal), there were no existing reviews of head and neck cancer risk prediction models and tools. (Anothaisintawee et al., 2012) This was identified as a significant gap in the literature and a primary aim of this thesis.

1.9 Summary of Literature Review Findings

This Chapter has reviewed the background and literature on the epidemiology of head and neck cancer, covering the disease burden and trends alongside risk factors. Additionally, the potential prevention approaches for head and neck cancers have been reviewed.

Head and neck cancer is a devastating disease with an increasing global burden. Major risk factors include HPV, low individual socioeconomic status, smoking and alcohol behaviours, while minor risk factors include poor oral health, occupational exposures, BMI, dietary and metabolic factors. Prevention remains a challenge with limited use of risk prediction support tools in this effort. With an aging population and an increasing proportion of disease diagnosed at advanced stage, there is a growing need for an emphasis on prevention and early detection strategies, as described by the Scottish Cancer Strategy. (Scottish Government, 2023a)

This chapter has also highlighted some evidence gaps in the literature around the availability, quality and use of risk prediction models in primary care for both primary and secondary prevention of head and neck cancer. Additionally, there is limited knowledge on the epidemiological pattern of trends over recent decades in the sociodemographic profile of head and neck cancers, especially by subsite, which warrants detailed investigation.

2 Chapter Two: Research Aims and Objectives

Chapter Two describes the overarching aims and objectives of this doctoral thesis. Based upon the review of existing literature of HNC, risk determinants and cancer risk prediction, this thesis aimed to answer the following research questions: Could a head and neck cancer risk prediction model be developed from a case-control study and validated in a UK population cohort, with a view to implement in a primary dental care setting. Additionally, what is the sociodemographic profile of HNC, and is this changing?

2.1 Overarching Aims

The overarching aims of this thesis were to assess the epidemiology and sociodemographic profile of HNC and to develop a clinical risk prediction model to predict individual HNC risk. These aims will be accomplished by fulfilling the thesis research chapter aims and objectives described below.

2.2 Chapter Aims and Objectives

Chapter Three: Risk prediction models for head and neck cancer: a rapid review.

The aims of this chapter were to:

i. Identify the range of head and neck cancer risk prediction models from around the world.

ii. Understand their uses and quality appraise the methods deployed to develop and validate them.

The objectives of this chapter were to:

- Identify and collate any existing HNC risk prediction models using a systematic search of multiple databases.

- Describe each model's study and performance characteristics.

- Conduct quality assessment of each model using the PROBAST quality assessment tool and further assessment of validation methods and clinical applicability.

- Undertake a narrative synthesis and discussion of the models, following an independent quality assessment.

Chapter Four: Head and neck cancer incidence is rising but the sociodemographic profile is unchanging: A population epidemiological study (2001-2020).

The aims of this chapter were to:

i. Investigate contemporary trends of the incidence of HNC in Scotland.

ii. Investigate whether the sociodemographic profiles of head and neck cancer and subsites are changing over time.

The objectives of this chapter were to:

- Describe and analyse trends in the incidence of HNC and key subsites over the study period using data from the Scottish Cancer Registry.

- Assess the sociodemographic profile in the incidence of HNC and main subsites by age, sex and area-based socioeconomic status.

- Assess whether the sociodemographic of people with HNC is changing over time.

Chapter Five: Socioeconomic status and HPV-related cancers of the oropharynx in the ARCAGE Study

The aim of this chapter was to:

Assess the relationship of oropharynx cancer with sociodemographics and behaviours, considering human papillomavirus (HPV) serostatus in an international multicentre case-control study (ARCAGE – Alcohol Related Cancers and Genetic-susceptibility in Europe).

The objectives of this chapter were to:

- Identify and stratify oropharyngeal cancer cases in the ARCAGE study according to their Human Papilloma Virus (HPV) serostatus.

- Describe and quantify any associations among the sociodemographics (socioeconomic status, along with age, sex) and behaviours (tobacco smoking and alcohol consumption) of **HPV-positive** oropharynx case participants versus **controls participants**.

- Describe and quantify any associations among the sociodemographics (socioeconomic status, along with age, sex) and behaviours (tobacco smoking and alcohol consumption) of **HPV-negative** oropharynx case participants versus **controls participants**

- Describe and quantify any associations among the sociodemographics (socioeconomic status, along with age, sex) and behaviours (tobacco smoking and alcohol consumption) of **HPV-negative oropharynx case participants** case versus **HPV-positive oropharynx case participants.**

Chapter Six: Development and external validation of a head and neck cancer risk prediction model.

The aims of this chapter were to:

i. Develop a head and neck cancer risk prediction model using the ARCAGE case-control study.

ii. Externally validate the head and neck cancer risk prediction model in the UK Biobank cohort study.

The objectives of this chapter were to:

- Conduct a descriptive analysis of data on exposures associated with head and neck cancer from the ARCAGE case-control study.

- Using data from the ARCAGE case-control study, develop a multivariable head and neck cancer risk prediction model that can accurately predict and quantify an individual's risk of head and neck cancer.

- Validate this model in the UK Biobank cohort and assess model performance.

Chapter Seven: Human Papillomavirus (HPV) serostatus and burden in the UK Biobank Study

The aims of this chapter were to:

i. Explore the burden of HPV seropositivity among HNC cases and non-events in the UK Biobank.

ii. Assess the feasibility of multiple imputation to address missing HPV tumour data.

The objectives of this chapter were to:

- Describe the existing HPV biomarker data within the UK Biobank in the context of head and neck cancer.

- Test for associations between HPV serostatus and sexual behaviour data or other demographic data.

- Discuss and quantify the strength of any associations, with a view to assess the suitability of multiple imputation as a technique to address missing HPV data in the study.

Chapter Eight: Discussion.

The objectives of this chapter were to:

- Summarise the findings of this thesis.
- Discuss the findings of this thesis in the context of the wider literature.
- Discuss the strengths and limitations of the thesis studies.
- Draw conclusions on the key findings and propose recommendations for practice, policy and research.

3 Chapter Three: Risk Prediction Models for Head and Neck Cancer: a Rapid Review

The contents of this chapter were published to the journal Laryngoscope Investigative Otolaryngology on the 28th of November 2022.

doi: 10.1002/lio2.982 PMCID: PMC9764804 PMID: 36544947
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9764804/
"Risk prediction models for head and neck cancer: A rapid review"

Craig D. L. Smith (corresponding author)[1, 2], Alex D. McMahon[1], Alastair Ross[1], Gareth J. Inman[2,3], David I. Conway[1,2]

Affiliations

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

3.1 Abstract

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

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

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

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

3.2 Background / Aims

Head and neck squamous cell carcinomas – generally defined as aerodigestive squamous cancers of the oral cavity, larynx, and pharynx – are a growing challenge for healthcare systems across the world: they are the 8th most common cancer, accounting for an estimated 878,348 new cases and 444,347 deaths globally in 2020. (Bray et al., 2018) (Global Cancer Observatory, 2020) The risk profile of head and neck cancer is also changing – with oropharyngeal cancer increasingly associated with human papillomavirus (HPV) infection (Chaturvedi, 2012), and inequalities in the burden of HNC associated with socioeconomic status. (Conway et al., 2010)

Overall head and neck cancer (HNC) survival varies greatly by subsite and stage of diagnosis. Despite advancement in treatments, 5-year survival has seen no major improvements observed in recent decades.(Hoffman et al., 1998) (Pulte and Brenner, 2010b) (Carvalho et al., 2005) Furthermore, marginal improvements in prognosis may be undercut by the overall increased disease burden, particularly due to the changing epidemiology of HPV-associated oropharyngeal cancer.(Guo et al., 2021) (Argirion et al., 2019)

As with all cancers, prognosis is worse with advanced stage disease at presentation. Thus, a major challenge posed by head and neck cancer is its traditionally late presentation with over half of cases diagnosed at stage III or IV, when locally advanced or regional or distant metastases are present.(Sanderson and Ironside, 2002a) (Northern Ireland Cancer Registry, 2019) (Lagiou et al., 2009) (Abrahão et al., 2018) (Abrahão et al., 2020)

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

There has already been some success and clinical adoption of other cancer prediction models in primary care; for example, the Q-series risk prediction models or cancer Risk Assessment Tool (RAT).(Cancer Research UK, 2020a) (Risk,

2013) These models have been well evaluated, demonstrating the potential of "personalised medicine" to identify and stratify those at risk.(Jackson and Chester, 2015) (Usher-Smith et al., 2015) (Medina-Lara et al., 2020) (Collins and Altman, 2013) (Kostopoulou et al., 2022) However, they do not assess for head and neck cancer risk and there seem to be few risk prediction models for head and neck cancer developed or adopted for clinical use. Furthermore, there have been no comprehensive reviews of head and neck risk cancer prediction models or tools published.

The aim of this study is to undertake such a review – via systematically searching and identifying models in the international literature, describing their characteristics and performance, quality appraising these models, and performing a narrative synthesis to compare and contrast risk prediction models for head and neck cancer.

3.3 Methods

A rapid review methodology was employed, following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.(Page et al., 2021) The review was also based on similar reviews on risk prediction models of other cancer sites / diseases.

3.3.1 Search Strategy

An electronic literature search of Ovid MEDLINE(R), (and In-Process, In-Data-Review & Other Non-Indexed Citations), and Embase (1947-Present, updated daily) databases was conducted using the following combination of key headings and search terms:

[(((risk prediction or risk factor or risk model or risk assessment or risk calculator or risk tool or risk score) and (cancer* or tumour or tumor or neoplas* or malignan* or squamous cell carcinoma) and ((head and neck) or oral cancer or oral cavity cancer or oropharynx cancer or oropharyngeal cancer or larynx cancer or laryngeal cancer leukoplakia or erythroplakia or submucous fibrosis or OPMD or Oral Potentially Malignant Disorder)) not (prognos* or survival))

Abstract + title (tw), no filters date: 22:03 23/09/2021 Results: Embase: 1787 results, Medline: 1258 results]

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

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

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

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

3.3.2 Screening and Study Selection

Two reviewers (CS, DIC) independently screened search results at a title / abstract level and then at a full-text level using the eligibility criteria. In the

event of a disagreement, articles were discussed and included or excluded by mutual consensus.

3.3.3 Data Extraction and Quality Assessment

Following full-text screening, data extraction was undertaken by two reviewers (CS, DIC) using a customised form containing pre-defined fields including items on: study characteristics (location, study design, cancer site/subsite and risk factors included. The data extraction form also assessed the requirement of clinician input (based on whether reported or the nature of the data required to run the model – for instance a patient would not be able to use machine learning tools or conduct HPV serology analysis), along with items on predictive performance (discrimination, sensitivity / specificity, calibration, PPV/ NPV and risk threshold cut-offs) and the method of validation (if undertaken). Measures of discrimination were considered to be "acceptable" if an Area Under the Curve (AUC) value over 0.7 was reported, and "excellent" if a value greater than 0.8 was reported.(David W. Hosmer Jr., 2013) Measures of calibration were assessed by the Expected/ Observed ratio or gradient of a calibration slope to the ideal value of 1 and of its intercept to the value of 0.(Richard D Riley, 2019) (Van Calster et al., 2019)

Two reviewers (CS, DIC) also examined the risk of bias of each model using PROBAST, a tool specifically designed to appraise clinical risk prediction models.(Wolff et al., 2019) Risk of bias ("high", "low", or "unclear") and applicability of the risk models was assessed using 20 questions across four domains (participants, predictors, analysis and outcomes).

An overall quality assessment was also given to each model considering model validation, and the risk of bias, and applicability concerns assessment (from PROBAST). This was categorised as "High", "Moderate", or "Low". If the model had a (i) low risk of bias, (ii) a low or unclear applicability concern, and (iii) a robust method of validation then it was considered "High" overall quality. Model performance was also considered separately by evaluating each model's discriminative ability. If a model achieved an "excellent" AUC over 0.8 it was classed as high performing (green in table).(Mandrekar, 2010) Models that achieved acceptable discrimination between 0.7 and 0.8 were classed as

moderate (amber in table) and discrimination less than 0.7 was classed as poor (red in table). These classifications also reflected for confidence intervals for model AUC (where reported).

3.3.4 Synthesis

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

3.4 Results

Following the removal of 100 duplicates, 2945 studies were identified by the search. Of these, 2900 were excluded by title or abstract screening. Of the remaining 45 studies, 34 were excluded following full-text assessment, with reasons for exclusion noted (Figure 3-1). The most common reasons for exclusion were studies did not use a statistical method to develop a risk model, or did not consider multiple risk factors together, or on further examination were a duplication of model already included. One conference abstract paper was excluded because the full text was not available despite attempts to contact the author.

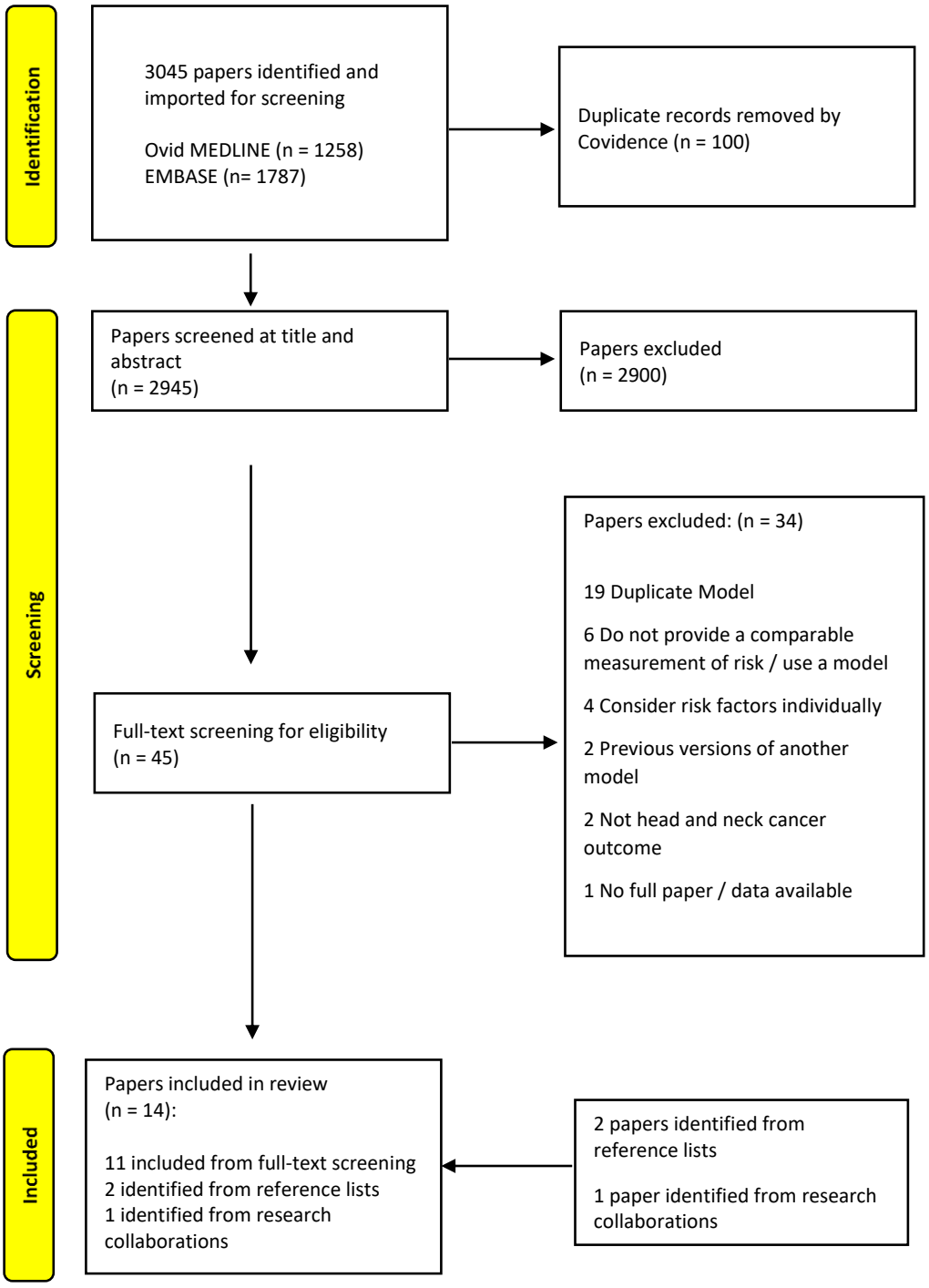


Figure 3-1: Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow diagram

A further three articles were identified – two of these were identified from reviewing the reference lists and the third (at the time of writing yet to be published) was identified from one of reviewer's research collaborations (DIC). Thus, in total, 14 papers were ultimately included in this review (Figure 3-1). (Amarasinghe et al., 2010) (Budhathoki1 et al., 2021) (Chen et al., 2018)

(Cheung et al., 2021) (Koyanagi et al., 2017) (Krishna Rao et al., 2016) (Lau et al., 2018) (Lee et al., 2020) (Liu et al., 2017) (McCarthy et al., 2020) (Sharma and Om, 2015) (Sun et al., 2019) (Tikka et al., 2020) (Tota et al., 2019c) Within these studies, three of the 14 models featured “sub-models”, using broadly similar methods but stratifying models by subsite or sex.(Budhathoki1 et al., 2021) (Chen et al., 2018) (Lee et al., 2020) These have been reported and considered accordingly, where reported.

3.4.1 Study / Model Characteristics

A summary of the 14 studies including model characteristics and performance information is shown in Tables 3-1 and 3-2 respectively. Of the 14 models, three were conducted in India, one in Sri Lanka, three in China, three in the USA, three in the UK, and one in multiple centres across Northern America / Europe. The majority of the studies (11 out of 14) used a case-control design or built upon previous case-controls for model development. These were mostly of a hospital-based case-control design. One study used data collected from a randomised control trial.(Cheung et al., 2021) The other two used a cross-sectional and prospective cohort study method respectively. There were 11 models which utilised a form of logistic regression analysis approach to evaluating head and neck cancer risk. Two of the three remaining models used machine learning methods, while one used a cox-regression approach to evaluate risk.

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

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

	Study Characteristics		Components of Model		
Study ID	**Country / Year / Study design**	**Participants**	**Cancer outcome / site + sub models**	**Method + Factors Included in model**	**Clinician Input**
Amarasinghe et al 2010	Sri Lanka / Community-based case-control November 2006 - November 2007	101 OPMD cases and 728 controls (men and women)	OPMD excluding lichen planus	Multivariate Logistic regression was used to develop the model. Effect estimates of each factor on the risk of OPMD were derived. Gradients for each factor were given a score derived from adjusted odds ratio and ultimately used to develop a 12-point score cut off model. Predictors used included Age, Socioeconomic Status, Betel-quid chewing, Alcohol drinking, smoking.	No
Budhathoki et al (Unpublished)	North America/ Europe - 4 Case-control studies, 1 prospective cohort - a mixture of hopsital and population-based case-controls. Recruitment dates - not specified	10,126 head and neck cancer cases and 5,254 controls	Head and neck cancer (10126), oral cavity cancer (2431) and oropharyngeal cancer (3727) (6 sub models for 3 sites including separate models for men and women)	Predictors were selected based on those that were statistically significant from univariate logistic regression via backwards stepwise selection. The final models used multivariate logistic regression to assess risk with separate models for men and women. Predictors include Smoking status, Drinking status, Education, HPV serostatus and Polygenic risk score	Yes
Chen et al 2018	China - Hospital based Case-control Conducted from September 2010 to March 2017	978 cases and 2646 controls SEPARATE models - 1924 men and 1700 women.	Oral cancer (tongue, buccal, gingiva, floor of mouth, palate, lip, and unspecified or overlapping) (380 tongue, 135 buccal, 128 gingival, 72 floor of mouth, 69 palate, 34 lip, and 160	Unconditional logistic regression was used, independently significant variables being included in the final nomogram models. Different sets of predictors were used to create separate nomograms for males and females. Men; (Smoking (pack-year), Alcohol drinking, Tea consumption, Fish, Seafood, Vegetables, Fruits, Teeth loss, Regular dental visits, Repetitive dental ulcer)	No

			unspecified or overlapping)	Women; (Passive smoking, Cooking oil fume exposure, Tea consumption, Vegetables, Fruits, Beans, Teeth loss, Regular dental visits, Repetitive dental ulcer, Age of first intercourse)	
Cheung et al 2021	India - Analysis of a previous Oral Cancer Randomised Control Trial conducted Screening was conducted in 3 waves with a 3 year interval between each induvial screening - 1996-1998, 1999-2001, and 2002-2004	95,354 control arm 96,516 screening arm (Male and female)	Oral cancer	Model was developed using a Cox regression–based risk prediction model for 7-year oral cancer incidence. Follow up time was used for a time scale and covariates were selected prior to analysis. Predictors included Sex, Age, Education, BMI, Tobacco chewing, Tobacco smoking, Chewing-smoking interaction, and Alcohol use.	No
Koyanagi et al 2017	Japan - Hospital-based case control 2001 - December 2005 365 HNC cases and 1260 controls (Male and female)	365 HNC cases and 1260 controls (Male and female)	3 separate models - Head and neck cancer (model of interest), Upper aerodigestive tract cancer, Oesophageal cancer	Three models were constructed for each subsite: a genetic, an environmental and an inclusive model. The latter, categories of alcohol and ALDH2 were coded as dummy variables. Interaction terms were used to assess the combined impact of alcohol and ALDH2 interaction. The models were derived using conditional logistic regression models. Predictors included Age, sex, ALDH2 genotype, cumulative smoking and alcohol consumption.	Yes
Krishna Rao et al 2016	India - Hospital based, unmatched Case-control Conducted between July 2011 and August 2012. 180 cases and 272 controls (men and women)	180 cases and 272 controls (men and women)	Oral cavity and oropharynx cancer	Multivariable logistic regression was used to identify significant predictors. Those predictors were included in the final model. This and ROCs were used to develop the risk score and cut offs required for further referral. Predictors included - Smoking, chewing tobacco, chewing quid with tobacco, alcohol, spiciness of food, fruit consumption, family history of UADT cancer, rinsing mouth with water after eating/chewing.	No

Study ID	Country / Year / Study design	Participants	Cancer outcome / site + sub models	Method + Factors Included in model	Clinician Input
Lau et al 2018	UK - 2 hospital-based case-control studies (01/07/2009–01/07/2010 (622 patients)) (1/4/2013–31/8/2013 (453 patients)) 73 cases and 932 controls (men and women)	73 cases and 932 controls (men and women)	Head and Neck Cancer	Regularised logistic regression was performed on 60% of the data to identify key predictors. Using information form this, the model was refined and underwent split sample (20%) and cross validation (20%). Predictors include Age, Sex, Smoking, Alcohol and Symptoms.	Yes
Lee et al 2020	USA - Pooled analysis of 14 US-case control studies from INHANCE consortium. Conducted between 1981 and 2010	7,299 HNC cases and 10,301 controls (Male and Female)	Cancer of the oral cavity, oropharynx, hypopharynx, or larynx. (By subsite) 2,388 oral cavity, 2,820 oropharynx, 459 hypopharynx, and 1,632 larynx (Separate models for men and women considering 4 different subsites - 8 sub models)	Logistic Regression Models developed using 70% of the dataset. Hazard and incidence rates from the Surveillance, Epidemiology, and End Results (SEER) Program were applied to the model. Competing risk models were used to ascertain risk by individual subsite but also an overall measure of absolute HNC risk. Models considers Age, sex, race/ethnicity, education, cigarette smoking duration and intensity, and/or alcohol drinking intensity; the second set of models additionally included family history of HNC, except for oropharyngeal cancer in both sexes and laryngeal cancer in men.	No

Liu et al 2017	China - Hopsital-based Cohort study March 2008 to July 2016.	28 OLK, 41 OSCC and 18 controls (men and women)	Risk of developing Oral cancer from Oral Leukoplakia	Peaks RF method was used for model development. Split sample testing was used to evaluate the model in addition to 10-fold cross validation. Subsequently the training set was used to train the model which was used to test the validation set. Predictors included Age, Sex, Site, Smoking and Drinking.	Yes
McCarthy et al 2020	UK - Nested case-control conducted in the UK Biobank 2006 - 2016	702 HNC cases and 423,050 controls (men and women)	Head and Neck cancer excluding laryngeal cancer - no reporting of numbers by subsite	The model was developed using multivariable logistic regression, final predictors being selected upon clinical significance in literature and consultation with a patient and public involvement group - Predictors included Age, Sex, Smoking status, Townsend deprivation index, Body Mass Index, Alcohol consumption, Moderate Exercise and Fruit and vegetable intake.	No
Sharma et al 2015	India - "retrospective chart review" - case control June 2004 to June 2009	1025 patients - not specified if men and women	Diagnosis of Oral cavity cancer - (no reporting of numbers of cases by subsite)	Dataset underwent filter reduction to select attributes for a PNN/GRNN model - Probabilistic and General Regression Neural Network. A leave one out method was then subsequently used for internal cross-validation. Attributes included sex, socioeconomic status, clinical symptom, history of addiction, comorbid condition, gross examination, site, predisposing factor, neck nodes, and tumour size.	Yes

Study ID	Country / Year / Study design	Participants	Cancer outcome / site + sub models	Method + Factors Included in model	Clinician Input
Sun et al 2019	China - Cross sectional study Conducted from August 2016 to May 2018	269 patients: OPMD (n=192) and OSCC (n=77), (men and women)	Risk of developing Oral cancer from OPMD (leukoplakia or oral lichen planus)	Risk factors included in the final model were developed from univariate logistic regression. Multivariate logistic regression was then used to evaluate the risk factors in a model, the beta-coefficient being used to assign a risk score to each factor. Cut-offs were determined using the ROC curve considering sensitivity and specificity. Risk factors included Gender, Age, Lesion site, Local stimulus, and Alcohol drinking	Yes
Tikka et al 2020 (v.2)	UK - Prospective data collection, building upon previous version of model developed using case-control. January 2017 until December 2018	307 HNC cases and 3224 controls (men and women)	Head and Neck cancer - "all primary cancers to the HaN regions (n = 247), metastatic cancers to the HaN from other regions, including lymphoma (n = 48) and cancers in neighbouring regions that manifested with HaN symptoms (n = 12)"	The final multivariate logistic regression was developed using univariate logistic regression and backwards elimination of non-statistically significant variables. The final model was then used for bootstrap validation on the final model. Variables included Age, Gender, Unintentional weight loss, Smoking, Alcohol, Positive and negative symptoms / Signs of HNC.	Yes
Tota et al 2019	USA - Synthetic Case-control study	241 Cases (unweighted) and 9327 controls (unweighted) 12,656 vs 154,532,508 (weighted) - men and women	Oropharyngeal Cancer	The model was developed using multivariable Weighted binary logistic regression. Cases and controls were propensity weighted according to incidence rates in the population identified by National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program. Using this, a one-year absolute risk was calculated. The model predictors include Age, sex, race, smoking, alcohol use, lifetime sexual partners, and oral oncogenic human papillomavirus (HPV) status.	Yes

Table 3-1: Model characteristics

	Development Model Performance			Validation	
Study ID	**AUC (95% CI) and Sensitivity / Specificity (where reported)**	**Calibration - E/O (95%CI)** **PPV, NPV (%)**	**Cut-offs**	**Method of Validation**	**Performance**
Amarasinghe et al 2010	OPMD excl. lichen planus: 0.84 (0.81, 0.87) SENS / SPEC: 93.7% / 67.7%, All OPMDs: 0.78 (0.75, 0.81) SENS / SPEC 81.1% / 67.7%	Not reported Exc. Lichen Planus 27.5%, 98.8% (All) PPV 50.9%, NPV 89.6%	AUC of 0.87 (95% CI: 0.83 – 0.91), SENS 95.5%, SPEC of 75.9%,	External (Different setting) Phase 2: Suburban population of the Colombo district and in a rural population in selected PHM areas of the Bulathkohupitiya MOH area in the Kegalle district of Sri Lanka.	AUC of 0.87 (95% CI: 0.83 –0.91), SENS 95.5%, SPEC of 75.9%,
Budhathoki et al (Unpublished)	Not reported for development - see (internal) validation SENS / SPEC - Not reported	Plots are provided but not quantified - good calibration Not reported	Not reported	Internal The model was internally validated using a split sample approach; data was randomly split into a training (70%) and testing set (30%) Total N = 4,601 (3,030 HNC cases/1,571 controls)	HNC [0.72, 95%CI=0.69-0.75 in men and 0.75, 95%CI=0.71-0.79 in women] [OCC - (AUC=0.73, 95%CI=0.69-0.77 in men and AUC=0.79, 95%CI=0.74-0.83 in women).] [Model + HPV serology for OPC- (AUC=0.94, 95%CI=0.91-0.95 in men and AUC=0.89, 95%CI=0.82-0.92 in women)]

Chen et al 2018	0.768 (0.723, 0.813) for men, 0.700 (0.635, 0.765) for women SENS / SPEC – Not reported	Plots are provided but not quantified in supplementary table. Male model had better calibration (closer to 1).	Not reported	Internal Model was internally validated with 1000 repeat samples	N.A. - see Development
Cheung et al 2021	Not reported for development - see (internal) validation SENS / SPEC - Not reported	Not reported for development - see (internal) validation Not reported	Not reported	Internal Internally validated using 5-fold cross validation of the development population evaluated for discrimination and calibration and accounting for right censoring.	AUC = 0.84; (95% CI, 0.77 - 0.90) Calibration = 1.08; (95% CI, 0.81 - 1.44)
Koyanagi et al 2017	(Most extensive model) 0.72 (0.69, 0.75) SENS / SPEC - Not reported	1.00 Not reported	Not reported	External validation was carried out with a second case control study (HERPACC-3) conducted between November 2005 and March 2013 309 head and neck cancer cases and 654 matched controls were recruited.	AUC = 0.73 (0.70–0.77) Calibration= 0.97

Study ID	AUC (95% CI) and Sensitivity / Specificity (where reported)	Calibration - E/O (95%CI) PPV, NPV (%)	Cut-offs	Method of Validation	Performance
Krishna Rao et al 2016	0.866 SENS / SPEC: 0.746 (0.682, 0.810), 0.846 (0.802, 0.890)	Not reported PPV 0.767 (0.704–0.831), NPV 0.830 (0.785–0.875)	Development SENS - 0.928 (0.890–0.966 SPEC -0.603 (0.545–0.661) PPV - 0.607 (0.550–0.665) NPV - 0.927 (0.888–0.965)	Internal Internally validated from original dataset with 200 bootstrap samples.	AUC = 0.865, SENS 0.744 (0.740–0.750), SPEC 0.851 (0.848–0.854) PPV 0.773 (0.768–0.777), NPV 0.830 (0.827–0.833)
Lau et al 2018	0.79 SENS / SPEC - Not reported	Not reported Not reported	Not reported	Internal Model underwent cross-validation and temporal validation using a further 235 patients from same hospital	SENS / SPEC - 31%, 92%

Lee et al 2020	Not reported for development - (See Validation) SENS / SPEC - Not reported	Provides plots for each subsite (male and female) but not quantified - mostly good calibration bar male and female hypopharyngeal cancer Not reported	Not reported	Internal Internally validated using random split sample (30%)	AUC lowest to highest = 0.643 - 0.820, A) Male oral cavity cancer (AUC = 0.752); B) female oral cavity cancer (AUC = 0.718); C) male oropharyngeal cancer (AUC = 0.643); D) female oropharyngeal cancer (AUC = 0.745); E) male hypopharyngeal cancer (AUC = 0.784); F) female hypopharyngeal cancer (AUC = 0.820); G) male laryngeal cancer (AUC = 0.794); H) female laryngeal cancer (AUC = 0.870).
Liu et al 2017	1 SENS / SPEC: 100.00%, 99.02%	Not reported PPV 98.94%, NPV 100.00%	A cut-off of 0.5 was used (50% risk)	Unclear - poor reporting 102 controls, 82 OLK, 93 OSCC	AUC = 1, SENS 100.00% , SPEC 100.00% PPV 100.00%, NPV 100.00%
McCarthy et al 2020	0.69 (0.66, 0.71) SENS / SPEC - Not reported	Plots but not quantified for development population - CIs are also included - good calibration Not reported	Not reported	Internal Model was internally validated (split sample) with 60,240 individuals Cohort from North West England UK Biobank	Discrimination = 0.64 (0.60-0.68), Calibration = 0.83

Study ID	AUC (95% CI) and Sensitivity / Specificity (where reported)	Calibration - E/O (95%CI) PPV, NPV (%)	Cut-offs	Method of Validation	Performance
Sharma et al 2015	0.9974 SENS / SPEC: 98.01%, 98.68%	Not reported PPV 99.35%, NPV 98.01%	Not reported	Internal Internally cross validated with data from original sample	(Unclear if this for same PNN/GRNN model) AUC = 0.821 SENS 87.67%, SPEC 69.46% PPV 62.86%, NPV 88.17%
Sun et al 2019	0.83 (0.77, 0.88) SENS / SPEC: 67.53%, 81.25%	Not reported PPV 59.09%, NPV 86.19%,	Cut-off score of 3 SENS, 67.53%, SPEC 81.25% PPV 59.09%, NPV 86.19%,	N.A Development only	N.A.
Tikka et al 2020 (v.2)	0.8856 (0.8818, 0.8879) USOC- SENS / SPEC: 85%, 78.3% 2ndary 4 week referral - SENS 97.1%, SPEC 52.9%	Not reported PPV 20.7%, NPV 98.6%.	USOC referral = 0.071 probability 4-week 2nd clinic classification = 0.022 probability	Internal only internal validation in v.2 with 1000 bootstrap samples BUT Previous version of risk model did externally validate.	See development metrics
Tota et al 2019	0.94; (0.92, 0.97) SENS / SPEC: Not reported	1.01 (0.70, 1.32) Not reported	Not reported	External External validation was conducted on a historical series of 116 oropharynx cancer cases recruited at the Johns Hopkins University,	AUC = 0.87;(95% CI, 0.84-0.90) Calibration - 1.08 (0.77-1.39)

Table 3-2: Model Performance Information

3.4.2 Discrimination

Discrimination, the ability of a model to discern between a positive and a negative result for disease, is a crucial performance metric of a risk model. All 14 models provided measurements of discriminatory accuracy in either their development, validation populations, or both. Ten of these models described the statistical uncertainty of their findings. Many models (n=9) reported AUC values (and intervals where reported) greater than 0.7, achieving “acceptable” or “excellent” discrimination. One model with the highest AUC value reportedly achieved a “perfect” model discrimination of 1.0, however, this model was constructed from a very small sample size, in addition to other key limitations and bias concerns such as a failure to report statistical uncertainty and any missing data.(Liu et al., 2017)

3.4.3 Accuracy

Measurements of accuracy in the form of sensitivity and specificity were described in seven of the 14 studies. These ranged from 67.53% to 100% (Sun et al., 2019) (Liu et al., 2017), and 67.7% to 100% for each measure respectively. (Amarasinghe et al., 2010) (Liu et al., 2017) The model that reported the highest sensitivity and specificity achieved 100% in both of these metrics in their validation population.(Liu et al., 2017)

3.4.4 Calibration

Calibration, the degree of correspondence between the estimated probability of an outcome predicted by a model vs the outcome observed is an important measurement to consider when evaluating model performance, in order to minimise overfitting. Despite this, calibration is often overlooked in favour of model discrimination (AUC also known as the C-statistic).(Van Calster et al., 2019) Measurements of model calibration, either in the form of an expected / observed ratio or calibration plot were described in seven out of the 14 models, however, statistical uncertainty was reported in only three of these.

Calibration was evaluated in light of hierarchy definitions described by Van Calster and colleagues.(Van Calster et al., 2016) Calibration was good in most, if not all, of the models where this metric was reported, with the exception of male and female hypopharyngeal cancer models in one study where calibration was sub-optimal in these particular calibration plots.(Lee et al., 2020) Most of the models that did report calibration presented graphs or statistics that were close to the ideal calibration slope (expected / observed) value of 1, with some models slightly above this value indicating some over-prediction.(Richard D Riley, 2019)

3.4.5 Positive Predictive Value (PPV) / Negative Predictive Value (NPV)

PPV and NPV are defined as the proportion of patients who actually have the disease that test positive and the proportion of patients without the disease that test negative respectively. Six models reported measurements of PPV and NPV.(Parikh et al., 2008) As such only one model reported the statistical uncertainty of this.(Krishna Rao et al., 2016) The PPV and NPV values reported ranged from 20.7% to 100% and 83% to 100%. Again, Liu and colleagues(Liu et al., 2017) achieved the highest PPV and NPV values of 100%.

3.4.6 Model Risk Cut-offs

Of the 14 models, only five reported model risk cut-offs during development. Two models used a risk probability as a cut-off.(Liu et al., 2017, Tikka et al., 2020) Three models reported cut-offs using performance metrics including AUC, sensitivity and specificity and PPV/NPV. (Amarasinghe et al., 2010) (Krishna Rao et al., 2016) (Sun et al., 2019)

3.4.7 Validation

Only three of the 14 models reported external validation - Amarasinghe et al(Amarasinghe et al., 2010), Koyanagi et al(Koyanagi et al., 2017) and Tota et

al. (Tota et al., 2019c) Three others reported robust methods of internal validation via split random sampling.(Budhathoki1 et al., 2021) (Lee et al., 2020) (McCarthy et al., 2020)

3.4.8 Risk Factors

Altogether, the 14 models considered over thirty various risk factors. The most common factors included were age (13 models), alcohol consumption (13 models), sex (12 models) and tobacco smoking (12 models). Notably, two models considered HPV serostatus in model development - Budhathoki et al(Budhathoki1 et al., 2021) and Tota et al.(Tota et al., 2019c) The number of risk factors included in models ranged between five(Liu et al., 2017) (Sun et al., 2019) and 13.(Chen et al., 2018)

3.5 PROBAST

The evaluation of each domain of the PROBAST risk of bias assessment tool is summarised in Table 3-3.

Of the 14 models, seven were deemed to have a “high” risk of bias in at least one domain. The “analysis” section was the most common domain where a high risk of bias was identified. Common reasons for these included low numbers of the outcome of interest, a lack of external validation and limited or no internal validation, failure to report statistical uncertainty of findings and no discussion of missing data (and procedures in the event of this). Five of the 14 models were reported to have an “unclear” applicability concern whereby aspects of the model may limit its applicability but as such these were not major limitations. The “participants” section was the most common domain where applicability concerns were classified as “unclear”. Reasons for these included limited generalisability owing to the outcome considered, limiting the analysis to those of one ethnicity, use of non-primary Head and Neck Cancer (HNC) cancer sites and low-quality reporting of methods.

Where models had a limitation but were otherwise fairly robust and well-developed the risk of bias was deemed as “low”. Overall, seven of the 14 models were deemed to have an overall low risk of bias(Amarasinghe et al., 2010) (Budhathoki1 et al., 2021) (Koyanagi et al., 2017) (Lee et al., 2020) (McCarthy et al., 2020) (Tikka et al., 2020) (Tota et al., 2019c).

	Participants		Predictors		Outcome		Analysis	Overall	
Study ID	Risk of Bias	Applicability Concern	Risk of Bias	Applicability Concern	Risk of Bias	Applicability Concern	Risk of Bias	Risk of Bias	Applicability Concern
Amarasinghe et al 2010 (28)	LOW	UNCLEAR	LOW	LOW	LOW	LOW	LOW	LOW	UNCLEAR
Budhathoki et al (Unpublished) (29)	LOW	UNCLEAR	LOW	LOW	LOW	UNCLEAR	LOW	LOW	UNCLEAR
Chen et al 2017 (30)	LOW	LOW	LOW	UNCLEAR	LOW	LOW	HIGH	HIGH	LOW
Cheung et al 2021 (31)	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW
Koyanagi et al 2017 (32)	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Krishna Rao et al 2016 (33)	UNCLEAR	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW
Lau et al 2018 (34)	UNCLEAR	LOW	UNCLEAR	LOW	UNCLEAR	LOW	HIGH	HIGH	LOW
Lee et al 2020 (35)	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Liu et al 2017 (36)	HIGH	LOW	UNCLEAR	UNCLEAR	LOW	LOW	HIGH	HIGH	UNCLEAR
McCarthy et al (37)	LOW	LOW	LOW	LOW	UNCLEAR	LOW	LOW	LOW	LOW
Sharma et al 2015 (38)	HIGH	UNCLEAR	HIGH	UNCLEAR	HIGH	UNCLEAR	HIGH	HIGH	UNCLEAR
Sun et al 2019 (39)	HIGH	UNCLEAR	LOW	LOW	UNCLEAR	LOW	HIGH	HIGH	UNCLEAR
Tikka et al 2020 (v.2) (40)	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Tota et al 2019 (41)	LOW	LOW	LOW	LOW	LOW	LOW	UNCLEAR	LOW	LOW

Table 3-3: PROBAST Performance by Model

3.6 Overall Quality Assessment

The overall quality assessment of the 14 models and the components considered in this quality assessment along with model predictive performance assessment can be seen in Table 3-4.

	Quality				Performance
Study ID	PROBAST bias	PROBAST applicability	Validation - external, internal, no	Overall Quality Assessment	AUROC
Amarasinghe et al 2010 (28)	LOW	UNCLEAR	External	HIGH	0.87 (95% CI: 0.83 - 0.91)
Budhathoki et al (Unpublished) (29)	LOW	UNCLEAR	Internal - large data 70/30 split	MODERATE	HNC [0.72, 95%CI=0.69-0.75 in men and 0.75, 95%CI=0.71-0.79 in women]
Chen et al 2018 (30)	HIGH	LOW	Internal - bootstrap	MODERATE	0.768 (0.723, 0.813) for men, 0.700 (0.635, 0.765) for women
Cheung et al 2021 (31)	HIGH	LOW	(RCT design), Internal - cross validation	MODERATE	0.84; (95% CI, 0.77 to 0.90)
Koyanagi et al 2017 (32)	LOW	LOW	External	HIGH	0.73 (0.70–0.77)
Krishna Rao et al 2016 (33)	HIGH	LOW	Internal - bootstrap	LOW	0.865,
Lau et al 2018 (34)	HIGH	LOW	Internal - cross and split validation	LOW	0.79
Lee et al 2020 (35)	LOW	LOW	Internal - large data 70/30 split	HIGH	Poorest model to best = 0.643 - 0.820,
Liu et al 2017 (36)	HIGH	UNCLEAR	Unclear	LOW	1
McCarthy et al 2020 (37)	LOW	LOW	Internal - large data split	HIGH	0.64 (0.60-0.68)
Sharma et al 2015 (38)	HIGH	UNCLEAR	Internal - cross validation	LOW	0.9974
Sun et al 2019 (39)	HIGH	UNCLEAR	No	LOW	0.83 (0.77, 0.88)
Tikka et al 2020 (v.2) (40)	LOW	LOW	Internal - bootstrap, BUT previous version did	HIGH	0.8856 (0.8818-0.8879)
Tota et al 2019 (41)	LOW	LOW	External	HIGH	0.87 (0.84-0.90)

Table 3-4: Overall Quality / Performance Assessment

Of the 14 models, six were assessed as “high” quality (Amarasinghe et al., 2010) (Koyanagi et al., 2017) (Lee et al., 2020) (McCarthy et al., 2020) (Tikka et al., 2020) (Tota et al., 2019c), three as “moderate” quality, and five as “low” quality. The main components, which impacted on quality were PROBAST risk of bias, applicability concern, and validation methods.

In terms of performance, eight of the 14 models were high performing, with AUCs greater than 0.8, ranging from 0.83 to 1.(Amarasinghe et al., 2010) (Cheung et al., 2021) (Krishna Rao et al., 2016) (Liu et al., 2017) (Sharma and Om, 2015) (Sun et al., 2019) (Tikka et al., 2020) (Tota et al., 2019c) Of the six high quality models, three had high predictive performance with good discriminative accuracy - Amarasinghe et al(Amarasinghe et al., 2010), Tikka et al(Tikka et al., 2020) and Tota et al. (Tota et al., 2019c)

3.7 Synthesis

All six of the high-quality models were more recently developed (since 2010). Despite the heterogenicity of the models, generally, those that were assessed as high quality shared common design aspects. All of the models were developed from case-control study data with some variation in design such as hospital-, community-, synthetic-controls, or a mixture of population and hospital controls. All six high quality studies also used a form of logistic regression to derive their risk models. These included binary, multivariate or conditional logistic regression. Three of the models required clinician input to use: two of these due to HPV or genotype information (Koyanagi et al., 2017) (Tota et al., 2019c), and one due to use of clinical examination information.(Tikka et al., 2020)

With regards to factors included in the high-quality models, all six had some sociodemographic factors - age was evaluated in all six of them and sex in five – one model did not analyse or adjust for sex which in turn resulted in reduced applicability. Four of the models adjusted for at least one additional sociodemographic factor (education, ethnicity, or socioeconomic

deprivation).(Amarasinghe et al., 2010) (Lee et al., 2020) (McCarthy et al., 2020) (Tota et al., 2019c) Two high-quality models evaluated socioeconomic deprivation, one synthesising educational and occupational status to define this, the other measured deprivation using an area based socioeconomic index.(Amarasinghe et al., 2010) (McCarthy et al., 2020) Similarly, all six also incorporated behavioural factors into their model – using both alcohol intake and smoking.

Notably, one model used betel quid chewing as an additional behavioural predictor, an important aetiological risk factor for the target population for this model.(Amarasinghe et al., 2010) One model also used exercise and fruit / vegetable intake as additional factors.(McCarthy et al., 2020) Two of the three models that used biomarker (genetic or HPV) data in their models were ultimately among the six assessed as high quality. One high quality model used HPV serostatus as a predictor.(Tota et al., 2019c) Another used DNA sampling to assess *ALDH2* genotype.(Koyanagi et al., 2017) Only one of the high-quality models reported family history as a predictor.(Lee et al., 2020) Finally, one of the six high quality models used clinical signs and symptoms as predictors to inform model design.(Tikka et al., 2020)

The critical design feature common to all of the high-quality models was robust validation methods. These included the use of external validation in another setting(Amarasinghe et al., 2010) (Koyanagi et al., 2017) (Tota et al., 2019c), a history of this (in a previous developmental version of the model)(Tikka et al., 2020), or the utilisation of well-conducted internal validation with a large split sample.(Lee et al., 2020) (McCarthy et al., 2020) All the high-quality models that did not use an external validation approach described this as a limitation and a next step in their model development.

In addition to the high quality models having a low risk of bias, five of the six also had a low applicability concern(Koyanagi et al., 2017) (Lee et al., 2020) (McCarthy et al., 2020) (Tikka et al., 2020) (Tota et al., 2019c), while the remaining model scored “unclear” for applicability concern assessment.(Amarasinghe et al., 2010) This was primarily due to the model evaluating the risk of OPMD only and not accounting for sex as a predictor during model development.

The high-quality models mostly had fair to good performance. One model had a sup-optimal AUC of 0.64(McCarthy et al., 2020), two high quality models reported a fair AUC over 0.7(Koyanagi et al., 2017) (Lee et al., 2020) and three of the high quality models achieved excellent AUC values over 0.8.(Amarasinghe et al., 2010) (Tikka et al., 2020) (Tota et al., 2019c)

The better the discriminative performance, the more accurately those at risk of disease can be identified. Two of the six high quality models did not report calibration metrics.(Amarasinghe et al., 2010) (Tikka et al., 2020) This was the main limitation of these models. Four models reported calibration, either numerically or via plots, illustrating that the models could accurately predict outcomes in line with the observed event of interest (head and neck cancer), with most of the calibration graphs or reported expected / observed values being close to the perfect prediction value of 1. Therefore, for the most part, head and neck cancer models, where reported, showed good calibration between the expected and observed risk of head and neck cancer.

Of the six high quality models, three were also high performing, achieving excellent discrimination with AUCs over 0.8.(Amarasinghe et al., 2010) (Tikka et al., 2020) (Tota et al., 2019c) The three models predicted the risk of OPMD, head and neck cancer, and oropharyngeal cancer respectively. Two of the models were externally validated in another cohort(Amarasinghe et al., 2010) (Tota et al., 2019c), whilst the other model had a history of external validation in its first version.(Tikka et al., 2020) This model has also seen some clinical use, being used to triage patients remotely during the COVID-19 pandemic.(Paleri et al., 2020) The three high-quality, high performing models all used similar sociodemographic and behavioural factors but where the three high performing models differed and ultimately excelled was in the choice of additional predictors used. These included the aforementioned use of betel quid in one model(Amarasinghe et al., 2010), clinical examination and symptoms in another model(Tikka et al., 2020), and the use of HPV serostatus and ethnicity in the third high quality model.(Tota et al., 2019c)

Those that were classified as moderate overall quality had at least one major methodological limitation, but generally had fair predictive performance. Studies that were classified as low quality had at least two significant

limitations; some of these reported very good model discrimination, although this needs to be interpreted with caution.

3.8 Discussion

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

The six high performing models incorporate the major head and neck cancer risk factors of tobacco smoking and alcohol consumption, in addition to the sociodemographic factors of age and sex. Additional factors are included contributing to improved performance. Four included socioeconomic factors, one: family history, one: betel quid chewing, one: HPV serology, one included a genetic marker, and one included clinical examination findings. These selected factors are consistent with the international analytical epidemiological literature which has identified tobacco smoking and alcohol drinking as the major risk factors (accounting for up to ~70% of the population attributable risk) (Dhull et al., 2018) (Cogliano et al., 2011) (Hashibe et al., 2009), the important role of demographics of age in cancer risk(Johnson et al., 2020b), and sex – particularly men being more predisposed to head and neck cancer. (Centers for Disease Control and Prevention, 2020a)

Moreover, the important role of HPV particularly in oropharyngeal cancer(Sabatini and Chiocca, 2020b) and betel quid chewing in oral cavity cancer in particular populations(Guha et al., 2014), and the increasingly refined role of genetic factors in head and neck cancer are reflected in the models. (Riaz et al., 2014) (Beck and Golemis, 2016)

Three of these high performing predictive models were of high quality and consistent methodological rigour – all included major behavioural and sociodemographic factors along with an additional factor. They were generally more specified models that were tailored to their target population or subsite

e.g to South Asia (the inclusion of betel quid)(Amarasinghe et al., 2010) or to oropharyngeal cancer (the inclusion of HPV serology).(Tota et al., 2019c)

Perhaps counterintuitively, some of the models that included many additional risk factors generally had lower predictive performance. This could be explained by a statistical phenomenon known as model overfitting, whereby a model becomes too tailored to a developmental dataset with unnecessary components. This violates the principle of parsimony, in turn limiting a model's generalisability when applied to another independent dataset.(Hawkins, 2004) The higher performing models and particularly the high performing, high-quality models generally required clinician input, reflecting the nature of the variables required.

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

Several similar reviews of risk prediction models have been undertaken for other cancer sites including colorectal (Usher-Smith et al., 2016) (Williams et al., 2016), breast (Anothaisintawee et al., 2012) (Louro et al., 2019), and lung. (Gray et al., 2016) The reviews of colorectal and lung cancer models both identified models with high performance (0.65-0.75, 0.76-0.96, and 0.57-0.879), while the breast cancer models generally reported poorer performance (0.56-0.63 and 0.56-0.71).

The poor performance of these breast cancer models was attributed in the reviews to limited knowledge and data on risk factors for breast cancer leading to sub-optimal prediction. The performance range of the head and neck cancer models was similar to the colorectal and lung cancer reviews and might in part be attributed to the growing epidemiological research base in the field.(Bravi et al., 2021) (Winn et al., 2015b)

The methodology of this review is similar to reviews of risk models for other cancers.(Usher-Smith et al., 2016) (Williams et al., 2016) (Anothaisintawee et

al., 2012) (Louro et al., 2019) (Gray et al., 2016) This review employed robust quality and methodological assessment including the PROBAST risk of bias and applicability concerns tool, and a focus on the nature of model development and validation approach. While validation is a domain of the PROBAST tool, this was explicitly assessed separately as external validation is gold standard methodology of risk prediction model development.(Ramspek et al., 2020) (Altman and Royston, 2000)

To our knowledge, this is the first review of head and neck cancer risk prediction models. This review has some strengths including searching multiple databases, dual article screening, as well as the comprehensive quality assessment. A detailed thematic narrative synthesis drew on the model quality and performance to identify key design characteristics.

There are some limitations to this review, including not publishing a protocol. The study started as a rapid review - but ultimately became more systematic in nature – particularly in term of quality assessment methods. However, it was not feasible to register the review retrospectively, hence the review was not registered with PROSPERO. The PICO / research question and search / inclusion criteria were developed a priori and did not change during the review.

The review was also conducted following PRIMSA guidelines and was advised by a subject librarian in the field. Secondly, the inclusion of papers published only in English may have excluded other pre-existing models. As with most reviews, the nature and limitations of available data can influence the overall quality of evidence synthesised - the source data of this review are largely from case-control studies which do have some potential recall and selection biases.(Tenny et al., 2022)

This review has also been conducted within the overarching objectives of developing a risk model for head and neck cancer and translating it to a clinical setting. Any findings of this review are intended to help inform model development.

3.9 Conclusions

This review illustrates that there is a limited but growing number of head and neck cancer risk prediction models. Some of the models reviewed do have the potential to identify and stratify those at risk of head and neck cancer. Model predictor selection should include, as a minimum, well established risk factors as well as sociodemographic predictors. Additional genetic, biomarker, or clinical factors have the potential to improve predictive performance. However, care should be taken to ensure a limited number of predicting factors are chosen to avoid model overfitting.

Such early identification of risk factors in the context of a HNC risk level could have important applications including using this “teachable moment” for behaviour change, directing patients to preventive care pathways (e.g. for smoking cessation), or in identifying the need for tailored frequencies in recall intervals for clinical examination (e.g. with primary care dental practitioners). These models could form the basis of a personalised approach to head and neck cancer prevention. Further work can be undertaken to refine, improve and validate these models and potentially trial in the clinical setting.

3.10 Additional Information and Author Contributions

3.10.1 Funding

This work was funded and supported by Cancer Research UK as part of the TRACC programme, Beatson Institute, Glasgow [Grant number 315941-01]

We would like to thank Paul Canon, University of Glasgow MVLS Librarian, for his support in helping us develop and check our review and search strategy.

We would also like to thank Catherine Winchester, research Manager at the Beatson Institute, Glasgow for her role in providing feedback and proofreading this review.

3.10.2 Author Contributions

Conceptualisation: CS AM AR GI DC

Data Curation: CS DC

Formal Analysis: CS

Investigation: CS

Methodology: CS DC

Project Administration: CS DC

Visualisation: CS

Writing - original draft: CS

Writing – review & editing: CS AM AR GI DC

4 Chapter Four: Head and Neck Cancer Incidence is Rising but the Sociodemographic Profile is Unchanging: A Population Epidemiological Study (2001-2020)

The contents of this chapter were accepted for publication by the British Journal of Cancer (BJC) Reports Journal as of the 7th of August 2024.

Authors:
Craig D L Smith[1,2,3] (corresponding author), Alex D McMahon[1], Mitana Purkayastha[4], Grant Creaney[1,3], Kelten Clements[1,3], Gareth J Inman[2,3,5], Lesley A Bhatti[6], Catriona M Douglas[1,3,7], Claire Paterson[3,8], David I Conway[1,2,3]

Affiliations:
1: School of Medicine, Dentistry, and Nursing, University of Glasgow, UK
2: School of Cancer Sciences, University of Glasgow, UK
3: Glasgow Head and Neck Cancer (GLAHNC) Research Group
4: Great Ormond Street Institute of Child Health, University College London, UK
5: Cancer Research UK Scotland Institute, Glasgow, UK
6: Public Health Scotland
7: Department of Otolaryngology/Head and Neck Surgery – Glasgow Royal infirmary and Queen Elizabeth University Hospital, Glasgow, UK
8: Beatson West of Scotland Cancer Centre, Glasgow, UK

4.1 Abstract

Background:
Increasing incidence of head and neck cancers (HNCs), driven by rising rates of oropharynx cancer (OPC), has been recorded around the world. This study examined trends in HNC and subsites (oral cavity, oropharynx, and larynx cancers) in Scotland focusing on assessing whether the sociodemographic profile has changed over the past 20 years.

Methods:
Scottish Cancer Registry data (2001-2020) including European Age Standardised Rates of HNC and subsites were analysed in multivariate Poisson regression by age, sex, area-based socioeconomic status, and year of diagnosis (with interaction tests).

Results:
Overall HNC and oral cavity cancer (OCC) incidence remained relatively stable. OPC incidence rates increased by 78%, while larynx cancer incidence declined by 27%. Over time, there were marginal shifts to a slightly older age profile for HNC ($p=0.001$) and OCC ($p=0.001$), but no changes in OPC ($p=0.86$) and larynx cancer ($p=0.29$). No shift in the sex profile of HNC was observed except for minor increases in female OCC rates ($p=0.001$), and the socioeconomic distribution remained unchanged across all HNC subsites.

Conclusions:
There have been no significant changes in the sociodemographic profile of HNC in Scotland over the last 20 years, despite the changing trends in HNCs with dramatically increasing incidence rates in OPC and reducing larynx cancer. This information can be used to target or stratify HNC prevention and control.

4.2 Introduction

Squamous cell carcinomas of oral cavity, larynx and pharynx are commonly known as head and neck cancers (HNC). As the seventh most common cancer globally, it is a growing and increasing public health challenge with over 800,000 incident cases and 400,000 deaths in 2020. (Johnson et al., 2020a, Sung et al., 2021, Global Cancer Observatory, 2020)

In the UK patterns are in keeping with this; since the early 1990s HNC incidence has risen by 37%, with approximately 12,400 cases and 4,100 deaths annually. (Gormley et al., 2022a) (Cancer Research UK) Incidence rates of HNC vary by geographical region and HNC subsite; large increases have been observed in oropharyngeal cancer rates in the UK as a whole (ranging from 45% to an 85% increase in Scotland). This has been accompanied by reports of modest increases

in oral cavity cancer rates and declines in larynx cancer rates. (Purkayastha et al., 2016) (Louie et al., 2015) (McCarthy et al., 2015) Wide socioeconomic inequalities in HNC incidence are observed, with the highest rates consistently found among those most socioeconomically deprived groups. (Purkayastha et al., 2016) HNC is more common among older people and men are typically at a two-four-fold increased risk of head and neck cancer compared with women. (Johnson et al., 2020a) (Park et al., 2022)

In recent years, the increase in oropharyngeal cancer rates has become a global phenomenon, and largely attributed to Human Papilloma Virus (HPV), particularly in the United States and Europe. (Gormley et al., 2022a) (Purkayastha et al., 2016) (Robinson and Macfarlane, 2003) (Louie et al., 2015) Within these trends, there are emerging reports of changes in the sociodemographic profile of people with head and neck cancer, especially oropharyngeal cancer. Some of these studies suggest the demographics of patients with OPC are younger, female, with higher socioeconomic status (SES) and presenting without a history of smoking or alcohol consumption, the traditional risk factors for HNC. (Donà et al., 2020) (Mariz et al., 2020) (Deschler et al., 2014b) (van Monsjou et al., 2013) (Dahlstrom et al., 2015a) (Tota et al., 2019a) (Gillison et al., 2008) (Mahal et al., 2019) (Liederbach et al., 2017) Here, we aimed to investigate the trends in incidence and sociodemographic profile (age, sex, and area-based socioeconomic deprivation) of cancers of the oral cavity, oropharynx, and larynx over time at a population-level.

4.3 Methods

Scottish Cancer Registry data for the years 2000-2021 were accessed including: Incident cases of HNC and its subsites - defined using international classification of disease (tenth edition ICD-10) codes and grouped as: oral cavity (inner lip - C00.3 - C00.9, dorsal, overlapping or NOS tongue - C02, gingivae - C03, floor of mouth - C04, soft palate, uvula, palate or NOS - C05 and cheek, other and NOS mouth - C06); oropharynx (base of tongue- C01, lingual tonsil - C02.4, tonsil - C09, oropharynx - C10, pharynx - C14.0,14.2); larynx (C32); and all HNC was defined as the above sites in addition to tumours of the nasopharynx (C11),

piriform sinus (C12), Hypopharynx (C13) and other overlapping sites (C14.8). (Public Heath Scotland, 2022) (World Health, 2004) (Conway et al., 2018b) Available sociodemographic factors of cases were also included: age (in five-year bands), sex (male / female), and an area-based index of multiple deprivation – the Scottish Index of Multiple Deprivation (SIMD). (Scottish Government, 2020) This small area-based deprivation index is calculated from several domains including income, employment, education, health, access to services, crime, and housing. In the Scottish Cancer Registry, this decile-based data were divided into fifths, where SIMD-1 was the most socioeconomically deprived and SIMD-5 was the least socioeconomically deprived. (National Records of Scotland (NRS), 2022)

European age-standardised rates were calculated for all subsites, age, sex, SIMD, and year of diagnosis. Using Poisson regression, rate ratios were calculated to describe and compare HNC incidence by subsite, age, sex, SIMD, and year of diagnosis. Age-standardised incidence rates were plotted by subsite to visualise trends in HNC incidence. In order to assess for temporal changes in incidence, interaction tests were also conducted for the 5-year aggregated data. All statistical analyses were conducted with SAS version 9.4.

4.4 Results

In Scotland there were 20,850 HNC cases identified by the Scottish Cancer Registry from 2001-2020. Of these cases, 70.5% (n = 14,706) were male and 29.5% (n = 6,144) were female. The crude number of cases and age standardised incidence rate per 100,000 for each subsite is summarised in Table 1 by age, sex, SIMD, region and year of diagnosis. Substantial increases in age standardised oropharynx cancer (OPC) incidence and declines in larynx cancer rates were observed. The age standardised incidence rate of HNC was greatest among 60–64-year-olds (3.41 cases per 100,000). OPC incidence peaked in 60–64-year-olds (1.00 per 100,000), whilst OCC incidence (1.17 per 100,000) and larynx cancer incidence (1.07 per 100,000) peaked among 65–69-year-olds.

Table 2 details Poisson regression rate-ratios and interaction tests for each subsite by age, sex, SIMD, and year of diagnosis in 5-year periods. From 2001-2005 to 2016-2020, the overall age-standardised HNC incidence rate increased by 5%, (RR = 1.05, 95% CI = 1.01 – 1.09). (Figure 1a) Within this same period, OPC incidence rates increased substantially by 78% (RR = 1.78, 95% CI = 1.65 – 1.93). Oral cavity cancer (OCC) incidence rates remained relatively stable, exhibiting a modest but non-statistically significant decrease during the overall study period by 23% (RR = 0.94, 95% CI = 0.88 – 1.01), while larynx cancer incidence rates decreased significantly by 27% (RR = 0.73, 95% CI = 0.68 – 0.79)

Table 4-1: Counts and Age-Standardised Incidence Rates for HNC and subsites by year, age, sex, SIMD (with peak incidence in bold)

	HNC		OPC		OCC		Larynx	
	N	EASR	N	EASR	N	EASR	N	EASR
Year								
2001	885	20.09	160	3.54	318	7.25	**318**	**7.27**
2002	881	19.81	162	3.58	327	7.37	291	6.59
2003	903	20.04	167	3.64	352	7.77	294	6.63
2004	933	20.57	192	4.20	335	7.39	310	6.90
2005	893	19.40	191	4.03	340	7.45	283	6.24
2006	968	20.84	207	4.34	336	7.27	321	6.98
2007	987	20.92	214	4.44	366	7.80	300	6.39
2008	937	19.56	216	4.42	357	7.50	281	5.92
2009	1058	21.74	288	5.71	371	7.72	288	6.02
2010	1044	21.23	258	5.11	379	7.76	309	6.40
2011	1062	21.35	275	5.41	368	7.42	296	6.06
2012	**1164**	**22.94**	341	6.61	401	7.95	283	5.64
2013	1139	22.31	330	6.30	412	8.14	289	5.75
2014	1111	21.48	304	5.75	**423**	**8.22**	297	5.81
2015	1164	22.17	339	6.32	409	7.85	297	5.70
2016	1112	21.02	318	5.92	400	7.63	285	5.39
2017	1153	21.40	369	6.76	397	7.42	272	5.08
2018	1217	22.39	399	7.23	407	7.55	288	5.32
2019	1126	20.45	390	7.03	342	6.25	276	5.04
2020	1113	20.07	**401**	**7.17**	368	6.65	228	4.14
Age								
Under 25	64	0.06	6	0.01	33	0.03	3	0.00
25-29	48	0.04	7	0.01	27	0.02	6	0.01
30-34	80	0.08	18	0.02	39	0.04	11	0.01
35-39	222	0.22	66	0.06	84	0.08	41	0.04
40-44	502	0.47	166	0.16	198	0.19	93	0.09
45-49	1095	1.00	417	0.38	370	0.34	215	0.20
50-54	2049	1.93	811	0.77	604	0.57	433	0.41
55-59	2990	2.81	1016	0.95	926	0.87	757	0.71
60-64	**3507**	**3.41**	**1030**	**1.00**	1149	1.12	984	0.96
65-69	3333	3.36	775	0.78	**1158**	**1.17**	**1064**	**1.07**
70-74	2816	3.06	563	0.61	1019	1.11	919	1.00
75-79	2029	2.27	370	0.41	766	0.86	697	0.78
80-84+	2115	2.28	277	0.30	1035	1.12	582	0.63
Sex								
Male	**14706**	**32.03**	**4135**	**8.63**	**4377**	**9.63**	**4593**	**10.25**
Female	6144	11.47	1387	2.59	3031	5.64	1212	2.28
SIMD								
SIMD: 1	**6290**	**36.04**	**1557**	**8.76**	**2005**	**11.34**	**2023**	**11.88**
SIMD: 2	4867	25.49	1193	6.15	1682	8.65	1492	8.03
SIMD: 3	3972	19.80	1107	5.41	1457	7.15	1023	5.27
SIMD: 4	3275	16.39	992	4.78	1249	6.20	741	3.92
SIMD: 5	2446	12.92	673	3.36	1015	5.33	526	2.97

HNC = All Head and Neck Cancer, OPC = Oropharyngeal Cancer, OCC = Oral Cavity Cancer, Larynx = Larynx Cancer,

N = Number, EASR per 100,000 = European Age-Standardised Rate per 100,000 persons, SIMD = Scottish Index of Multiple Deprivation, where SIMD-1 was the most socioeconomically deprived and SIMD-5 was the least socioeconomically deprived

Table 4-2: Adjusted Rate-Ratios and interaction tests for HNC and subsite incidence by Age, Sex, SIMD and Year of Diagnosis (5-year period)

	HNC				OPC				OCC				Larynx			
	RR	95%CI Lower	95%CI Upper	p	RR	95% CI Lower	95% CI Upper	p	RR	95% CI Lower	95%CI Upper	p	RR	95% CI Lower	95%CI Upper	p
Interaction																
Age*Year	.	.	.	**0.001**	.	.	.	0.86	.	.	.	**0.001**	.	.	.	0.29
Sex*Year	.	.	.	0.55	.	.	.	0.22	.	.	.	**0.01**	.	.	.	0.59
SIMD*Year	.	.	.	0.75	.	.	.	0.55	.	.	.	0.23	.	.	.	0.15
Year																
2001-2005	REF	REF	REF		REF	REF	REF		REF	REF	REF		REF	REF	REF	
2006-2010	1.04	1.00	1.08	0.05	1.27	1.16	1.38	**<.0001**	1.02	0.95	1.09	0.65	0.94	0.87	1.01	0.08
2011-2015	1.10	1.05	1.14	**<.0001**	1.59	1.47	1.73	**<.0001**	1.06	0.99	1.13	0.10	0.85	0.79	0.91	**<.0001**
2016-2020	1.05	1.01	1.09	**0.03**	1.78	1.65	1.93	**<.0001**	0.94	0.88	1.01	0.08	0.73	0.68	0.79	**<.0001**
Age																
Under 25	0.01	0.01	0.02	**<.0001**	0.00	0.00	0.01	**<.0001**	0.02	0.02	0.03	**<.0001**	0.00	0.00	0.01	**<.0001**
25-29	0.05	0.03	0.06	**<.0001**	0.02	0.01	0.04	**<.0001**	0.08	0.05	0.12	**<.0001**	0.03	0.01	0.06	**<.0001**
30-34	0.08	0.06	0.10	**<.0001**	0.05	0.03	0.07	**<.0001**	0.11	0.08	0.16	**<.0001**	0.05	0.03	0.10	**<.0001**
35-39	0.21	0.18	0.25	**<.0001**	0.17	0.13	0.22	**<.0001**	0.24	0.19	0.30	**<.0001**	0.20	0.14	0.27	**<.0001**
40-44	0.47	0.42	0.52	**<.0001**	0.41	0.34	0.49		0.54	0.46	0.65	**<.0001**	0.43	0.34	0.55	**<.0001**
45-49	REF	REF	REF		REF	REF	REF		REF	REF	REF		REF	REF	REF	
50-54	1.92	1.79	2.07	**<.0001**	1.98	1.76	2.23	**<.0001**	1.68	1.48	1.92	**<.0001**	2.09	1.77	2.46	**<.0001**
55-59	3.01	2.81	3.23	**<.0001**	2.66	2.37	2.98	**<.0001**	2.77	2.46	3.12	**<.0001**	3.91	3.36	4.55	**<.0001**
60-64	3.97	3.71	4.24	**<.0001**	3.02	2.70	3.39	**<.0001**	3.85	3.42	4.33	**<.0001**	5.72	4.93	6.62	**<.0001**
65-69	4.30	4.01	4.60	**<.0001**	2.59	2.30	2.92	**<.0001**	4.41	3.92	4.96	**<.0001**	7.07	6.11	8.19	**<.0001**
70-74	4.37	4.08	4.69	**<.0001**	2.28	2.01	2.58	**<.0001**	4.64	4.12	5.23	**<.0001**	7.39	6.37	8.57	**<.0001**
75-79	4.17	3.87	4.48	**<.0001**	2.00	1.74	2.30	**<.0001**	4.54	4.01	5.14	**<.0001**	7.48	6.42	8.71	**<.0001**
80-84+	3.65	3.39	3.92	**<.0001**	1.25	1.08	1.46	**0.0035**	4.97	4.41	5.60	**<.0001**	5.47	4.68	6.40	**<.0001**
Sex																

Male	2.83	2.74	2.91	**<.0001**	3.35	3.16	3.57	**<.0001**	1.74	1.66	1.82	**<.0001**	4.57	4.29	4.87	**<.0001**
Female	REF	REF	REF		REF	REF	REF		REF	REF	REF		REF	REF	REF	
SIMD																
SIMD: 1 (20% most deprived)	2.87	2.73	3.00	**<.0001**	2.60	2.38	2.85	**<.0001**	2.18	2.02	2.35	**<.0001**	4.28	3.89	4.71	**<.0001**
SIMD: 2	2.02	1.92	2.12	**<.0001**	1.84	1.67	2.02	**<.0001**	1.66	1.53	1.79	**<.0001**	2.85	2.58	3.15	**<.0001**
SIMD: 3	1.56	1.48	1.64	**<.0001**	1.60	1.45	1.76	**<.0001**	1.37	1.27	1.49	**<.0001**	1.85	1.67	2.06	**<.0001**
SIMD: 4	1.28	1.22	1.35	**<.0001**	1.42	1.28	1.56	**<.0001**	1.18	1.09	1.28	**<.0001**	1.34	1.20	1.50	**<.0001**
SIMD: 5 (20% least deprived)	REF	REF	REF		REF	REF	REF		REF	REF	REF		REF	REF	REF	

HNC = All Head and Neck Cancer, OPC = Oropharyngeal Cancer, OCC = Oral Cavity Cancer, Larynx = Larynx Cancer,

*RR = Rate Ratio (derived from Poisson Regression), 95% CI = 95% Confidence Interval, * = interaction test, REF = Reference Category*

SIMD = Scottish Index of Multiple Deprivation, where SIMD-1 was the most socioeconomically deprived and SIMD-5 was the least socioeconomically deprived

Overall, males were observed to have significantly higher HNC incidence rates (Figure 1c) (RR = 2.83, 95% CI 2.74 - 2.91). This effect was also observed across all subsites (Table 2); male incidence rates were over three times that of females for OPC (RR = 3.35, 95% CI 3.16 - 3.57), over one and a half times greater for OCC (RR = 1.74, 95% CI 1.66 - 1.82) and over four times greater times for larynx cancer (RR = 4.57, 95% CI 4.30 - 4.87).

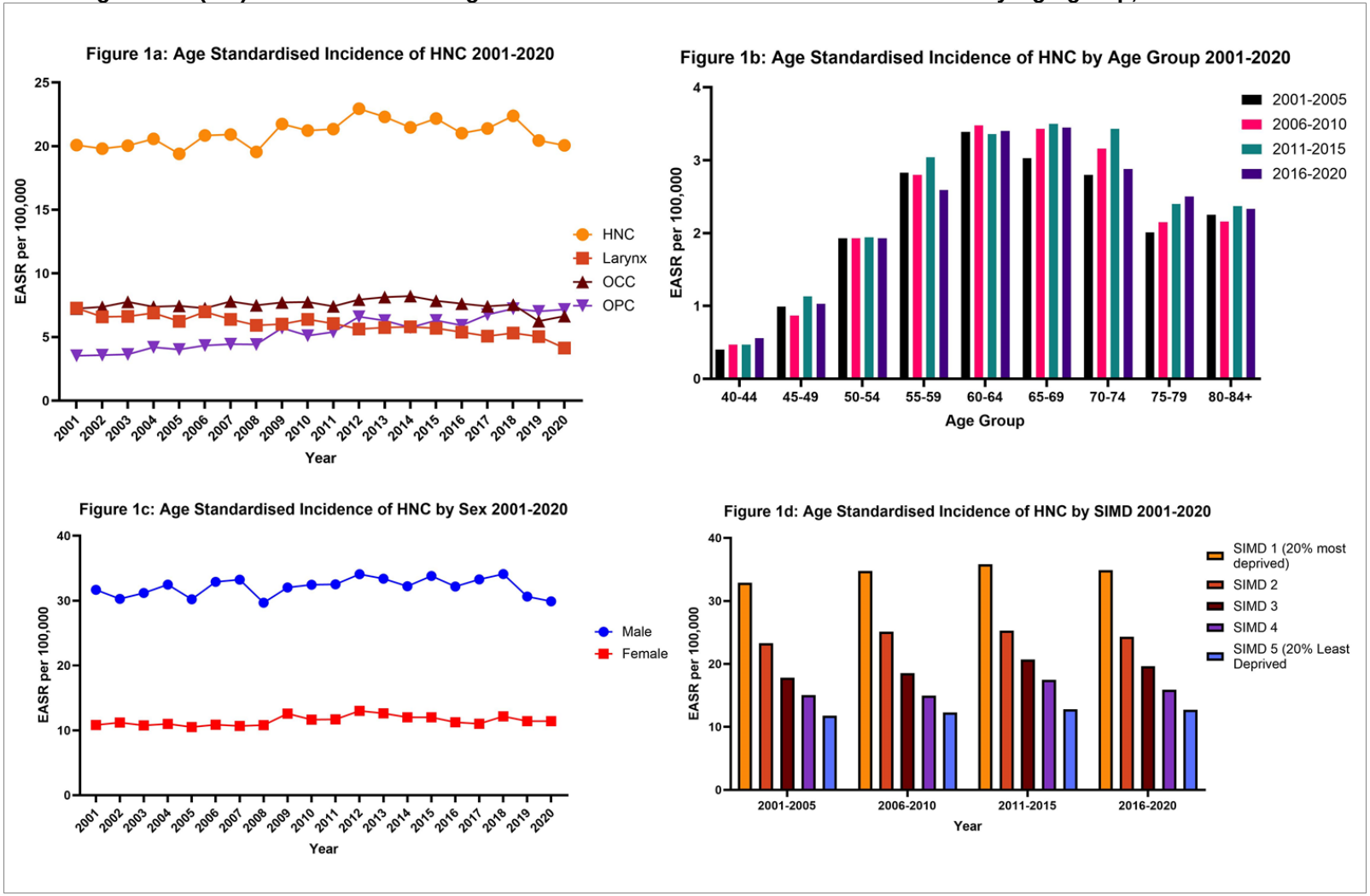
Temporally, HNC incidence rates by SIMD have remained stable. Large inequalities persist with the highest incidence rates of HNC observed among those most socioeconomically deprived (Figure 1d). (RR = 2.87, 95% CI 2.74 – 2.91). The incidence rate trends by SIMD reflect the trends of each subsite (for example, increasing rates across all SIMD quintiles for OPC) but within each sub-site, the same strong inequality gradient remained true.

As can be seen from the Figures 1a-d and Supplementary Figures 1-9, there have been no major shifts in the peak incidence rates by age-groups, SIMD, or sex. Interaction tests conducted within the Poisson models were suggestive of no temporal associations in the relationship between sex and HNC incidence (p=0.55). This remained true for cancers of the oropharynx (p = 0.22) and larynx (p = 0.59). A significant interaction was observed when assessing OCC incidence by sex over time, with small increases and declines in the burden of female and male cases, respectively (p = 0.01).

There were no temporal changes or associations observed when assessing HNC incidence by SIMD over time (p = 0.75). This remained true for all subsites (OPC: p = 0.55, OCC: p =0.23, Larynx: p= 0.15). Temporal associations of increasing cancer incidence with age over time varied by subsite. No temporal interactions were observed for OPC cases (p=0.86) or larynx cancer by age-group over time (p = 0.29). The interactions tests were suggestive of fluctuations in OCC incidence by age over time (p = 0.001). However, upon interpretation of plots and table values, there was no consistent pattern change in OCC incidence, although there was a marginal shift to a slightly older peak age (65-69 years) (Supplementary Figure 2). At an overall level, a significant temporal interaction was associated with age and

HNC incidence (p = 0.001). This could be explained by a similar, but also marginal, shift to a slightly older peak-age group (65-69 years). A consistent increase in patients aged 75-79 years old was also observed over time (Figure 1b).

Figure 4-1: (a-d) Combined Plot of Age-Standardised Overall HNC Incidence Trends by Age-group, Sex and SIMD



HNC = All Head and Neck Cancer, OPC = Oropharyngeal Cancer, OCC = Oral Cavity Cancer, Larynx = Larynx Cancer, EASR per 100,000 = European Age-Standardised Rate per 100,000 persons, SIMD = Scottish Index of Multiple Deprivation, where SIMD-1 was the most socioeconomically deprived and SIMD-5 was the least socioeconomically deprived

4.5 Discussion

This study shows that the predominant sociodemographic profile of people with all types of HNC is consistently males in their early to mid-60s (and older) from low socioeconomic backgrounds and this has remained unchanged over the last 20 years. Over this period, OPC incidence rates have risen while rates of larynx cancer have declined, and OCC rates have remained stable. In this changing makeup of head and neck cancer diagnoses and despite differing aetiologies, the underlying sociodemographic profile is unchanged.

Our results contrast with previous reports which have suggested a changing sociodemographic profile of HNC. (Donà et al., 2020) We found no evidence of changing sex distribution of HNCs overall, with incidence rates consistently greater among men for all subsites and across the study period – consistent with previous studies. (Gillison et al., 2015) (Sonawane et al., 2017) (Conway et al., 2016) A significant interaction was observed for sex over time for OCC cancer incidence – relating to only marginal increases in female OCC incidence, although incidence rates among males remained at least 1.5 times higher than females across the entire study duration. Increasing OCC cancer among women has previously been reported across the world potentially explained by latent uptake of smoking and alcohol behaviours in women (Gillison et al., 2008) (Chaturvedi et al., 2013) (Renou et al., 2023)

Low socioeconomic Status (SES) is an established determinant of HNCs. (Conway et al., 2021b) (Johnson et al., 2008) (Conway et al., 2015) (Creaney et al., 2022). A strong inequality gradient was consistently observed across all HNC subsites, including OPC, over the study period. Furthermore, interaction tests showed no temporal interactions with time for HNC incidence. In contrast with previous reports that suggested people with OPC tended to be from higher socioeconomic groups, we found a consistently strong socioeconomic gradient and greater disease burden among those from lower socioeconomic groups. (Deschler et al., 2014b) (Dahlstrom et al., 2015a)

These previous studies evaluating the sociodemographic profile of OPC were smaller clinical cohort or case-control studies mainly conducted in the USA. In

these studies, the socioeconomic profile reported may have been skewed towards more affluent participants with access to insurance and healthcare by the nature of recruiting cancer centres. Prior research has shown that high-risk sexual behaviours (which are also associated with an increased risk of HPV infection and HPV-positive OPC), have a similar socioeconomic pattern to the other major HNC risk factors smoking and alcohol consumption – i.e., greater among lower socioeconomic groups. (Wellings et al., 2001) (Jackson et al., 2012)

Ultimately, this study contradicts suggestions that the socioeconomic pattern of OPC differs from other HNC subsites and that this pattern is not changing over time. These findings are consistent with previous studies of HNC stage and mortality in Scotland, where similar inequality patterns have been observed. (Creaney et al., 2022) (Ingarfield et al., 2018) The strong inequality gradient observed in this study highlights the need for targeted primary and secondary prevention strategies to reach at risk groups in socioeconomically deprived areas.

Our study revealed that there been no substantial shift in the age profile of people with HNC. While other reports suggest people with HNC are increasingly of a younger age (Deschler et al., 2014b) (Chaturvedi et al., 2013) we found that rates generally increased with age-group and there was no consistent shift to younger age profile over time across all HNC subsites. This was supported by the interaction test results which revealed no temporal change in the age-profile of people with OPC and larynx cancer and marginal increases in the peak age of OCC and overall HNC. These findings could be explained by an aging population observed in Scotland . (Scotland's Census, 2022)

The incidence trends observed in this study are in accordance with a global trend of increases in HNC driven by OPC, accompanied by declines in smoking-related cancers (larynx) among higher-income nations. (Gormley et al., 2022a) Prior epidemiological analyses of UK and international cancer registry data have also shown similar findings of increasing HNC incidence, with rapidly rising OPC rates, declining larynx cancer incidence, and plateauing OCC rates. (Purkayastha et al., 2016) (Louie et al., 2015) (Mourad et al., 2017) (Chaturvedi et al., 2013)

(Gillison et al., 2015) Our analysis was constrained by a lack of individual patient data but the findings were consistent with behavioural trends in the literature. The stabilisation of OCC rates may be explained by persisting alcohol behaviours and the stronger synergistic effects of alcohol and tobacco behaviours associated with cancers of the oral cavity. Increases in OPC are suggestive of a growing burden of HPV-mediated disease. (Lechner et al., 2022) (Wakeham et al., 2019b) (Gribb et al., 2023) The reduction of larynx cancer rates may be explained by declining tobacco smoking trends (Cabinet Secretary for NHS Recovery, 2021), while OCC – which is also associated with smoking and the HNC subsite most strongly associated with alcohol drinking (Lubin et al., 2009) – may not be changing due to persisting levels of harmful alcohol consumption. (Cabinet Secretary for NHS Recovery, 2021) (Hashibe et al., 2009)

This study had several strengths. Notably, the use of high-quality data recorded and kept by the Scottish cancer registry which allows for a robust, population-wide analysis of HNC trends over time. The registry uses a thorough verification process when recording cases, in addition to quality and accuracy checks, with high levels of accuracy and limited numbers of data discrepancies reported. (Brewster et al., 1994) (Brewster et al., 2002) A robust analysis approach was employed to assess the sociodemographic profile of HNC and subsites including assessing interaction with time.

This analysis had some limitations; the Scottish Cancer registry does not collect tumour HPV data, meaning we were unable to compare the sociodemographics of people with HPV-positive and HPV-negative OPC directly. Future work should investigate the associations between the sociodemographics of people with HPV-positive and HPV-negative OPC in a similarly large study. The registry also does not possess behaviour (smoking / alcohol consumption), nor does the registry include individual-level socioeconomic metrics, e.g., education or income. While postcode-derived (data zone) Indexes of Multiple Deprivation can provide some insight into SES, they are limited by the assumption that individuals within areas are all socioeconomically homogenous. Nevertheless, SIMD is considered a comparable and powerful measure of socioeconomic deprivation at the community level. (Scottish Government, 2020) (Bradford et al., 2023)

4.6 Conclusions

We have shown that the sociodemographic profile in terms of age, sex, and socioeconomic background of the incidence burden of head and neck cancers has remained largely unchanged in the last 20 years despite increasing incidence rates which have been driven by rises in oropharynx cancer. These findings are important to help inform efforts to stratify and target prevention, early detection, and cancer services to reach those at the greatest risk. The stark inequality gradient in HNC incidence across all subsites reinforces the importance of upstream tobacco and alcohol regulation, with universal and proportionate prevention strategies to tackle such wide socioeconomic inequalities.

4.7 Additional information and author contributions

4.7.1 Funding Information

This work was funded and supported by Cancer Research UK as part of the CRUK Scotland Centre clinical academic training (TRACC) programme [Grant number 315941-01]

DC, AM and GC are funded by (HEADsPACE) European Union's Horizon 2020 research and innovation programme under grant agreement No 825771.

GI is funded by Cancer Research UK core funding to the CRUK Scotland Institute (A31287) and by a Cancer Research UK core programme award (A29802).

4.7.2 Author contributions

Conceptualisation – DC, CS, AM, CP, CD, GI

Data Curation – CS, KC, LB

Formal Analysis – CS, AM

Methodology - CS, AM, MP, DC

Project Administration - CS, LB, DC

Data visualisation - CS, DC, GC

Writing: original draft - CS

Writing: review & editing - All authors

The manuscript was also critically reviewed by Catherine Winchester (CRUK Scotland Institute).

4.7.3 Ethical Approval

Ethical approval was obtained from both the Scottish Cancer Registry and the University of Glasgow MVLS ethics committee (Project ID Number 200220043).

4.7.4 Data Availability

The datasets analysed during the current study are not publicly available due to the nature of Cancer Registry data but are available from Public Health Scotland (PHS) on request and ethical approval.

5 Chapter Five: Socioeconomic Status and HPV related Cancers of the Oropharynx in the ARCAGE Study

Chapter Four observed a clear inequality gradient with those from lower socioeconomic areas having the greatest incidence of Oropharynx Cancer (OPC). However, the exact relationship between individual-level socioeconomic status (SES) and Human papillomavirus (HPV)-mediated OPC remained unclear due to individual SES and HPV tumour data not being captured by the Scottish Cancer Registry. Chapter Five is a preliminary analysis using data from the ARCAGE case-control study to investigate and quantify any associations between individual socioeconomic status and HPV-mediated OPC risk.

5.1 Introduction

Oropharyngeal cancer (OPC) is defined as squamous cell carcinomas of the soft palate, base of the tongue, tonsils, uvula and posterior pharyngeal wall. As with all head and neck cancers, smoking and alcohol are major risk factors for OPC. However, in more recent years, persistent infection with Human Papillomavirus (primarily strain 16 and to a lesser degree, 18) has been confirmed by the WHO as a causative agent for OPC. (WHO, 2023) (D'Souza et al., 2007)

The incidence of oropharyngeal cancer has dramatically increased, particularly in high-income countries in North America and Europe. (Lorenzoni et al., 2022) In 2022 there were an estimated 106,400 novel cases and 52,305 deaths, globally. (GLOBOCAN, 2022 -a) As consistent with many European nations, OPC incidence in the UK and Scotland has risen dramatically in recent years; OPC incidence in Scotland increased by 85% from 2001-2012. (Purkayastha et al., 2016) Globally, HPV is estimated to be attributable to 33% of OPC cases. (Carlander et al., 2021) There is wide Geographic variation, however, with the highest rates of HPV-positive OPC in the USA and Europe. (Lechner et al., 2022)

The rise of HPV-mediated OPC has led to HPV-negative OPC occurring alongside growing numbers of HPV-positive OPC. HPV-positive and negative OPC also differ clinically in terms of disease progression and prognosis. HPV-positive disease is more responsive to treatments and associated with significantly improved

survival. This improved prognosis for HPV-positive OPC has been attributed to several factors including fewer alterations in the genetic environment, increased sensitivity to oncological therapies, and people with HPV-positive OPC reportedly being of a younger age and better performance status. (Posner et al., 2011, Fakhry et al., 2014) Despite the improved prognosis associated with HPV-positive OPC, the long-term effects of treatment can still have significant impacts on quality of life. (Ang et al., 2010)

As such, the diseases are now staged and treated differently, with the American Joint Committee on Cancer (AAJC) releasing a separate staging system for HPV-positive tumours of the oropharynx in 2018. (Lydiatt et al., 2018) The rationale for this can be explained by the biological and clinical differences exhibited by the tumour types and the necessity to differentiate by HPV status. For instance, nodal involvement, whilst massively detrimental and associated with poor prognosis for people with HPV-negative tumours, is less significant for people with HPV-positive tumours. The introduction of the AAJC 8th edition staging system has been shown to be beneficial for the staging and management of people with OPC. (Valero and Shah, 2021)

Global epidemiological research has indicated that a greater proportion of HPV-positive OPC cases are among males (4:1 Male-Female ratio). (Menezes et al., 2021, Lorenzoni et al., 2022) The registry findings of Chapter 4 (RR = 3.35, 95% CI 3.16 - 3.57) were also consistent with these observations. It has been hypothesised that males may be more predisposed to HPV-infection and subsequent OPC via various behavioural and sexual factors. Males are predisposed to HPV-positive OPC due to sexual and anatomical factors, are more likely to have a greater number of sexual partners, participate in oral sex and have subsequent higher oral HPV prevalence. Females may also be at a lower risk due to existing HPV antibodies from prior cervical infections. (Menezes et al., 2021)

Socioeconomically, HPV-positive OPC patients are reportedly more likely to be from less deprived socioeconomic backgrounds and subsequently will have a better prognosis than HPV-positive patients from more deprived backgrounds. (Dahlstrom et al., 2015b) (Marks et al., 2021) This evidence, however, largely

originates from data gathered in the USA, with a different socioeconomic, ethnic and healthcare landscape to that of Europe. The former study was also constrained by limited numbers of HPV-negative patients. In contrast, a large analysis of SEER and CDC data conducted in the United States found a lower SES was associated with increased incidence of OPC and OCC. However, this analysis was limited by no tumour HPV data and the combined grouping of the two subsites. (Benard et al., 2008) The registry analysis conducted earlier in this thesis consistently observed the highest rates of OPC amongst those most deprived. This strong socioeconomic gradient was observed across the entire study duration. However, that analysis was constrained by a lack of tumour HPV status data. With this unclear relationship between individual socioeconomic status and HPV-positive disease, there is a gap in the literature to explore this.

5.2 Aims

The primary aim of this analysis was to assess and quantify the association between individual-level socioeconomic status (educational attainment) and HPV-positive oropharynx cancer. This analysis also aimed to assess other demographic and behavioural associations with HPV-positive OPC.

5.3 Methods

5.3.1 The ARCAGE Study and data collection

Data from the ARCAGE study was used for this sociodemographic analysis. The ARCAGE study is a multicentre European case-control study that was coordinated by the International Agency for Research on Cancer (IARC). (Lagiou et al., 2009) Participants were recruited across 15 different centres in 11 counties from 2003-2005, with cases and controls matched by five-year age bands and sex.

Ethical approval was obtained via the University of Glasgow Medicine, Veterinary and Life Sciences (MVLS) Ethics System (Project no. 200210024, approved 2/11/21) The majority of the participating centres gave consent to

share study data apart from the Norwegian centre (Oslo). A description of cases and controls by recruitment centre is detailed in Table 5-1. Population-based controls were employed in the UK centres (recruited from matched general medical practices), while all other study centres utilised hospital-based controls. Control participants were inpatients or outpatients recruited from the same hospitals as participating cases. Controls that were admitted for alcohol, tobacco or dietary related reasons were excluded. Recruitment was designed to ensure no single diagnostic group constituted more than a third of the study controls.

The study defined UADT cancer cases as the following ICD-10 codes: "C00 Lip; C01 base of tongue; C02 Other and unspecified parts of the tongue; C03 Gum; C04 Floor of the mouth; C05 Palate; C06 Other and unspecified parts of the mouth; C09 Tonsil; C10 Oropharynx; C12 Pyriform sinus; C13 Hypopharynx; C14.0 Pharynx, wall of pharynx, lateral wall of pharynx, posterior wall of pharynx, retropharynx, throat (all these sites non-otherwise specified); C14.8 Overlapping lesion of lip, oral cavity and pharynx; C15.0 Cervical oesophagus; C15.3 Upper third of oesophagus; C15.4 Middle third of oesophagus; C15.5 Lower third of oesophagus; C15.8 Overlapping lesion of oesophagus; C15.9 oesophagus unspecified; and C32 Larynx. Neoplasms were assigned to the subcategory that included the apparent point of origin of the tumor. Data on C07 (Parotid gland) and C08 (Other and unspecified major salivary glands) cancer cases were also recorded in a separate database for future analyses." (Lagiou et al., 2009) In this thesis, cases of the oesophagus and salivary glands were excluded from analyses.

Table 5-1: List of ARCAGE study centres and participants

Table of Centre by Case			
Centre	**HNC Cases**		
Frequency Col %	**Cases**	**Controls**	**Total**
Prague (Czech Republic)	128 6.7%	187 9.2%	315
Bremen (Germany)	276 14.3%	328 16.1%	604
Athens (Greece)	222 11.5%	194 9.5%	416
Aviano (Italy)	133 6.9%	151 7.4%	284
Padavo (Italy)	118 6.1%	130 6.4%	248
Turin (Italy)	150 7.8%	198 9.7%	348
Dublin (Ireland)	24 1.3%	19 0.9%	43
Glasgow (UK)	97 5.0%	91 4.5%	188
Manchester (UK)	151 7.8%	186 9.1%	337
Newcastle (UK)	65 3.4%	113 5.5%	178
Barcelona (Spain)	185 9.6%	166 8.1%	351
Zagreb (Croatia)	54 2.8%	46 2.3%	100
INSERM (France)	323 16.8%	234 11.5%	557
Total	1926	2043	3969

Detailed sociodemographic, behavioural, and dietary histories were recorded using detailed standardised questionnaires delivered via interviews by trained research teams. Various medical and dental health data were also recorded. Biological and genetic samples were collected; blood samples were largely collected at the interview stage or otherwise at the most convenient opportunity. If possible, fresh tumour samples were collected, with careful recording of handling methods and pathology reports.(Lagiou et al., 2009)

Participant cases and controls were given identical questionnaires (translated into the corresponding language for each centre) and underwent standardised

interviews, where interviewers were trained to minimise missing data by "probing" missing questionnaire responses during interviews. Interviewers were trained according to a standard protocol defined by IARC. Participants were blinded to the research aims of the study to reduce recall biases (e.g., a patient attributing their diagnosis to certain exposures). These methods were informed by a previously conducted feasibility study.

5.3.2 OPC Analysis

Cases in this analysis were defined as tumours of the oropharynx and re-staged according to AJCC 8th edition definitions for tumours of the oropharynx. (Lydiatt et al., 2018) Due to ARCAGE using an older TNM staging classification, the pathological staging system was used to allow for the updated staging of cases. Stages I and II, were classified as "early-stage disease, while stages III and IV were defined as "advanced-stage disease". The cases were stratified into HPV-positive and HPV-negative disease using HPV serology data collected from patient blood collected at the respective study sites. This was done using previously established and described bead-based multiplex serology testing techniques at the DKFZ, Heidelberg, Germany. (Waterboer et al., 2005a, Anantharaman et al., 2013) The HPV-16 seropositivity of cases was derived using definitions detailed by Holzinger et al. (Holzinger et al., 2017b) This was defined as a test of HPV-16E6 MFI > 1000, or three out of four E-proteins greater than threshold values (HPV16 E1 > 200 MFI, HPV16 E2 > 679 MFI, HPV16 E6 > 484 MFI, HPV16 E7 > 548 MFI).

To account for multinational heterogeneity, a standardised education level variable was used by investigators to assess socioeconomic status. Individual educational level data was categorised as "Finished primary school / Worker", "Finished further school / clerk" and "University degree/Manager". Participant smoking and alcohol status was described as "never", "former", "current" or "missing". Measures of smoking frequency and duration were captured using smoking pack-years, where an individual's typical daily number of packs of cigarettes smoked is multiplied by the participant's years of smoking. This was formatted categorically (Never, < 20, 20-39, 40-59, 60+ Pack years, Missing), along with alcohol frequency of consumption data, which was defined using

standardised measures across all sites (Never, less than once per month, 1-4 per month, 1-3 per week, Most days, Every day, Missing, Not in INSERM).

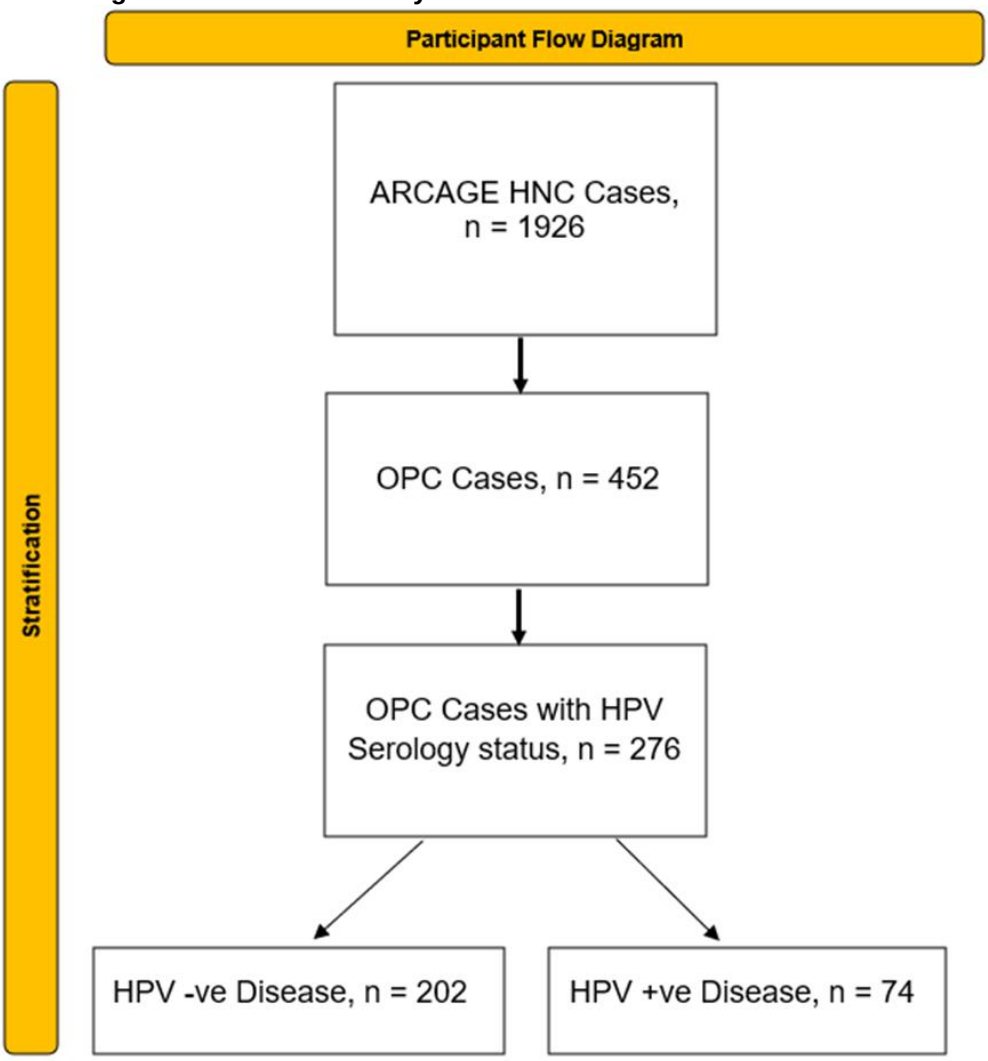
5.3.3 Statistical analysis

Descriptive statistics were calculated for controls and OPC cases further stratified by HPV-serostatus. Where variables were continuous, two sample t-tests were calculated. Chi-squared, or Fisher's tests were calculated for categorical variables. In order to assess associations between socioeconomic status, behaviours and HPV-mediated OPC risk, logistic regression models were constructed for each variable. Unadjusted and fully adjusted multivariable models were constructed, the latter adjusting for age, sex, smoking pack years and alcohol status (owing to data availability). Associations were described using odds ratios and 95% confidence intervals. The descriptive and modelling results were detailed in summary tables. These analyses were conducted for HPV-positive and negative cases, where they were compared directly. HPV-positive and negative cases were also compared indirectly using controls as a sensitivity analysis. All of these analyses were conducted using SAS v9.4.

5.4 Results

There were 452 OPC cases and 2043 controls available for analysis. Of the 452 cases, 61% (n=276) had blood HPV serology data. Of these cases, 74 (26.8%) tested positive for HPV-16, while 202 cases (73.2%) tested negative for HPV-16. Of the HPV-positive and HPV-negative OPC cases, 53 and 161 patients respectively had TNM data that allowed for revised staging according to AJCC eighth edition definitions. This process is summarised in Figure 5-1. Descriptive statistics and associations for these participants are detailed in Tables 5-2, 5-3 and 5-4.

Figure 5-1: ARCAGE Study OPC Case Identification Flowchart



5.4.1 HPV-Positive OPC

The results of descriptive statistics and tests of association for the HPV-Positive OPC Cases and controls are summarised in Table 5-2. A notable proportion of HPV-positive patients (28.4%, n = 21) had missing stage data. Among the HPV-positive OPC cases with staging data, many presented at an early stage, with 48.97% of cases presenting at stages 0, 1 or 2 (n = 36).

Table 5-2: HPV-Positive disease and controls - descriptive statistics and model results

Variable	Status		Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval		p
	Controls (n = 2043)	HPV+ve OPC (n = 74)		Lower	Upper		Lower	Upper	
Age*, Years (mean ± SD)	59.3 (±11.6)	58.8 (± 9.5)	1.00	0.98	1.02	1.00	0.98	1.03	0.77
Sex *									0.20
Female	491 (24.0%)	18 (24.3%)	REF	REF	REF	REF	REF	REF	
Male	1552 (76.0%)	56 (75.7%)	0.98	0.57	1.69	0.68	0.38	1.22	
Disease									
Early	n.a.	36 (48.7%)	n.a.	n.a.	n.a.	.	.	.	
Advanced	.	17 (23.0%)	.	.	.	.	.	.	
Missing		21 (28.4%)							
Stage									
0	n.a.	2 (2.7%)	n.a.	n.a.	n.a.	.	.	.	
1	.	22 (29.7%)	.	.	.	.	.	.	
2	.	12(16.2%)	.	.	.	.	.	.	
3	.	12 (16.2%)	.	.	.	.	.	.	
4	.	5 (6.8%)	.	.	.	.	.	.	
Missing		21 (28.4%)							
Smoking Status									
Never	664 (32.5%)	12 (16.2%)	REF	REF	REF	.	.	.	
Former	700 (34.3%)	27 (36.5%)	**2.85**	**1.47**	**5.54**	.	.	.	
Current	679 (33.2%)	35 (47.3%)	**2.13**	**1.07**	**4.25**	.	.	.	
Smoking Pack Years									**0.003**
Never	664 (32.5%)	12 (16.2%)	REF	REF	REF	REF	REF	REF	
< 20 Pack Years	545 (26.7%)	24 (32.4%)	**2.44**	**1.21**	**4.92**	**2.53**	**1.22**	**5.22**	0.95
20-39 Pack Years	458 (22.4%)	30 (40.5%)	**3.62**	**1.84**	**7.15**	**3.95**	**1.93**	**8.09**	0.93
40-59 Pack Years	224 (11.0%)	5 (6.8%)	1.24	0.43	3.54	1.34	0.45	3.95	0.97

60+ Pack years	148 (7.2%)	3 (4.1%)	1.12	0.31	4.03	1.26	0.34	4.68	0.97
Missing	4 (0.2%)	0	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999	0.96
Alcohol Status									0.38
Never	258 (12.6%)	5 (6.8%)	REF	REF	REF	REF	REF	REF	
Former	184 (9.0%)	4 (5.4%)	1.12	0.30	4.23	0.83	0.48	1.44	0.98
Current	1600 (78.3%)	65 (87.8%)	2.10	0.84	5.26	1.17	0.52	2.67	0.97
Missing	1 (0.1%)	0	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999	0.98
Alcohol Frequency									
Never	238 (11.7%)	5 (6.8%)	REF	REF	REF	.	.	.	
Less 1 per month	107 (5.2%)	4 (5.4%)	1.78	0.47	6.76	.	.	.	
1-4 per month	179 (8.8%)	3 (4.1%)	0.80	0.19	3.38	.	.	.	
1-3 per week	408 (20.0%)	21 (28.4%)	2.45	0.91	6.58	.	.	.	
Most days	185 (9.1%)	10 (13.5%)	2.57	0.87	7.66	.	.	.	
Every day	608 (29.8%)	27 (36.5%)	2.11	0.81	5.55	.	.	.	
Missing	84 (4.1%)	4 (5.4%)	2.27	0.60	8.64				
Not in INSERM	234 (11.5%)	0	<0.001	<0.001	>999.999				
Education (Years)	10.3 (±4.2)	11.4 (±3.6)	1.78	0.47	6.76	.	.	.	
Highest level of education attained									0.79
Finished primary school/Worker	548 (26.8%)	21 (28.4%)	REF	REF	REF	REF	REF	REF	
Finished further school/Clerk	1275 (62.4%)	44 (59.5%)	0.90	0.53	1.53	0.83	0.48	1.44	0.95
University degree/Manager	214 (10.5%)	9 (12.2%)	1.01	0.50	2.43	1.17	0.52	2.67	0.94
Missing	6 (0.3%)	0	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999	0.94

*HPV+ve = HPV-Positive, OPC = Oropharyngeal Cancer, 1.00 (REF) = 1.00 Reference Value, * = Matched, n.a. = not applicable*

Adjusted = Adjusted for Age, Sex, Smoking Pack Years, Alcohol Status, Highest level of education attained

Relative to never smokers, current and former smokers were at an increased risk of HPV-positive OPC. Similarly, relative to never-smokers, participants with a low to moderate smoking packyears were associated with an increased risk of HPV+ve OPC (<20 pack years OR = 2.53, 95% CI = 1.22-5.22); (20-39 pack years OR = 3.95, 95% CI = 1.93 – 8.09). At higher levels of smoking pack years, no associations were found at a univariable or multivariable level (40-59 pack years OR = 1.24, 95% CI = 0.43-3.54); (60+ pack years OR = 1.12, 95% CI = 0.31–4.03). Similarly, no associations were observed for alcohol consumption status and frequency. Relative to those with a low education level, there was no association with a higher socioeconomic status when comparing controls and HPV-Positive OPC cases (multivariable OR = 1.17, 95% CI = 0.52-2.67).

5.4.2 HPV-Negative OPC

A descriptive analysis comparing HPV-negative cases and controls is detailed in Table 5-3. In contrast with their HPV-positive counterparts, many patients with HPV-negative OPC presented with advanced-stage disease (58.4%, n = 118). As with the HPV-positive cases, a notable proportion of the HPV-negative patients had missing stage data (20.3%, n = 41) While non-significant univariably, an increasing age was associated with a small reduction in the risk of HPV-negative OPC at a multivariable level (OR = 0.98. 95% CI = 0.96 – 0.99). Similarly, sex was not associated with HPV-negative OPC risk univariably, but upon multivariable adjustment, male sex was shown to have a protective effect (OR = 0.49, 95% CI = 0.33 – 0.73).

Table 5-3: HPV-Negative disease and controls - descriptive statistics and model results

Variable	Status		Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval		p
	Controls (n = 2043)	HPV-ve OPC (n = 202)		Lower	Upper		Lower	Upper	
Age*, Years (mean ± SD)	59.3 (±11.6)	57.7 (±8.7)	0.99	0.98	1.00	**0.98**	**0.96**	**0.99**	**0.001**
Sex*									**0.001**
Female	491 (24.0%)	42 (20.8%)	REF	REF	REF	REF	REF	REF	
Male	1552 (76.0%)	160 (79.2%)	1.21	0.85	1.72	**0.49**	**0.33**	**0.73**	
Disease									
Early	n.a.	43 (21.3%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	
Advanced	.	118 (58.4%)	.	.	.	.	.	.	
Missing		41 (20.3%)							
Stage									
1	n.a.	21 (10.4%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	
2	.	22 (10.9%)	.	.	.	.	.	.	
3	.	39 (19.3%)	.	.	.	.	.	.	
4	.	79 (39.1%)	.	.	.	.	.	.	
Missing		41 (20.3%)							
Smoking Status									
Never	664 (32.5%)	12 (5.9%)	REF	REF	REF	.	.	.	
Former	700 (34.3%)	38 (18.8%)	**3.00**	**1.56**	**5.80**	.	.	.	
Current	679 (33.2%)	152 (75.3%)	**12.38**	**6.81**	**22.50**	.	.	.	
Smoking Pack Years									**< 0.0001**
Never	664 (32.5%)	12 (5.9%)	REF	REF	REF	REF	REF	REF	
< 20 Pack Years	545 (26.7%)	24 (11.9%)	**2.44**	**1.21**	**4.92**	**2.25**	**1.10**	**4.60**	**0.0004**
20-39 Pack Years	458 (22.4%)	82 (40.6%)	**9.90**	**5.34**	**18.36**	**9.21**	**4.85**	**17.49**	0.06

40-59 Pack Years	224 (11.0%)	50 (24.8%)	**12.35**	**6.46**	**23.60**	**12.56**	**6.37**	**24.75**	**0.003**
60+ Pack years	148 (7.2%)	33 (16.3%)	**12.33**	**6.22**	**24.45**	**13.92**	**6.75**	**28.72**	**0.002**
Missing	4 (0.2%)	1 (0.5%)	**13.82**	**1.44**	**133.11**	**12.01**	**1.16**	**124.91**	**0.47**
Alcohol Status									**< 0.0001**
Never	258 (12.6%)	5 (2.5%)	REF	REF	REF	REF	REF	REF	
Former	184 (9.0%)	42 (20.8%)	**11.78**	**4.57**	**30.34**	**8.32**	**3.11**	**22.24**	0.96
Current	1600 (78.3%)	155 (76.7%)	**5.00**	**2.03**	**12.30**	**4.21**	**1.66**	**10.69**	0.97
Missing	1 (0.1%)	0	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999	0.97
Alcohol Frequency									
Never	238 (11.7%)	5 (2.5%)	REF	REF	REF	.	.	.	
Less 1 per month	107 (5.2%)	8 (4.0%)	**3.56**	**1.14**	**11.13**	.	.	.	
1-4 per month	179 (8.8%)	4 (2.0%)	1.06	0.28	4.02	.	.	.	
1-3 per week	408 (20.0%)	22 (10.9%)	2.57	0.96	6.87	.	.	.	
Most days	185 (9.1%)	10 (5.0%)	2.57	0.87	7.66	.	.	.	
Every day	608 (29.8%)	125 (61.9%)	**9.79**	**3.95**	**24.22**	.	.	.	
Missing	84 (4.1%)	28 (13.9%)	**15.87**	**5.93**	**42.43**	.	.	.	
Not in INSERM	234 (11.5%)	0	<0.001	<0.001	>999.999	.	.	.	
Education (Years)	10.3 (±4.2)	9.2 (±3.4)	**0.93**	**0.90**	**0.97**				
Highest level of education attained									**0.01**
Finished primary school/Worker	548 (26.8%)	73 (36.1%)	REF	REF	REF	REF	REF	REF	
Finished further school/Clerk	1275 (62.4%)	121 (59.9%)	**0.71**	**0.52**	**0.97**	**0.64**	**0.46**	**0.90**	0.93
University degree/Manager	214 (10.5%)	8 (4.0%)	**0.28**	**0.13**	**0.59**	**0.34**	**0.16**	**0.73**	0.95
Missing	6 (0.3%)	0	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999	0.93

*HPV-ve = HPV-Negative, OPC = Oropharyngeal Cancer, 1.00 (REF) = 1.00 (REF)erence Value, * = matched, n.a. = not applicable*

Adjusted = Adjusted for Age, Sex, Smoking Pack Years, Alcohol Status, Highest level of education attained

Smoking status (former and current) was shown to be a significant predictor of HPV-negative disease, this increasing with frequency ($p < 0.001$). Relative to never-drinkers, alcohol usage status was associated with an increased risk of HPV-negative disease, particularly former alcohol consumption (multivariable OR = 8.32, 95% CI = 3.11 – 22.24). Relative to never drinkers, a high frequency of consumption was also associated with an increased risk of HPV-negative OPC (OR = 9.79, 95% CI = 3.95 – 24.22).

Educational level also exhibited a clear association with disease status, with an increased educational level associated with a protective effect relative to those who reported primary school level education, particularly where participants reported a university education or managerial occupational level (multivariable OR = 0.34, 95% CI = 0.16 – 0.73).

5.4.3 HPV-Positive and HPV-Negative OPC

Descriptive statistics and tests of association comparing HPV-positive and negative OPC cases are summarised in Table 5-4. Upon comparing HPV-negative and positive disease, no significant associations for age were observed in favour of HPV-positive or negative disease ($p = 0.35$). No associations with sex were observed at a univariable analysis (OR = 0.82, 95% CI = 0.44 – 1.53) and upon adjustment for other variables males, relative to females, had an increased but non-significant association with HPV-positive disease (OR = 2.04, 95% CI = 0.95 – 4.41).

Table 5-4: HPV-Positive and HPV-Negative OPC - descriptive statistics and model results

Variable	HPV-16 Status of OPC Cases		p	Odds Ratio	95% Confidence Interval		Adjusted Odds Ratio	95% Confidence Interval	
	Negative (n = 202)	Positive (n = 74)			Lower	Upper		Lower	Upper
Age,Years (mean ± SD)	57.7 (±8.7)	58.8 (±9.5)	0.35	1.01	0.98	1.05	1.03	0.99	1.06
Sex			0.53						
Female	42 (20.8%)	18 (24.3%)		REF	REF	REF	REF	REF	REF
Male	160 (79.2%)	56 (75.7%)		0.82	0.44	1.53	2.04	0.95	4.41
Disease			**<.0001**						
Early	43 (21.3%)	36 (48.7%)		REF	REF	REF	.	.	.
Advanced	118 (58.4%)	17 (23.0%)		**0.17**	**0.09**	**0.34**	.	.	.
Missing	41 (20.3%)	21 (28.4%)		0.61	0.31	1.22			
Stage			**<.0001**						
0	0	2 (2.7%)		>999.999	<0.001	>999.999	.	.	.
1	21 (10.4%)	22 (29.7%)		**16.55**	**5.60**	**48.92**	.	.	.
2	22 (10.9%)	12(16.2%)		**8.62**	**2.74**	**27.09**	.	.	.
3	39 (19.3%)	12 (16.2%)		**4.86**	**1.60**	**14.77**	.	.	.
4	79 (39.1%)	5 (6.8%)		REF	REF	REF	.	.	.
Missing	41 (20.3%)	21 (28.4%)		**8.09**	**2.84**	**23.03**			
Smoking Status			**<.0001**						
Never	12 (5.9%)	12 (16.2%)		REF	REF	REF	.	.	.
Former	38 (18.8%)	27 (36.5%)		0.71	0.28	1.82	.	.	.
Current	152 (75.3%)	35 (47.3%)		**0.23**	**0.10**	**0.56**	.	.	.
Smoking Pack Years			**<.0001**						
Never	12 (5.9%)	12 (16.2%)		REF	REF	REF	REF	REF	REF
< 20 Pack Years	24 (11.9%)	24 (32.4%)		1.00	0.38	2.66	1.03	0.35	3.05
20-39 Pack Years	82 (40.6%)	30 (40.5%)		**0.37**	**0.15**	**0.90**	0.41	0.14	1.17

40-59 Pack Years	50 (24.8%)	5 (6.8%)		**0.10**	**0.03**	**0.34**	**0.09**	**0.02**	**0.34**
60+ Pack years	33 (16.3%)	3 (4.1%)		**0.09**	**0.02**	**0.38**	**0.07**	**0.02**	**0.34**
Missing	1 (0.5%)	0		<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Alcohol Status			**0.004**						
Never	5 (2.5%)	5 (6.8%)		REF	REF	REF	REF	REF	REF
Former	42 (20.8%)	4 (5.4%)		**0.10**	**0.02**	**0.48**	**0.11**	**0.02**	**0.74**
Current	155 (76.7%)	65 (87.8%)		0.42	0.12	1.50	0.45	0.09	2.14
Missing									
Alcohol Frequency			**<.0001**						
Never	5 (2.5%)	5 (6.8%)		REF	REF	REF	.	.	.
Less 1 per month	8 (4.0%)	4 (5.4%)		0.50	0.09	2.81	.	.	.
1-4 per month	4 (2.0%)	3 (4.1%)		0.75	0.11	5.24	.	.	.
1-3 per week	22 (10.9%)	21 (28.4%)		0.96	0.24	3.78	.	.	.
Most days	10 (5.0%)	10 (13.5%)		1.00	0.22	4.56	.	.	.
Every day	125 (61.9%)	27 (36.5%)		**0.22**	**0.06**	**0.80**	.	.	.
Missing	28 (13.9%)	4 (5.4%)		**0.14**	**0.03**	**0.72**			
Education (Years)	9.2 (±3.4)	11.4 (±3.6)	**<.0001**	**1.19**	**1.10**	**1.29**	.	.	.
Highest level of education attained			**0.03**						
Finished primary school/Worker	73 (36.1%)	21 (28.4%)		REF	REF	REF	REF	REF	REF
Finished further school/Clerk	121 (59.9%)	44 (59.5%)		1.26	0.70	2.29	1.48	0.74	2.94
University degree/Manager	8 (4.0%)	9 (12.2%)		**3.91**	**1.34**	**11.39**	3.07	0.92	10.30

*OPC = Oropharyngeal Cancer, 1.00 (REF) = 1.00 (REF)erence Value, * = Test for Trend, ** = matched, n.a. = not applicable*

Adjusted = Adjusted for Age, Sex, Smoking Pack Years, Alcohol Status, Highest level of education attained,

Despite large quantities of missing stage data (20.3% HPV-ve, 28.4% HPV+ve), an association was observed between disease stage and the HPV-serostatus of OPC Cases ($p < 0.001$) (Figure 5-2). Early-stage disease (0, I or II) was significantly associated with HPV-positive disease, with HPV-positive OPC cases being 83% less likely to present with advanced-stage (III or IV) disease (OR = 0.17, 95% CI = 0.09 – 0.34). No significant association was observed among cases with missing stage data (OR = 0.61, 95% CI = 0.31 - 1.22).

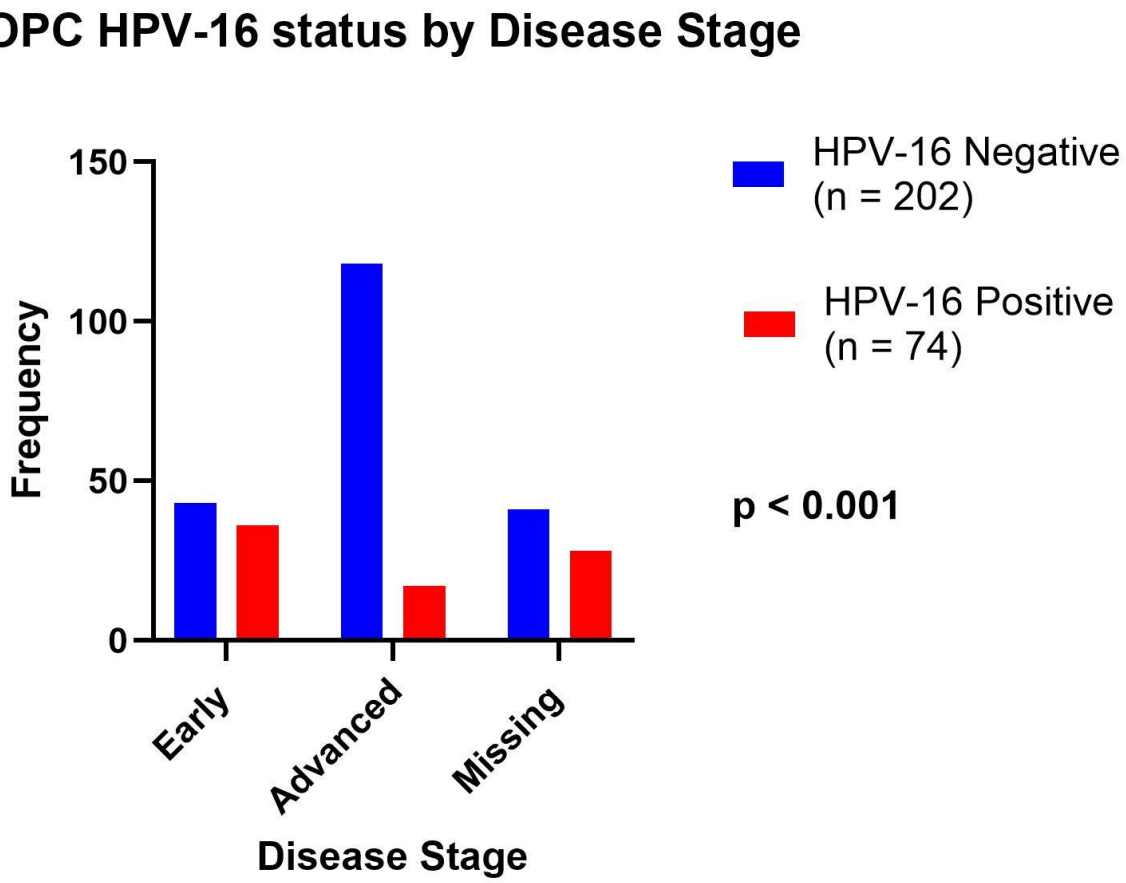


Figure 5-2: Histogram of Early and Advanced Stage OPC disease burden by HPV-seropositivity with Chi-squared test of association.

Participants that reported a current smoking status and high values of pack years were associated with HPV-negative disease at a multivariable level relative to never-smokers (40-59 pack years – OR = 0.09, 95% CI = 0.02 – 0.34); (60+ pack years OR = 0.07, 95% CI = 0.02 – 0.34). Former (multivariable OR = 0.11, 95% CI = 0.02 – 0.74) and daily alcohol consumption (univariable OR = 0.45, 95% CI = 0.09 – 2.14) were both associated with an increased risk of HPV-negative disease relative to never-drinkers, however, this association was only significant among former drinkers.

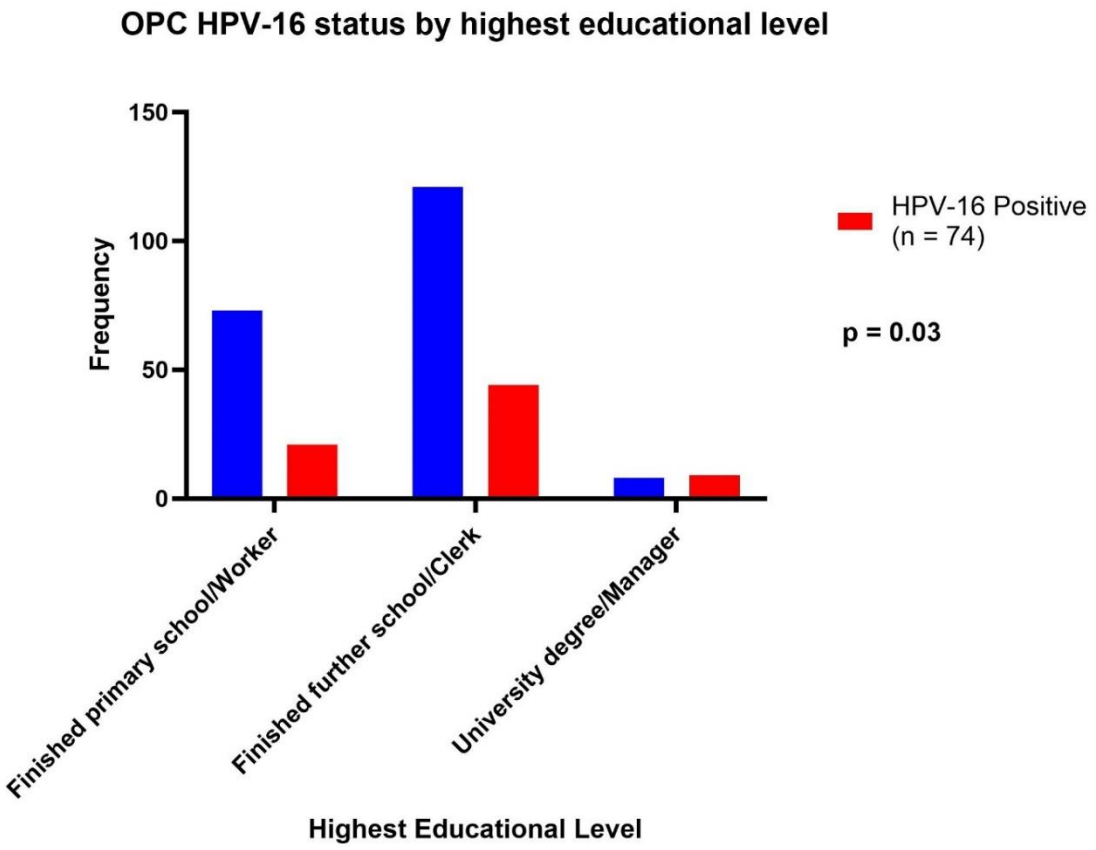


Figure 5-3: Histogram of educational level of OPC cases by HPV-seropositivity with Chi-squared test of association.

Descriptive tests revealed a (narrowly) significant association between individual educational level and HPV-mediated OPC (p = 0.04) (Figure 5-3). Compared with those that completed primary school alone (or workers), the highest category of educational level (University Degree / manager) showed an increased risk of HPV-Positive OPC versus HPV-negative disease at a univariable level (OR = 3.91, 95% CI = 1.34 - 11.39). However, upon multivariable adjustment, a higher

educational level was still associated with an increased risk in favour of HPV-positive disease, but this association lost statistical significance (OR = 3.07, 95% CI = 0.92 – 10.30) (Table 5-3).

5.5 Discussion

This analysis sought to describe the HPV-16 burden among OPC cases within the ARCAGE study and compare the socio-demographics of these patients with HPV-negative disease and controls.

There was some variation in the relationship between individual socioeconomic status (educational level) and the risk of OPC by HPV-serostatus. Upon assessment of HPV-positive OPC cases versus controls, there was no clear association between HPV-positive disease and educational level (p = 0.64). This was the case at both a univariable and a multivariable level. In contrast, earlier research has suggested HPV-positive OPC cases are of a higher SES. (Dahlstrom et al., 2015b)

The comparison of HPV-negative cases with controls revealed that an increased socioeconomic status in the form of educational level was associated with a protective effect, particularly if participants had a university education or held a managerial position (OR = 0.34, 95% CI = 0.16 - 0.73). This remained true across all categories and in both univariable and multivariable analyses. This phenomenon has been shown in several prior analyses, including large pooled international case-control studies, where socioeconomic status has been shown to be an independent predictor of head and neck cancers, even when adjusted for other demographics and behaviours. (Conway et al., 2008) (Conway et al., 2021b)

However, a direct comparison of HPV-positive and HPV-negative disease revealed more mixed findings. A university education or managerial occupation was associated with HPV-positive disease when compared with HPV-negative counterparts. However, upon adjustment, these findings lost significance (OR = 3.07, 95% CI = 0.92 - 10.30). The tendency of increased socioeconomic status

towards HPV-positive disease does correlate with some earlier research suggestive of a less-deprived socioeconomic profile associated with HPV-positive OPC. (Dahlstrom et al., 2015b) However, other analyses observed inequality gradients within OPC, irrespective of HPV status.(Semprini and Williams, 2023) The differences in findings could be attributed to study design and geographical heterogeneity; prior research on the SES of people with HPV-mediated OPC has primarily originated from the United States of America, with a different healthcare and socioeconomic landscape to that of Europe. These prior studies were clinical cohorts or smaller case-control studies; the more affluent socioeconomic profile reported may have been skewed towards higher SES individuals with access to insurance and healthcare. Ultimately, there was no strong evidence from this analysis to definitively suggest a higher socioeconomic status is linked with HPV-positive OPC.

HPV-negative disease was strongly associated with smoking and alcohol behaviours. The strength of these associations increased with frequency and duration. This association was not shared with HPV-positive disease, with only lower frequency smoking associated with statistically raised odds of disease. Upon direct comparison of HPV-positive and HPV-negative disease, smoking and alcohol behaviours were more strongly associated with HPV-negative disease. Our results suggest that tobacco smoking (and potentially alcohol) behaviours may still have an aetiological contribution to the development of HPV-positive OPC, but that persistent HPV infection is likely a larger aetiological factor of disease. These findings echo the existing literature on the role of tobacco and alcohol behaviours in the role of non-HPV-mediated OPCs. (Anantharaman et al., 2016b) Reverse causation could also explain some alcohol findings, where declining health or symptoms of undiagnosed cancer could cause cessation and artificially result in a higher risk associated with cessation than current consumers. (Gapstur et al., 2023)

Variations in associations by age and sex were observed. Despite reports of HPV-positive OPC patients being characteristically younger (median age of 53 years vs 66), there were no significant associations upon the direct comparison of HPV-positive OPC cases with controls nor HPV-negative disease. (Johnson et al., 2020b) HPV-negative patients were younger when compared with controls (p =

0.02). Whilst a non-significant univariable association was observed, when adjusted for other variables, age regained significance. These findings could potentially be attributed to the 5-year age and sex matching of cases and controls which may have attenuated the strength of any of these relationships with subsequent risk.

The relationship between sex and OPC risk was also subject to variation. No significant relationship was observed when comparing HPV-positive disease versus controls. Meanwhile, a protective effect was observed in a multivariable analysis for male sex against HPV-negative disease when compared with controls. (OR = 0.49, 95% CI = 0.33 - 0.73). The inconsistent findings of this analysis could be explained by the sex-matching of controls conducted in the ARCAGE study or by adjustment for smoking and alcohol behaviours.

Upon adjustment, the multivariable model comparing HPV-negative and HPV-Positive disease directly revealed that males were at an increased but non-significant risk of HPV-positive disease (OR = 2.04, 95% CI = 0.95 - 4.41). In the context of the wider literature, incidence rates of both HPV-positive and negative disease among males have been observed to be higher than among females. (Menezes et al., 2021) However, another systematic review and meta-analysis observed a higher proportion of overall OPC among males, but a similar proportion of HPV-positive OPC among male and female patients. (Mariz et al., 2020)

Our analysis had some limitations; the primary limitation of this study was the limited number of HPV-Positive OPC cases (n = 74), as evidenced by some of the wider confidence intervals reported when assessing these cases. The HPV prevalence among the OPC cases was reflective of the more limited HPV-driven disease burden in Europe during the early 2000s. (Lechner et al., 2022) The study also collected no sexual history data; systematic reviews and meta-analyses have shown that a greater number of sexual partners, a younger sexual debut and oral sex behaviours have all been shown to be significant risk factors for HPV infection and OPC. (Chancellor et al., 2017, Durrant et al., 2024) With the exception of UK centres in ARCAGE, the majority of sites in the ARCAGE study used hospital-recruited controls, which may not be as generalisable to the

population. (Sadetzki et al., 2003) Our analysis had some strengths; the ARCAGE study was a well-conducted study and a robust data source. It was a multicentre study conducted according to a standardised protocol from various European centres. Whilst numbers of HPV-positive cases were limited, precise and reliable HPV serology testing methods were used, and detailed behavioural histories were captured for all cases and controls. (Lagiou et al., 2009)

5.6 Conclusions

In conclusion, this analysis found no evidence to suggest a higher socioeconomic status (education) is linked with HPV-positive OPC. A higher educational level was associated with a protective effect against HPV-negative disease, whereas for HPV-positive disease, there was no clear association. While strong behavioural risk associations were observed for HPV-negative disease, limited or no associations were found for HPV-positive disease.

With no conclusive evidence to suggest HPV-OPC patients have a more affluent socioeconomic profile, this analysis, in conjunction with the prior registry work, supports our inclusion of individual socioeconomic status in the all-HNC risk prediction model. There is a further need to replicate this analysis with sufficient case numbers in different strata. This preliminary work lays the foundation for further socioeconomic analysis of OPC cases using pooled data from the VOYAGER and HEADSpAcE Consortiums. As larger and more contemporary studies, they could offer valuable comparisons with the findings of this analysis and allow for further assessment of the sociodemographics of OPC.

6 Chapter Six: Development and External Validation of a Head and Neck Cancer Risk Prediction Model

The contents of this chapter were published to the journal Head and Neck on the 8th of June 2024.

Smith CDL, et al. Development and external validation of a head and neck cancer risk prediction model. Head & Neck. 2024; 1-13.
https://doi.org/10.1002/hed.27834

Authors:

Craig D L Smith*[1,2,3] (corresponding author), Alex D McMahon[1], Donald M Lyall[4], Mariel Goulart[1], Gareth J Inman[2,3,5], Al Ross[6], Mark Gormley[7], Tom Dudding[7], Gary J Macfarlane[8], Max Robinson[9], Lorenzo Richiardi[10], Diego Serraino[11], Jerry Polesel[11], Cristina Canova[12], Wolfgang Ahrens[13], Claire M Healy[14], Pagona Lagiou[15], Ivana Holcatova[16], Laia Alemany[17], Ariana Znoar[18], Tim Waterboer[19], Paul Brennan[20], Shama Virani[18,20], David I Conway[1,3]

Affiliations:

[1]School of Medicine, Dentistry, and Nursing, University of Glasgow, UK

[2]School of Cancer Sciences, University of Glasgow, UK
[3]Glasgow Head and Neck Cancer (GLAHNC) Research Group
[4] School of Health & Wellbeing, University of Glasgow, UK
[5]Cancer Research UK Scotland Institute, Glasgow, UK
[6]School of Health, Science and Wellbeing, Staffordshire University, UK
[7]Bristol Dental School, University of Bristol, UK
[8]Epidemiology Group, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK.
[9]Centre for Oral Health Research, Newcastle University, Newcastle upon Tyne, UK.
[10]Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Turin, Italy.
[11] Unit of Cancer Epidemiology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, Italy.

[12]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardio-Thoraco-Vascular Sciences and Public Health, University of Padua, Padova, Italy.
[13]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[14]School of Dental Science, Trinity College Dublin, Dublin, Ireland.
[15]School of Medicine, National and Kapodistrian University of Athens, Athens, Greece.
[16]Institute of Hygiene and Epidemiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic.
[17]Catalan Institute of Oncology/IDIBELL, Barcelona, Spain.
[18]Cancer Surveillance Branch, International Agency for Research on Cancer, Lyon, France.
[19]Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany
[20]Genomic Epidemiology Group, International Agency for Research on Cancer, Lyon, France.

6.1 Abstract

Background: Head and Neck Cancer (HNC) incidence is on the rise, often diagnosed at late stage and associated with poor prognoses. Risk prediction tools have a potential role in prevention and early detection.

Methods: The IARC-ARCAGE European case-control study was used as the model development dataset. A clinical HNC risk prediction model using behavioural and demographic predictors was developed via multivariable logistic regression analyses. The model was then externally validated in the UK Biobank cohort. Model performance was tested using discrimination and calibration metrics.

Results: 1926 HNC cases and 2043 controls were used for the development of the model. The development dataset model including sociodemographic, smoking and alcohol variables had moderate discrimination, with an Area Under Curve (AUC) value of 0.75 (95% CI, 0.74 - 0.77); the calibration slope (0.75) and tests were suggestive of good calibration. 384,616 UK Biobank participants (with 1177 HNC cases) were available for external validation of the model. Upon external validation, the model had an AUC of 0.62 (95% CI, 0.61 - 0.64).

Conclusions: We developed and externally validated a HNC risk prediction model using the ARCAGE and UK Biobank studies, respectively. This model had moderate performance in the development population and acceptable performance in the validation dataset. Demographics and risk behaviours are strong predictors of HNC, and this model may be a helpful tool in primary dental care settings to promote prevention and determine recall intervals for dental examination. Future addition of HPV serology or genetic factors could further enhance individual risk prediction.

6.2 Introduction

Head and neck cancers (HNC), comprising of cancers of the oral cavity (OCC), pharynx and larynx, are the eighth most common cancer globally with over 800,000 cases and 400,000 deaths in 2020.(Global Cancer Observatory, 2020) (Bray et al., 2018) The incidence of HNC is increasing and projected to further rise by 30% by 2030.(Johnson et al., 2020a)

Key risk factors include tobacco smoking and alcohol consumption, both alone and synergistically in combination.(Hashibe et al., 2009) Additionally, socioeconomic factors are important with those from lower socioeconomic groups having a greater risk and burden of disease.(Conway et al., 2010) The incidence of oropharyngeal cancers (OPC) are the most rapidly rising, which has been attributed to human papillomavirus (HPV) infection.(Parkin and Bray, 2006) (Anantharaman et al., 2017) (Van Dyne et al., 2018) (Schroeder et al., 2020) HNC often presents late, with the majority of global HNC cases being diagnosed at advanced stage (III or IV), which is associated with poorer outcomes and prognosis.(Creaney et al., 2022) (Abrahão et al., 2018) (Abrahão et al., 2020) (Thompson-Harvey et al., 2020) (Ingarfield et al., 2021)

Given the concurrent challenges of growing incidence and late-stage presentation, there has been an increased emphasis on the need for primary and secondary prevention strategies. Risk prediction models and tools have been proposed as having a potential role to help improve earlier detection and promote preventive interventions, such as referrals to smoking cessation services.(Freedman et al., 2005) Risk prediction models for other diseases and cancer sites have already been utilised in primary care settings, for example the Q-risk and Q-cancer series of risk tools.(Hippisley-Cox et al., 2017) (Clin Risk, 2013) Assessment of clinical risk prediction tools has suggested they are beneficial in supporting clinical management and promoting behavioural change.(Walker et al., 2015a) (Kostopoulou et al., 2022)

However, there are a limited number of existing HNC risk tools that have been developed or translated into practice. A review of existing HNC risk models identified that many of the models did not undertake external validation (i.e.

testing the model in a dataset that is independent from that within which the model was developed).(Smith et al., 2022a) This external validation is now widely considered to be an essential feature of clinical risk model development, ensuring that the model is both reproducible and generalisable to other populations.(Ramspek et al., 2020) (Altman and Royston, 2000) (Collins et al., 2014) The growing number of large population-cohort studies offers new opportunities for developing and validating clinically applicable risk models.(Riley et al., 2016)

The aim of this research was to develop and validate a multivariable logistic-regression HNC risk prediction model that can accurately predict and quantify an individual's risk of overall HNC (OCC, OPC and larynx) in the population. This model was designed as part of a primary prevention strategy with the intention of later conducting a feasibility study in primary dental care settings. We hypothesised that a HNC risk model developed using a dedicated HNC case-control study and externally validated in a large population cohort could achieve good predictive performance and generalisability in the population.

6.3 Methods

6.3.1 Definitions and Data Sources

HNC cases were defined as squamous cell carcinomas of the oral cavity, pharynx, and larynx according to WHO International Classification of Disease-10 (ICD-10) codes and definitions (ICD- 10 codes C00.3 - C06, C09 - C14, C32).(World Health Organization, 2004) (Conway et al., 2018b) Cases of the salivary glands and the oesophagus were excluded.

The Alcohol Related Cancers And Genetic susceptibility in Europe (ARCAGE) study was selected as the training dataset for the model. (Lagiou et al., 2009) ARCAGE is a large European multi-centre case control study that was coordinated by the IARC, with 14 different sites across 11 nations.(WHO, 2022) The study recruited over 2000 Upper Aero-Digestive Tract (UADT) cancer cases

and controls (age and sex matched) from 2002 to 2005. Further descriptions of study methods and data collection are detailed in Chapter 5.

The UK Biobank cohort study was selected for model validation. It has over 500,000 participants recruited from 2006-2010. Data included sociodemographic, behavioural, clinical, and genetic information. The UK Biobank is also linked to national cancer and death registries, which allows for ready identification of newly diagnosed and existing cases within the cohort.(Sudlow et al., 2015) (Conroy et al., 2022) (Rory, 2007)

The HNC risk prediction model development and validation were conducted in accordance with the Transparent Reporting of a multivariable Prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines.(Collins et al., 2015a) Ethical approval for secondary data analysis was obtained from the MVLS college ethics committee of the University of Glasgow (Project no: 200210024). The ARCAGE study had original ethical approval from IARC and local research ethics boards, while the UK Biobank received ethical approval from the North West Multi-centre Research Ethics Committee (MREC).

6.3.2 Model Development in ARCAGE

Logistic regression modelling was used to compute odds ratios (ORs) and 95% Confidence Intervals (95% CI). Model discrimination was reported using the Area Under Curve (AUC) values with 95% CIs and the calibration was reported using Spiegelhalter's Z statistic.(Walsh et al., 2017) The model was designed for practicality in a primary care setting. Frequencies and means were also calculated for each variable. In addition, univariable logistic regression analysis was conducted and AUCs and ORs with 95% CIs were reported.

Three sequential strategies were used for variable predictor selection. Firstly, a "black box" approach was used, which was essentially an agnostic logistic regression of all available variables in the ARCAGE study to identify key statistically significant variables. Following this, the logistic regression model was refined based upon existing evidence of HNC risk factors; models were

constructed using HNC risk factors established by pre-existing literature including original ARCAGE study analyses.(Gormley et al., 2022a, Shaw and Beasley, 2016, Hashim et al., 2019a, Hashibe et al., 2013, Bravi et al., 2012, Bravi et al., 2021, Macfarlane et al., 2010, Anantharaman et al., 2013, Ahrens et al., 2014) The third and final strategy entailed finalising the model informed by the black box and literature with variables that were (i) available in the UK Biobank, and (ii) would be feasible for recording in a clinical setting (for example, behaviours such as smoking and alcohol behaviours are relatively easy to assess and are recorded routinely at new patient or check-up examinations in primary dental care, while a food frequency questionnaire might prove difficult to include in such routine appointments). Forward selection was used to select variables, with backward selection also used as a quality check.

Descriptive statistics and univariable associations were described for the key variables considered for the model at the literature-informed stage. Variables that were not ultimately selected were excluded for the following reasons: failure to survive stepwise selection in a multivariable model; considered impractical to test in a primary dental care setting; or the variable lacked sufficient data in the validation dataset. ORs and 95% CIs were calculated in multivariable logistic regression for the variables that were ultimately selected in the final model. A complete case analysis approach was adopted; variables with 10% or more missing data were categorised or removed from model development altogether.(Hughes et al., 2019) (Roderick J. A. Little, 2002)

6.3.3 Model Validation in UK Biobank

Cases were identified in the UK Biobank by adopting previous methods and code used by Burrows and colleagues, but matched according to our pre-defined list of ICD codes.(Burrows and Haycock, 2021) Cases with cancer diagnoses prior to 1st April 2007 (baseline assessment) were excluded (n=61). If more than one cancer was diagnosed, the first chronological instance was taken to avoid duplication of cases (n=285). Non-cancer patients were defined as individuals with no cancer diagnosis (n=383,442). Variables were formatted to match the formatting of the ARCAGE study variables that were selected during model development. A later attempt to stratify models by sex was also made.

Once the HNC cases within the cohort were identified, descriptive analysis, probability calculations and the subsequent logistic regression using the coefficients from ARCAGE were conducted. The methods used for reporting the performance during model development were repeated for the validation process. Model discrimination was reported using the Area Under the Receiver Operating Curve (AUC) with 95% CIs and calibration was reported using Spiegelhalter's Z statistic.

Frequencies, means and descriptive tests were also calculated for each variable, using two sample or Welch's two sample T-tests and Chi-square or continuity corrected score tests for categorical data.

Model training and validation analyses were conducted using SAS v9.4 and R version 4.2.2

6.4 Results

6.4.1 Model Development in ARCAGE

The ARCAGE study had 1926 HNC cases and 2043 controls for model development. A summary of the descriptive results of the study are summarised in Table 6-1. Cases and controls were broadly similar in terms of age and sex, as expected with the ARCAGE study matching controls by age-group and se

Table 6-1 Descriptive, univariable and multivariable results for key model development variables in the ARCAGE Study

Variable	Cases (n = 1926)	Controls (n =2043)	p	Univariable Odds Ratio (95% CI)	Final Model Multivariable Odds Ratio (95% CI)
Oral Cavity	490 (25.4%)	.	.	.	.
Oropharynx	452 (23.5%)	.		.	.
Larynx	670 (34.8%)	.		.	.
Hypopharynx	184 (9.6%)	.		.	.
Overlapping	130 (6.7%)	.		.	.
Age, Mean (± SD)	58.8 (±10.2) years	59.3 (±11.6) years	0.18	1.00 (0.99 - 1.00)	**1.01 (1.002 - 1.02)**
Sex:			**<0.0001**		
Male	1584 (82.2%)	1552 (76.0%)		**1.47 (1.26 - 1.71)**	**0.61 (0.50 - 0.75)**
Female	342 (17.8%)	491 (24.0%)		REF	REF
Years of Education:			**<0.0001**		
16+ years	98 (5.1%)	196 9.6%)		REF	REF
No education	35 (1.8%)	34 (1.7%)		**2.06 (1.21 - 3.50)**	1.61 (0.87 - 3.00)
1-3 years	53 (2.8%)	34 (1.7%)		**3.12 (1.90 - 5.11)**	**2.00 (1.13 - 3.54)**
4-6 years	285 (14.8%)	255 (12.5%)		**2.24 (1.66 - 3.00)**	**1.49 (1.06 - 2.09)**
7-9 years	439 (22.8%)	433 (21.2%)		**2.03 (1.54 - 2.67)**	**1.41 (1.03 - 1.92)**
10-12 years	491 (25.5%)	599 (29.3%)		**1.64 (1.25 - 2.15)**	1.23 (0.91 - 1.67)
13-15 years	165 (8.6%)	236 (11.6%)		**1.40 (1.02 - 1.91)**	1.27 (0.90 - 1.81)
Missing	360 (18.7%)	256 (12.5%)		**3.36 (1.88 - 6.01)**	1.32 (0.95 - 1.83)

Highest Educational Level:			0.32		
Finished primary school	684 (35.5%)	548 (26.8%)		**2.23 (1.73 - 2.86)**	Not included*
Finished further school/clerks	1117 (58.0%)	1275 (62.4%)		**1.56 (1.23 - 1.98)**	Not included*
University degree/Manager	120 (6.2%)	214 (10.5%)		REF	Not included*
Missing	5 (0.3%)	6 (0.3%)		1.48 (0.44 - 4.98)	Not included*
Smoking Status:			**<0.0001**		
(Never)	158 (8.2%)	664 (32.5%)		REF	REF
(Former)	452 (23.5%)	700 (34.3%)		**2.71 (2.20 - 3.35)**	**1.96 (1.54 - 2.50)**
(Current)	1316 (68.3%)	679 (33.2%)		**8.15 (6.69 - 9.92)**	**5.27 (4.11 - 6.76)**
Smoking, Pack years Mean (± SD)	41.2 (±34.5)	21.6 (±33.7)	**<0.0001**	**1.03 (1.02 - 1.03)**	**1.01 (1.004 - 1.01)**
Alcohol drink status:			**<0.0001**		
(Never)	111 (5.8%)	258 (12.6%)		REF	REF
(Former)	309 (16.0%)	184 (9.0%)		**3.90 (2.93- 5.20)**	Not included*
(Current)	1505 (78.1%)	1600 (78.3%)		**2.19 (1.73 - 2.76)**	Not included*
(Missing)	1 (0.1%)	1 (0.1%)		2.32 (0.14 - 37.5)	Not included*
Alcohol Daily Drink Frequency (Measures):			**0.02**		
Never	111 (5.8%)	258 (12.6%)		**0.26 (0.20 - 0.35)**	0.05 (0.00 - 1.09)
<1	414 (21.5%)	733 (35.9%)		**0.35 (0.28 - 0.42)**	**0.47 (0.37 - 0.59)**
1 to 2	522 (27.1%)	647 (31.7%)		**0.49 (0.40 - 0.60)**	**0.60 (0.48 - 0.74)**
3 to 4	367 (19.1%)	224 (11.0%)		REF	REF
5 to 6	233 (21.1%)	73 (3.6%)		**1.95 (1.43 - 2.66)**	**1.83 (1.32 - 2.53)**
7+	236 (12.3%)	67 (3.3%)		**2.15 (1.56 - 2.96)**	**1.81 (1.29 - 2.53)**
Missing	99 (2.2%)	41 (2.0%)		0.64 (0.41 - 1.01)	0.72 (0.43 - 1.20)

BMI kg/m², Mean (± SD)	24.3 (±4.5)	26.2 (±4.4)	**<0.0001**	**0.91 (0.89 - 0.92)**	Not Included**
Fruit Consumption:			0.54		
Never	69 (3.6%)	33 (1.6%)		REF	Not Included**
Once per month or less	69 (3.6%)	35 (1.7%)		0.94 (0.53 - 1.69)	Not Included**
Several times per month	43 (2.2%)	17 (0.8%)		1.21(0.60 - 2.43)	Not Included**
Once per week	192 (10.0%)	104 (5.1%)		0.88 (0.55 - 1.43)	Not Included**
Several times a week	550 (28.6%)	421 (20.6%)		**0.63 (0.41 - 0.96)**	Not Included**
Once per day	481 (25.0%)	618 (30.3%)		**0.37 (0.24 - 0.57)**	Not Included**
Several times per day	492 (25.6%)	799 (39.1%)		**0.30 (0.19 - 0.45)**	Not Included**
Missing	30 (1.6%)	16 (0.8%)		0.90 (0.43 - 1.87)	Not Included**
Frequency of Dental Attendance:			**<0.0001**		
Never	298 (15.5%)	174 (8.5%)		REF	Not included***
Less than every 5 years	501 (26.0%)	389 (19.0%)		**0.75 (0.60 - 0.95)**	Not included***
Every 2 to 5 years	348 (18.1%)	402 (19.7%)		**0.51 (0.40 - 0.64)**	Not included***
At least every year	423 (22.0%)	820 (40.1%)		**0.30 (0.24 - 0.38)**	Not included***
Missing	356 (0.18%)	258 (0.13%)		0.80 (0.46 - 1.40)	Not included***
Denture Use:			**<0.0001**		
Never	770 (40.0%)	1065 (52.1%)		REF	Not Included*
Ever	830 (43.1%)	739 (36.2%)		**1.55 (1.36 - 1.78)**	Not Included*
Missing	3 (0.2%)	5 (0.2%)		0.83 (0.20 - 3.48)	Not Included*
Not in INSERM	323 (16.8%)	234 (11.5%)		**1.91 (1.58 - 2.31)**	Not Included*

HPV-16 Negative	1078 (56.0%)	1252 (61.3%)	**<0.0001**	REF	Not included***
HPV-16 +ve Positive	85 (4.4%)	5 (0.2%)		**19.73 (7.99 - 48.78)**	Not included***
Missing	763 (39.6%)	786 (38.5%)		1.13 (0.99 - 1.28)	Not included***

Note: Statistical significance is highlighted in **bold** ($p < 0.05$). Model intercept = - 0.18

* - not included for statistical reasons / variable inclusion

** - not included for clinical practicality reasons

*** - not included due to no comparable variable or insufficient data in Biobank dataset

As consistent with the “black box” model and evidence from existing literature, male sex, increasing age, lower educational attainment (and virtually synonymous years of education), smoking, alcohol consumption frequency and HPV-16 seropositivity, (defined as a test of HPV-16E6 MFI > 1000, or 3 out of 4 E-proteins greater than threshold values (HPV16 E1 > 200 MFI, HPV16 E2 > 679 MFI, HPV16 E6 > 484 MFI, HPV16 E7 > 548 MFI)) were associated with an increased HNC risk.(Holzinger et al., 2017a) Regular dental visits, fruit and vegetable consumption, and increased BMI were associated with modest protective effects.

The ARCAGE study collected a detailed food frequency history which was deemed impractical to replicate in a clinically applied model, resulting in the decision to drop dietary variables from the final model. The BMI variable offered only a marginal improvement in prediction (AUC of 0.76), and there were concerns about conflicting evidence on the relationship between BMI and HNC risk from case-control studies and cohort studies – such that validating case-control derived data in a cohort would not improve prediction. Moreover, BMI was considered more challenging to accurately measure in some primary care settings (e.g. dental practices) where scales and stadiometers may not always be routinely available. This could also lead to potential recall biases and metric conversion challenges, impeding this variables utility. For these reasons, BMI was not included in further modelling.

Frequency of attendance at a dental practice was a statistically significant predictor of HNC risk. However, the UK Biobank had no comparable variable and participant data for dental practice attendance frequency, which was consequently dropped from the model selection.

HPV-16 serostatus data provided an increase in HNC prediction (AUC of 0.80, 95 % CI = 0.79 - 0.82) and excellent calibration (Supplementary Figure 6-3). However, a number of the ARCAGE study participants lacked HPV serology (n = 1549, 39.0%). Furthermore, at the time of writing, 9,695 UK Biobank participants were randomly sampled for HPV testing and subsequently, only a proportionately small proportion of our validation UK Biobank dataset sample (n = 7238, 1.9%) had HPV serology data available for analysis. Of these participants with HPV

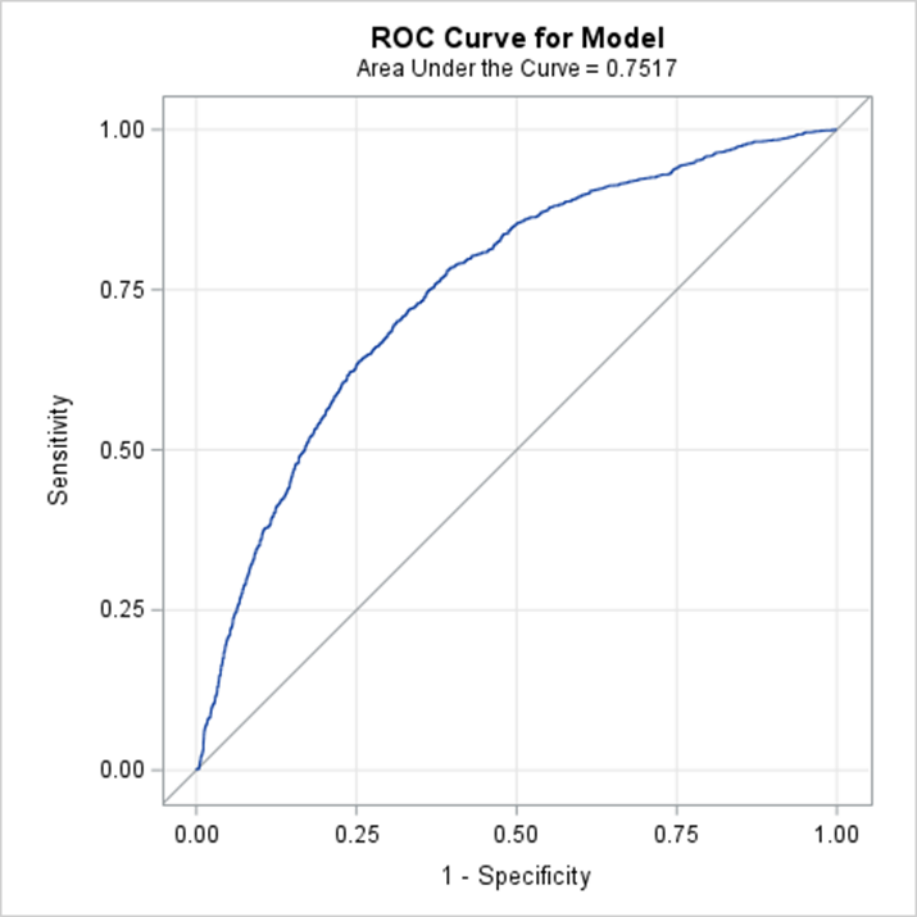
serology data, only four of the 1177 HNC cases had an HPV-positive serology test (0.3%) - making effective validation with this variable non-viable.

Thus, the final prediction model included age, sex, socioeconomic status via categories of years of education, smoking status, smoking pack years, alcohol consumption status, and alcohol consumption frequency (Table 6-1).

Age was associated with an increased risk association with each year. Females were at an increased risk of HNC versus their male counterparts, which may be attributable to matching. Increased risks for HNC were observed for: low relative to a high number of years in education; current (and former) smoker relative to never smoking status; increased number of pack years relative to zero; and a high frequency of alcohol consumption relative to never drinking alcohol.

The final risk prediction model for development (Figure 6-1) had an AUC of 0.75 (95% CI, 0.74 – 0.77). The results of Spiegelhalter's Z-test for calibration (-0.603, p = 0.55) suggest the model was calibrated.

Figure 6-1: Receiver Operating Curve for final ARCAGE Development Model



6.4.2 Model Validation in UK Biobank

Table 6-2: UK Biobank Cohort study Descriptive Statistics Results

Variable	UK Biobank Cohort (%) n = 384,616	HNC Cases, (%) n = 1177	Univariable Odds Ratio (95% CI)	p
Oral Cavity	n.a	490 (25.4%)	.	.
Oropharynx	n.a	452 (23.5%)	.	.
Larynx	n.a	670 (34.8%)	.	.
Hypopharynx	n.a	184 (9.6%)	.	.
Overlapping	n.a	130 (6.7%)	.	.
Age, years, mean (± SD)	55.6 (±8.1)	58.3 (±7.1)	**1.05 (1.04 - 1.05)**	**<0.0001**
Sex				**<0.0001**
Female	208740 (54.3%)	313 (26.6%)	REF	.
Male	175876 (45.7%)	864 (73.4%)	**3.29 (2.89 - 3.74)**	.
Years of Education (%)				**<0.0001**
0: No Education	376 (0.1%)	0	<0.001 (<0.001 - >999.999)	.
1: <1-3 Years	320 (0.1%)	1 (0.1%)	1.28 (0.18 - 9.15)	.
2: 4-6 Years	243 (0.1%)	2 (0.2%)	3.39 (0.84 - 13.70)	.
3: 7-9 Years	3258 (0.8%)	20 (1.7%)	**2.56 (1.61 - 3.97)**	.
4: 10-12 Years	180575 (46.9%)	646 (54.9%)	**1.47 (1.29 - 1.67)**	.
5: 13-15 Years	44908 (11.7%)	130 (11.1%)	1.19 (0.97 - 1.45)	.
6: 16+ Years	154936 (40.3%)	378 (32.1%)	REF	.
Missing	0	0	N.A.	.
Smoking Status (%)				**<0.0001**
Never	214642 (55.8%)	326 (27.7%)	REF	.
Former	127382 (33.1%)	524 (44.5%)	**2.72 (2.36 - 3.12)**	.
Current	40355 (10.5%)	316 (26.9%)	**5.19 (4.44 - 6.06)**	.
Missing	1060 (0.3%)	11 (0.9%)	**3.25 (1.78 - 5.93)**	.

Smoking Pack years, mean (± SD)	6.5 (±14.2)	19.8 (±26.4)	**1.03 (1.03 - 1.03)**	**<0.0001**
Alcohol Drink Frequency (%)				**0.02**
0: Never	31494 (8.2%)	127 (10.8%)	REF	.
1: 1/2 x a week or Special Occasions only	44257 (11.5%)	86 (7.3%)	**0.48 (0.37 - 0.63)**	.
2: 1-3 x a month	43477 (11.3%)	78 (6.6%)	**0.44 (0.34 - 0.59)**	.
3: 1/2x a week	99803 (26.0%)	290 (24.6%)	**0.72 (0.58 - 0.89)**	.
4: 3/4 x a week	88527 (23.0%)	233 (19.8%)	**0.65 (0.53 - 0.81)**	.
5: Daily or almost daily	75820 (19.8%)	358 (30.4%)	1.17 (0.96 - 1.44)	.
Missing	331 (0.1%)	5 (0.4%)	1.00 (0.41 - 2.45)	.

Note: Statistical significance is highlighted in **bold** ($p < 0.05$).

Descriptive statistics of the validation population are summarised in Table 6-2. Following data management procedures (Figure 6-2) there were 384,616 participants that were available for model validation. Within this, there were 1177 HNC cases, of which the largest proportion were cases of the oropharynx (n = 453, 38.5%). Upon external validation, the final risk prediction model (Figure 6-3) had an AUC of 0.62 (95% CI, 0.61 - 0.64). The results of Spiegelhalter's z-test (-0.013, p = 0.99) suggested that the model has acceptable calibration.(Walsh et al., 2017)

Figure 6-2: ARCAGE and UK Biobank Participant Flowchart

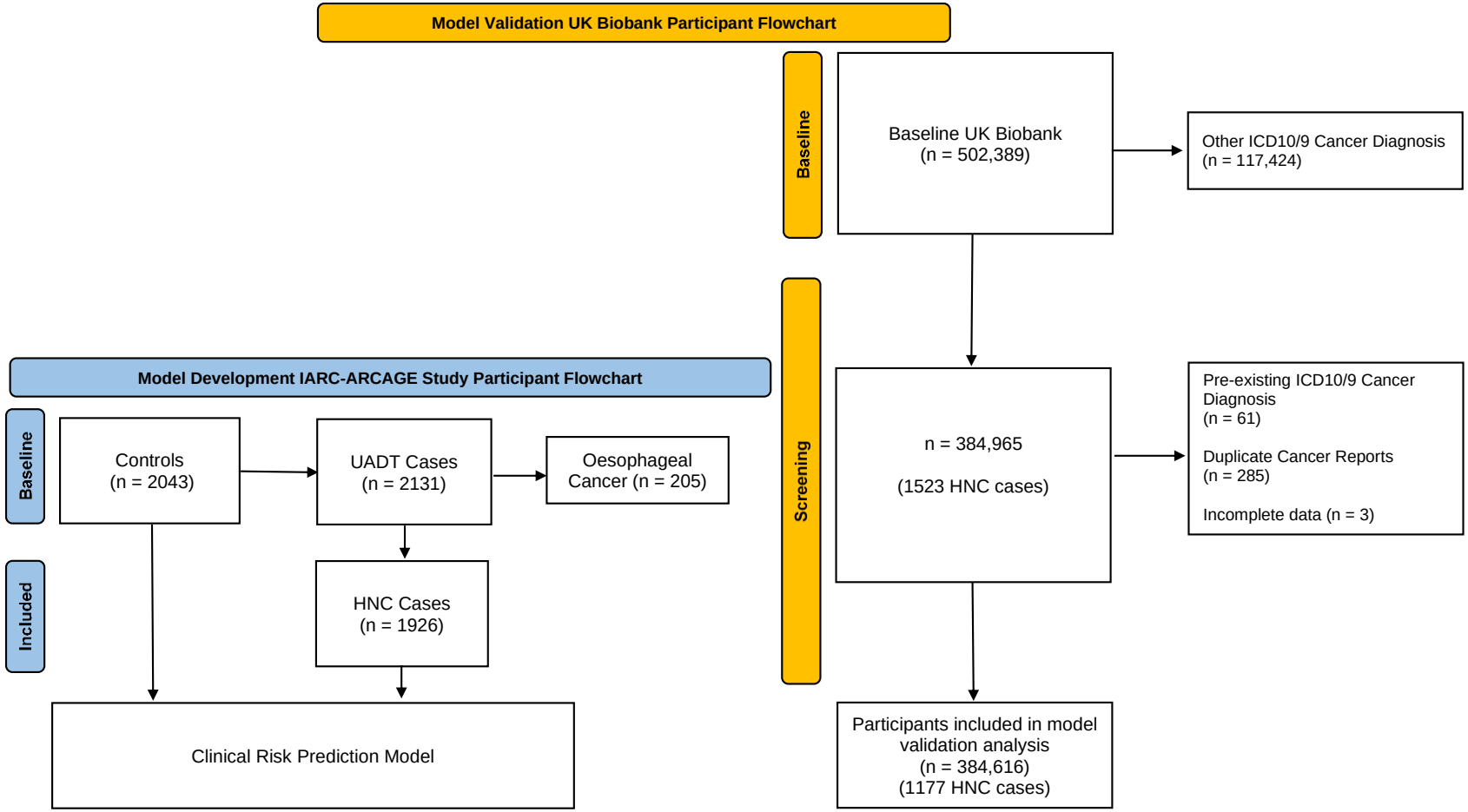
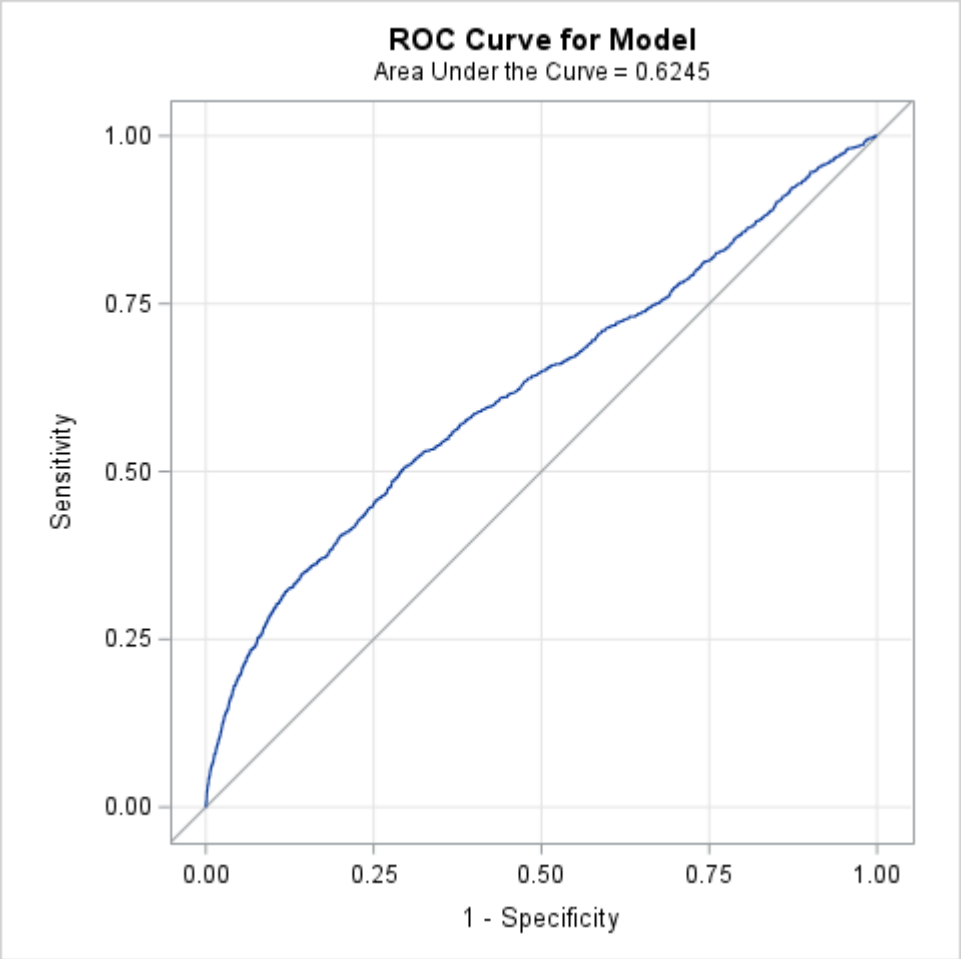


Figure 6-3: Receiver Operating Curve for the Validation of the HNC Risk Prediction Model in the UK Biobank



6.4.3 Sensitivity Analysis

An attempt to account for potential HPV-associated OPC cases was made by trialling the same model for OCC and Laryngeal cases only (Supplementary Figure 6-1). However, this only yielded a marginal improvement in discriminative performance in the validation dataset with an AUC of 0.63 (95% CI, 0.60 - 0.65).

Similarly, a model was created using exclusively UK participants in ARCAGE to account for potential heterogeneity associated with the multi-national nature of the study; ARCAGE UK centres used population controls, while other centres used hospital patient controls. On validation, this also offered limited discriminative performance (supplementary Figure 6-2) with an AUC of 0.52 (95% CI, 0.51 - 0.54) in the UK Biobank. Another model, including denture use offered little improvement in performance (AUC of 0.77, 95% CI, 0.75 - 0.78) and had limited discrimination upon validation (AUC = 0.61, 95% CI, 0.59 - 0.63). (Supplementary Figures 6-4 and 6-5).

6.5 Discussion

We developed a risk prediction model for all HNC sites using two separate sources – a European multicentre HNC case-control study for model development and a UK population-based cohort study for model validation. The model performed well in the developmental dataset. Upon validation, the AUC results show that while the model can predict individual risk of HNC, its discriminative ability is acceptable, but more limited, in the UK Biobank. Similar findings were observed when the model was developed from OCC and laryngeal subsites, and exclusively UK participants. The models were calibrated, with non-significant results for Spiegelhalter's Z-test suggestive that we can accept the null hypothesis that models were well calibrated.

Two other HNC risk models have made use of these study datasets. Budhathoki and colleagues recently developed multiple models stratified by subsite using pooled data from five separate studies including data from the ARCAGE and UK Biobank studies. The models included epidemiological risk factors, HPV

serostatus, polygenic risk scores (PRS) and combinations of these.(Budhathoki et al., 2023) Our model took a different approach, opting ultimately for feasibility and practicality of use by predicting overall HNC risk using epidemiological predictors that could readily be captured in a clinical setting, as opposed to the site and gender specific models created by Budhathoki and colleagues.(Budhathoki et al., 2023) These epidemiology models performed marginally better than our model using demographic and behavioural factors. However, interestingly, the variable selection was largely similar. The models were also well calibrated. The discrepancy in performance could perhaps be explained by the larger sample size from the pooled studies used by Budhathoki et al for both model development and validation via randomly splitting the dataset rather than using an independent external validation dataset as conducted by our study.(Budhathoki et al., 2023) The models using HPV serostatus were highly predictive of OPC but seemingly less predictive for overall HNC risk, as consistent with our findings. Similarly, the models using a combination of epidemiological and PRS had good predictive performance. However, the use of models incorporating these factors is not as feasible to replicate in primary care and community settings at present.

Another HNC risk prediction model developed by McCarthy and colleagues using the UK Biobank, split the dataset geographically for development and validation. (McCarthy et al., 2020) This model used demographic predictors, in addition to smoking and alcohol consumption status, BMI, exercise levels and daily fruit / vegetable consumption. (McCarthy et al., 2020) The model had good calibration and marginally improved, but relatively limited, discriminative performance with an AUC of 0.64. The performance of this model, like ours, could perhaps be explained by the sole use of the UK Biobank as a validation dataset and use of epidemiological predictors.

Notably, following our evaluation of this, we opted to exclude BMI due to temporal variability on its risk relationship in the literature. The relationship between BMI and HNC risk may be subject to temporal variation depending on the time point assessed. There is an existing body of evidence derived from case-control analyses, including that of the ARCAGE study, suggestive of a lower BMI being associated with an increased HNC risk.(Chen et al., 2019) (Park et al.,

2011) (Gaudet et al., 2010b) While some of these studies assessed BMI estimate at mid-life (e.g. at age 30-years), longitudinal cohort studies show either an increased or no clear HNC risk with higher BMI over a longer period(Gaudet et al., 2012) (Maasland et al., 2015a) (Gormley et al., 2023) (Recalde et al., 2023), which is more similar to the risk relationship for many other cancers where an increased BMI is associated with increased inflammatory burden, various comorbidities, and subsequent cancer risk.(Recalde et al., 2023) (Bhaskaran et al., 2014)

Studies developing risk models for other cancers (including colorectal and renal cancers) in the UK Biobank have shown variability in performance.(Harrison et al., 2022) (Usher-Smith et al., 2018) Most models showed limited to reasonable (AUC > 0.60) levels of discrimination within the UK Biobank, with similarly varying levels of calibration. The variables selected for our model and those considered for selection, but not ultimately chosen, chime with the existing literature on HNC epidemiology. Demographic factors including age, sex, and socioeconomic status are well established predictors of HNC.(Gormley et al., 2022a) (Park et al., 2022) (Purkayastha et al., 2016) Our model performance metrics are also supportive of these findings. Similarly, smoking and alcohol consumption have also been shown to be highly predictive of HNC both in the literature and within our model, these also having clear dose relationships.(Hashibe et al., 2009)

The extent of missing HPV data in the UK Biobank (98.1%) and limited number of HPV-Positive HNC cases meant accurate validation of an HPV model was not viable. Furthermore, our model was designed with the intention to be non-invasive for feasibility testing in a clinical setting, so our variables were chosen with the practicalities of this in mind. However, there is undoubtedly future scope for HPV status to be used for risk prediction of OPC, especially as technology and testing methods continue to improve. This is evidenced by a recent cohort study that was undertaken in Hamburg, where population HPV antibody testing and follow-up were used in order to inform risk stratification and investigations. This allowed the investigators to detect HPV positive OPC cases at an earlier stage.(Busch et al., 2022a) Suggestions for any further modelling would be to utilise HPV serology status data (ideally with the

development of a “rapid” test that could be used in primary care) for use in a separate OPC risk model, as conducted by Budhathoki et al or Tota et al.(Budhathoki et al., 2023) (Tota et al., 2019b) However, as stated, this was out of the scope of this project. While genomic or HPV biomarkers could improve the predictive accuracy of a risk model, their inclusion limits the utility of a tool in primary care settings with limited time/resources. There may be evidence to suggest that given the heterogeneity of subsites (and their relevant risk factors) included in HNC, future models should stratify by subsite. However, high-risk behaviours (such as smoking and alcohol) and sociodemographic predictors are generalisable across all HNC subsites, even among people with HPV-positive tumours. This approach also becomes more challenging for less common subsites where fewer cases for analysis are available (e.g. hypopharynx).

Another notable HNC risk prediction model was the HANRC V.2 tool developed by Tikka and colleagues. (Tikka et al., 2020) This model had excellent performance and has seen use in secondary and tertiary care settings.(Tan et al., 2024) (Banerjee et al., 2021) However, the model focuses on clinical signs and symptoms of HNC, many of which are associated with existing or advanced stage disease. In contrast, our model was designed with the specific complementary intention of primary prevention activity in a dental care setting, where at the time of writing, no such tool exists. Thus, what sets this model apart was its deliberate design with ease of use in primary care at the forefront, achieved through the integration of robust yet readily accessible predictors that could inform preventive dialogues and inform recall activity.

This study had some strengths. We used a large, multinational HNC-focused case-control study that allowed for the selection of robust and predictive variables to assess HNC risk. The pooling of participants allowed for more accurate estimates of risk and greater generalisability, than a UK-based study alone. The use of a large UK-based population cohort as a validation dataset allowed for high-quality model validation. All of these analyses were also conducted in accordance with TRIPOD guidelines.

Our model also faced some limitations. Firstly, the cases were age-category and sex matched in the ARCAGE case-control study. This could have potentially

weakened or altered the associations between two key demographic predictors (as observed with sex, where being female was associated with an increased HNC risk multi-variably). However, despite this, the model fared well in the development stages. Stratification by sex yielded modest improvements in male predictive performance upon validation but at the cost of female predictive performance (Data not shown).

There is also evidence to suggest that, comparatively, participants in the UK Biobank are less socioeconomically deprived than the general population.(Fry et al., 2017) Thus, the "healthy volunteer" effect associated with large volunteer cohorts may have also attenuated the performance of the model in the validation dataset, as previously observed leading to underestimation of the strength of associations between exposures and outcomes.(Lyall et al., 2022)

Finally, one of the major limitations and challenges of this analysis was the matching of predictor variables between the studies – first the variable (or a similar variable) had to exist and second it had to exist in sufficient quantity within both datasets. In some instances, this resulted in the exclusion of otherwise potentially viable variables, most notably HPV-16 serostatus and frequency of dental attendance. However, the comparative heterogeneity of the studies also served as a strength – it ensured the total number of variables was kept minimal and truly served to test the generalisability of the model.

6.6 Conclusions

We have developed and externally validated a HNC risk model using the ARCAGE and UK Biobank studies respectively. This model had good performance in the development study and had a fair level of performance in the UK Biobank validation dataset. Ultimately, demographics and behaviours are strong predictors of HNC, however, these factors alone cannot reliably predict individual risk with a high degree of accuracy. Future incorporation of further biomarkers such as HPV-16 serostatus or high-risk genetic variants, could enhance the model prediction. The developed model still has potential to be feasibility tested and adapted for use as a clinical decision support tool in the

primary care settings (including dental practices) – informing patient recall intervals and prompting preventive interventions.

6.7 Additional Information and Author Contributions

6.7.1 Funding

Funding for the ARCAGE study was received from European Community (5th Framework Programme) grant no. QLK1-CT-2001-00182. The Athens and Turin centre received funding from the University of Athens Medical School) and Compagnia San Paolo, AIRC, respectively. M. Nelis and A. Metspalu were also supported by the Estonian Ministry of Education and Research grant 0180142s08. This research has been conducted using the UK Biobank Resource under Application Number 79789.

The UK Biobank study was established and funded by the Welcome Trust medical charity, Medical Research Council (MRC), Department of Health, Scottish Government and the Northwest Regional Development Agency. It also received funding from the Welsh Government, British Heart Foundation, Cancer Research UK and Diabetes UK. The study is supported by the National Health Service (NHS). Currently, core funding is received from the Wellcome Trust, the MRC, Cancer Research UK and NIHR.

This study and CS were funded and supported by Cancer Research UK, as part of the to Train and retain Academic Cancer Clinicians (TRACC) programme, CRUK Beatson Institute (now CRUK Scotland Institute), Glasgow [Grant number 315941-01]
Author DC is funded by (HEADsPACE) European Union's Horizon 2020 research and innovation programme under grant agreement No 825771.
DS and JP were partly supported by Italian Ministry of Health “Ricerca Corrente” [no grant number].

6.7.2 Authors Contribution

Study concepts: CS, AM, AR, GI, DC
Study design: CS, AM, AR, MG, TD, DC, ARCAGE Study Group
Data acquisition: DC, DL, ARCAEG Study Group
Quality control of data and algorithms: CS, AM, MaG, DC
Data analysis and interpretation: CS, AM, MaG, DC
Statistical analysis: CS, AM, MaG
Manuscript preparation: CS
Manuscript editing: CS, AM, DC
Manuscript review: All Authors

7 Chapter Seven Human Papillomavirus (HPV)Serostatus and Burden in the UK Biobank Study

Chapter Six detailed the process for the development and validation of a head and neck cancer risk prediction model. However, a lack of Human Papillomavirus (HPV) data in the UK Biobank cohort meant the validation of an HPV model was not viable. This additional thesis analysis sought to explore and describe the available HPV data in the UK Biobank, to assess the feasibility of using other variables to impute HPV serostatus across the cohort.

7.1 Background

7.1.1 Definitions

Human Papillomavirus (HPV), is a double-stranded DNA virus of the Papovaviridae family, and a common Sexually Transmitted Infection (STI). (Centres for Disease Control and Prevention (CDC), 2021) As an STI, HPV is primarily transmitted through sexual behaviours including vaginal, penile, anal and oral sex. Transmission can also occur via skin-to-skin contact. Risk factors for infection with HPV include an earlier sexual debut, an increased number of sexual partners, smoking behaviours and long-term oral-contraceptive usage. (Vinodhini et al., 2012) (Chelimo et al., 2013) Men who have sex with men (MSM) are also at an increased risk of anal HPV transmission and subsequent anogenital cancer. (Goldstone et al., 2011)

HPV infection typically occurs during adolescence or early adulthood, shortly after an individual's sexual debut. Most individuals' immune systems will naturally clear an HPV infection within a one to two-year period. (Rodríguez et al., 2008) In a minority of patients, HPV infection will be persistent. Infection with HPV is often asymptomatic, although infection with some strains (namely 6 and 11) can cause genital warts. (Greer et al., 1995)

There are over 200 different strains of HPV that have been identified by genomic sequencing. (Centres for Disease Control and Prevention (CDC), 2021) The majority of these are harmless. However, some strains have been identified to be "high-risk", possessing carcinogenic potential. The following fourteen HPV strains are classified as high-risk: HPV strains 16, 18, 31, 33, 35, 39, 45, 51, 52,

56, 58, 59, 66, and 68. Where individuals have a persistent infection with a high-risk oncogenic strain, there is an increased risk of developing a pre-malignant or malignant condition. (Okunade, 2020) (Cuschieri et al., 2005)

High-risk HPV strains have been shown to play a key causative role in the development of cervical, anogenital and oropharynx cancers, and as such are classified by the WHO as grade I carcinogens. (World Health Organisation (WHO), 2023) A large global cross-sectional study showed high-risk HPV strains were attributable to some 71% of invasive cervical cancers and 94% of cervical adenocarcinomas. (de Sanjose et al., 2010)

7.1.2 OPC and HPV-mediated OPC Epidemiology

Oropharynx cancer is defined as squamous cell carcinomas (SCC) of the soft palate, base of tongue, tonsils and uvula. Classically, OPC has been linked to smoking and alcohol behaviours. However, in more recent years, persistent infection with HPV (strains 16 and 18) has been confirmed by the WHO as a risk factor for OPC. Briefly, persistent HPV infection in the stratified squamous epithelium typical of surfaces of the oropharynx allows for conditions whereby abnormal cells can proliferate and differentiate. Subsequent oncogenic changes result in an eventual progression to malignancy. (Lechner et al., 2022)

As discussed in previous chapters, the epidemiology of HNC is changing. There has been a global shift in HNC disease patterns; the incidence of HPV-mediated oropharyngeal cancer has dramatically increased, particularly in high-income countries. (Gormley et al., 2022a) Globally, HPV is estimated to be attributable to 33% of OPC cases, with wide geographic variations in prevalence; the highest rates of HPV-positive OPC in the USA and Europe. (Lechner et al., 2022)

In the UK, an estimated 52% of OPC cases are HPV-positive and these numbers are expected to rise. (Schache et al., 2016) As with many Western nations, OPC incidence in the UK and Scotland has risen dramatically in recent years; OPC incidence in Scotland increased by 85% from 2001-2012. (Purkayastha et al., 2016)

The rise of HPV-mediated OPC has led to HPV-negative OPC existing alongside growing numbers of HPV-positive OPC. HPV-positive and HPV-negative OPC also differ in terms of disease progression and prognosis. HPV-positive OPC is more responsive to treatment modalities and is subsequently associated with significantly improved survival. (Ang et al., 2010, Fakhry et al., 2014). In 2018, the American Joint Committee on Cancer (AAJC) released a separate staging system for HPV-positive tumours of the oropharynx. (Lydiatt et al., 2018)

7.1.3 Prevention

Preventive methods against HPV infection and associated cancers include safe-sex practices such as condom use, regular cervical screening and immunisation against HPV. (Chelimo et al., 2013) HPV-vaccination is a key aspect of the WHO's strategy for the elimination of cervical and other HPV-related cancers. (World Health Organization, 2022)

The Gardasil®9 vaccine is now the main HPV vaccine used in UK national immunisation programmes. (Department of Health and Social Care, 2023) (Public Health Scotland, 2023c) The vaccine protects against strains 16 and 18 (which previous bivalent vaccines targeted), in addition to strains 11, 31, 33, 45, 53 and 58. (Cheng et al., 2020) In Scotland, school-age girls (12-13) have been offered the HPV vaccine since 2008. (White, 2008) In recent years, the HPV immunisation programme has been expanded to school age boys (12-13) since 2019. The vaccine is also offered to MSM up to the age of 45. (Public Health Scotland, 2023c)

HPV vaccination has already been shown to be effective in the reduction of HPV prevalence and subsequent cervical and cervical carcinoma in situ rates among women in the UK and Scotland. (Palmer et al., 2019, Falcaro et al., 2021) Although there is a larger latent period between HPV infection and subsequent oncogenesis associated with oropharyngeal cancer (OPC) than cervical and other anogenital cancers, the hope is HPV vaccination will also be effective in reducing OPC rates in the future.

7.1.4 HPV Serology in the UK Biobank Study

The UK biobank cohort study collected blood, urine, and saliva samples from study participants with the objective of assessing infectious agents and biomarkers within the population. Complete details of these are previously described in the UK Biobank study protocol. (Rory, 2007)

One of the study objectives was to estimate the seroprevalence of various infectious diseases in the study population using various antibody response levels and biomarkers. A working committee selected 20 key pathogens for an infectious disease panel. These were selected for their known role as risk factors for cancers, cardiovascular or neurodegenerative diseases or as agents of interest. Among these 20 pathogens, HPV-16 and HPV-18 were chosen to be included in the disease panel.

A sample of UK Biobank participants (n = 9,724) was selected at random as part of a pilot study assessing the seroprevalence of these pathogens. The samples were assayed in July 2016 at the Deutsches Krebsforschungszentrum (DKFZ) in Heidelberg, Germany, using Multiplex serology. The methodology and validation of this was described in detail by Waterboer and colleagues. (Waterboer et al., 2005a) (Waterboer et al., 2006) The majority of samples (99.7%, n = 9,625) passed subsequent quality and validity checks. The outputs of the Multiplex serology were quantified using median fluorescence intensity values for each disease-specific antigen. Cut-offs for seropositivity were defined using percentile plots and prior validation work.

HPV-16 antigens L1, E6 and E7 were included in the infectious disease assay panel. HPV-16 Seropositivity was defined as either (i) an MFI > 175 for the L1 antigen or (ii) an MFI > 120 for the E6 antigen and / or MFI > 150 for the E7 antigen. The seroprevalence of HPV-16 within the pilot study (n = 9,695) was calculated to be 4.4%. These were consistent with other studies' estimates of seroprevalence within UK populations. (Tanton et al., 2017). (Jit et al., 2007)

Further details of the sample collection, methodology and cut-offs are described in detail by Mentzer and colleagues and the UK Biobank infectious disease

serology pilot study information document. (Mentzer et al., 2022) (UK Biobank, 2019)

7.1.5 Existing HPV Research in the UK Biobank

Currently, there is little existing literature exploring HPV serology in the context of HNC research within the UK Biobank. At the time of writing, the only publication which specifically investigates HPV serostatus within the UK Biobank is the work by Brenner and colleagues. (Brenner et al., 2020)

Following the exclusion of HPV-associated malignancies, Brenner et al. described the HPV serology data included in the pilot study sample within the UK Biobank. Brenner et al. assessed various risk factors for the prediction of HPV-seropositivity for several antigens using logistic regression analysis. Behaviours were largely shown to be non-significant predictors, whilst sexual factors (age of sexual debut, number of sexual partners and same-sex intercourse) were associated with HPV-antigen seropositivity.

7.2 Aims

A limited quantity of HPV data is available for analysis within the UK Biobank. Building upon the work of Brenner and colleagues, this analysis aimed to map out existing HPV biomarker data within the UK Biobank in the context of HNC, then test the strength of any associations with sexual data or other demographic data and HPV serostatus.

The first aim of this analysis was to assess the strength of associations between sexual data and other relevant fields with HPV serostatus. If associations were strong, the next aim would be to consider the imputation of HPV serostatus across the dataset using data collected from these predictors, to validate an existing ARCAGE HPV model. Whilst multiple imputation is typically reserved for use in the context of limited numbers of missing data, if sexual (or other) factors were shown to be very strongly associated with HPV serostatus, there may be grounds to attempt multiple imputation to assess an estimated seroprevalence and compare with existing estimates. Notably, the samples were selected at

random, meaning that the mechanism of missing data for HPV data was non-systematic.

7.3 Methodology

Following the external validation of the thesis risk model in Chapter Six, a small number of participants (n= 43) withdrew their consent to partake in the study. The records for these individuals were removed (using corresponding anonymised ID numbers) from the dataset prior to this analysis. HNC cases and non-events were identified using the same methodology to describe cases and exclude other cancers as described in Chapter Six.

The sociodemographic data captured in the study included age, sex, ethnicity ("White, Asian, Black, Chinese, Mixed, or Other / Unknown") and socioeconomic status. Both area-based (Townsend Index) and individual-level socioeconomic predictors were collected. This included educational attainment (College or University degree, A levels/AS levels, O levels/GCSEs, CSEs or equivalent, NVQ or HND or HNC, Other professional Qualification, Prefer not to Say, or None of the above) and income (Missing / Prefer Not to Say, Less than £18,000, £18,000 to £30,999, £31,000 to £51,999, £52,000 to £100,000, or > £100000).

Smoking alcohol and sexual behaviour data were also collected. Smoking data included smoking status (Never, Former, Current, or Missing / Prefer Not to Say) while individual smoking frequency and duration data were captured using smoking packyears, where an individual's typical daily number of packs of cigarettes smoked is multiplied by the participant's years of smoking. This was treated as a continuous variable. Individual alcohol status (Never, Former, Current, Missing / Prefer Not to say) and frequency of alcohol consumption (Prefer not to say, Daily or almost daily, 3/4 x a week, 1/2 x a week, 1-3 x a month, Special occasions only, Never) variables were included. Sexual behaviour data included the practice of same-sex intercourse (Never, Ever, Prefer Not to Say), age of sexual debut (Never had sex, 17 or younger, 18-19, 20 or older, Missing / Prefer Not to Say) and number of sexual partners (Never had sex, 1, 2-3, 4-5, 6-10, 11 or more, Missing / Prefer Not to Say).

Descriptive statistics of the existing HPV data, sexual behaviours, sociodemographics and behaviours were calculated and reported. Continuous variables (age, area-based deprivation (Townsend index), smoking packyears), were described using two-sample t-tests, and Welch's t-test where equal variance could not be assumed and categorical variables were described using Chi-squared or Fisher's exact test, where appropriate. Univariable and multivariable logistic regression models were used to assess the associations between sexual, sociodemographic and behavioural factors with HPV serostatus. Multivariable logistic regression models were adjusted for age and sex. Odds ratios (OR) with corresponding 95% confidence intervals (CI) were reported for each variable.

7.4 Descriptive Results – UK Biobank

7.4.1 HPV-16 Status Definition I. Results

Descriptive statistics and tests of association for participants with HPV serology Definition I (MFI > 175 for the L1 antigen) are detailed in Table 7-1. Following the application of prior eligibility criteria described in Chapter 6, 7238 participants of the 384,616 included in the study had HPV-serology data. According to definition one of HPV-seropositivity, there were 6913 participants with HPV-negative serology results and 325 participants with positive HPV-16 MFI tests.

Table 7-1: HPV-16 Definition (i) Descriptive statistics and associations

Variable	HPV-16 Status Def. 1			Odds Ratio	95% CI		Adjusted Odds Ratio	95% CI	
	Negative	Positive	p		Upper	Lower		Upper	Lower
Status			**0.01**						
Non-events (%)	6896 (99.8%)	321 (98.8%)		REF	REF	REF	REF	REF	REF
HNC (%)	17 (0.3%)	4 (1.2%)		**5.06**	**1.69**	**15.11**	**6.46**	**2.11**	**19.79**
Age, (mean ± SD)	55.7 (±8.2)	52.8 (±7.5)	**<.0001**	**0.96**	**0.94**	**0.97**	**0.96**	**0.94**	**0.97**
Sex			**<.0001**						
Female	3851 (55.7%)	233 (71.7%)		REF	REF	REF	REF	REF	REF
Male	3062 (44.3%)	92 (28.3%)		**0.50**	**0.39**	**0.64**	**0.49**	**0.38**	**0.63**
Ethnicity			0.09						
White	6457 (93.4%)	299 (92.0%)		REF	REF	REF	REF	REF	REF
Asian	172 (2.5%)	4 (1.2%)		0.50	0.19	1.36	0.48	0.17	1.29
Black	111 (1.6%)	10 (3.1%)		**1.95**	**1.01**	**3.75**	1.53	0.79	2.98
Chinese	30 (0.4%)	1 (0.3%)		0.72	0.10	5.30	0.53	0.07	3.91
Mixed	1 (0.0%)	0		<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Other / Unknown	142 (2.1%)	11 (3.4%)		1.673	0.896	3.123	1.49	0.80	2.81
Townsend, (mean ± SD)	-1.32 (±3.08)	-0.88 (±3.19)	**0.01**	**1.05**	**1.01**	**1.08**	**1.04**	**1.004**	**1.08**
Educational Attainment			0.22						
College or University degree	2242 (32.5%)	115 (35.5%)		1.25	0.88	1.77	0.93	0.65	1.34
A levels/AS levels	773 (11.2%)	28 (8.6%)		0.88	0.55	1.42	0.64	0.39	1.04
O levels/GCSEs	1498 (21.7%)	72 (22.2%)		1.17	0.80	1.70	0.88	0.60	1.30
CSEs or equivalent	407 (5.9%)	25 (7.7%)		1.50	0.91	2.46	0.98	0.58	1.64
NVQ or HND or HNC	445 (6.4%)	13 (4.0%)		0.71	0.38	1.33	0.62	0.33	1.17
Other professional Qualification	329 (4.8%)	20 (6.2%)		1.48	0.86	2.53	1.33	0.77	2.28
Prefer not to Say	71 (1.0%)	4 (1.2%)		1.37	0.48	3.91	1.20	0.42	3.47

None of the above	1144 (16.6%)	47 (14.5%)		REF	REF	REF	REF	REF	REF
Income			0.22						
Missing / Prefer Not to Say	993 (14.4%)	30 (9.4%)		**0.60**	**0.39**	**0.93**	**0.56**	**0.36**	**0.87**
Less than £18,000	1286 (18.7%)	65 (20.3%)		REF	REF	REF	REF	REF	REF
£18,000 to £30,999	1452 (21.1%)	69 (21.6%)		0.94	0.67	1.33	0.89	0.63	1.26
£31,000 to £51,999	1590 (23.1%)	77 (24.1%)		0.96	0.68	1.34	0.80	0.56	1.13
£52,000 to £100,000	1246 (18.1%)	60 (18.8%)		0.95	0.67	1.37	0.73	0.51	1.07
> £100000	330 (4.8%)	19 (5.9%)		1.14	0.67	1.93	0.94	0.55	1.60
Smoking Status			0.84						
Never	3950 (57.1%)	179 (55.1%)		REF	REF	REF	REF	REF	REF
Former	2244 (32.5%)	111 (34.2%)		1.09	0.86	1.39	**1.29**	**1.00**	**1.64**
Current	684 (9.9%)	34 (10.5%)		1.10	0.75	1.60	1.17	0.80	1.71
Missing / Prefer Not to Say	35 (0.5%)	1 (0.3%)		0.63	0.09	4.63	0.74	0.10	5.48
Smoking Pack Years, (mean ± SD)	6.46 (±14.48)	6.38 (±13.26)	0.91	1.00	0.99	1.01	1.01	1.00	1.01
Alcohol Status			0.64						
Never	323 (4.7%)	18 (5.4%)		REF	REF	REF	REF	REF	REF
Former	244 (3.5%)	10 (3.0%)		2.04	0.99	4.21	**2.20**	**1.06**	**4.56**
Current	6327 (91.6%)	302 (91.2%)		1.16	0.66	2.05	1.27	0.72	2.25
Missing / Prefer Not to say	13 (0.2%)	1 (0.3%)		1.94	0.24	15.98	2.32	0.27	19.57
Alcohol Frequency			0.14						
Prefer not to say	8 (0.1%)	0		<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Daily or almost daily	1375 (19.9%)	66 (20.4%)		0.85	0.55	1.31	1.05	0.67	1.62
3/4 x a week	1579 (22.9%)	70 (21.6%)		0.78	0.51	1.20	0.85	0.55	1.32
1/2 x a week	1762 25.5%)	92 (28.4%)		0.92	0.61	1.39	0.98	0.64	1.48
1-3 x a month	785 (11.4%)	21 (6.5%)		**0.47**	**0.27**	**0.83**	0.44	0.25	0.77
Special occasions only	836 (12.1%)	43 (13.3%)		0.91	0.57	1.45	0.84	0.52	1.34

Never	564 (8.2%)	32 (9.9%)		REF	REF	REF	REF	REF	REF
Same Sex Intercourse			**0.03**						
Never	6007 (96.3%)	291 (93.9%)		REF	REF	REF	REF	REF	REF
Ever	206 (3.3%)	19 (6.1%)		**1.90**	**1.17**	**3.09**	**1.80**	**1.09**	**2.95**
Prefer Not to Say	27 (0.4%)	0		<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Age Sexual Debut			**0.0003**						
Never had sex	74 (1.2%)	1 (0.3%)		0.29	0.04	2.10	0.32	0.04	2.31
17 or younger	2296 (36.4%)	150 (48.2%)		**1.39**	**1.05**	**1.85**	**1.38**	**1.04**	**1.85**
18-19	1576 (25.0%)	74 (23.8%)		REF	REF	REF	REF	REF	REF
20 or older	2162 (34.2%)	77 (24.8%)		0.76	0.55	1.05	0.88	0.63	1.22
Missing / Prefer Not to Say	206 (3.3%)	9 (2.9%)		0.93	0.46	1.89	1.09	0.53	2.22
Number of Sexual Partners			**<.0001**						
Never had sex	74 (1.2%)	1 (0.3%)		0.73	0.10	5.44	0.81	0.11	6.06
1	1681 (26.6%)	31 (10.0%)		REF	REF	REF	REF	REF	REF
2-3	1367 (21.7%)	38 (12.2%)		1.51	0.93	2.44	1.49	0.92	2.41
4-5	911 (14.4%)	64 (20.6%)		**3.81**	**2.46**	**5.89**	**3.78**	**2.43**	**5.88**
6-10	982 (15.6%)	72 (23.2%)		**3.98**	**2.59**	**6.10**	**4.11**	**2.65**	**6.38**
11 or more	731 (11.6%)	62 (19.9%)		**4.11**	**2.56**	**6.58**	**5.45**	**3.44**	**8.62**
Missing / Prefer Not to Say	568 (9.0%)	43 (13.8%)		**4.60**	**2.96**	**7.14**	**4.58**	**2.84**	**7.39**

HPV = Human PapillomaVirus, Definition (i): MFI > 175 for the L1 antigen, HNC = Head and Neck Cancer, SD = Standard Deviation, REF = Reference Value, Adjusted = Adjusted for Age and Sex

Of the participants with HPV-16 definition I (L1) serology data, 21 HNC cases were identified. Four of these patients tested positive for HPV-16. Participants with HPV-seropositivity were, on average, younger with a mean age of 52.8 (±7.5) versus controls with a mean age of 55.7 (±8.2) (p <.0001). The majority of HPV-positive participants were female, constituting 71.7% (n =233) of positive tests. Relative to females, males were less likely to be HPV seropositive (adjusted OR = 0.49, 95% CI = 0.38 -0.63). There were no significant associations between ethnicity and HPV status (p = 0.09).

Some associations were observed between individual SES and HPV-seropositivity. HPV-negative patients had, on average, a lower Townsend Score of Deprivation with a mean value of -1.32 (±3.08) versus HPV-positive participants, suggesting a higher socioeconomic status. (-0.88 (±3.19)) (p = 0.01). This was also reflected in logistic regression analysis, whereby an increase in Townsend score was (narrowly) associated with an increased risk of HPV-seropositivity (multivariable OR =1.04, 95% CI =1.004 -1.08). No associations were observed for individual educational level. Income was largely not associated with HPV-serostatus; the only exception was participants who had not reported their income, which was associated with a protective effect relative to those who earned less than £18,000 (OR = 0.56, 95% CI = 0.36 - 0.87).

Limited associations were found between alcohol and smoking behaviours and HPV-serostatus. At a multivariable level, participants who reported a former smoking or alcohol use status were observed to have an increased risk relative to never smokers (adjusted OR = 1.29. 95% CI = 1.00 -1.64) and never drinkers (adjusted OR = 2.20, 95% CI = 1.06 - 4.56). These could perhaps be explained by past high-risk behaviours and a quitting effect associated with malignancy or declining health. No significant associations were observed for smoking pack-years or alcohol frequency at a multivariable level.

Several associations were observed between sexual behaviours and HPV serostatus. A younger sexual debut (age 17 or younger) was associated with an increased risk of HPV seropositivity at both a univariable and multivariable level relative to those aged 18-19 (OR = 1.38, 95% CI = 1.04 -1.85). Same-sex intercourse was also

associated with an increased risk ever vs. never - adjusted OR = 1.80, 95% CI = 1.09 -2.95). Finally, an increased number of sexual partners was also significantly associated with an increased risk of HPV-seropositivity. HPV serostatus was strongly associated with an individual's number of sexual partners (p <.0001). Relative to those with one sexual partner, participants who reported intercourse with 11 or more people were over five times more likely to be HPV-seropositive (adjusted OR = 5.45, 95% CI = 3.44 - 8.62).

7.4.2 HPV-16 Status Definition II Results

Results from the descriptive analysis of HPV-16 Definition II (MFI > 120 for the E6 antigen and / or MFI > 150 for the E7 antigen) serology data are described in Table 7-2. Of the 7238 participants with HPV-16 Definition II serology, 6907 participants tested HPV-seronegative, while 331 participants had a positive serology test. A small number of HNC cases were identified (n = 21) and only seven were HPV-positive. No significant difference was found between the mean age of seronegative and seropositive patients (55.5 years (±8.2) and 56.4 years (±8.2) respectively, p = 0.06). Significant associations were observed for participant sex (p = 0.004). In contrast with L1 seropositivity definitions, male sex was associated with an increased risk of definition (ii) seropositivity (adjusted OR = 1.39, 95% CI = 1.12 - 1.73). No associations between HPV status and participant ethnicity were found (p = 0.84).

No significant differences nor associations were observed between the participants' socioeconomic status and HPV serostatus. HPV-negative and HPV-positive participants had, on average, similar Townsend scores. HPV-positive participants reported a slightly higher (more affluent) Townsend score (-1.31 (± 3.09) and -1.18 (± 3.08) respectively, p =0.46). Participants reported similar levels of educational level (p = 0.98); the largest proportion of seronegative and seropositive participants reported a college or university-level education. These findings were also observed for participant income, where no significant differences were shown (p = 0.65).

Table 7-2: HPV-16 Definition (ii) Descriptive statistics and associations

Variable	HPV-16 Status Def. 2			Odds Ratio	95% CI		Adjusted Odds Ratio	95% CI	
	Negative	Positive	p		Upper	Lower		Upper	Lower
Status									
Non-events (%)	6893 (99.8%)	324 (97.9%)	**<.0001**	REF	REF	REF	REF	REF	REF
HNC (%)	14 (0.2%)	7 (2.1%)		**10.64**	**4.27**	**26.54**	**9.87**	**3.94**	**24.73**
Age, (mean ± SD)	55.51 (±8.18)	56.39 (±8.22)	0.06	1.01	1.00	1.03	1.01	1.00	1.03
Sex			**0.004**						
Female	3923 (56.8%)	161 (48.6%)		REF	REF	REF	REF	REF	REF
Male	2984 (43.2%)	170 (51.4%)		**1.39**	**1.11**	**1.73**	**1.39**	**1.12**	**1.73**
Ethnicity			0.84						
White	6445 (93.3%)	311 (93.7%)		REF	REF	REF	REF	REF	REF
Asian	170 (2.5%)	6 (1.8%)		0.73	0.32	1.66	0.73	0.32	1.67
Black	117 (1.7%)	4 (1.2%)		0.71	0.26	1.93	0.77	0.28	2.09
Chinese	30 (0.4%)	1 (0.3%)		0.69	0.09	5.08	0.78	0.11	5.75
Mixed	1 (0.0%)	0		<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Other / Unknown	144 (2.1%)	9 (2.7%)		1.30	0.65	2.56	1.34	0.67	2.65
Townsend, (mean ± SD)	-1.31 (± 3.09)	-1.18 (± 3.08)	0.46	1.01	0.98	1.05	1.02	0.98	1.05
Educational Attainment			0.98						
College or University degree	2251 (32.6%)	106 (32.1%)		0.97	0.70	1.36	1.04	0.74	1.47
A levels/AS levels	769 (11.1%)	32 (9.7%)		0.86	0.55	1.34	0.94	0.60	1.48
O levels/GCSEs	1493 (21.6%)	77 (23.3%)		1.07	0.75	1.52	1.16	0.81	1.66
CSEs or equivalent	411 (6.0%)	21 (6.4%)		1.06	0.63	1.77	1.20	0.70	2.03
NVQ or HND or HNC	439 (6.4%)	19 (5.8%)		0.89	0.53	1.52	0.90	0.53	1.54
Other professional Qualification	332 (4.8%)	17 (5.2%)		1.06	0.61	1.85	1.09	0.62	1.91
Prefer not to Say	72 (1.0%)	3 (0.9%)		0.86	0.26	2.82	0.87	0.27	2.85

None of the above	1136 (16.5%)	55 (16.7%)		REF	REF	REF	REF	REF	REF
Income			0.65						
Missing / Prefer Not to Say	979 (14.2%)	44 (13.4%)	.	1.04	0.69	1.56	1.07	0.72	1.61
Less than £18,000	1295 (18.8%)	56 (17.0%)	.	REF	REF	REF	REF	REF	REF
£18,000 to £30,999	1452 (21.1%)	69 (21.0%)	.	1.10	0.77	1.58	1.11	0.78	1.60
£31,000 to £51,999	1595 (23.2%)	72 (21.9%)	.	1.04	0.73	1.49	1.11	0.77	1.60
£52,000 to £100,000	1238 (18.0%)	68 (20.7%)	.	1.27	0.88	1.83	1.40	0.96	2.05
> £100000	329 (4.8%)	20 (6.1%)	.	1.41	0.83	2.38	1.51	0.88	2.58
Smoking Status			0.30						
Never	3956 (57.3%)	173 (52.3%)	.	REF	REF	REF	REF	REF	REF
Former	2234 (32.3%)	121 (36.6%)	.	1.24	0.98	1.57	1.17	0.92	1.49
Current	682 (9.9%)	36 (10.9%)	.	1.21	0.84	1.74	1.16	0.80	1.68
Missing / Prefer Not to Say	35 (0.5%)	1 (0.3%)	.	0.65	0.09	4.80	0.61	0.08	4.50
Pack Years, (mean ± SD)	6.38 (±14.34)	8.12 (±16.11)	0.05	**1.01**	**1.001**	**1.01**	1.01	1.00	1.01
Alcohol Status			0.84						
Never	323 (4.7%)	18 (5.4%)	.	REF	REF	REF	REF	REF	REF
Former	244 (3.5%)	10 (3.0%)	.	0.74	0.33	1.62	0.71	0.32	1.56
Current	6327 (91.6%)	302 (91.2%)	.	0.86	0.53	1.40	0.81	0.50	1.33
Missing / Prefer Not to say	13 (0.2%)	1 (0.3%)	.	1.38	0.17	11.14	1.24	0.15	10.05
Alcohol Frequency			0.14						
Prefer not to say	8 (0.1%)	0	.	<0.001	<0.001	>999.999	<0.001	<0.001	>999.999
Daily or almost daily	1363 (19.8%)	78 (23.6%)	.	1.16	0.75	1.81	1.08	0.69	1.68
3/4 x a week	1566 (22.7%)	83 (25.2%)	.	1.08	0.69	1.67	1.03	0.67	1.61
1/2 x a week	1771 (25.7%)	83 (25.2%)	.	0.95	0.61	1.47	0.93	0.60	1.44
1-3 x a month	774 (11.2%)	32 (9.7%)	.	0.84	0.50	1.41	0.85	0.51	1.43
Special occasions only	853 (12.4%)	26 (7.9%)	.	0.62	0.36	1.07	0.64	0.37	1.10

Never	568 (8.2%)	28 (8.5%)	.	REF	REF	REF	REF	REF	REF
Same Sex Intercourse			**0.0001**						
Never	6025 (96.4%)	273 (91.6%)	**.**	REF	REF	REF	REF	REF	REF
Ever	202 (3.2%)	23 (7.7%)	**.**	**2.51**	**1.61**	**3.93**	**2.51**	**1.59**	**3.96**
Prefer Not to Say	25 (0.4%)	2 (0.7%)	**.**	1.77	0.42	7.49	1.67	0.39	7.12
Age Sexual Debut			0.63						
Never had sex	72 (1.1%)	3 (1.0%)	**.**	0.96	0.29	3.11	0.91	0.28	2.97
17 or younger	2322 (36.7%)	124 (41.2%)	**.**	1.22	0.91	1.65	1.21	0.90	1.64
18-19	1581 (25.0%)	69 (22.9%)	**.**	REF	REF	REF	REF	REF	REF
20 or older	2144 (33.9%)	95 (31.6%)	**.**	1.02	0.74	1.39	0.97	0.70	1.33
Missing / Prefer Not to Say	205 (3.2%)	10 (3.3%)	**.**	1.12	0.57	2.20	1.05	0.53	2.08
Number of Sexual Partners			0.24						
Never had sex	72 (1.1%)	3 (1.0%)	.	0.98	0.30	3.20	0.99	0.31	3.23
1	1643 (26.0%)	69 (22.9%)	.	REF	REF	REF	REF	REF	REF
2-3	1349 (21.3%)	56 (18.6%)	.	1.00	0.70	1.44	0.99	0.69	1.42
4-5	936 (14.8%)	39 (13.0%)	.	1.01	0.67	1.51	0.99	0.67	1.48
6-10	993 (15.7%)	61 (20.3%)	.	**1.48**	**1.03**	**2.13**	**1.46**	**1.03**	**2.08**
11 or more	752 (11.9%)	41 (13.6%)	.	1.28	0.84	1.93	1.30	0.39	4.32
Missing / Prefer Not to Say	579 (9.2%)	32 (10.6%)	.	1.29	0.84	1.99	1.32	0.86	2.02

HPV = Human PapillomaVirus, Definition (ii): MFI > 120 for the E6 antigen and / or MFI > 150 for the E7 antigen, HNC = Head and Neck Cancer, SD = Standard Deviation, REF = Reference Value,

Adjusted = Adjusted for Age and Sex

There were largely no associations between HPV serostatus and smoking or alcohol behaviours. No differences in HPV serostatus by smoking and alcohol status were observed (p = 0.30 and p = 0.84 respectively). Similarly, no associations between smoking pack years and HPV status were observed (adjusted OR = 1.01, 95% CI = 1.00 - 1.01). Alcohol behaviours were also not shown to be associated with HPV serostatus.

Sexual behaviours showed weaker associations with HPV definition II serostatus. Same-sex intercourse remained significantly associated with an increased risk of seropositivity (adjusted OR = 2.51, 95% CI = 1.59 - 3.96), which was consistent with findings from the definition I serology analysis. However, no associations were found with a younger sexual debut (p = 0.63) and limited associations were found with an individual's number of sexual partners (p = 0.24); relative to those with a single sexual partner, only those who reported 6-10 sexual partners were found to be at a statistically significant increased risk (adjusted OR = 1.46, 95% CI = 1.03 - 2.08).

7.5 UK Biobank Discussion

This analysis sought to describe HPV serology and HNC data in the UK Biobank and to potentially impute HPV data to validate a clinical risk model if the findings were supportive of it.

7.5.1 Associations

Associations between participant demographics and HPV serostatus were limited. A younger age was associated with an increased risk of HPV definition (i) seropositivity but not definition (ii). Participant sex associations varied; male sex was associated with a reduced risk of HPV definition (i) seropositivity, however, the opposite finding was observed for definition (ii) seropositivity. Socioeconomically, HPV-positive and negative participants were comparatively similar, with few to no associations observed. Similarly, there was no evidence to suggest any associations between ethnicity and HPV serostatus. This does not chime with existing research, which has reported higher HPV seroprevalence

among those of black and minority ethnic backgrounds. (Lin et al., 2015) (Berenson et al., 2021) However, it is worth acknowledging that the UK Biobank participant population (and subsequent HPV serology sample) was largely white Caucasian, with over 93% of those with HPV serology data of a white Caucasian ethnicity.

Weak or non-existent associations were observed between alcohol, smoking behaviours, and subsequent HPV-16 serostatus. This remained true for status and measures of frequency or duration. This finding also contrasts with existing literature, which suggested an association of smoking and high-frequency alcohol behaviours with an increased risk of HPV infection. (Schabath et al., 2012) (Schabath et al., 2015) (Chung et al., 2015)

Besides participant sex and a diagnosis of HNC, sexual behaviours were the only variables which exhibited significant associations with HPV serostatus. However, even within these, some variation was exhibited; only same-sex intercourse remained significant across both seropositive definitions. Whilst age of sexual debut and number of partners were significantly associated with definition (i) HPV seropositivity, this was not the case for definition (ii).

The variation of associations observed for participant sex and sexual behaviours could potentially be explained by the differing roles of the threshold proteins used to classify seropositivity. The L1 protein is associated with the outer capsule of the HPV virus (also known as the viral capsid). (Buck et al., 2013) Raised antigen levels associated with this protein are suggestive of exposure to the outer shell of the virus – which would be more readily associated with sexual transmission and behaviours.

HPV-16 E6 and E7 are oncoproteins responsible for the regulation of cells and viral replication that are typically associated with higher-risk HPV strains. Thus, E6 and E7 proteins are associated with the virus's ability to persist in the cell lining and, with time, subsequent oncogenesis. (Narisawa-Saito and Kiyono, 2007) Raised antigen levels for L1 protein may be indicative of more recent viral exposure that is associated with sexual activity, whereas raised E6 / E7 antigen levels could be indicative of a persistent infection, rather than more immediate

sexual activity and transmission. This is consistent with the weaker associations with sexual behaviours that were observed for definition (ii) positivity. Furthermore, this hypothesis is also reflected in the stronger association between HNC and E6 or E7 seropositivity versus L1 seropositivity. Similarly, it has been hypothesised that, in contrast with females, males may be at an increased risk of prolonged HPV infection (and subsequent oncogenesis) due to a variety of mechanisms including a lack of protection conferred by prior cervical HPV infection.

7.5.2 Multiple Imputation

Multiple imputation is a technique that can be used to handle missing data. It works by replacing missing data using plausible values derived from known data or associations. (Li et al., 2015, Rubin, 2018) Multiple imputation is typically performed in two steps. The first of these entails computing replacement values for the missing data (i.e. imputation) and the generation of multiple subsequent datasets. The second step entails the synthesis and analysis of these generated datasets. Multiple datasets are created with estimated plausible values for the variable with missing data. This allows for the quantification of uncertainty when describing plausible values for an imputed variable, as opposed to a single imputation, which can lead to incorrect levels of precision of estimates.

There are no explicit guidelines for the use of multiple imputation. Some authors have suggested that estimates from multiple imputation are at an increased risk of bias where over 10% of data is missing. (Bennett, 2001) Whilst multiple imputation is listed as a strategy for the handling of missing data in the TRIPOD model development guidelines, there is no explicit mention (at the time of writing) of the use of multiple imputation in the validation guidelines, nor a defined acceptable quantity of missing data. (Collins et al., 2015b)

Even though the HPV data was missing completely at random (MCAR) due to the random selection of participants for the disease panel, the large volume of missing data far exceeded the maximum quantity recommended by some for multiple imputation. As can be seen from the descriptive tables, analysis of HPV serostatus and burden in the UK Biobank revealed some associations between

participant behaviours, sexual history and HPV serostatus. However, the large volume of missing HPV data (>98%) and the lack of (consistently) strong associations meant that any attempts to conduct multiple imputation using sexual behaviours could not have been considered reliable nor of a robust standard.

7.5.3 Limitations and Strengths

This analysis had some limitations; the primary limitation of this analysis was the limited number of participants with HPV serology data. It has also been suggested that, socioeconomically, UK Biobank participants are less deprived relative to the general population. (Fry et al., 2017) This may have potentially limited the representativeness of the results. Similarly, most participants with HPV serostatus data were of a white ethnicity, potentially further limiting the generalisability of findings to other ethnic groups. This analysis also had some strengths. The quality of the data collected was high, with the vast majority of HPV serology samples passing quality checks. (UK Biobank, 2019) Similarly, the questionnaires used for data capture were thorough and supported by trained interviewers. (Rory, 2007)

7.6 Conclusions

While a large and robust dataset, this analysis has shown that in the context of HPV-driven HNCs, the utility of the UK Biobank is ultimately hindered by a lack of HPV data. If HPV serology testing were to be expanded beyond the limited numbers of randomly selected pilot-study participants in the cohort, there may be scope for future studies in the UK Biobank to improve our understanding of the aetiology of HNC.

8 Chapter Eight – Discussion

In this chapter, the key findings of this thesis are discussed in the context of the wider literature, in addition to the strengths and limitations of the thesis studies included. Conclusions and recommendations for future policy, practice, and research are also presented.

8.1 Findings

This thesis has described the changing epidemiology of head and neck cancer and the potential of risk prediction modelling for HNC. While risk prediction modelling is a more common phenomenon for cancers of other sites and comorbidities, it remains an underutilised concept in the prevention of HNC.

What is the current evidence base of head and neck cancer risk prediction models and tools?

The development of the HNC risk prediction model discussed in Chapter Six was informed by undertaking a comprehensive review of the international literature, in which existing HNC models were identified, assessed and their findings synthesised using the PROBAST framework and an additional quality assessment. (Wolff et al., 2019) The review identified a gap in available fully validated risk prediction models and models translated into clinical tools. Three of the models in the published literature were identified to be high quality and high performing (AUC > 0.80). A narrative synthesis of the models highlighted that simpler HNC risk models and those that were tailored to their target population or subsite had better predictive performance. It also informed methodologically rigorous approaches to developing the risk prediction model of this thesis.

What is the sociodemographic profile of head and neck cancer and is it changing over time?

The analysis of the Scottish Cancer Registry in Chapter Four highlighted the changing burden of disease in Scotland but that the sociodemographic profile (defined by age, sex and area-based socioeconomic status) of HNC is unchanging over time. Relative to the 2001-2005 period, significant increases in the

incidence of OPC were observed by the 2016 to 2020 period (RR = 1.78, 95% CI = 1.65-1.93). Over the same period, oral cavity cancer incidence remained stable (RR = 0.94, 95% CI = 0.88-1.01) and larynx cancer incidence rates declined (RR = 0.73, 95% CI = 0.68-0.79). The registry analysis also revealed that while the burden increased at an overall level, the sociodemographic profile of people with HNC has remained largely unchanged – both overall and within subsites where there were no major shifts in terms of age, sex and socioeconomic status. A consistent and strong socioeconomic pattern was observed, with the highest rates among SIMD 1 (20% most socioeconomically deprived RR = 2.87, 95% CI = 2.73-3.00) compared with SIMD 5 (20% least socioeconomically deprived).

The primary limitation of the registry analysis in Chapter four was a lack of tumour HPV data. In the global context of increasing trends of HPV-positive OPC incidence, the increases in OPC observed in the Scottish Cancer Registry analysis were likely HPV-mediated; an estimated 60% of OPC cases in Scotland test HPV-positive. (Wakeham et al., 2019b, Gormley et al., 2022a) The analysis would have been enhanced if there was the ability to distinguish between HPV-positive and HPV-negative OPC cases. Using data from the ARCAGE study, an analysis of case participants with OPC and HPV serology data was conducted in Chapter Five. The sociodemographics (age, sex and educational level) and behaviours of **HPV-positive** and **HPV-negative** were compared directly and with controls as a sensitivity analysis. There was no strong evidence to suggest people with **HPV-positive** OPC possess a different sociodemographic profile from people with **HPV-negative** OPC. Relative to controls, **HPV-negative** participants reported a lower socioeconomic status (University / manager vs primary education / worker - OR = 0.34, 95% CI = 0.16 – 0.73). Smoking and alcohol behaviours were more strongly associated with HPV-negative OPC. When compared with controls, there was no association with a higher educational level and **HPV-positive** OPC (OR = 1.17, 95% CI = 0.52-2.67).

When participants with **HPV-positive** and **negative** cases of the oropharynx were compared directly, there was an unclear association with a higher, relative to a lower, educational level and HPV serology status (University / manager vs primary education / worker) (OR = 3.07, 95% CI = 0.92 – 10.30). These findings contrasted with the limited existing literature which was suggestive of a higher

SES associated with HPV-positive OPC patients. (Dahlstrom et al., 2015b) The smaller numbers of OPC cases is highlighted by the imprecise estimates (wide 95% CI) in this analysis. Further analyses of pooled international studies are warranted.

Could a head and neck cancer risk prediction model be developed and externally validated, with a view to implementing in a primary dental care setting?

A risk prediction model for head and neck cancer is described in Chapter Six. Using data from the ARCAGE case-control study, a clinical risk prediction model was developed using established and readily ascertainable risk factors in the clinical setting including patient sociodemographics and high-risk behaviours. The model was designed with consideration of the applicability of use in the primary care setting. While performing moderately well in the developmental stages (AUC = 0.75, 95% CI = 0.74–0.77; Spiegelhalter's Z test = 0.603, p = 0.55), the model had a reduced but still acceptable performance level in the UK Biobank cohort validation dataset (AUC = 0.62, 95% CI = 0.61–0.64; Spiegelhalter's Z test = −0.013, p = 0.99)).(Hosmer Jr et al., 2013) Attempts to revise the model (using further variables (denture status, BMI), stratifying by the risk of OCC and larynx cancer and using only UK centre data) yielded no or minimal improvements.

Due to small numbers of participants in the UK Biobank with HPV data, it was not feasible to effectively validate a model which included HPV serostatus. It is likely that the inclusion of HPV serostatus would have improved the performance of the model based on its promising performance in the developmental dataset (AUC = 0.80, 95% CI = 0.79–0.82) and the larger proportion of OPC cases in the UK Biobank dataset. While subject to some limitations, the UK Biobank validation analysis has shown that epidemiological risk factors can offer generalised HNC prediction but biomarkers may enhance individual risk prediction and provide the opportunity for a precision medicine approach. (Ginsburg and Phillips, 2018)

Chapter Seven explored the availability of HPV data in the UK Biobank and the feasibility of imputing HPV serostatus across the dataset for the testing of an HPV model. Limited demographic and behavioural (smoking and alcohol) associations with HPV serostatus were observed. Some associations were observed with sexual behaviours but even these were not consistent across definitions with the only exception being the practice of same-sex intercourse. The limited strength of associations and the large volume of missing HPV serology data (>98%) meant that imputing HPV serostatus across the dataset was impractical. If serology testing were to be expanded to a larger cohort sample, the UK Biobank could be a useful resource in furthering the understanding of head and neck cancer epidemiology.

Ultimately, the performance of the risk prediction model was considered to be fair but could be improved. If the model were to be used for secondary screening and major clinical decision making, e.g. deciding whether to conduct a biopsy, this level of performance may not be sufficient. However, for the purposes of primary prevention, the model could still have utility in prompting behavioural discussions and helping to inform patient recall decisions in the dental setting. A study protocol was also developed with a future view to pilot the feasibility of a model in primary dental care.

8.2 Comparisons with the International Literature

Over the course of this thesis, there have been several additions to the literature since the initial summary of the literature in Chapter 1 and published rapid review of risk prediction models. This section explores these and the key thesis findings in this wider context.

8.2.1 Reviews of Head and Neck Cancer Risk Models

The published review of HNC risk models (Chapter Three) compares with a recently published systematic review of oral cancer risk prediction models which has expanded upon some of the research conducted and reported in this thesis with a detailed review of OCC models. (Espressivo et al., 2024) This systematic review identified and appraised 23 different models using a similar methodology to this thesis, including the use of the PROBAST tool, and identified many models

in common with those described in Chapter Three (n = 11). The review also identified several models that were not included in the rapid review, including six models that used genetic factors. Upon examination, these can primarily be attributed to models being published at a date after the conclusion of the thesis review's literature searches, the exclusion of genetic models or publications not explicitly using the word "model" or "tool" in the title or abstract.

The systematic review's assessment largely chimed with the findings of the thesis rapid review. The authors identified several models that have potential utility. (Espressivo et al., 2024) Many of the models made use of at least one behavioural factor, many including key predictors such as smoking and alcohol, as consistent with the thesis review. They noted the importance of the feasibility of implementation, highlighting the merits of epidemiological models that can use readily accessible variables such as demographics or smoking. The review also discussed the predictive potential of HPV as a biomarker for models. Crucially, the authors also noted that many of the models lacked external validation and clinical testing and recommended that future research should focus on the validation of existing models.

Espressivo et al's (2024) systematic review had several strengths, including the publication of a protocol, robust appraisal methodology, and a detailed synthesis and discussion. It also had some limitations in comparison to the Thesis review as it was confined to only OCC risk models. The review also only included papers published in English, excluding several papers from regions with high OCC incidence. Notable differences were also observed in the use of the PROBAST tool, primarily harsher judgements and use of domains; several studies were classified as having a high risk of bias owed to the use of a case-control study for the development of a model. While case-control studies are subject to some limitations, they remain an important epidemiological tool in furthering the understanding of disease aetiology and often offer detailed exposure data not always captured by larger cohort studies, especially for lower volume diseases such as HNC and OCC.

8.2.2 Epidemiology

The key novel finding of the registry analysis was the unchanging sociodemographic profile of HNC in spite of the substantial increases in OPC rates. This was not consistent with reports of significant changes in the sociodemographics of people with HNC driven by HPV-related OPC. (Gillison et al., 2008, Dahlstrom et al., 2015b) These studies were smaller clinical case-controls or cohorts constrained by limited numbers of people with HPV-positive OPC. These studies were also conducted in the United States, where the SES differences observed could be explained by a socioeconomic skew towards higher SES patients with insurance and access to healthcare. (Hoffman and Paradise, 2008)

The peak age incidence of HNC remained relatively stable with a modest increase in age at an overall level and for OCC. These findings make sense in the context of an aging population in Scotland. (Scotland's Census, 2022) Although OPC patients were slightly younger (60-64) than other sites (65-69), the age distribution of people with OPC remained stable. Similarly, observed rates were consistently higher among males across all subsites, mirroring the findings of previous studies. A strong socioeconomic gradient was also observed in HNC incidence, with the highest rates consistently among the most deprived areas. Purkayastha et al. observed a similar inequality gradient across all HNC subsites in Scotland. (Purkayastha et al., 2016) Although the English Cancer Registry data used in a study by Mcarthy et al. did not include socioeconomic status data, regional data suggested a similar trend in England; the highest rates of HNC were observed in the North of England, which has higher levels of socioeconomic deprivation than the South of England. (McCarthy et al., 2015, Office for National Statistics, 2023)

This finding was also observed in OPC incidence trends and remained unchanging across the study duration. With approximately 60% of Scottish OPC cases estimated to be HPV-positive, the clear inequality gradient observed is not coherent with the suggestion of a higher SES associated with HPV-positive OPC cases observed in studies conducted in the USA. (Wakeham et al., 2019b) Prior studies have suggested that high-risk sexual behaviours, risk factors for HPV

infection and HPV-positive OPC, are similarly socioeconomically patterned with these more common among those of a lower socioeconomically status. (Wellings et al., 2001, Jackson et al., 2012) Moreover, these findings suggest that a similar inequality pattern exists in both HPV-positive and HPV-negative OPC incidence.

The findings of an unchanging sociodemographic profile in the context of growing overall HNC incidence from the Registry study offer an important contribution to the literature. The incidence trends observed (substantial increases in OPC and declines in larynx cancer rates) in Scotland also correlate with existing global incidence trends of increases in OPC and declines in larynx cancer rates in some high-income countries. (Gormley et al., 2022b) Similarly, the findings were consistent with other robust UK registry analyses. (McCarthy et al., 2015, Purkayastha et al., 2016) In a previous Scottish Cancer Registry analysis by Purkayastha et al., it was forecast that OPC rates would supersede those of the OCC rates. The study findings of this thesis have confirmed these projections. (Purkayastha et al., 2016) The large increases in OPC incidence rates were also consistent with trends observed in England by McCarthy et al. (McCarthy et al., 2015) In their analysis, OCC rates increased and larynx cancer rates stabilised, while stabilisations in OCC rates and declines in larynx cancer rates were observed in this thesis's Scottish Cancer Registry analysis.

The findings are consistent with a global trend of an increasing burden of (likely HPV-mediated) OPC and declines or stabilisations of smoking and alcohol-mediated disease. (Gormley et al., 2022b) Scottish Health survey data has suggested a decline in population smoking trends, but harmful alcohol drinking levels persist. (Cabinet Secretary for NHS Recovery, 2021) This may explain the findings of a decline in larynx cancer rates and a stabilisation in OCC rates, where it has the strongest association with alcohol consumption. (Hashibe et al., 2009)

Similarly, the findings of the sociodemographic analysis of people with OPC in the ARCAGE case-control study in Chapter Five did not chime with existing literature. While some prior studies had suggested an increased risk of HPV-positive OPC among high socioeconomic groups, no significant associations were observed between a higher socioeconomic status and HPV-positive OPC in this

chapter. This and the clear inequality gradient observed among all OPC cases in the registry analysis contrasts with existing research suggestive that people with HPV-positive OPC are of a higher SES. (Dahlstrom et al., 2015b) However, like prior studies, the analysis was constrained by limited numbers of HPV-positive OPC cases. Further investigation of the SES profile of HPV-positive OPC is warranted.

8.2.3 Head and Neck Cancer Risk Models

The model developed in this thesis (Chapter Six) was informed by a methodical review of the literature (Chapter Three) and used two large and robust studies, one for development and the second for external validation. The key finding of an unchanging sociodemographic profile in the Scottish Cancer Registry analysis supported the use of the ARCAGE study (2002-2004) to develop a risk prediction model that could be validated in a more contemporary dataset (2006 – present). When compared and contrasted with other studies in accordance with the rapid review, it would likely be considered as a high-quality model due to its robust methodology, but it would not be considered high-performing due to the drop in model performance exhibited on external validation (AUC < 0.80). Nevertheless, the model still has a reasonable level of performance sufficient for primary prevention activity. The most comparable risk prediction models to that of this thesis are the works of McCarthy and colleagues (2020) and Budhathoki et al (2023). (McCarthy et al., 2020, Budhathoki et al., 2023) The authors developed separate models using data from the UK Biobank and the ARCAGE studies, respectively, while this thesis developed and validated a model with these studies.

The study by Budhathoki et al. (2023) created a series of high-performing models for all HNC and several subsites. The authors pooled several studies from the VOYAGER (Human Papillomavirus, Oral and Oropharyngeal Cancer Genomic Research) Consortium including the ARCAGE, Carolina Head and Neck Cancer Epidemiology (CHANCE), Pittsburgh and Toronto case-control studies and the Head and Neck 5000 cohort. (World Health Organisation (WHO), 2024b)A split sample approach was conducted, where data were divided into a training and hold-out (internal validation) datasets and the UK Biobank study was used to

estimate absolute population risks. These models stratified by sex and used both epidemiological (age, smoking packyears, alcohol drinking intensity and socioeconomic status (education) and biomarkers (polygenic risk scores, HPV status) predictors. This study was well conducted, and the large number of cases was a major strength. The models had good performance, particularly the HPV and epidemiology models for the oropharynx (Male - AUC = 0.92, 95% CI = 0.90-0.94, Female - 0.91, 95% CI = 0.86-0.94). A key limitation of this study was that the authors did not employ external validation. Comparatively, the models performed better than the thesis model (Chapter Six). However, the epidemiology models for HNC had better, but still modest, performance (All HNC Male - AUC = 0.69, 95% CI = 0.67-0.71: Female - AUC = 0.75, 95% CI = 0.72-0.78). The combined epidemiology and biomarker (polygenic risk scores, HPV) models, especially the OPC models which utilised HPV serostatus, had significantly improved performance characteristics. However, the clinical utility of this biomarker approach would currently be restricted by limited (HPV or genomic) testing capabilities in primary care. The findings of this approach suggest that epidemiological models have predictive potential but that future models, where practical, should strive to incorporate biomarkers. This correlates with the findings of the model development study in this thesis, where HPV serostatus was highly predictive in HNC model.

The model developed by McCarthy et al. (2020) was designed for use in primary care settings. In this model, the UK Biobank dataset was split geographically into a development and validation dataset using the North of England. The model used similar predictors to those chosen in the ARCAGE model. This included demographic factors (age, sex, and area-based socioeconomic (Townsend) Deprivation score), behaviours (fruit consumption, exercise, smoking, and alcohol status) and BMI. Smoking packyears / duration data were also not included due to missing data. The model had several strengths; it was a well-developed and robust model, which followed TRIPOD guidelines. The variables chosen were largely clinically feasible to capture. The authors also utilised patient and public involvement in the model's design. Like the risk prediction model described in this thesis (Chapter Six), the model's performance dropped to acceptable but limited levels of discrimination (AUC = 0.64, 95% CI = 0.60-0.68) upon validation. Collecting BMI from patients may have proven challenging

in a dental setting, with currently no evidence assessing it's feasibility. BMI is also subject to potential reverse causation in HNC risk, as discussed in this thesis and acknowledged by the authors.

The model in this thesis shares several similarities with this study. (McCarthy et al., 2020) Both of the models used epidemiological predictors and were explicitly designed for use in a clinical setting. Following the findings of the rapid review, whereby simpler models tended to report better performance, and in consistency with the statistical principle of parsimony, the thesis model was intentionally designed to include a minimal number of variables. At the time of development, it was theorised that the drop in performance between the development and validation phases observed in the McCarthy et al. (2020) model was attributable to overfitting and the use of additional, non-major risk factors. However, given the similar drop in performance observed upon validation of the thesis model, it may suggest that the UK Biobank dataset may not be necessarily representative of the general population. The UK Biobank cohort population has been shown to be, on average, more socioeconomically affluent than the general population, which may have limited it's suitability for the risk prediction of HNC which disproportionately affects those from lower socioeconomic groups. (Fry et al., 2017) Alternatively, it may suggest that epidemiological predictors are insufficient to precisely stratify those at risk and that biomarker adjuncts are needed. In the discussion, McCarthy et al (2020) also acknowledged that HPV serostatus could improve the prediction of a model but at the cost of limited clinical utility, in agreement with the suggestions of this thesis. Another key theme in the discussion by McCarthy and colleagues was the potential role for a model in a dental setting for preventive activity which were the implications and conclusions also drawn in this thesis.

Since the publication of the rapid review, a risk model which was developed into a clinical tool (the Head and Neck Cancer Risk Calculator, HANCRC v.2 by Tikka et al. (2020)) has since undergone further testing. This model was included and described in the thesis rapid review. Work by Simpson and colleagues evaluated the accuracy of HANCRC v.2 against a cohort of urgent suspected cancer referrals to an Oral and Maxillofacial Surgery unit for the first time. (Simpson et al., 2023) The calculator was shown to be effective in assessing referrals, with

good accuracy, recommending 76% of malignancies and only 41% of non-malignancies for urgent referrals. The model also had a high Negative Predicted Value (NPV) of 99.2% (i.e. 99.2% of all negative results will be truly negative). The authors also concluded that the tool could improve the triaging of referrals but could also be improved for OCC specific signs or symptoms (e.g. oral erythroleukoplakia). The study was subject to some limitations, namely smaller numbers of patients with confirmed malignancy and a binary classification between desirable and inappropriate referrals which does not fully consider many of the OPMDs included, which can carry a high risk of malignant transformation and would also need referral and further investigations, e.g. biopsy. (Iocca et al., 2020) This work confirms prior evaluations of this risk tool that highlighted its effectiveness in the secondary screening of symptomatic patients. (Hardman et al., 2021a) Ultimately, the work by Tikka et al (2020) has a different clinical function to the potential role of the thesis model developed in this thesis.

8.2.4 Risk Factors for HNC

Further studies on HNC risk factors have also recently been published. Work from the INHANCE consortium by Goyal et al. revealed that risk associations with behaviours vary internationally with higher risks observed in lower-income countries. (Goyal et al., 2023) This study observed that longer term smoking behaviours was associated with a higher risk of OCC and larynx cancer in higher income countries, whereas long-term smoking was associated with an increased risk of OPC and hypopharynx cancer in lower income nations. Similarly, long-term alcohol behaviours were associated with an increased risk of OPC, larynx and hypopharynx cancer in lower income countries. These differences were attributed to geographical variations in tobacco and alcohol products available and the burden of HPV-mediated disease. These differences suggest that there are global differences in exposure profiles, both geographically and economically, but reinforce the universal importance of regulation.

A recent study using data for the INHANCE consortium by Sassano et al. observed a significant protective effect associated with aspirin against HNC, particularly for tumours of the oropharynx and larynx. (Sassano et al., 2024) A dose response

effect was also observed with duration. The authors theorised that the anti-inflammatory effects of the drug could have an anti-neoplastic effect, but also cautioned against potential recall biases and confounding associated with hospital-based controls. This exposure was explored in the development of the model (Chapter Six) but lost significance at the black box stage, suggesting it would likely have had a limited impact on the risk model, at best.

Although not feasible to include in the thesis model due to data limitations (the ARCAGE study had no recreational drug use data), several publications assessing recreational drug usage as an exposure have also emerged. A Scottish analysis among laryngeal patients, although limited by a smaller sample size, found an increased morbidity in outcomes among patients which used recreational drugs. (Woodley et al., 2022) Similarly, analysis of the INHANCE consortium found a weak but increased association between cocaine inhalation and HNC. However, this sample was primarily limited by limited numbers of cocaine users. (Zhang et al., 2024) In consistency with this, a recently conducted systematic review suggested strong associations with opioid usage and HNC risk, especially for tumours of the larynx. (Mohebbi et al., 2024) Future models could explore the inclusion of recreational drug usage.

8.2.5 Primary Prevention and Risk Prediction

The thesis research including the review, registry analysis and development and validation of a risk prediction model are comparable to similar research on other cancers. The thesis review was robust and well-conducted, learning from and adapting methodologies to previous reviews of colorectal, breast and lung cancers. (Usher-Smith et al., 2016, Louro et al., 2019, Gray et al., 2016) These reviews also made use of the PROBAST tool and undertook a detailed synthesis. The performance of the models identified for HNC were broadly comparable to those of other cancers. A lack of external validation and testing of models in the clinical setting was a commonly identified theme.

The evaluation of patient demographics and risk characterises is crucial for prevention. The Scottish Cancer Registry analysis conducted in Chapter Four is of a similar nature and quality to that of other registry analyses such as that of

melanoma, breast and thyroid cancers. (Reynolds et al., 2005, Brewster et al., 2007, Mesa-Eguiagaray et al., 2020) The analysis was also similar to previous head and neck cancer registry analyses. (Purkayastha et al., 2016) Large socioeconomic inequalities were also observed among cancer other cancers in Scotland. (Tweed et al., 2018)

The general findings of this thesis are in accordance with the literature on other cancer risk prediction models. The potential preventive benefits and resource-efficiency of cancer risk models have been described, including positive behavioural change effects, no increased cancer worry and improved patient knowledge of risk . (Walker et al., 2015b) Despite this, cancer risk prediction tools are under-utilised especially in the primary care setting where there is inherent preventive potential. (Usher-Smith et al., 2015)

There is an abundance of prognostic models for HNC, many of these having undergone review. (Tham et al., 2019, Russo et al., 2021, Aly et al., 2023) Many of the challenges described with prognostic modelling were similar to those of risk prediction models, particularly a lack of external validation and testing. With the success of the Tikka et al. (2020) referral model, there has been an increased focus on secondary detection studies as well as contemporary changes to referral guidelines such as the new Scottish “Optimal Head and Neck Cancer Diagnostic Pathway”. (NHS Scotland, 2023)

Comparatively, the literature focused on the primary prevention risk modelling of HNC is much more limited. There are currently no studies or reviews of the utilisation of HNC cancer risk models in primary dental care, nor could any studies of testing risk prediction models for HNC in dental settings be identified. This is consistent with the literature suggesting that cancer risk models are an underutilised tool in the primary care setting. (Price et al., 2019)

8.3 Limitations of Thesis Approaches

8.3.1 Review Methodology

The studies of this thesis were subject to some limitations. The primary limitation of the rapid review (Chapter Three) was a lack of a published protocol which would clearly set out the review methods a-priori, as per PRISMA systematic review guidelines. Despite this, the review was very thorough (having been informed by a subject librarian, included dual screening and a rigorous quality assessment with the PROBAST framework and a defined PICO question). While there was no published protocol, the methods of systematic search (search terms, and multiple databases), quality / performance appraisal (PROBAST) and synthesis were decided a-priori.

8.3.2 Data Limitations

The primary limitation of the analysis of the Scottish Cancer Registry was related to the limitations of the data available, including a lack of tumour HPV status. Some trends, e.g. declines in predominately smoking-driven laryngeal rates, could be somewhat explained by smoking and alcohol data reported by sources such as the Scottish Health survey. (Scottish Government, 2021) With an estimated 60% of Scottish OPC tumours diagnosed as HPV-positive, it was not feasible to differentiate between HPV-positive and negative tumours and further assess for sociodemographic changes. (Wakeham et al., 2019a) This is increasingly important because HPV-positive and HPV-negative OPC are commonly considered as distinct diseases. (Elrefaey et al., 2014)

Attempts to account for this limitation were made in the subsequent Chapter (Chapter 5), where the demographics and behaviours of people with HPV-positive and negative tumours of the oropharynx were investigated. However, this analysis was constrained by a limited number of HPV-positive OPC study participants. This chapter, however, serves as a preliminary analysis for future international collaboration with partners in the HEADSpAcE consortium. With a much larger pool of people with oropharynx cancer, this future analysis could have more power to help answer this research question.

8.3.3 Modelling Data and Approaches

The risk prediction model analyses (Chapter Six) were also subject to some limitations. While the model underwent external validation in an independent dataset (UK Biobank population cohort) and was constructed with a large and robust developmental dataset (ARCAGE case-control study), the use of two heterogenous datasets meant the matching of variables for inclusion was more limited. This resulted in the exclusion of otherwise potentially relevant risk predictors such as an individual's frequency of dental attendance, which would have made particular sense in the context of the risk prediction tool potentially influencing decisions on future recall intervals.

While using an all-site approach improved the generalisability and ease of use of the thesis model, this grouped approach was also less specific to HNC subsites and may have limited the predictive utility of the model. Some of the site-specific models developed by Budhathoki et al., (2023) exhibited a significantly higher performance. For example, the HPV and epidemiology model for OPC. Although there was a high number of events per variable (EPV) and variables were carefully selected to minimise overfitting, further internal validation strategies to minimise overfitting such as bootstrapping or cross validation were not utilised. Given the pre-approved methodology of external validation in the UK Biobank (and time sensitivities of the project), further internal validation was deemed to be unnecessary since external validation is regarded as more robust. This could be revisited in future modelling and validation studies.

The UK Biobank is also subject to the "healthy volunteer" effect whereby, on average, participants in the study are of a higher SES than the general population. (Fry et al., 2017) This coupled with the (age-group and sex) matching conducted in the ARCAGE study may have attenuated the strength of key sociodemographic predictors in the model, and in turn led to the underrepresentation of the predictive effects of these variables. No patient and public involvement (PPI) was utilised for variable selection; however PPI engagement (via Radnet) was planned for the feasibility testing of the model (Appendix 8).

While the use of a robust case-control study meant there was strong behavioural data capture, the use of heterogenous datasets may have also limited the utility of the model for longitudinal estimates, e.g. 10-year risk. Most of the studies assessed in the rapid review used the same study type for development and validation (i.e. case-control data and case-control datatset). Longitudinal estimates would either require closely matching recall estimates for risk behaviours (if one of the studies was a case-control) or longitudinal follow-up with outcomes, e.g. development and validation with two cohort studies (for example, the Q-series / cancer tools make use of this). (Hippisley-Cox and Coupland, 2015, Hippisley-Cox et al., 2017) While this increases the complexity of modelling, there may be grounds to suggest this should become standardised practice. Population cohorts, while a good data source, also require large participant numbers and long follow up time for "low-volume" outcomes such as HNC. Equally, a "snapshot" of risk may be sufficient for preventive prompts and informing recall intervals as part of a "brief advice" intervention. (Omaña-Cepeda et al., 2016, Department of Health and Social Care, 2021c) Nevertheless, this thesis has shown that it is possible to build and validate a model using two heterogenous study types, and still possess reasonable prediction performance.

The limited quantity of HPV data in the UK Biobank study (>98% missing) also meant an HPV model could not be explored in the context of HNC risk prediction (Chapter Seven). This lack of data was so extensive that even multiple imputation methods were deemed inappropriate. However, this could be re-explored if HPV testing were to be expanded to a larger proportion of the study.

8.4 Strengths of Thesis Studies and Approaches

8.4.1 Review Methodology

The studies included in this thesis had a number of strengths. The rapid review undertaken in Chapter Three offered a robust, in-depth summary of existing HNC risk models, using the established PROBAST quality-assessment framework. (Wolff et al., 2019) It made use of a pre-defined research question, a systematic search supported by a subject librarian, dual screening and an independent quality assessment with narrative synthesis. The review was the first of its kind

for HNC at the time of the publication and offered an important contribution to collating the international evidence base in the field.

8.4.2 Data Strengths

The analysis of the Scottish Cancer Registry in Chapter Four was informed by registry data which has a high reporting accuracy and data completeness. (Brewster et al., 1994, Brewster et al., 2002) This national population-level data allowed for the examination of trends beyond a local or study level and answer an important question in HNC research in terms of patient demographics. The analysis was undertaken in a robust manner following methods of peer-reviewed publications. (Purkayastha et al., 2016)

The data from the ARCAGE study, used to assess the sociodemographics of OPC case participants and for development of the risk model in Chapters Five and Six respectively, is another major strength of this thesis. The ARCAGE study was a large and well-powered international multicentre case-control study with 14 sites across 11counties in Europe. Detailed clinical, behavioural and sociodemographic data were captured by trained standardised interviewers following a standardised protocol. (Lagiou et al., 2009) HPV tumour status data was informed by sensitive HPV serology techniques. (Waterboer et al., 2005b)

The UK Biobank cohort data used in chapters Six and Seven was another strength of the thesis analyses. The study was well powered with over half a million participants, with comprehensive data collection via standardised interviews and provided a dataset resource well-suited to cancer research with linked data to cancer and death registries. (Conroy et al., 2022)

8.4.3 Modelling Approaches

The development of the model discussed in Chapter Six also had several merits. It was developed in accordance with TRIPOD guidelines, which advises on the reporting of the rationale, methodology, reporting of results and evaluation of models. The developmental dataset (ARCAGE) was a large, multi-centre case-control study coordinated by IARC with standardised procedures, detailed data capture, via trained interviewers. It also offered a reliable dataset to begin

exploring sociodemographic and behavioural associations within OPC (Chapter Five) with harmonised educational data according to UNSECO classifications. (Conway et al., 2015) The model itself was carefully designed using several approaches to ensure the validity of variables included, with practicality of use at the forefront. The UK Biobank study also served as a large, reliable and well conducted study to validate the ARCAGE model in.

The external validation of the model was a large strength of this study, a lack of validation being a common pitfall of many other models. (Wolff et al., 2019, Smith et al., 2022b) Another common pitfall in the model "pipeline" is a lack of a plan for clinical implementation and testing. (Markowetz, 2024) Clinical applicability was considered throughout the development and validation process, with the preparation of an entire study protocol for further piloting (Appendix 8).

8.5 Conclusions and Recommendations

8.5.1 Conclusions

The burden of HNC is changing, but crucially the sociodemographics of people with HNC in Scotland remain unchanged. This phenomenon remains true across all subsites including OPC cases and challenges prior studies which have suggested the demographics of HNC are changing, driven by increasing numbers 0f HPV-positive OPC. Wide socioeconomic inequalities in incidence were observed across the study duration. Moreover, in contrast with the limited existing literature, analysis of the albeit small sample of OPC case participants in the ARCAGE study detected no significant socioeconomic status differences among HPV-positive case participants with OPC versus other OPC cases and controls.

The model analyses have shown that it is feasible to develop a HNC risk prediction model using epidemiological predictors with acceptable performance. Although a large and robust population cohort, analysis of the UK Biobank was hindered by a lack of HPV data. Future modelling could potentially be enhanced

with the use of biomarkers, primarily HPV-16 serology data, especially with the growing incidence of OPC observed in this thesis and in the global literature. However, these recommendations would also necessitate improvements in testing capabilities for the primary care setting. Expansion of HPV serology testing may improve the study's utility for HNC research.

The HNC risk prediction model developed in this thesis still has merit for generalised prevention activity in the primary dental care setting. Risk thresholds could be adjusted such that the model may offer a clinician "prompt" and a "teachable moment" for those presenting with high-risk behaviours but yet to develop disease (e.g., alcohol consumption, tobacco smoking). Clinical prompts have been shown to improve patient safety, follow up and documentation. (Sutton et al., 2020) In the context of the recent shift to risk-based "dental check-ups" in Scotland, this model could also help inform recall decisions, rather than clinical judgement, alone. (NICE, 2020, NHS Scotland, 2024)

8.5.2 Practice Recommendations – a Primary Prevention Model

The risk tool referral developed by Tikka and colleagues (2020) is a robust and well-developed model. At the time of writing, it is arguably still one of the best-performing and clinically tested HNC models. It is well suited for secondary prevention and the triaging referrals, i.e. symptomatic patients potentially with head and neck malignant disease. However, for primary prevention purposes, recall and longitudinal risk assessment, the model is less suitable and there is still a gap for a HNC risk prediction tool in primary care settings.

Moreover, there is very limited evidence on the utility of risk prediction support tools in dental settings. There is currently no evidence evaluating the use of a head and neck cancer risk prediction model in primary dental care. At the time of writing, regular dental check-ups have moved from a six-monthly to annual frequency in Scotland to a risk-based recall interval with lower-risk patients suggested to be reviewed every two years. Recall frequency is based upon clinical judgment and knowledge of risk factors,

alone, with no aids to support these decisions. (NHS Scotland, 2024) Some evidence from the NICE clinical recall guidelines and from a primary care dental trial on recall intervals support the use of risk-based recall informed by clinical judgement. (NICE, 2020, Clarkson et al., 2021) In these circumstances, a clinical risk prediction model could be used to help support and inform said clinical judgments.

A worthy continuation of the thesis research would be the feasibility testing of the HNC risk prediction tool in primary dental care. An early version of a protocol was developed for such a study (see appendix 8). Unfortunately, due to the time constraints of an intercalated PhD, this was not feasible. However, the work of this thesis and protocol development lays the foundations for future feasibility testing of a HNC risk model in primary dental care.

Improvements in earlier detection and prevention cannot be achieved in a vacuum; it will require greater efforts to incorporate patients and primary care physicians' views and experiences into discussions as to how best to implement and utilise clinical risk prediction models, in addition to barriers and facilitators.(Taylor et al., 2023)

8.5.3 Policy Recommendations

Smoking and alcohol remain two of the biggest preventable risk factors for head and neck cancer. (WHO, 2017) Smoking regulation policies have been shown to reduce smoking uptake and related disease. (Frazer et al., 2016, Flor et al., 2021) Although smoking behaviours may be on the decline in some high-income countries, they are still common-place and strongly socioeconomically patterned with the highest rates amongst those most deprived. (Gormley et al., 2022b, Scottish Government, 2023b) Future restrictions or a complete generational ban of tobacco products are warranted; a modelling study conducted by the UK government predicted substantial reductions in uptake, smoking related disease and death in England. (UK Government, 2023)

Alcohol regulation has also been shown to be effective; for example, despite fierce industry opposition and delays, alcohol minimum unit pricing regulation in Scotland was shown to have reduced alcohol-related harms and deaths, with the greatest impact in lower SIMD areas. (Wyper et al., 2023) Further alcohol control is required including reviewing minimum unit pricing, other taxation strategies, licensing and advertising restrictions. (Scottish Government, 2018)

HPV-vaccination is an effective primary prevention strategy against many HPV-related cancers including those of the oropharynx. (World Health Organization, 2014) The HPV vaccination programme shows promising results, particularly with the recent report of no novel cervical cancer cases among those fully vaccinated at age 12-13 years in Scotland. (Palmer et al., 2024a) In the context of OPC, it will take several deacdes decades before the full protective effects of the vaccine become apparent. (Zhang et al., 2021) Vaccination efforts should be continued and carefully monitored to ensure there is wide population coverage and protection against infections and subsequent HPV-mediated cancers in the future; disparities in vaccine uptake by area-based socioeconomic status and by sex have been observed in Scotland. (Public Health Scotland, 2023b)

Access to dental care is another crucial policy consideration, which is currently a topic of political interest. In the UK, NHS dentistry faces a plethora of challenges including a retention crisis, delays in contract reform, underfunding and access challenges associated with widening socioeconomic inequalities. (Evans et al., 2023) Although registration rates with a dental practice are high in Scotland, dental participation rates have exhibited widening inequalities among children and adults. (Scotland, 2023) Ensuring sufficient and equitable access to dental care is essential to improve the primary prevention of HNC and other dental diseases in practice, which is potentially offered through translating the work of this thesis.

While improvements in treatment and early detection strategies will inevitably yield some benefits for patients, it is crucial to acknowledge and tackle the underlying role of socioeconomic inequalities in these cancers. Future policy will need to look carefully at social determinants and develop strategies to help address the inequalities and barriers facing those most vulnerable with universal, but proportionate, interventional strategies.

8.5.4 Research Recommendations

The research recommendations of this are described below.

8.5.4.1 Feasibility Testing

The risk model developed in this thesis should also be feasibility tested in dental practice, assessing the viability of a risk prediction model in the dental setting and clinician / patient acceptability. A draft protocol for such a study has been developed as part of this thesis and is described (Appendix 8). This would have assessed the acceptability and barriers / facilitators of implementing a model with patients and clinicians.

Future research could also assess the feasibility of opportunistic examination and risk prediction for HNC among high-risk individuals in other screening settings, for example lung cancer screening where smoking behaviours are a major risk factor in common. (Walser et al., 2008, Cavers et al., 2022)Similarly, with improvements in testing technologies, HNC risk assessment and investigations could be supported by multi-cancer blood tests such as those currently being evaluated in the Galleri trial. (Klein et al., 2021, NHS, 2021)

8.5.4.2 Data and Modelling Approaches

While the Scottish Cancer Registry offers various demographic and clinical information fields, the registry currently has no data on tumour HPV status. P16 testing is now the clinical standard for suspected oropharyngeal tumours and is provided on all pathology reports. The

addition of HPV status information to the Scottish cancer registry would help inform the assessment of the epidemiological burden of HPV-mediated cancers, the efficacy of the vaccine rollout and other interventions. This work could also, in theory, be undertaken via data linkage but there is currently no single national pathology database for Scotland, and therefore collating multiple health board databases would be a substantial undertaking.

Model validation with the ARCAGE HPV model that was created in Chapter Six should be conducted in another external study or dataset which possesses sufficient HPV data. Tests run in Biobank suggested an HPV model could have been predictive but due to the limited amount of HPV data, it was not possible to draw definitive conclusions.

The future of risk prediction modelling and answers to the challenges described in this thesis may lie with large population / Biobank datasets which are linked to health records. For example, the FinnGen consortium of studies where some 12% of the Finnish population are participants. (Kurki et al., 2023) Models like that of McCarthy and colleagues (2020) and the model of this thesis made use of epidemiological predictors and a large population dataset (the UK Biobank) are steps in the right direction. (McCarthy et al., 2020) There must be an emphasis on data completeness, particularly for exposures of interest such as alcohol usage, individual socioeconomic status, and HPV.

Similarly, it will be important to ensure study populations are representative of the general population - a potential limitation of the UK Biobank. The linkage of health data to datasets for research would have significant financial and ethical considerations for both policy makers and the public, but the insights afforded from such datasets would have immense research benefits for HNC and many other diseases.

8.5.4.3 Research in Risk Factors

Tobacco, alcohol and more recently, HPV, are recognised as the major risk factors for HNC. (Gormley et al., 2022b) However, they do not

entirely explain the burden of disease and risk, with only marginal progress made in further understanding the aetiology of HNC in recent years. It is recommended that future epidemiological research on HNC and exposures must be increasingly agnostic in nature and look further to the surrounding environment and determinants. (Davey Smith et al., 2023) Future work should continue to explore the molecular epidemiology of HNC and markers of subsite beyond HPV with international collaboration and sharing of resources / expertise. Some other potential exposures that may merit further investigation are listed in this section. Studies should strive to assess how these exposures interact with HNC risk, which in turn could help inform management and prevention strategies.

Poor oral health and dental attendance has been established as a modest risk factor for HNC. (Hashim et al., 2016b) However, the relationship between the oral microbiome and HNC risk is not fully understood. (Dorobisz et al., 2023) Similarly, although some genetic variants of interest have been identified, there is currently no singular major genetic risk factor that explains HNC susceptibility. (Gormley et al., 2022b) Further research is needed in these areas as markers of risk.

Dietary factors are described as a minor risk factor for HNC. (Chuang et al., 2012b) However, dietary factors are notoriously hard to accurately measure, much of the existing evidence originating from case-control studies which are subject to their own recall and reporting biases. (Clinton et al., 2020) Further research on diet and HNC risk is needed, particularly with the growing quantity and availability of processed foods. Processed food consumption has already been attributed to an increased mortality risk. (Fang et al., 2024) Prospective population cohorts may help offer future insights into diet and HNC risk.

Recreational drug usage is a relatively unexplored exposure in terms of HNC risk. Reviews and pooled studies have suggested opioid and cocaine usage is associated with an increased risk. (Zhang et al., 2024, Mohebbi et al., 2024) Moreover, recreational drug usage presents a major public health challenge in Scotland and, like HNC, is strongly socioeconomically

patterned. (Scottish Government, 2021) Future research and modelling on these exposures is warranted.

8.5.4.4 Socioeconomic Inequalities

Socioeconomic inequalities can be described as inequalities in economic and social affluence. These can be described using individual (e.g. education, income, and occupation) or area-based measures. (Conway et al., 2019b) Socioeconomic inequalities impact across the HNC continuum, from risk to prevention / early detection, diagnosis / treatment, and outcomes. Future research should continue to monitor and quantify socioeconomic inequalities in head and neck cancer, with a view to developing interventions such as targeted early detection or vaccination programmes. Lessons will also need to be learned from research in other cancers where barriers and challenges have been identified for early detection / screening among marginalised and vulnerable groups. (Cuypers et al., 2024) (Kotzur et al., 2022) It will also be crucial to ensure that relevant HNC clinical trials and population studies are representative of the patient / general population. (Bibbins-Domingo, 2022)

8.5.4.5 OPC and Inequalities Research

Primarily driven by HPV, OPC incidence has increased significantly in recent years and differs from other subsites in terms of biology, epidemiology, management and subsequent outcomes. (Gormley et al., 2022a) In Chapter Four a 78% rise in OPC rates was observed from 2001-2005 to 2016-2020 in Scotland; OPC is one of a few cancers which has exhibited an alarming increase in incidence over a short period, along with cancers of the thyroid and liver. (Cancer Research UK, 2021) This necessitates a need for new approaches and strategies to improve prevention, early detection and subsequent disease outcomes. The preliminary analysis of HPV-OPC cases by socioeconomic status in the ARCAGE study sets the stage for future work with the HEADSpAcE and NIH-funded VOYAGER consortiums. (World Health Organisation (WHO), 2024a) (World Health Organisation (WHO), 2024b) A large pool of OPC cases and controls, with HPV data could help overcome the limitations identified by

these analyses and help elucidate the relationship between individual SES and HPV-mediated OPC. This subsequent analysis could also identify any international variations in patterns of HPV-OPC by SES and analyse whether there are any other potential explanatory factors (e.g. smoking and alcohol behaviours). This, in turn, could help better stratify individuals for early detection interventions.

The growing burden of OPC is, and will continue to be, a major challenge for healthcare systems. Primary dental care remains an ideal and relatively untapped setting for HNC early detection. Although intra-oral examination is routine in the dental setting, there may be potential for the early detection of OPC via HPV testing in the dental setting, where GDPs have ready access to the oral cavity and oropharynx. The concept of HPV testing in the dental setting has already been demonstrated by the HOPSCOTCH study, where oral rinses were shown to be a viable method of testing the prevalence of HPV. (Conway et al., 2016) The results of this study were consistent with the international literature on oral HPV prevalence. (Kreimer et al., 2010) As evidenced by the Hamburg population cohort study by Busch et al., HPV-informed serology testing can allow for the early detection of OPC at stage I. (Busch et al., 2022b) Using HPV serology data taken from participants' blood in a population cohort, the investigators followed up those with positive HPV tests. They identified three early-stage OPC cases among the nine HPV-positive participants who attended for follow-up over two years. This study highlights that HPV serology testing has the potential to identify and stratify those at risk in a population.

Management strategies for people who test and remain HPV-positive will also be needed. For example, this could entail an enhanced monitoring pathway in primary or secondary care with additional endoscopy or ultrasound testing, similar to further testing for those with a family history of breast cancer and the BRCA gene. (NIH National Cancer Institute, 2020) Although HPV-informed early detection strategies show promise, these will necessitate improvements in point-of-care testing capabilities.

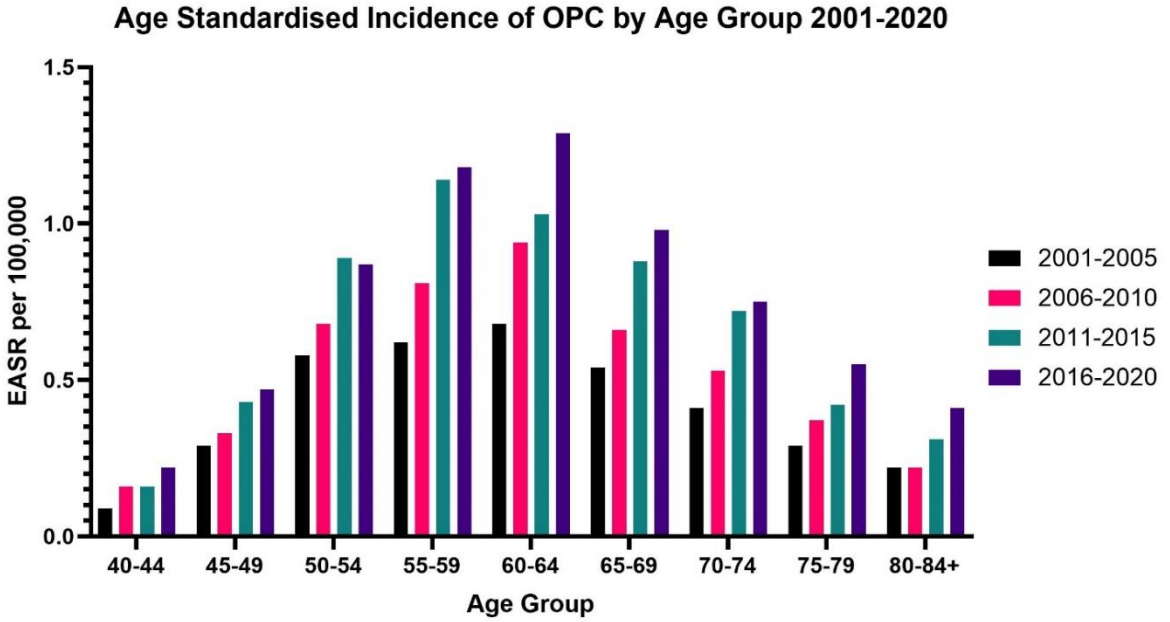
8.5.5 Final Remarks

This thesis started with identifying gaps in the literature, so, to close, I will use an anecdote on a gap. Throughout the course of my PhD, the theme of World Cancer Day (2022-2024) has been “Close the Care Gap”, emphasising the importance of global cancer care inequities, working together to address these and make a difference. To the unacquainted, epidemiology could seem somewhat removed from this theme. However, the goal of epidemiology is not just to quantify and describe disease, but to ultimately cure it. I hope this thesis embodies this theme and it will serve as one step, however small, in this direction. I hope I can continue to work with colleagues near and afar on “Close(ing) the Gap”. Our mission is not over yet.

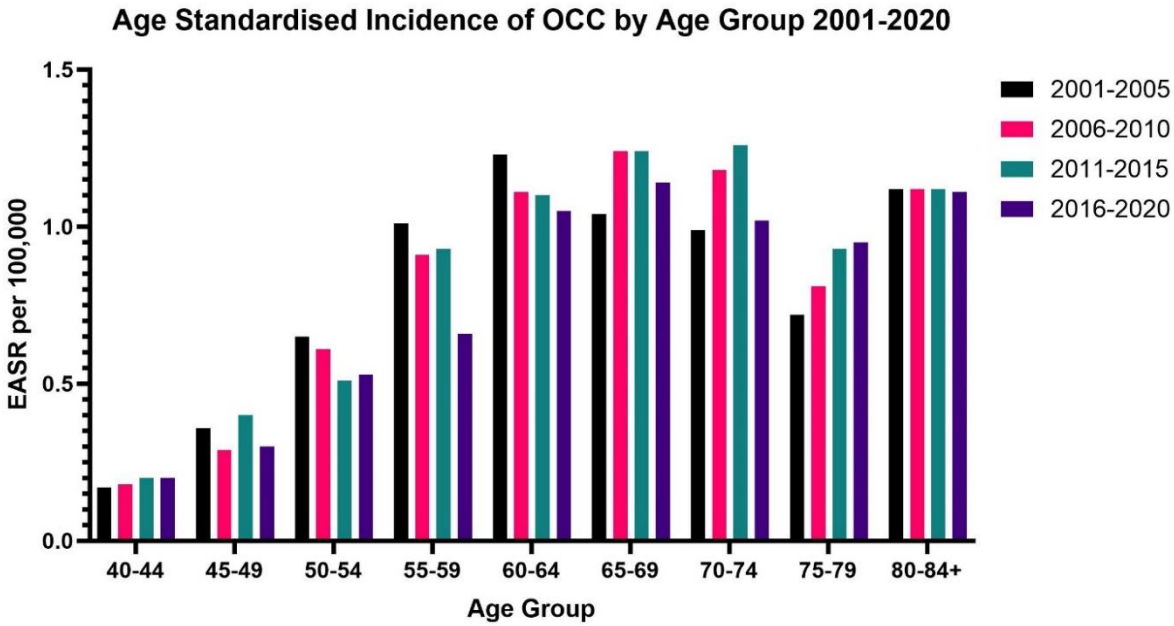
Fin

Appendices

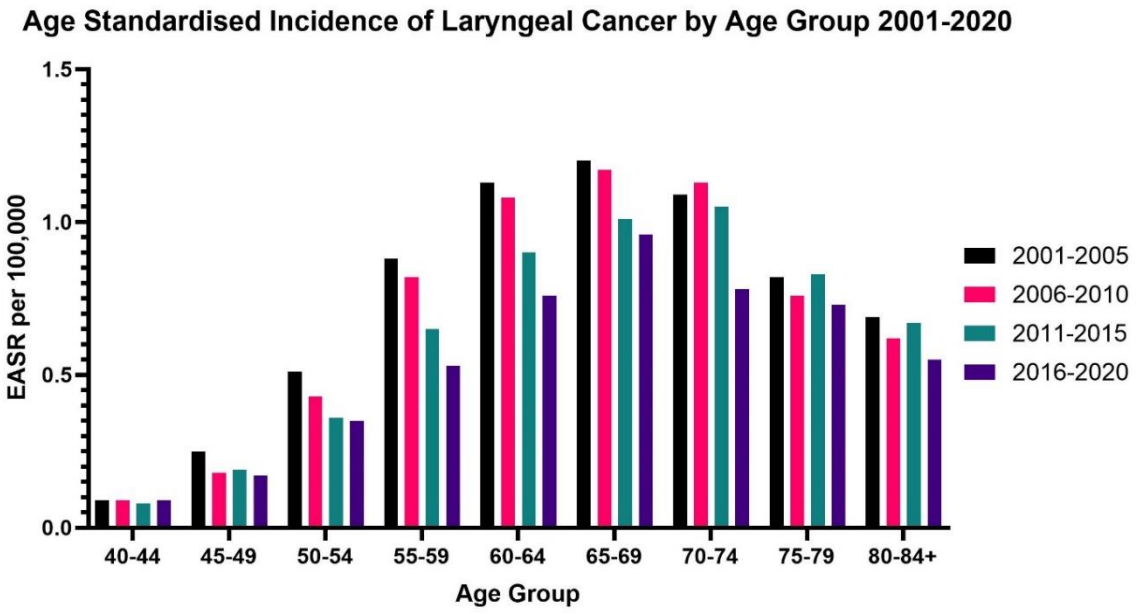
Appendix 4: Scottish Cancer Registry Analysis Supplementary Material (Chapter 4)



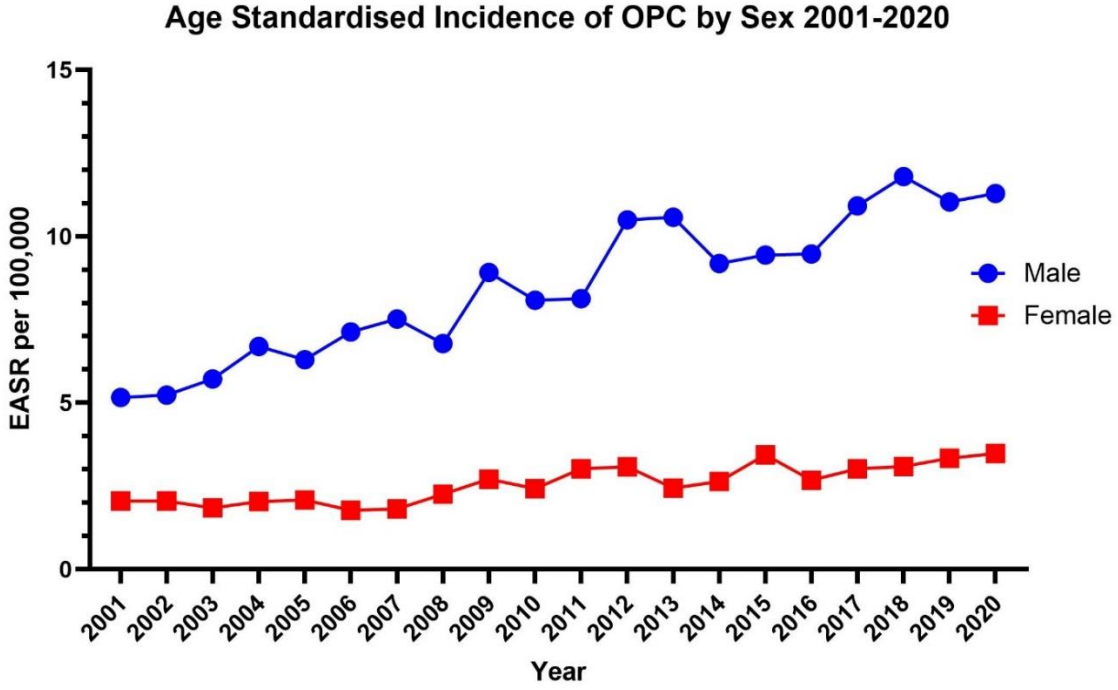
Supplementary Figure 4-1: Plot of Scottish Age-Standardised OPC Incidence Trends by Age-group and 5-year period (2001-2020)



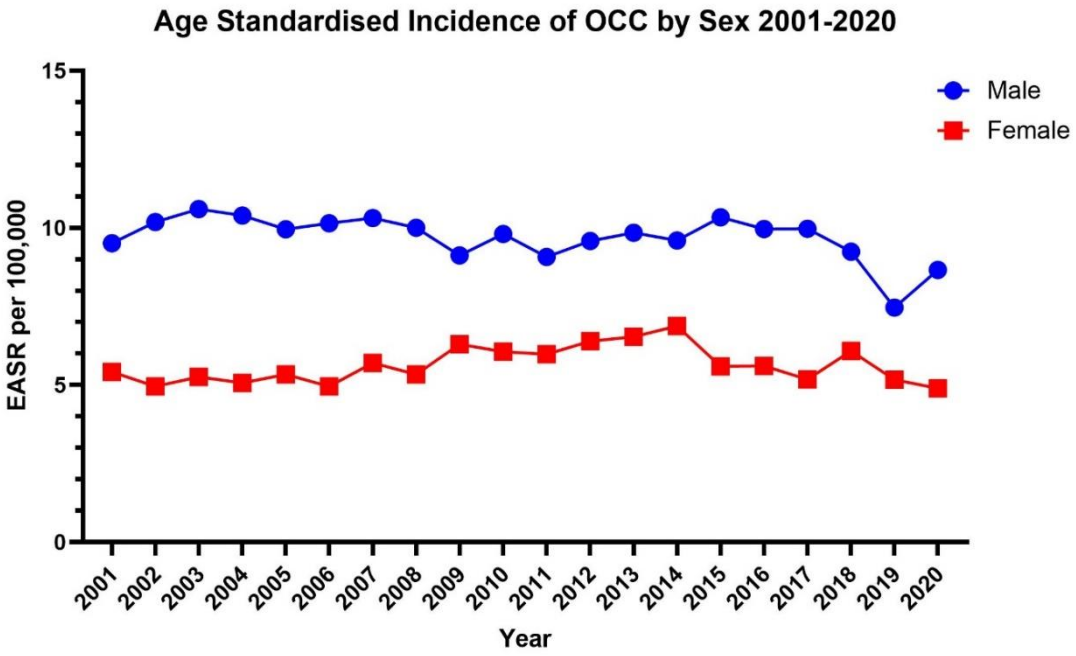
Supplementary Figure 4-2: Plot of Scottish Age-Standardised OCC Incidence Trends by Age-group and 5-year period (2001-2020)



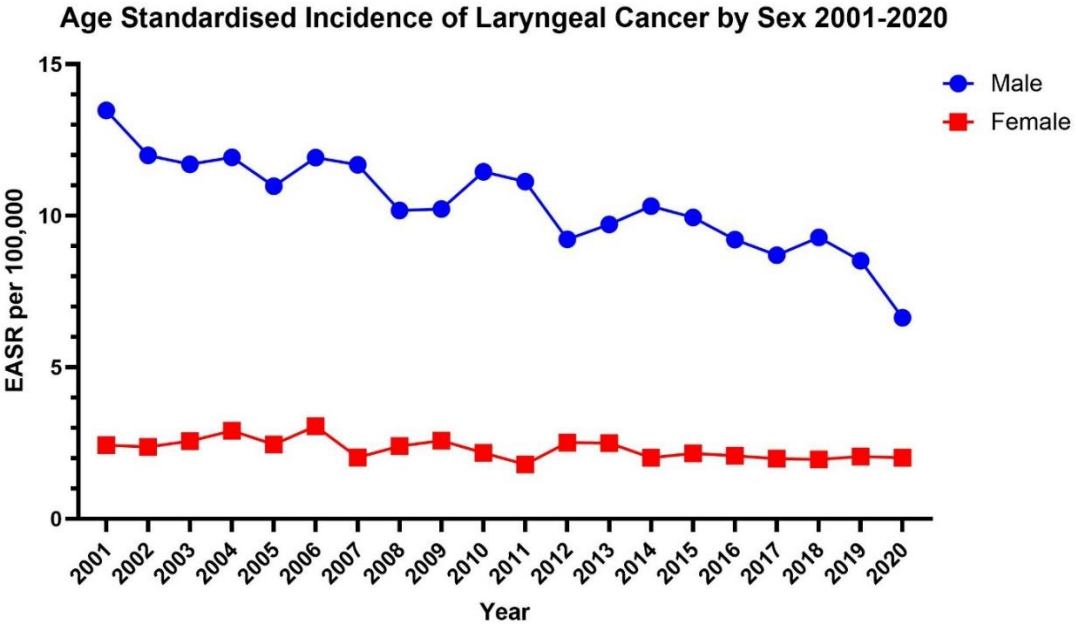
Supplementary Figure 4-3: Plot of Scottish Age-Standardised Laryngeal Cancer Incidence Trends by Age-group and 5-year period (2001-2020)



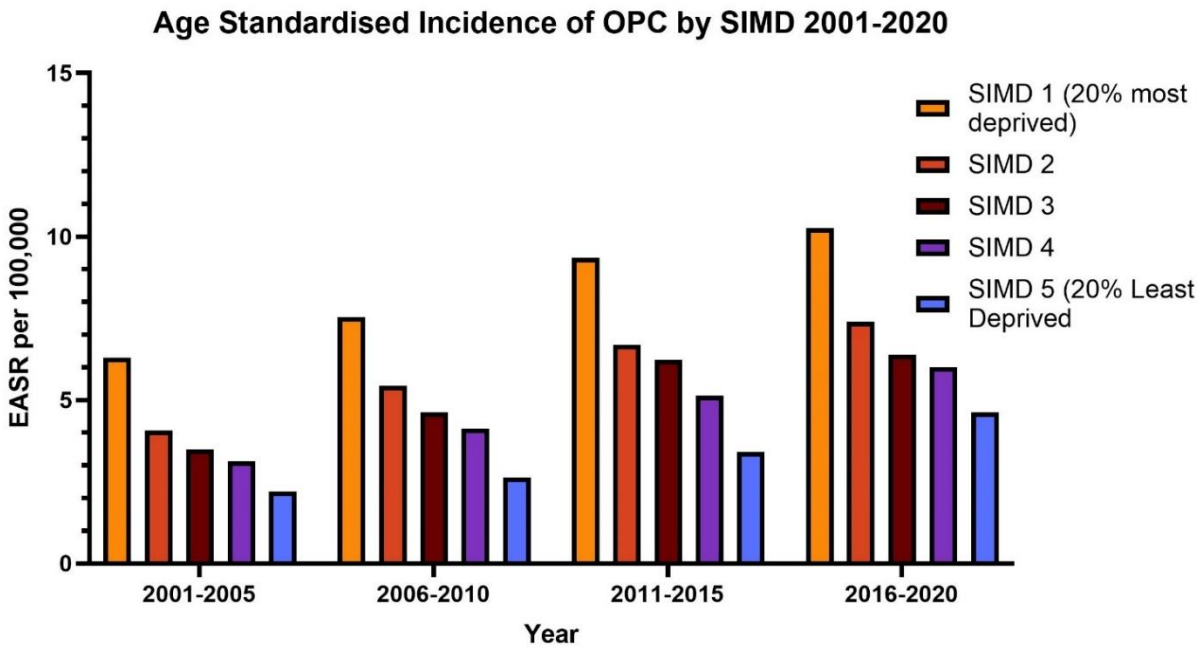
Supplementary Figure 4-4: Plot of Scottish Age-Standardised OPC Incidence Trends by Sex



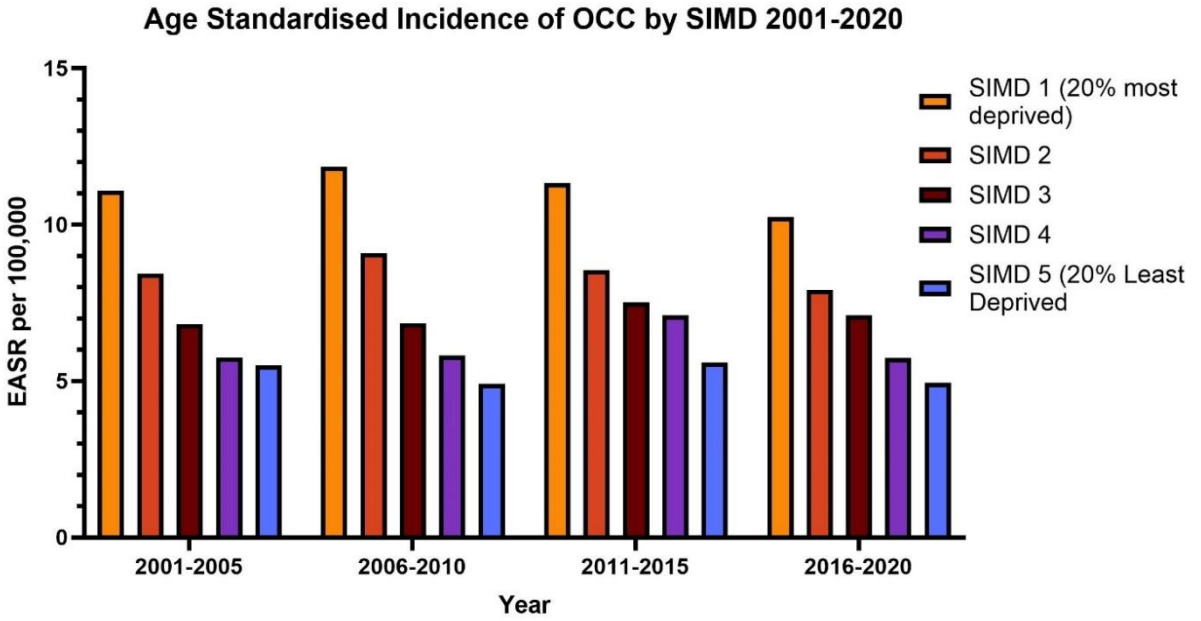
Supplementary Figure 4-5: Plot of Scottish Age-Standardised OCC Incidence Trends by Sex



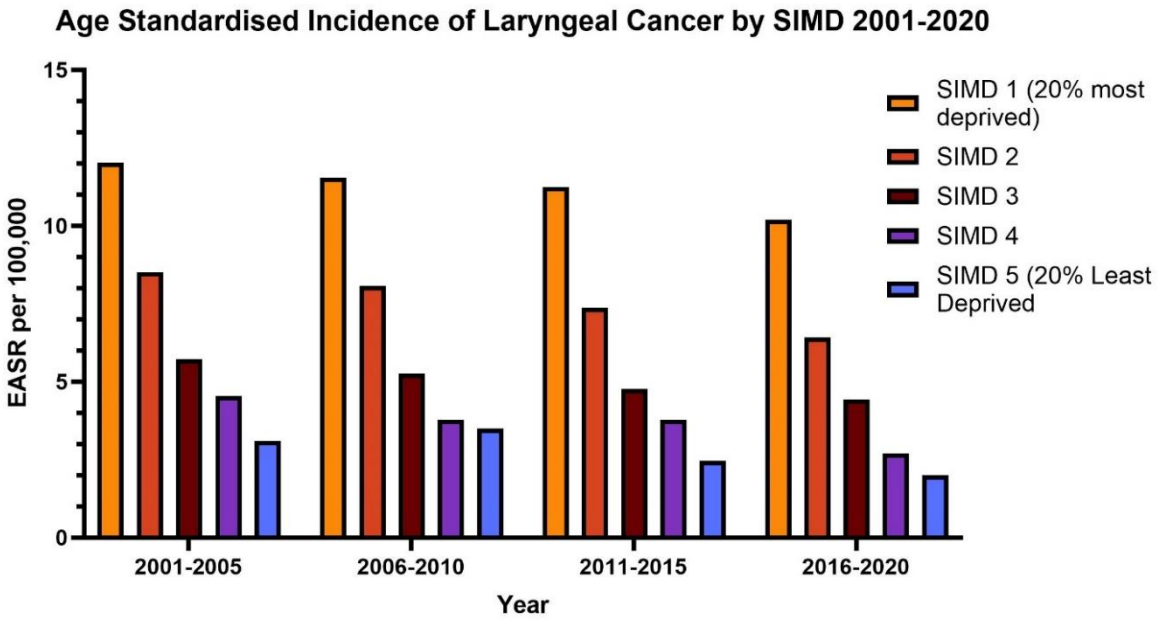
Supplementary Figure 4-6: Plot of Scottish Age-Standardised Laryngeal Cancer Incidence Trends by Sex



Supplementary Figure 4-7: Plot of Scottish Age-Standardised OPC Incidence Trends by SIMD Quintile and 5-year period (2001-2020)

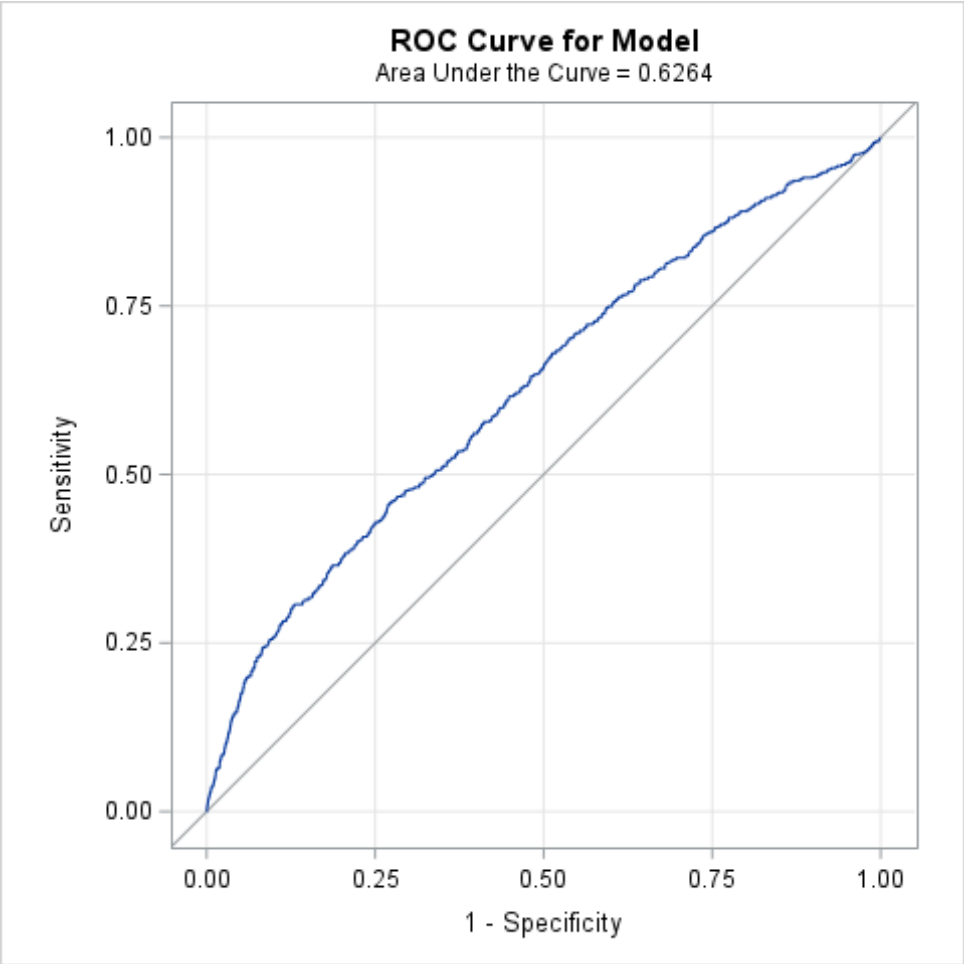


Supplementary Figure 4-8: Plot of Scottish Age-Standardised OCC Incidence Trends by SIMD Quintile and 5-year period (2001-2020)

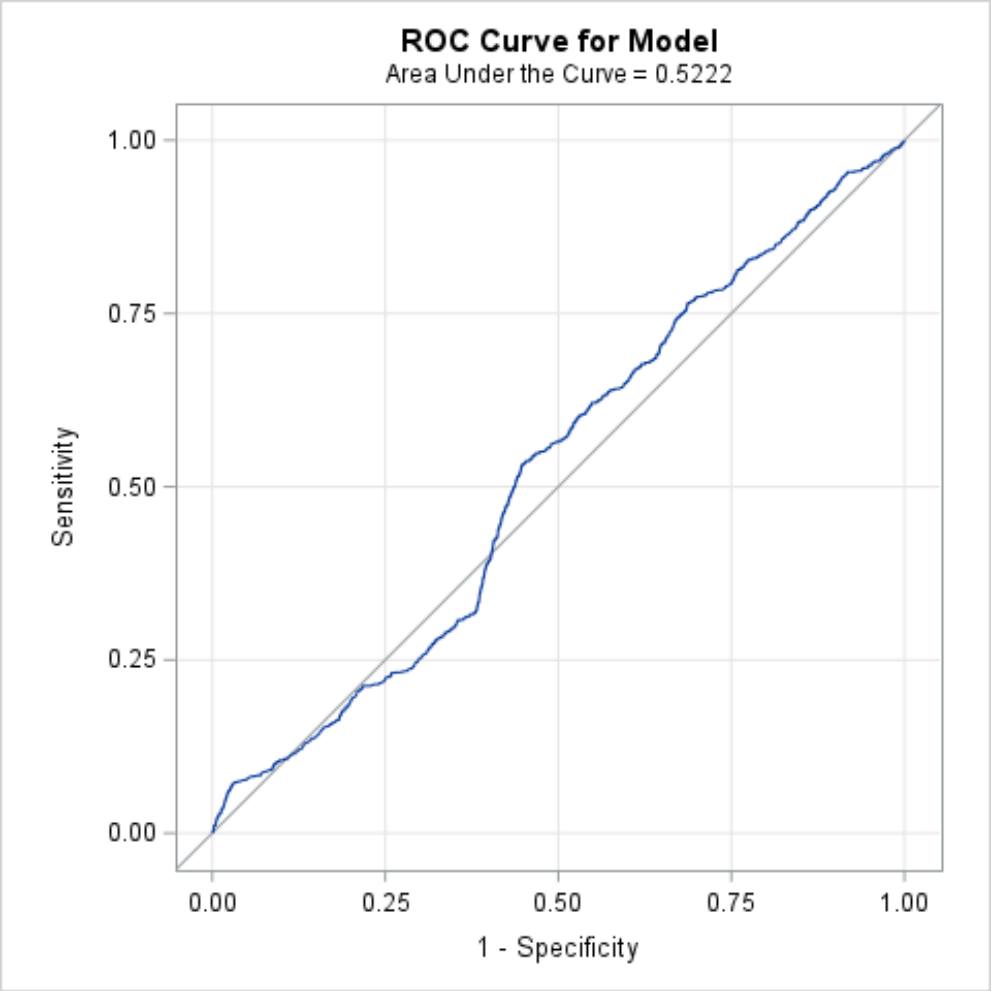


Supplementary Figure 4-9: Plot of Scottish Age-Standardised Laryngeal Cancer Incidence Trends by SIMD Quintile and 5-year period (2001-2020)

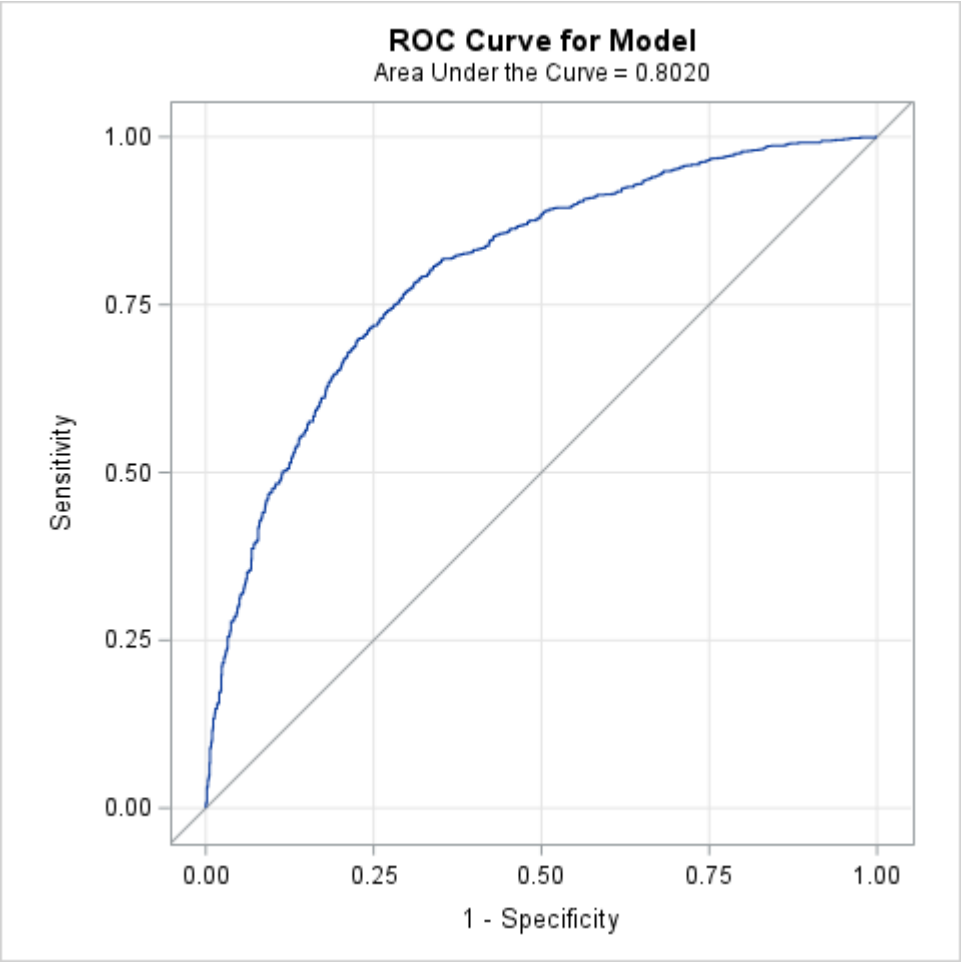
Appendix 6-1: Development and External Validation of a Head and Neck Cancer Risk Prediction Model - Supplementary Material (Chapter 6)



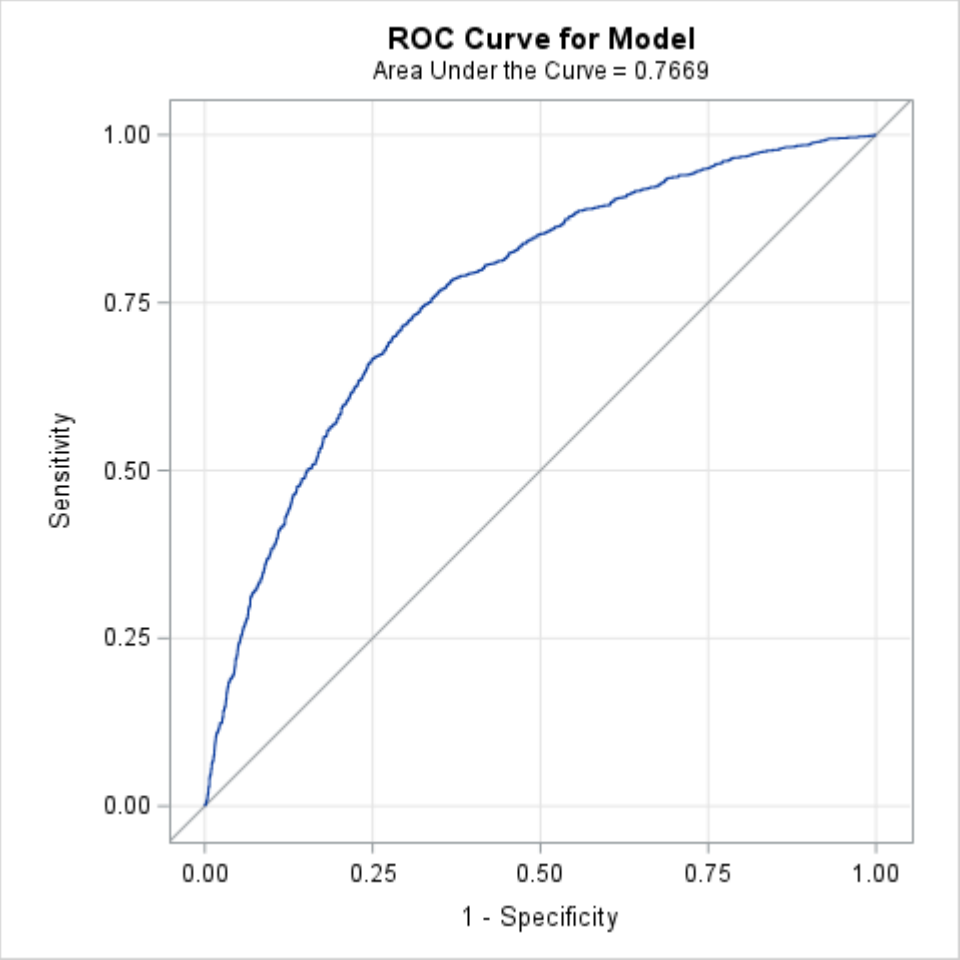
Supplementary Figure 6-1: OCC + Larynx Model Validation Result



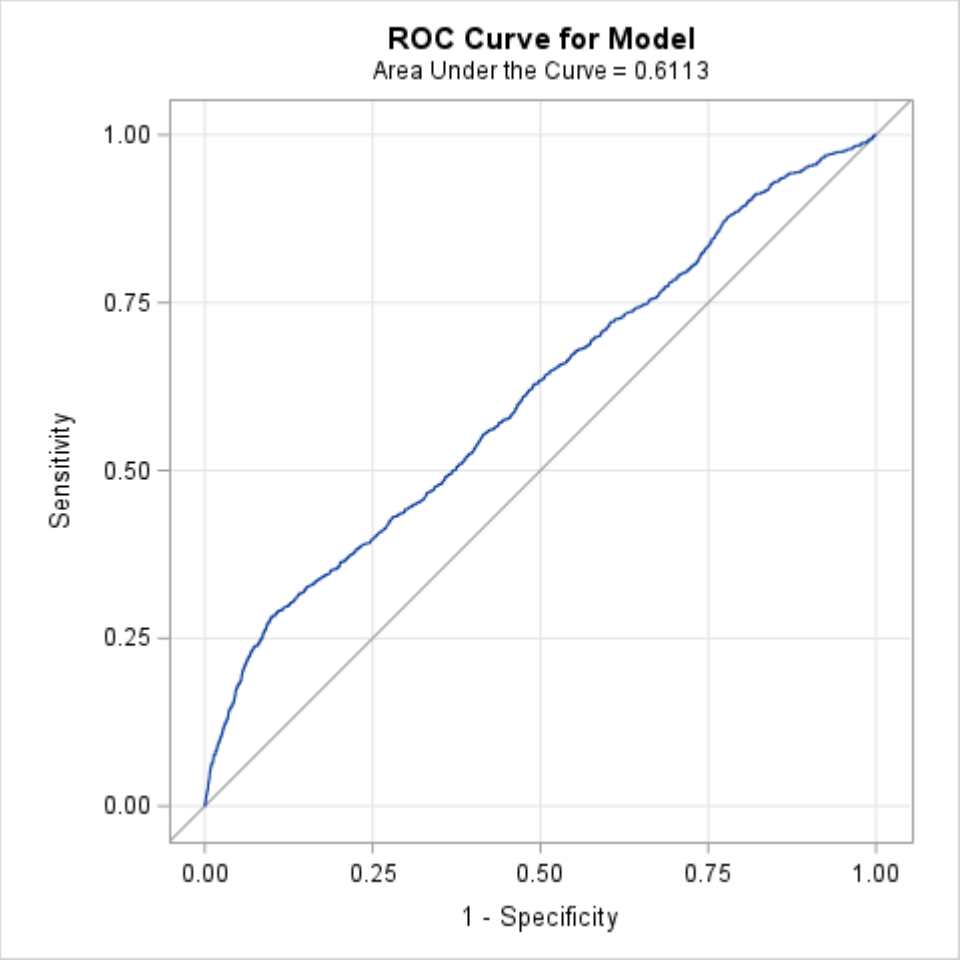
Supplementary Figure 6-2: UK-only model Validation Results



Supplementary Figure 6-3: ARCAGE HPV model development Results



Supplementary Figure 6-4: ARCAGE Denture model development Results



Supplementary Figure 6-5: Denture Model Validation results

Appendix 6-2: TRIPOD Guidelines

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	118
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	119
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	120,121
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	121
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	121-124
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	"
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	"
	5b	D;V	Describe eligibility criteria for participants.	"
	5c	D;V	Give details of treatments received, if relevant.	N.A.
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	121
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N.A.
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	121 -132
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N.A.
Sample size	8	D;V	Explain how the study size was arrived at.	121-133
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	123
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	122,123
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	122,123
	10c	V	For validation, describe how the predictions were calculated.	122-124
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	122,124
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N.A.
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N.A.
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	131,132
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	133
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	122-132
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	122-132
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	122-133
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	125-128
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	125-128
	15b	D	Explain how to the use the prediction model.	N.A.
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	122-132
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N.A.
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	138, 139
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	135
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	135-139
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	135-139
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Appendix 6-1
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	140

Appendix 8: Head and Neck Cancer Risk Prediction Tool - Feasibility Study Protocol

Research Study Protocol

Running title:	**Feasibility study of a Head and Neck Cancer Risk Prediction Tool**
Protocol Version:	**2.0**
Date:	**4/4/23**
REC Reference Number:	**UGN23ON131**
ISRCTN/Clinical trial.gov:	**N/A**
Sponsor's Protocol Number:	
Sponsor:	**NHS Greater Glasgow & Clyde**
Funder:	**Cancer Research UK [Grant number 315941-01]**

Amendment number	Date	Protocol version
1.	4/4/23	2.0

STUDY GROUP / CONTACTS

Principal Investigators

(Student)

Mr. Craig DL Smith
BDS-PhD Candidate, TRACC Programme
University of Glasgow, Dental School / Beatson Institute
378 Sauchiehall Street
Glasgow G2 3JZ
Tel: 07803 419158
E-mail: xxxxxxxx@student.gla.ac.uk

(PhD Supervisors)

Prof David I Conway
Chief Investigator, Professor / Honorary Consultant in Dental Public Health
University of Glasgow School of Medicine, Dentistry and Nursing
378 Sauchiehall Street
Glasgow G2 3JZ
Tel: 0141 211 9750
E-mail: david.conway@glasgow.ac.uk

Dr. Alex D McMahon

Reader in Epidemiology (Dental School)
University of Glasgow School of Medicine, Dentistry and Nursing
378 Sauchiehall Street
Glasgow G2 3JZ
Tel:
E-mail: Alex.McMahon@glasgow.ac.uk

Dr. Al Ross

Senior Lecturer in Human Factors in Healthcare (Dental School)
University of Glasgow School of Medicine, Dentistry and Nursing
378 Sauchiehall Street
Glasgow G2 3JZ
Tel: 0141 211 9811
E-mail: Alastair.Ross@glasgow.ac.uk

Prof Gareth J Inman

Director of Research (School of Cancer Sciences)
(Beatson Institute for Cancer Research)
Switchback Rd, Bearsden,
Glasgow G61 1BD
E-mail: Gareth.Inman@glasgow.ac.uk

Sponsor: NHS Greater Glasgow and Clyde
Sponsor Representative:
Ms. Emma-Jane Gault
Research Regulation & Compliance Manager
Room 327, Wolfson Medical School Building, University Avenue,
Glasgow, G12 8QQ
Tel: 0141 330 5519
E-mail: EmmaJane.Gault@glasgow.ac.uk

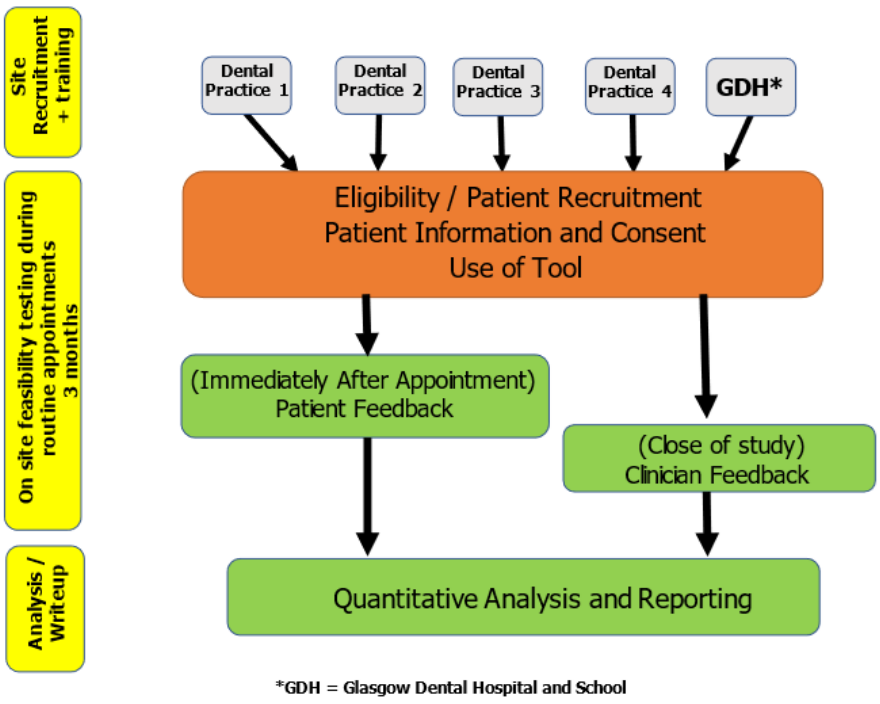
General Dental Practioner Representative (TBC)

Dr. Ailsa Creaney
GDP Representative
General Dental Practioner
Address:
Tel:
E-mail: ailsa.woodley@nhs.scot

Patient and Public Involvement Representative (PPI) (TBC)

Linda Galbraith
Patient and Public Involvement Representative (PPI)
Patient
Address: TBC
Tel: TBC
E-mail: TBC

STUDY FLOW CHART



ABBREVIATIONS

BDA	British Dental Association
COH	Community Oral Health
CI	Chief investigator
ENT	Ear Nose Throat
EPV	Events Per Variable
GDH	Glasgow Dental Hospital
GDP	General Dental Practitioner
GUSS	Glasgow University Software Services
HNC	Head and Neck Cancer
HPV	Human Papilloma Virus
NHS GG&C	Greater Glasgow and Clyde Health Board
SES	Socioeconomic Status
SIMD	Scottish Index of Multiple Deprivation
RAT	Risk Assessment Tool
REC	Research Ethics Committee
TAU	Treatment As Usual
UofG	University of Glasgow

STUDY SYNOPSIS

Title of Study:	Feasibility study of a Head and Neck Cancer Risk Prediction Tool
Study Centre:	NHS Dental Practices + Glasgow Dental Hospital and School, NHS Greater Glasgow & Clyde,
Duration of Study:	3 months feasibility + 3 months write up / synthesis
Primary Objective:	To assess the feasibility (rate of completion) of a HNC risk prediction tool in a dental setting.
Secondary Objectives:	• To gain patient and clinician feedback on the acceptability of the HNC risk prediction tool. • To assess if the HNC risk prediction tool increased patient awareness of HNC risk

	factors, promoted discussions of prevention or had any negative effects. • To assess if clinicians found the tool helpful in communicating HNC risk to patients and identify any key barriers / facilitators to its success.
Primary Endpoint:	Completion of use of the HNC risk prediction tool. This will be measured by participation completion rates.
Secondary Endpoint:	Patient and Clinician perceptions on the acceptability, the feasibility and process of the HNC risk tool, including barriers / facilitators to use / implementation. This will be measured by patient and clinician questionnaires.
Rationale:	Characterised by growing incidence and common late stage-presentation, it has been suggested that there needs to be an increased emphasis on primary prevention strategies for HNC. Risk prediction tools have been proposed to have a potential role to help clinicians identify those at risk earlier and where possible, improve prevention pathways and promote behavioural change. However, there are relatively few HNC risk models and even fewer that have seen clinical testing. There exists a gap in the literature and crucially, scope to test the feasibility of a HNC risk tool in a dental care setting where the preventative potential of a HNC risk model could be fully realised. Our team at the University of Glasgow Dental School / School of Cancer sciences have developed and validated a risk prediction tool for HNC for use in primary care settings.
Methodology:	Cross-sectional Feasibility study
Sample Size:	150-200 patients; one clinician per setting
Screening:	Dental team / clinician screening of patient lists on selected study appointment days – approval and inclusion / exclusion based on clinical or personal circumstances of patient
Registration:	All eligible patients (subject to dental practitioner approval) with routine dental examination / checkup appointments at practice research sites during course of the study will be invited to participate.
Main Inclusion Criteria:	• Adults aged 18 and above registered with a NHS general dental practice.
Main Exclusion Criteria:	• Participants unable to give consent • Participants unable to read English (and thus complete questionnaire)

	• Participants with a history of HNC, undergoing treatment for HNC or carcinoma in situ. • Patients attending for emergency dental appointments or attending for treatment.
Intervention:	Following completion of standard care, (treatment as usual – clinical oral exam and prevention advice) a non-invasive clinical risk tool will be used to assess individual patient risk of Head and Neck Cancer. Delivered **after** treatment as usual (TAU) i.e standard dental history, examination and preventative advice. Questionnaire issued at end of patient appointment and to clinicians at the end of patient recruitment.
Duration of Intervention:	• Information and informed consent process obtained before patient appointment. (10 minutes) • Brief use of risk tool after (with no change to) standard care (5 minutes at most). • Patient Questionnaire post consultation (5 minutes) in close room
Statistical Analysis:	Recruitment, completion, retention rates and confidence intervals will be calculated. Descriptive analysis of the questionnaire responses will be carried out through analysis of tables with X^2 tests and logistic or ordinal regression. Thematic analysis will be used for the open-ended / semi-structured questions in the patient and clinician questionnaires, respectively.

INTRODUCTION

1.1 Background

Head and Neck Cancer (HNC), defined as squamous cell cancers of the oral cavity, pharynx and larynx, is the 7th most common cancer globally. (Johnson et al., 2020b) With some estimated 848,000 global incident cases in 2020 alone, incidence has risen by a third over the last decade. (Bray et al., 2018) In Scotland there are approximately 1200 incident cases each year; there were 1182 incident cases recorded from April 2020 to March 2021. (Public Health Scotland, 2022) Incidence rates are especially high in the West of Scotland, where some 50% of cases are diagnosed. (Conway et al., 2006)

Major risk factors for HNC include smoking and alcohol consumption, these having a synergistic effect. (Hashibe et al., 2009) Other notable risk factors include low socioeconomic status (SES), HPV (for cancers of the oropharynx) and a diet low in

fruit and vegetables. (Conway et al., 2015, Spence et al., 2016, Chuang et al., 2012b)
Despite technological advances in treatment, there have only been marginal improvements in overall survival; most cases are detected and diagnosed at later stages where treatment is typically more complex, and subsequent outcomes are poorer. For example, analysis of the Scottish Cancer Registry from 2009-2018 revealed that 65% of HNC cases were diagnosed at advanced stages (3 or 4). (Creaney et al., 2022) Overall 5-year HNC survival in the UK is poor varying between 28-67% by subsite. (Cancer Research UK, 2016)

1.2 Rationale

Considering these challenges, it has been suggested that there needs to be an increased emphasis on primary prevention strategies for HNC. Risk prediction tools have been postulated to have a potential role to help clinicians identify those at risk earlier and where possible, improve prevention pathways and promote behavioural change. Patient-clinician interaction has been suggested to be a key component in creating "teachable moments" and promoting behavioural change. The use of a risk prediction tool provides a quantifiable output. Such an output has scope for creating opportunities for discussion and a potential "teachable moment". (Lawson and Flocke, 2009)
Non-invasive clinical risk tools such as Q-cancer or Cancer Risk Assessment Tools (RATs) have been developed for other cancers including colorectal, breast, and lung cancer. (Q-Cancer, 2017) (Cancer Research UK, 2020a) Some of these models have been evaluated extensively and seen some success. The former, for example, being integrated into the General Medical Practice management system Egton Medical Information Systems (EMIS). Such tools have been shown to be underutilised in practice but also to possess great potential in aiding decision making. (Price et al., 2019)
Despite this, there are relatively few HNC risk models and even fewer that have had clinical testing. We have previously reviewed the international literature on HNC risk prediction models. (Smith et al., 2022c) Of the few existing models, only one model developed by Tikka and ENT colleagues in Glasgow had seen any significant clinical piloting being used with good effect in a secondary care setting to triage patients during the pandemic. (Hardman et al., 2021a)
Dental care settings have inherent primary prevention and secondary preventative potential. The dental setting has been shown to be effective in various screening and preventative efforts, for example SARS-CoV-2 testing, the detection of HPV in dental settings and most famously, the world-renowned Child Smile programme. (Conway et al., 2021a, Conway et al., 2016, Turner et al., 2010)
Whilst some online resources exist to help support oral cancer screening - for example the BDA oral cancer toolkit - to our knowledge, no HNC risk tools have been utilised in dental settings. (British Dental Association, 2015) Thus, dental settings offer an untapped area of potential - there still remains a gap in the literature and crucially, scope to test the feasibility of a HNC risk model in a dental care setting where the preventative potential of a HNC risk model could be fully realised.
The risk tool, (HANS?) is a risk prediction tool developed in collaboration with Glasgow University Software Services (GUSS). The tool was derived using data captured and analysed via logistic regression from the Alcohol Related Cancers

and Genetic Susceptibility in Europe (ARCAGE) multi-centre case control study and subsequently externally validated in a large UK wide cohort study, the UK Biobank. During this process we selected various sociodemographic and behavioral variables for our final model.

None of these variables are physically invasive in nature i.e. they **do not** require examination information or biological sample collection e.g., blood or stool sample. All of these variables are routinely collected as part of the clinical history conducted by dentists as per standard clinical examination.

The tool calculates an individual's probability (or risk) of developing HNC according to the answers supplied in the data fields. This is not dissimilar other to risk tools available and used in primary care general practices (e.g Q-Cancer). Final variable selection (is to be confirmed) includes as demographic data (age, sex, SIMD) and behavioral data (alcohol and smoking behaviours). By design, the tool is simple and compact allowing for ease of use and conciseness in already busy clinical settings.

2 STUDY OBJECTIVES

Aim:
To undertake an initial feasibility study of using a head and neck cancer risk prediction tool for patients attending a clinical dental setting.

Objectives:

Primary Objective
To assess the feasibility (rate of completion) of a HNC risk prediction tool in a dental setting.

Secondary Objectives
To gain patient and clinician feedback on the acceptability of the HNC risk prediction tool.
To assess if the HNC risk prediction tool increased patient awareness of HNC risk factors, promoted discussions of prevention or had any negative effects.
To assess if clinicians found the tool helpful in communicating HNC risk to patients and identify any key barriers / facilitators to its success.

- **Primary Endpoint**
    - Completion of use of the HNC risk prediction tool. This will be measured by participation completion rates.

- **Secondary endpoints**
    - Patient and Clinician perceptions on the acceptability, the feasibility and process of the HNC risk tool, including barriers / facilitators to use / implementation. This will be measured by patient and clinician questionnaires.

3 STUDY POPULATION

Participants will be continuously recruited from general dental practices or dental hospital clinics, with a target population of 50 patients (and a minimum of 40) per

setting. One clinician in each setting will also complete feedback on the tool at the close of the study. Sites are:

- Four dental practices
- One dental hospital / secondary care setting

Participating staff will constitute of one GDP per site. One nurse per site will also be part of the study team for participant recruitment.
Practices in low or mixed socioeconomically deprived areas will be purposively selected to capture different SES risk profiles. SIMD is a common area-based multiple deprivation index, calculated from factors such as income, education, housing and local crime scores. ranging from low (1 most deprived) to high (10 least deprived).
As discussed in the background, low SES is an important predictor of HNC. For this reason, practices in low SIMD areas (SIMD 1/2/3) will be selected (for more detail please see section 3.3).

3.1 Inclusion criteria

- Eligible participants shall include adults aged 18 and above registered with a general dental practice attending for an examination / appointment.
- Patient recruitment methods should, by nature, allow for age, sex, SES and ethnic diversity during the recruitment process.

1.2 Exclusion criteria

- Participants unable (do not have capacity) to give consent shall be excluded.
- Participants with a history of HNC, undergoing treatment for HNC or carcinoma in situ shall also be excluded.
- Patients unable to read English (and thus complete questionnaire) - resources available for feasibility study preclude interpretation / translation.
- Patients attending for emergency appointments (e.g trauma, bleeding etc) or scheduled treatment
- Clinician knowledge of prohibitive personal circumstances for either clinical (e.g. medical history, adverse mental health) or personal reasons (e.g recent bereavement)

Reason for non-participation in study, if provided, will also be recorded using a Participation Log. This shall be used to collect the age / sex of excluded patients, the reason for exclusion and will be completed by a dental practitioner or a member of the research team.

3.3 Recruitment and consent of dental team

For this feasibility study, a target number of 4 practices was defined, in addition to a general clinic in Glasgow Dental Hospital. We will fully assess their suitability in relation to location, patient profile, and practical issues (such as space, spare room for questionnaire completion, sufficient throughput and diversity of patients.). Practices in low SIMD will be selected using existing research networks linked to the UofG Dental School COH department.
Participating clinicians and staff will be approached and consented prior to the start of the study. Each potential site centre and staff participant will be sent the

study information sheet and other relevant documentation before being contacted by telephone to enable discussion of the study in greater detail and confirm suitability to take part. (Please see attached 'Clinician Information Form 1.0' and 'Clinician Consent Form V1.0')
Staff will be provided relevant training and manuals, be required to undertake GCP training and attend a training session in order to be familiar with use of the risk tool, gaining informed consent and all study procedures. This will ensure participating staff are calibrated to the tool's intended use / function. (For more details on patients / staff see participants section). Clinicians will be provided standard CPD guild rates for the training session.
Clinicians or practice staff can withdraw their support from the study at any time.

3.4 Identification of (patient) participants, recruitment and consent

Participating dentists / dental nurses will screen lists of patients attending for examination / checkup appointments during the course of the study.

Eligibility will be assessed based on inclusion criteria and (where applicable) the practitioner's knowledge of any prohibitive personal circumstances (See sections 3.2 and 3.3). All eligible patients will be invited to participate on arrival. Informed verbal and written consent will be sought by NHS research staff / nurses at each study site, with patients guided through pre-defined information and consent forms explaining the research procedures in simple terms. The information and consent forms will be in English. Patients will be made aware that participation is voluntary and that they can leave the study with no effect on their care.

Patients will be asked to provide consent for the following:

(a) To attend for an appointment as normal with the use of the risk prediction tool at the end of a routine consultation.
(b) Answering the risk prediction tool questions as honestly and accurately as possible (commonly asked during routine histories e.g smoking status)
(c) To complete a brief study questionnaire at the end of the appointment, and leave this in the practice (see Appendix).
(d) To be approached with opportunities to participate in further studies

Only consent to (a) and (b) (c) are pre-requisites to participation, non-consent to (d) and do not preclude participation.

(See attached forms 'Patient Information Form v1.0' and 'Patient Consent Form V1.0').

3.5 Withdrawal of Participants

Participants, both patient and clinician can freely withdraw from the feasibility study at any point for any reason.

The investigator can also withdraw patients from the study in the event of illness, protocol violations or any other relevant reasons. If the participant consents, all data (e.g questionnaire) up to the point of withdrawal will be retained. If consent is not given to retain data in the event of withdrawal, data collected up to that point will be destroyed.

Reason for non-participation / withdrawal, if provided will also be recorded using a non-Participation Log (see section 3.2 and appendix) This would contain demographic data and reason for withdrawal / exclusion. Postcodes would be used to identify patient SIMD and then immediately deleted.

4 STUDY PROCEDURES

4.1 Proposed Overall Study schedule

- September 2023: Once ethical and management approvals have been obtained. We will look to begin participant recruitment.

- Prior to September 2023 Practice Recruitment (Identify eligible and willing sites)

- September 2023 (one week prior to beginning of study): Clinician recruitment and consent with Information / Consent forms. Training session on use of tool / research practices (See appended Information Forms Consent forms)

- September 2023 – February 2024 (latest): Patient recruitment on sites for simultaneously trialling the tool and gaining feedback in each practice using the patient questionnaire for approx. 3 months each or until sufficient collection of data has occurred. At the conclusion of the study, we will also collect feedback from clinicians using the clinician questionnaire. The details of these processes are summarised in the study flow chart.

- February 2024: Conclusion of study to allow for analysis and thesis writeup.

4.2 In Practice Procedures

Prior to appointment

- Identification of potentially eligible patients on lists attending
- Patient Participant Information form
- Consent form (written) – completed in private room

End of appointment

- Use of tool with no alteration to standard care
- Generation of unique 5-digit identifier

Following appointment

- Patient completion of Questionnaire in practice (with identifier)
- Completed questionnaire collection in practice

Close of the study

- Clinician completion of Questionnaire
- Collection of questionnaires
- Analysis + Write up

1.3 Study Outcome Measures

- Consent rates for the study
- Completeness and return rates of patient feedback questionnaire
- Completeness and return rates of clinician feedback questionnaire

- Preliminary estimates of feasibility / acceptability of a HNC cancer risk tool in primary care

1.3.1 Primary Outcome Measure

- To assess the feasibility (rate of completion) of a HNC risk prediction tool in a dental setting.

1.3.2 Secondary Outcome Measure (Questionnaire)

- Patient and clinician feedback on the acceptability of the HNC risk prediction tool.
- Patient awareness of HNC risk factors, promotion of discussions of prevention or any negative effects.
- How helpful clinicians found the tool in communicating HNC risk to patients
- Identification of any key barriers / facilitators to its success.
- Any other comments

4.4 Delivery of Risk Tool

The aim is for the tool to be delivered to patients after routine treatment and thus will not interfere or alter the clinical pathway / standard of care. Clinicians will be trained and calibrated with the tool prior to commence of the study. (See section 3.3) Use of the tool will generate a unique five-digit identifier which will correspond to each patient questionnaire (and setting). This will help provide information on setting which will aid qualitative synthesis. Data collected by the tool will be securely stored on a secure University of Glasgow One-Drive Cloud e.g. demographics and smoking status.

The study and tool are by design, not an alteration of routine clinical practice / care. Routine social history taking, discussions of behavioural change and communication of risk should be standard practice. Thus, the tool should serve as aid for clinicians and in no way alter the standard of care. If a patient were identified to be high-risk, we will encourage discussion of risk, smoking cessation, careful monitoring and earlier recall times as per good clinical practice guidelines. For example, chapters 3,6,11 and 12 of Delivering Better Oral Health. (Department of Health and Social Care, 2021a)
We have a number of points of contacts for advice for research staff, dentists (or their team), and research participants to access appropriate advice. E.g smoking cessation.

This study is a low-risk study from the point of view of patient safety, with the study design conferring no immediate risks of harm to any participants. However, as a

tool intended to evaluate and promote discussion of the risk of HNC, this may have the potential to induce some worry or anxiety. We do not anticipate that this would be over and above the typical nature of concerns that the dental team are highly skilled in managing. Staff involved in the intervention would have additional training (see section 3.3) in the use of the tool including emphasising to patients that this is a theoretical tool to predict risk and that it is not a definitive diagnostic tool in any form.

If in the event a patient is presenting with signs or symptoms or presentation suggestive of HNC, patients should be referred for urgent further investigations in secondary care within 2 weeks (USOC referral) as per standard clinical guidelines. (NICE, 2015) (NHS Scotland, 2019)

4.5 Questionnaires

The questionnaires have been developed to answer the research questions posed in the secondary objectives, assessing acceptability, clarity etc. The majority of the questions use Likert scales ranging from strongly disagree (1) to not sure (3) to strongly agree (5). The final question is an open-ended question for any further feedback and qualitative / narrative analysis.

Patient questionnaires will be filled with a corresponding 5-digit identifier to identify practices and link to risk data. These will be completed immediately after consultation so as to minimise any potential loss to follow-up. Results / feedback shall be transcribed verbatim by the research team for subsequent qualitative and thematic analysis. Completeness and return rates will be measured.

The questionnaires will include items on:

- Understanding the need for the tool
- Clarity / appropriateness of the questions / variables
- Trust in the results
- Whether the tool caused anxiety
- HNC awareness (Patient)
- Likelihood to change behaviours
- Overall Acceptability of the tool
- Strengths and Limitations
- Any further feedback

Participants will be asked to complete the questionnaires on paper forms either in the waiting area or in a private room, if desired. As per the number generated use of the tool, a unique five-digit identifier will correspond to a patient's questionnaire.

At the conclusion of the study, each clinician will also complete a separate clinician questionnaire evaluating the tool. These will also address the primary and secondary research questions posed. This questionnaire includes a mixture of Likert scales, ranging from strongly disagree (1) to not sure (3) to strongly agree (5), and semi-structured questions, culminating with an open-ended question.

Both questionnaires are included in appendix A. Synthesis will be conducted using a thematic analysis approach to identify key barriers and facilitators to implementation and strengths / weaknesses as per discussed primary and secondary objectives.

2. STATISTICS AND DATA ANALYSIS

1.1 Statistical analysis plan

Questionnaires are the main data collection method. Clinician interviews were considered for initial study proposals. However, this was ultimately reconsidered due to time, resource and staff constraints in favour of questionnaires to allow for high-quality qualitative analysis.

Data collected from questionnaires will be transcribed to Microsoft Excel / Word documents for synthesis with the aims of identifying key barriers and facilitators to implementation and strengths/weaknesses as per discussed primary and secondary objectives. Those with missing data shall be excluded unless for good reason which will be thoroughly documented.

Analysis: Confidence intervals for recruitment, completion of questionnaires and retention will be calculated using binomial proportions and the Wilson method. In small subgroups, exact methods, using the Binomial Distribution may have to be used. When comparing subgroups, rate ratios will be used with tests and confidence intervals based upon the large sample normal approximation, if appropriate, or exact small sample methods otherwise.

Descriptive analysis of the questionnaire responses will be carried out through analysis of tables with X2 tests and logistic or ordinal regression.

Thematic analysis will be used for the open-ended / semi-structured questions in the patient and clinician questionnaires, respectively.

5.2 Primary efficacy analysis

Not efficacy study but one variable explored is the proportion of patients willing to take part in the study and whether this proportion will vary across sites.

5.3 Secondary efficacy analysis

Again, this is not an efficacy study. Acceptability of the tool questionnaire will include analysis of consent rates, response completeness of individual questions, and qualitative analysis of the open-ended feedback question at the end of the questionnaire.

5.4 Software for statistical analysis

(SAS Version 9.4)

5.5 Sample size

Feasibility studies are generally used for hypothesis construction and may ultimately inform subsequent (pilot) studies which do have power calculations.

Given that feasibility studies are also largely qualitative in nature, power calculations are often deemed unnecessary. However, it was agreed that there must be a sufficient sample size to ensure there were distinct risk profiles among patients, e.g. high vs low risk, heavy smokers vs non-smokers.

For logistic regression risk model development, a value of 10 Events Per Variable (EPV) is often quoted as a rule-of-thumb value (Moons et al., 2014)This was satisfied during model development. For pilot studies, a minimum of 50 patients is a commonly quoted number, derived by Sim and colleagues (Sim and Lewis, 2012) An audit of feasibility and pilot studies in the UK found the median number of participants in a feasibility study was 36, ranging from 10 to 300 (Billingham et al., 2013).

Public Health Scotland data for NHS dental service patient registration and participation were also examined to help inform target recruitment numbers (Public Health Scotland, 2023a). According to data as of the 30th of September 2022, 95.4%% of the adult population were registered with a dentist. When measured by the Scottish Index of Multiple Deprivation (SIMD) quintile, this translates to 42.7% of people living in the most deprived areas and 53.5 % of those from the least deprived areas attending for an appointment within the prior 2 years.

In consideration of the above, (with purposive selection of practices from deprived SIMD quintiles) a target sample size of 30-40 patients per practice, was deemed sufficient to meet the study objectives. Assuming a recruitment rate of at least 5 patients per week (or one a day), this provides a projected total target sample size of 150-200 patients and 5 clinicians over the 2.5 - 3-month study period.

5.6 Management and delivery

The statistical / qualitative analysis of the study data will be led by Mr. Craig DL Smith, Prof David I Conway, Dr Alex J McMahon and Dr Al Ross in an advisory role.

6.0 STUDY CLOSURE / DEFINITION OF END OF STUDY

The study will end when the investigators agrees that one or more of the following situations applies:

i. The planned sample size has been achieved;
ii. The maximum planned duration of the study has been reached;
iii. Recruitment is so poor that completion of the study cannot reasonably be anticipated;

7. DATA HANDLING

7.1 Consent/study registration forms and Questionnaire paper forms

Data confidentiality will be safeguarded. Data collected by the risk prediction tool will be transferred securely onto a University of Glasgow Cloud as per the study Data Management Plan (sections 2, 3, 5) and DPIA.

The paper consent forms will also be the study registration form. These will be stored safely in the dental practice before being securely transferred by a member of the research team to the University of Glasgow Dental Hospital and School (Community Oral Health department), where the consent forms will be stored. The participation logs will also be analysed to assess any differential completion rates.

The paper questionnaire forms will be the primary method of data collection in this study. Data collected from questionnaires will be transcribed verbatim by a study PI with the support of a secretary as a quality check, to Microsoft excel / word documents for synthesis. There are no specific patient identifiers in analysis / synthesis. However, participants will be assigned a unique identification number for recording of questionnaire analysis/ qualitative synthesis. Data will be saved to a secure Dental Hospital J-drive.

The paper copies of questionnaires will be secured in locked drawers within the Dental Hospital. The questionnaires will contain no identifiable data. Digital Transcribed (digital) data will be stored on the Dental Hospital J drive hosted by official university server as per the COH data security protocol (for more information please see appended Data Management Plan).

Monitoring will be continuous and ongoing throughout the study. This will be conducted by study PIs. Any temporary or permanent halt to the study will be reported to the sponsor representative, in the first instance. Additionally, any deviation form protocol or delays e.g due to illness or administration delays will also be reported to the sponsor representative, for further guidance. For further details of data management and monitoring please see the relevant Data Management Plan (appendix).

7.2 Record Retention

Data will be retained and archived with the University of Glasgow for 10 years (following University of Glasgow Data and Community Oral Health Security Protocols).

8 STUDY GROUP

The study will be coordinated from the University of Glasgow Dental School COH department by the Study Management Group. The study management group comprises the Research Team, the PPI representative and the GDP advisor. The study NHS-University Research Governance Officer would be advised of any relevant information arising during the course of the study.

The role of the group is to monitor all aspects of the conduct and progress of the study, ensure that the protocol is adhered to in accordance with the principles of GCP and the relevant regulations. NHS GG&C, the study sponsor, may audit the conduct and progress of the study.

Decisions about continuation or termination of the study or any amendments to the protocol will be the responsibility of the Study Group.

9. AMENDMENTS

Change to any approved element of the study will require an amendment. Any proposed amendments will be initiated by the CI following discussion with the Study Steering Committee and any required amendment forms will be submitted to the sponsor representative, in the first instance, before being submitted for any necessary approvals.

10. ETHICAL CONSIDERATIONS

Favourable ethical opinion will be sought from an NHS research ethics committee (REC) and NHS Research & Innovation (R&I) management approval from the participating health board before participants are entered into this study. Participants will only be allowed to enter the study once they have provided fully informed and written consent.

The PI will be responsible for updating the sponsor representative and ethics committee of any new information related to the study.

11. INSURANCE AND INDEMNITY

This study has been submitted for approval and sponsorship by NHS Greater Glasgow & Clyde. NHS indemnity is provided under the Clinical Negligence and Other Risks Indemnity Scheme (CNORIS). Protocol authors substantively employed by University of Glasgow will be covered by the organisation's Clinical Trials insurance policy.

The NHS has a duty of care to patients treated, whether or not the patient is taking part in a study, and the NHS remains liable for clinical negligence and other negligent harm to patients under its duty of care.

12. FUNDING

This work was funded and supported by Cancer Research UK as part of the TRACC programme, Beatson Institute, Glasgow [Grant number 315941-01]

13. CONFLICTS OF INTEREST

None to declare.

14. DISSEMENIATION AND PUBLICATION PLANS

The methodology and findings of this work will be included in CS' PhD thesis – "development and validation of a head and neck cancer risk model". This will be published on the UofG Enlighten service

Findings will be published on the study website, a link for which will be added to the Participant Information Sheet.
In addition to research reports, we may look to disseminate our study findings via submitting papers for publication in peer review journals and present our findings at national / international academic and clinical conferences across our multidisciplinary networks.
Such work will likely also be presented at university academic or Community Oral Health departmental meetings in addition to stakeholders in the (To train and Retain Academic Cancer Clinicians) TRACC programme or CRUK.

15. REFERENCES

BILLINGHAM, S. A., WHITEHEAD, A. L. & JULIOUS, S. A. 2013. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. *BMC Med Res Methodol,* 13**,** 104.

BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin,* 68**,** 394-424.

BRITISH DENTAL ASSOCIATION. 2015 *Oral Cancer Recognition Toolkit* [Online]. Available: https://www.doctors.net.uk/eClientopen/CRUK/oral_cancer_toolkit_2015_open/home.html [Accessed].

CANCER RESEARCH UK. 2016 *Head and neck cancers survival statistics* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/survival#heading-Zero [Accessed].

CANCER RESEARCH UK. 2020. *Cancer decision support tools overview.* [Online]. Available: https://www.cancerresearchuk.org/health-professional/diagnosis/suspected-cancer-referral-best-practice/clinical-decision-support-tools-overview#CDS_Overview0 [Accessed].

CHUANG, S. C., JENAB, M., HECK, J. E., BOSETTI, C., TALAMINI, R., MATSUO, K., CASTELLSAGUE, X., FRANCESCHI, S., HERRERO, R., WINN, D. M., LA VECCHIA, C., MORGENSTERN, H., ZHANG, Z. F., LEVI, F., DAL MASO, L., KELSEY, K., MCCLEAN, M. D., VAUGHAN, T., LAZARUS, P., MUSCAT, J., RAMROTH, H., CHEN, C., SCHWARTZ, S. M., ELUF-NETO, J., HAYES, R. B., PURDUE, M., BOCCIA, S., CADONI, G., ZARIDZE, D., KOIFMAN, S., CURADO, M. P., AHRENS, W., BENHAMOU, S., MATOS, E., LAGIOU, P., SZESZENIA-DABROWSKA, N., OLSHAN, A. F., FERNANDEZ, L., MENEZES, A., AGUDO, A., DAUDT, A. W., MERLETTI, F., MACFARLANE, G. J., KJAERHEIM, K., MATES, D., HOLCATOVA, I., SCHANTZ, S., YU, G. P., SIMONATO, L., BRENNER, H., MUELLER, H., CONWAY, D. I., THOMSON, P., FABIANOVA, E., ZNAOR, A., RUDNAI, P., HEALY, C. M., FERRO, G., BRENNAN, P., BOFFETTA, P. & HASHIBE, M. 2012. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer Causes Control,* 23**,** 69-88.

CONWAY, D. I., BRENNER, D. R., MCMAHON, A. D., MACPHERSON, L. M., AGUDO, A., AHRENS, W., BOSETTI, C., BRENNER, H., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., CURIONI, O. A., DAL MASO, L., DAUDT, A. W., DE GOIS FILHO, J. F., D'SOUZA, G., EDEFONTI, V., FABIANOVA, E., FERNANDEZ, L., FRANCESCHI, S., GILLISON, M., HAYES, R. B., HEALY, C. M., HERRERO, R., HOLCATOVA, I., JAYAPRAKASH, V., KELSEY, K., KJAERHEIM, K., KOIFMAN, S., LA VECCHIA, C., LAGIOU, P., LAZARUS, P., LEVI, F., LISSOWSKA, J., LUCE, D., MACFARLANE, T. V., MATES, D., MATOS, E., MCCLEAN, M., MENEZES, A. M., MENVIELLE, G., MERLETTI, F., MORGENSTERN, H., MOYSICH, K., MÜLLER, H., MUSCAT, J., OLSHAN, A. F., PURDUE, M. P., RAMROTH, H., RICHIARDI, L., RUDNAI, P., SCHANTZ, S., SCHWARTZ, S. M., SHANGINA, O., SIMONATO, L., SMITH, E., STUCKER, I., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., THOMSON, P., VAUGHAN, T. L., WEI, Q., WINN, D. M., WUNSCH-FILHO, V., YU, G. P., ZHANG, Z. F., ZHENG, T., ZNAOR, A., BOFFETTA, P., CHUANG, S. C., GHODRAT, M., AMY LEE, Y. C., HASHIBE, M. & BRENNAN, P. 2015. Estimating and explaining the

effect of education and income on head and neck cancer risk: INHANCE consortium pooled analysis of 31 case-control studies from 27 countries. *Int J Cancer,* 136**,** 1125-39.

CONWAY, D. I., CULSHAW, S., EDWARDS, M., CLARK, C., WATLING, C., ROBERTSON, C., BRAID, R., O'KEEFE, E., MCGOLDRICK, N., BURNS, J., PROVAN, S., VANSTEENHOUSE, H., HAY, J. & GUNSON, R. 2021. SARS-CoV-2 Positivity in Asymptomatic-Screened Dental Patients. *Journal of Dental Research,* 100**,** 583-590.

CONWAY, D. I., ROBERTSON, C., GRAY, H., YOUNG, L., MCDAID, L. M., WINTER, A. J., CAMPBELL, C., PAN, J., KAVANAGH, K., KEAN, S., BHATIA, R., CUBIE, H., CLARKSON, J. E., BAGG, J., POLLOCK, K. G. & CUSCHIERI, K. 2016. Human Papilloma Virus (HPV) Oral Prevalence in Scotland (HOPSCOTCH): A Feasibility Study in Dental Settings. *PLOS ONE,* 11**,** e0165847.

CONWAY, D. I., STOCKTON, D. L., WARNAKULASURIYA, K. A. A. S., OGDEN, G. & MACPHERSON, L. M. D. 2006. Incidence of oral and oropharyngeal cancer in United Kingdom (1990–1999)—recent trends and regional variation. *Oral Oncology,* 42**,** 586-592.

CREANEY, G., MCMAHON, A. D., ROSS, A. J., BHATTI, L. A., PATERSON, C. & CONWAY, D. I. 2022. Head and neck cancer in the UK: what was the stage before COVID-19? UK cancer registries analysis (2011-2018). *British Dental Journal,* 233**,** 787-793.

DEPARTMENT OF HEALTH AND SOCIAL CARE 2021. Delivering better oral health: an evidence-based toolkit for prevention. *In:* CARE, D. O. H. A. S. (ed.).

HARDMAN, J. C., TIKKA, T. & PALERI, V. 2021. Remote triage incorporating symptom-based risk stratification for suspected head and neck cancer referrals: A prospective population-based study. *Cancer,* 127**,** 4177-4189.

HASHIBE, M., BRENNAN, P., CHUANG, S. C., BOCCIA, S., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., DAL MASO, L., DAUDT, A. W., FABIANOVA, E., FERNANDEZ, L., WÜNSCH-FILHO, V., FRANCESCHI, S., HAYES, R. B., HERRERO, R., KELSEY, K., KOIFMAN, S., LA VECCHIA, C., LAZARUS, P., LEVI, F., LENCE, J. J., MATES, D., MATOS, E., MENEZES, A., MCCLEAN, M. D., MUSCAT, J., ELUF-NETO, J., OLSHAN, A. F., PURDUE, M., RUDNAI, P., SCHWARTZ, S. M., SMITH, E., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., WEI, Q., WINN, D. M., SHANGINA, O., PILARSKA, A., ZHANG, Z. F., FERRO, G., BERTHILLER, J. & BOFFETTA, P. 2009. 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,* 18**,** 541-50.

IARC, I. A. F. R. O. C. 2020. *Cancer Today - Estimated number of new cases in 2020, worldwide, both sexes, all ages, Estimated number of deaths in 2020, worldwide, both sexes, all ages,* [Online]. Available: https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1

https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=1&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1 [Accessed].

JOHNSON, D. E., BURTNESS, B., LEEMANS, C. R., LUI, V. W. Y., BAUMAN, J. E. & GRANDIS, J. R. 2020. Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers,* 6**,** 92.

LAWSON, P. J. & FLOCKE, S. A. 2009. Teachable moments for health behavior change: a concept analysis. *Patient Educ Couns,* 76**,** 25-30.

MOONS, K. G., DE GROOT, J. A., BOUWMEESTER, W., VERGOUWE, Y., MALLETT, S., ALTMAN, D. G., REITSMA, J. B. & COLLINS, G. S. 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med,* 11**,** e1001744.

NHS SCOTLAND 2019. Scottish Referral Guidelines for Suspected Cancer.

NICE, T. N. I. F. H. A. C. E. 2015. Suspected cancer: recognition

and referral
PRICE, S., SPENCER, A., MEDINA-LARA, A. & HAMILTON, W. 2019. Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care. *British Journal of General Practice,* 69**,** e437-e443.
PUBLIC HEALTH SCOTLAND 2022. Head and neck cancer Quality Performance Indicators.
PUBLIC HEALTH SCOTLAND 2023. Dental statistics - registration and participation

NHS registration and participation.
Q-CANCER. 2017. *Q-Cancer* [Online]. Available: https://www.qcancer.org/ [Accessed].
SIM, J. & LEWIS, M. 2012. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *Journal of Clinical Epidemiology,* 65**,** 301-308.
SMITH, C. D. L., MCMAHON, A. D., ROSS, A., INMAN, G. J. & CONWAY, D. I. 2022. Risk prediction models for head and neck cancer: A rapid review. *Laryngoscope Investigative Otolaryngology,* n/a.
SPENCE, T., BRUCE, J., YIP, K. W. & LIU, F. F. 2016. HPV Associated Head and Neck Cancer. *Cancers (Basel),* 8.
TURNER, S., BREWSTER, L., KIDD, J., GNICH, W., BALL, G. E., MILBURN, K., PITTS, N. B., GOOLD, S., CONWAY, D. I. & MACPHERSON, L. M. 2010. Childsmile: the national child oral health improvement programme in Scotland. Part 2: Monitoring and delivery. *Br Dent J,* 209**,** 79-83.

List of References

ABRAHÃO, R., ANANTHARAMAN, D., GABORIEAU, V., ABEDI-ARDEKANI, B., LAGIOU, P., LAGIOU, A., AHRENS, W., HOLCATOVA, I., BETKA, J., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., SERRAINO, D., POLESEL, J., SIMONATO, L., ALEMANY, L., AGUDO TRIGUEROS, A., MACFARLANE, T. V., MACFARLANE, G. J., ZNAOR, A., ROBINSON, M., CANOVA, C., CONWAY, D. I., WRIGHT, S., HEALY, C. M., TONER, M., CADONI, G., BOCCIA, S., GHEIT, T., TOMMASINO, M., SCELO, G. & BRENNAN, P. 2018. The influence of smoking, age and stage at diagnosis on the survival after larynx, hypopharynx and oral cavity cancers in Europe: The ARCAGE study. *Int J Cancer,* 143**,** 32-44.

ABRAHÃO, R., PERDOMO, S., PINTO, L. F. R., NASCIMENTO DE CARVALHO, F., DIAS, F. L., DE PODESTÁ, J. R. V., VENTORIN VON ZEIDLER, S., MARINHO DE ABREU, P., VILENSKY, M., GIGLIO, R. E., OLIVEIRA, J. C., MINEIRO, M. S., KOWALSKI, L. P., IKEDA, M. K., CUELLO, M., MUNYO, A., RODRÍGUEZ-URREGO, P. A., HAKIM, J. A., SUAREZ-ZAMORA, D. A., CAYOL, F., FIGARI, M. F., OLIVER, J., GABORIEAU, V., KEOGH, R. H., BRENNAN, P. & CURADO, M. P. 2020. Predictors of Survival After Head and Neck Squamous Cell Carcinoma in South America: The InterCHANGE Study. *JCO Glob Oncol,* 6**,** 486-499.

AHRENS, W., POHLABELN, H., FORAITA, R., NELIS, M., LAGIOU, P., LAGIOU, A., BOUCHARDY, C., SLAMOVA, A., SCHEJBALOVA, M., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., AGUDO, A., CASTELLSAGUE, X., MACFARLANE, T. V., MACFARLANE, G. J., LEE, Y. C., TALAMINI, R., BARZAN, L., CANOVA, C., SIMONATO, L., THOMSON, P., MCKINNEY, P. A., MCMAHON, A. D., ZNAOR, A., HEALY, C. M., MCCARTAN, B. E., METSPALU, A., MARRON, M., HASHIBE, M., CONWAY, D. I. & BRENNAN, P. 2014. Oral health, dental care and mouthwash associated with upper aerodigestive tract cancer risk in Europe: the ARCAGE study. *Oral Oncol,* 50**,** 616-25.

AKANUWE, J. N. A., BLACK, S., OWEN, S. & SIRIWARDENA, A. N. 2021. Barriers and facilitators to implementing a cancer risk assessment tool (QCancer) in primary care: a qualitative study. *Primary health care research & development,* 22**,** e51-e51.

AL-HELOU, N. 2021. The extra oral and intra oral examination. *BDJ Team,* 8**,** 20-22.

ALTMAN, D. G. & ROYSTON, P. 2000. What do we mean by validating a prognostic model? *Stat Med,* 19**,** 453-73.

ALY, F., HANSEN, C. R., AL MOUIEE, D., SUNDARESAN, P., HAIDAR, A., VINOD, S. & HOLLOWAY, L. 2023. Outcome prediction models incorporating clinical variables for Head and Neck Squamous cell Carcinoma: A systematic review of methodological conduct and risk of bias. *Radiotherapy and Oncology,* 183**,** 109629.

AMARASINGHE, H. K., JOHNSON, N. W., LALLOO, R., KUMARAARACHCHI, M. & WARNAKULASURIYA, S. 2010. Derivation and validation of a risk-factor model for detection of oral potentially malignant disorders in populations with high prevalence. *British Journal of Cancer,* 103(3)**,** 303-309.

ANANTHARAMAN, D., ABEDI-ARDEKANI, B., BEACHLER, D. C., GHEIT, T., OLSHAN, A. F., WISNIEWSKI, K., WUNSCH-FILHO, V., TOPORCOV, T. N., TAJARA, E. H., LEVI, J. E., MOYSES, R. A., BOCCIA, S., CADONI, G., RINDI, G., AHRENS, W., MERLETTI, F., CONWAY, D. I., WRIGHT, S., CARREIRA, C., RENARD, H., CHOPARD, P., MCKAY-CHOPIN, S., SCELO, G., TOMMASINO, M., BRENNAN, P. & D'SOUZA, G. 2017. Geographic heterogeneity in the prevalence of human papillomavirus in head and neck cancer. *Int J Cancer,* 140**,** 1968-1975.

ANANTHARAMAN, D., GHEIT, T., WATERBOER, T., ABEDI-ARDEKANI, B., CARREIRA, C., MCKAY-CHOPIN, S., GABORIEAU, V., MARRON, M., LAGIOU, P., AHRENS, W., HOLCÁTOVÁ, I., MERLETTI, F., KJAERHEIM, K., TALAMINI, R., SIMONATO, L., CASTELLSAGUE, X., MACFARLANE, T. V., BIGGS, A.-M., THAKKER, N., ZNAOR, A., THOMSON, P., CANOVA, C., CONWAY, D. I., HEALY, C. M., TOMMASINO, M., PAWLITA, M. & BRENNAN, P. 2013.

Human Papillomavirus Infections and Upper Aero-Digestive Tract Cancers: The ARCAGE Study. *JNCI: Journal of the National Cancer Institute,* 105**,** 536-545.

ANANTHARAMAN, D., MARRON, M., LAGIOU, P., SAMOLI, E., AHRENS, W., POHLABELN, H., SLAMOVA, A., SCHEJBALOVA, M., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., CASTELLSAGUE, X., AGUDO, A., TALAMINI, R., BARZAN, L., MACFARLANE, T. V., TICKLE, M., SIMONATO, L., CANOVA, C., CONWAY, D. I., MCKINNEY, P. A., THOMSON, P., ZNAOR, A., HEALY, C. M., MCCARTAN, B. E., HASHIBE, M., BRENNAN, P. & MACFARLANE, G. J. 2011. Population attributable risk of tobacco and alcohol for upper aerodigestive tract cancer. *Oral Oncol,* 47**,** 725-31.

ANANTHARAMAN, D., MULLER, D. C., LAGIOU, P., AHRENS, W., HOLCÁTOVÁ, I., MERLETTI, F., KJÆRHEIM, K., POLESEL, J., SIMONATO, L., CANOVA, C., CASTELLSAGUE, X., MACFARLANE, T. V., ZNAOR, A., THOMSON, P., ROBINSON, M., CONWAY, D. I., HEALY, C. M., TJØNNELAND, A., WESTIN, U., EKSTRÖM, J., CHANG-CLAUDE, J., KAAKS, R., OVERVAD, K., DROGAN, D., HALLMANS, G., LAURELL, G., BUENO-DE-MESQUITA, H., PEETERS, P. H., AGUDO, A., LARRAÑAGA, N., TRAVIS, R. C., PALLI, D., BARRICARTE, A., TRICHOPOULOU, A., GEORGE, S., TRICHOPOULOS, D., QUIRÓS, J. R., GRIONI, S., SACERDOTE, C., NAVARRO, C., SÁNCHEZ, M.-J., TUMINO, R., SEVERI, G., BOUTRON-RUAULT, M.-C., CLAVEL-CHAPELON, F., PANICO, S., WEIDERPASS, E., LUND, E., GRAM, I. T., RIBOLI, E., PAWLITA, M., WATERBOER, T., KREIMER, A. R., JOHANSSON, M. & BRENNAN, P. 2016a. Combined effects of smoking and HPV16 in oropharyngeal cancer. *International Journal of Epidemiology,* 45**,** 752-761.

ANANTHARAMAN, D., MULLER, D. C., LAGIOU, P., AHRENS, W., HOLCÁTOVÁ, I., MERLETTI, F., KJÆRHEIM, K., POLESEL, J., SIMONATO, L., CANOVA, C., CASTELLSAGUE, X., MACFARLANE, T. V., ZNAOR, A., THOMSON, P., ROBINSON, M., CONWAY, D. I., HEALY, C. M., TJØNNELAND, A., WESTIN, U., EKSTRÖM, J., CHANG-CLAUDE, J., KAAKS, R., OVERVAD, K., DROGAN, D., HALLMANS, G., LAURELL, G., BUENO-DE-MESQUITA, H. B., PEETERS, P. H., AGUDO, A., LARRAÑAGA, N., TRAVIS, R. C., PALLI, D., BARRICARTE, A., TRICHOPOULOU, A., GEORGE, S., TRICHOPOULOS, D., QUIRÓS, J. R., GRIONI, S., SACERDOTE, C., NAVARRO, C., SÁNCHEZ, M. J., TUMINO, R., SEVERI, G., BOUTRON-RUAULT, M. C., CLAVEL-CHAPELON, F., PANICO, S., WEIDERPASS, E., LUND, E., GRAM, I. T., RIBOLI, E., PAWLITA, M., WATERBOER, T., KREIMER, A. R., JOHANSSON, M. & BRENNAN, P. 2016b. Combined effects of smoking and HPV16 in oropharyngeal cancer. *Int J Epidemiol,* 45**,** 752-61.

ANG, K. K., HARRIS, J., WHEELER, R., WEBER, R., ROSENTHAL, D. I., NGUYEN-TÂN, P. F., WESTRA, W. H., CHUNG, C. H., JORDAN, R. C., LU, C., KIM, H., AXELROD, R., SILVERMAN, C. C., REDMOND, K. P. & GILLISON, M. L. 2010. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med,* 363**,** 24-35.

ANOTHAISINTAWEE, T., TEERAWATTANANON, Y., WIRATKAPUN, C., KASAMESUP, V. & THAKKINSTIAN, A. 2012. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Res Treat,* 133**,** 1-10.

ARGIRION, I., ZARINS, K. R., DEFEVER, K., SUWANRUNGRUANG, K., CHANG, J. T., PONGNIKORN, D., CHITAPANARUX, I., SRIPLUNG, H., VATANASAPT, P. & ROZEK, L. S. 2019. Temporal Changes in Head and Neck Cancer Incidence in Thailand Suggest Changing Oropharyngeal Epidemiology in the Region. *J Glob Oncol,* 5**,** 1-11.

BAAN, R., STRAIF, K., GROSSE, Y., SECRETAN, B., EL GHISSASSI, F., BOUVARD, V., ALTIERI, A. & COGLIANO, V. 2007. Carcinogenicity of alcoholic beverages. *The Lancet Oncology,* 8**,** 292-293.

BAI, X., CUI, C., YIN, J., LI, H., GONG, Q., WEI, B. & LU, Y. 2023. The association between oral hygiene and head and neck cancer: a meta-analysis. *Acta Odontologica Scandinavica,* 81**,** 374-395.

BAKER, E. H. 2014. Socioeconomic Status, Definition. *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society.*

BANERJEE, S., PARVIN, L., RAMA, S. P. K. & VOLIOTIDIS, D. 2021. Telephone triage of suspected head and neck cancer patients during the coronavirus disease 2019 pandemic using the

Head and Neck Cancer Risk Calculator version 2. *The Journal of Laryngology & Otology,* 135**,** 241-245.

BECK, T. N. & GOLEMIS, E. A. 2016. Genomic insights into head and neck cancer. *Cancers of the Head & Neck,* 1**,** 1.

BENARD, V. B., JOHNSON, C. J., THOMPSON, T. D., ROLAND, K. B., LAI, S. M., COKKINIDES, V., TANGKA, F., HAWKINS, N. A., LAWSON, H. & WEIR, H. K. 2008. Examining the association between socioeconomic status and potential human papillomavirus-associated cancers. *Cancer,* 113**,** 2910-2918.

BENNETT, D. A. 2001. How can I deal with missing data in my study? *Aust N Z J Public Health,* 25**,** 464-9.

BERENSON, A. B., HIRTH, J. M. & CHANG, M. 2021. Prevalence of Genital Human Papillomavirus by Age and Race/Ethnicity Among Males. *Clin Infect Dis,* 73**,** 1625-1633.

BERTHILLER, J., STRAIF, K., AGUDO, A., AHRENS, W., BEZERRA DOS SANTOS, A., BOCCIA, S., CADONI, G., CANOVA, C., CASTELLSAGUE, X., CHEN, C., CONWAY, D., CURADO, M. P., DAL MASO, L., DAUDT, A. W., FABIANOVA, E., FERNANDEZ, L., FRANCESCHI, S., FUKUYAMA, E. E., HAYES, R. B., HEALY, C., HERRERO, R., HOLCATOVA, I., KELSEY, K., KJAERHEIM, K., KOIFMAN, S., LAGIOU, P., LA VECCHIA, C., LAZARUS, P., LEVI, F., LISSOWSKA, J., MACFARLANE, T., MATES, D., MCCLEAN, M., MENEZES, A., MERLETTI, F., MORGENSTERN, H., MUSCAT, J., OLSHAN, A. F., PURDUE, M., RAMROTH, H., RUDNAI, P., SCHWARTZ, S. M., SERRAINO, D., SHANGINA, O., SMITH, E., STURGIS, E. M., SZESZENIA-DABROWSKA, N., THOMSON, P., VAUGHAN, T. L., VILENSKY, M., WEI, Q., WINN, D. M., WÜNSCH-FILHO, V., ZHANG, Z. F., ZNAOR, A., FERRO, G., BRENNAN, P., BOFFETTA, P., HASHIBE, M. & LEE, Y. C. 2016. Low frequency of cigarette smoking and the risk of head and neck cancer in the INHANCE consortium pooled analysis. *Int J Epidemiol,* 45**,** 835-45.

BHASKARAN, K., DOUGLAS, I., FORBES, H., DOS-SANTOS-SILVA, I., LEON, D. A. & SMEETH, L. 2014. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet,* 384**,** 755-65.

BIBBINS-DOMINGO, K., HELMAN A 2022. *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups.*, National Academies Press (US).

BIEN, T. H., MILLER, W. R. & TONIGAN, J. S. 1993. Brief interventions for alcohol problems: a review. *Addiction,* 88**,** 315-336.

BILLINGHAM, S. A., WHITEHEAD, A. L. & JULIOUS, S. A. 2013. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. *BMC Med Res Methodol,* 13**,** 104.

BOCCIA, S., HASHIBE, M., GALLÌ, P., DE FEO, E., ASAKAGE, T., HASHIMOTO, T., HIRAKI, A., KATOH, T., NOMURA, T., YOKOYAMA, A., VAN DUIJN, C. M., RICCIARDI, G. & BOFFETTA, P. 2009. Aldehyde dehydrogenase 2 and head and neck cancer: a meta-analysis implementing a Mendelian randomization approach. *Cancer Epidemiol Biomarkers Prev,* 18**,** 248-54.

BOUVARD, V., NETHAN, S. T., SINGH, D., WARNAKULASURIYA, S., MEHROTRA, R., CHATURVEDI, A. K., CHEN, T. H.-H., AYO-YUSUF, O. A., GUPTA, P. C., KERR, A. R., TILAKARATNE, W. M., ANANTHARAMAN, D., CONWAY, D. I., GILLENWATER, A., JOHNSON, N. W., KOWALSKI, L. P., LEON, M. E., MANDRIK, O., NAGAO, T., PRASAD, V. M., RAMADAS, K., ROITBERG, F., SAINTIGNY, P., SANKARANARAYANAN, R., SANTOS-SILVA, A. R., SINHA, D. N., VATANASAPT, P., ZAIN, R. B. & LAUBY-SECRETAN, B. 2022. IARC Perspective on Oral Cancer Prevention. *New England Journal of Medicine,* 387**,** 1999-2005.

BRADFORD, D. R. R., ALLIK, M., MCMAHON, A. D. & BROWN, D. 2023. Assessing the risk of endogeneity bias in health and mortality inequalities research using composite measures of multiple deprivation which include health-related indicators: A case study using the Scottish Index of Multiple Deprivation and population health and mortality data. *Health & Place,* 80**,** 102998.

BRAVI, F., EDEFONTI, V., RANDI, G., FERRARONI, M., LA VECCHIA, C. & DECARLI, A. 2012. Dietary patterns and upper aerodigestive tract cancers: an overview and review. *Annals of Oncology,* 23**,** 3024-3039.

BRAVI, F., LEE, Y. A., HASHIBE, M., BOFFETTA, P., CONWAY, D. I., FERRARONI, M., LA VECCHIA, C. & EDEFONTI, V. 2021. Lessons learned from the INHANCE consortium: An overview of recent results on head and neck cancer. *Oral Dis,* 27**,** 73-93.

BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians,* 68**,** 394-424.

BRENNER, H., STOCK, C. & HOFFMEISTER, M. 2014. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ : British Medical Journal,* 348**,** g2467.

BRENNER, N., MENTZER, A. J., HILL, M., ALMOND, R., ALLEN, N., PAWLITA, M. & WATERBOER, T. 2020. Characterization of human papillomavirus (HPV) 16 E6 seropositive individuals without HPV-associated malignancies after 10 years of follow-up in the UK Biobank. *EBioMedicine,* 62**,** 103123.

BREWSTER, D., CRICHTON, J. & MUIR, C. 1994. How accurate are Scottish cancer registration data? *British Journal of Cancer,* 70**,** 954-959.

BREWSTER, D. H., BHATTI, L. A., INGLIS, J. H. C., NAIRN, E. R. & DOHERTY, V. R. 2007. Recent trends in incidence of nonmelanoma skin cancers in the East of Scotland, 1992–2003. *British Journal of Dermatology,* 156**,** 1295-1300.

BREWSTER, D. H., STOCKTON, D., HARVEY, J. & MACKAY, M. 2002. Reliability of cancer registration data in Scotland, 1997. *European Journal of Cancer,* 38**,** 414-417.

BRITISH DENTAL ASSOCIATION. 2015 *Oral Cancer Recognition Toolkit* [Online]. Available: https://www.doctors.net.uk/eClientopen/CRUK/oral_cancer_toolkit_2015_open/home.html [Accessed].

BROCKLEHURST, P., KUJAN, O., O'MALLEY, L., OGDEN, G. R., SHEPHERD, S. & GLENNY, A. M. 2013. Screening programmes for the early detection and prevention of oral cancer. *Cochrane Database of Systematic Reviews*.

BROWN, D., CONWAY, D. I., MCMAHON, A. D., DUNDAS, R. & LEYLAND, A. H. 2021. Cancer mortality 1981-2016 and contribution of specific cancers to current socioeconomic inequalities in all cancer mortality: A population-based study. *Cancer Epidemiol,* 74**,** 102010.

BUCK, C. B., DAY, P. M. & TRUS, B. L. 2013. The papillomavirus major capsid protein L1. *Virology,* 445**,** 169-74.

BUDHATHOKI1, S., , B. D., , G. L., 4, , A. O., , A. N., 7, , T., WATERBOER8, , S. V., , P. B., NOEMI BENDER8, , N. B., , T. D., , HAYES11, N., , A. H., , S. H. H., , K. H., , B. K., , MCKAY9, J. D., , M. P., , S. T., , K. W., SERA THOMAS1, , Y., BRHANE1, , A. A., , W. A., , L. A., , A. L., , L. B., , CRISTINA CANOVA20, D. I. C., , C. M. H., , I. H., , P. L., , R., LOWRY25, , G. J. M., , T. V. M., , J. P., , L. R., , M., ROBINSON29, , A. Z. & , P. B. A. R. J. H., 4 2021. A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV

serology and genetic markers.

BUDHATHOKI, S., DIERGAARDE, B., LIU, G., OLSHAN, A., NESS, A., WATERBOER, T., VIRANI, S., BASTA, P., BENDER, N., BRENNER, N., DUDDING, T., HAYES, N., HOPE, A., HUANG, S. H., HUENIKEN, K., KANTEREWICZ, B., MCKAY, J. D., PRING, M., THOMAS, S., WISNIEWSKI, K., THOMAS, S., BRHANE, Y., AGUDO, A., ALEMANY, L., LAGIOU, A., BARZAN, L., CANOVA, C., CONWAY, D. I., HEALY, C. M., HOLCATOVA, I., LAGIOU, P., MACFARLANE, G. J., MACFARLANE, T. V., POLESEL, J., RICHIARDI, L., ROBINSON, M., ZNAOR, A., BRENNAN, P. & HUNG, R. J. 2023. A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV serology and genetic markers. *International Journal of Cancer,* 152**,** 2069-2080.

BURROWS, K. & HAYCOCK, P. 2021. Genome-wide Association Study of Cancer Risk in UK Biobank. *University of Bristol*.

BUSCH, C.-J., HOFFMANN, A. S., VIARISIO, D., BECKER, B. T., RIECKMANN, T., BETZ, C., BENDER, N., SCHROEDER, L., HUSSEIN, Y., PETERSEN, E., JAGODZINSKI, A., SCHÄFER, I., BURANDT, E.,

LANG KUHS, K., PAWLITA, M., WATERBOER, T. & BRENNER, N. 2022a. Detection of stage I HPV-driven oropharyngeal cancer in asymptomatic individuals in the Hamburg City Health Study using HPV16 E6 serology – A proof-of-concept study. *eClinicalMedicine,* 53.

BUSCH, C. J., HOFFMANN, A. S., VIARISIO, D., BECKER, B. T., RIECKMANN, T., BETZ, C., BENDER, N., SCHROEDER, L., HUSSEIN, Y., PETERSEN, E., JAGODZINSKI, A., SCHÄFER, I., BURANDT, E., LANG KUHS, K., PAWLITA, M., WATERBOER, T. & BRENNER, N. 2022b. Detection of stage I HPV-driven oropharyngeal cancer in asymptomatic individuals in the Hamburg City Health Study using HPV16 E6 serology - A proof-of-concept study. *EClinicalMedicine,* 53**,** 101659.

CABINET SECRETARY FOR NHS RECOVERY, H. A. S. C. 2021. The Scottish Health Survey Main Report

CADONI, G., BOCCIA, S., PETRELLI, L., DI GIANNANTONIO, P., ARZANI, D., GIORGIO, A., DE FEO, E., PANDOLFINI, M., GALLÌ, P., PALUDETTI, G. & RICCIARDI, G. 2012. A review of genetic epidemiology of head and neck cancer related to polymorphisms in metabolic genes, cell cycle control and alcohol metabolism. *Acta otorhinolaryngologica Italica : organo ufficiale della Societa italiana di otorinolaringologia e chirurgia cervico-facciale,* 32**,** 1-11.

CANCER RESEARCH UK. *Head and neck cancers statistics* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers#heading-One [Accessed 28/11/23 2023].

CANCER RESEARCH UK. 13th July 2020. *Cancer decision support tools overview* [Online]. Available: https://www.cancerresearchuk.org/health-professional/diagnosis/suspected-cancer-referral-best-practice/clinical-decision-support-tools-overview#CDS_Overview0 [Accessed].

CANCER RESEARCH UK. 2016 *Head and neck cancers survival statistics* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/survival#heading-Zero [Accessed].

CANCER RESEARCH UK. 2020a. *Cancer decision support tools overview.* [Online]. Available: https://www.cancerresearchuk.org/health-professional/diagnosis/suspected-cancer-referral-best-practice/clinical-decision-support-tools-overview#CDS_Overview0 [Accessed].

CANCER RESEARCH UK 2020b. Head and neck cancers incidence statistics.

CANCER RESEARCH UK. 2021. *Cancer incidence for common cancers - Incidence trends over time for common cancers* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Three [Accessed].

CANCER RESEARCH UK. 2022 *Head and neck cancers mortality statistics - Head and Neck Cancer (C00-C14, C30-C32), Annual Average Number of Deaths, Crude and European Age-Standardised (AS) Mortality Rates per 100,000 Persons Population, UK, 2017-2019* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/mortality#heading-Zero [Accessed].

CARLANDER, A. F., JAKOBSEN, K. K., BENDTSEN, S. K., GARSET-ZAMANI, M., LYNGGAARD, C. D., JENSEN, J. S., GRØNHØJ, C. & BUCHWALD, C. V. 2021. A Contemporary Systematic Review on Repartition of HPV-Positivity in Oropharyngeal Cancer Worldwide. *Viruses,* 13.

CARVALHO, A. L., NISHIMOTO, I. N., CALIFANO, J. A. & KOWALSKI, L. P. 2005. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer,* 114**,** 806-16.

CAVERS, D., NELSON, M., ROSTRON, J., ROBB, K. A., BROWN, L. R., CAMPBELL, C., AKRAM, A. R., DICKIE, G., MACKEAN, M., VAN BEEK, E. J. R., SULLIVAN, F., STEELE, R. J., NEILSON, A. R. & WELLER, D. 2022. Optimizing the implementation of lung cancer screening in Scotland: Focus group participant perspectives in the LUNGSCOT study. *Health Expect,* 25**,** 3246-3258.

CENTERS FOR DISEASE CONTROL AND PREVENTION. 2020a. *Head and Neck Cancers* [Online]. Available: https://www.cdc.gov/cancer/headneck/index.htm [Accessed].

CENTERS FOR DISEASE CONTROL AND PREVENTION, D. O. C. P. A. C. 2020b. *Head and Neck Cancers* [Online]. Available: https://www.cdc.gov/cancer/headneck/index.htm [Accessed].

CENTRES FOR DISEASE CONTROL AND PREVENTION (CDC), E. M., MD, MPH; JULIANNE GEE, MPH; ELIZABETH UNGER, PHD, MD; AND LAURI MARKOWITZ, MD 2021. Human Papillomavirus.

CHANCELLOR, J. A., IOANNIDES, S. J. & ELWOOD, J. M. 2017. Oral and oropharyngeal cancer and the role of sexual behaviour: a systematic review. *Community Dent Oral Epidemiol,* 45**,** 20-34.

CHANG, C.-P., CHANG, S.-C., CHUANG, S.-C., BERTHILLER, J., FERRO, G., MATSUO, K., WÜNSCH-FILHO, V., TOPORCOV, T. N., DE CARVALHO, M. B., LA VECCHIA, C., OLSHAN, A. F., ZEVALLOS, J. P., SERRAINO, D., MUSCAT, J., STURGIS, E. M., LI, G., MORGENSTERN, H., LEVI, F., DAL MASO, L., SMITH, E., KELSEY, K., MCCLEAN, M., VAUGHAN, T. L., LAZARUS, P., RAMROTH, H., CHEN, C., SCHWARTZ, S. M., WINN, D. M., BOSETTI, C., EDEFONTI, V., GARAVELLO, W., NEGRI, E., HAYES, R. B., PURDUE, M. P., BOCCIA, S., CADONI, G., SHANGINA, O., KOIFMAN, R., CURADO, M. P., VILENSKY, M., SWIATKOWSKA, B., HERRERO, R., FRANCESCHI, S., BENHAMOU, S., FERNANDEZ, L., MENEZES, A. M. B., DAUDT, A. W., MATES, D., SCHANTZ, S., YU, G.-P., LISSOWSKA, J., BRENNER, H., FABIANOVA, E., RUDNAI, P., BRENNAN, P., BOFFETTA, P., ZHANG, Z.-F., HASHIBE, M. & LEE, Y.-C. A. 2019. Age at start of using tobacco on the risk of head and neck cancer: Pooled analysis in the International Head and Neck Cancer Epidemiology Consortium (INHANCE). *Cancer Epidemiology,* 63**,** 101615.

CHATURVEDI, A. K. 2012. Epidemiology and clinical aspects of HPV in head and neck cancers. *Head and neck pathology,* 6 Suppl 1**,** S16-S24.

CHATURVEDI, A. K., ANDERSON, W. F., LORTET-TIEULENT, J., CURADO, M. P., FERLAY, J., FRANCESCHI, S., ROSENBERG, P. S., BRAY, F. & GILLISON, M. L. 2013. Worldwide Trends in Incidence Rates for Oral Cavity and Oropharyngeal Cancers. *Journal of Clinical Oncology,* 31**,** 4550-4559.

CHELIMO, C., WOULDES, T. A., CAMERON, L. D. & ELWOOD, J. M. 2013. Risk factors for and prevention of human papillomaviruses (HPV), genital warts and cervical cancer. *Journal of Infection,* 66**,** 207-217.

CHEN, F., LIN, L., YAN, L., LIU, F., QIU, Y., WANG, J., HU, Z., WU, J., BAO, X., LIN, L., WANG, R., CAI, G., AOYAGI, K., CAI, L. & HE, B. 2018. Nomograms and risk scores for predicting the risk of oral cancer in different sexes: a large-scale case-control study. *Journal of Cancer,* 9**,** 2543-2548.

CHEN, Y., LEE, Y.-C. A., LI, S., LI, Q., CHEN, C.-J., HSU, W.-L., LOU, P.-J., ZHU, C., PAN, J., SHEN, H., MA, H., CAI, L., HE, B., WANG, Y., ZHOU, X., JI, Q., ZHOU, B., WU, W., MA, J., KAWAKITA, D., BOFFETTA, P., ZHANG, Z.-F., DAI, M. & HASHIBE, M. 2019. Body mass index and the risk of head and neck cancer in the Chinese population. *Cancer Epidemiology,* 60**,** 208-215.

CHENG, L., WANG, Y. & DU, J. 2020. Human Papillomavirus Vaccines: An Updated Review. *Vaccines (Basel),* 8.

CHEUNG, L. C., RAMADAS, K., MUWONGE, R., KATKI, H. A., THOMAS, G., GRAUBARD, B. I., BASU, P., SANKARANARAYANAN, R., SOMANATHAN, T. & CHATURVEDI, A. K. 2021. Risk-Based Selection of Individuals for Oral Cancer Screening. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 39(6)**,** 663-674.

CHOI, S.-Y., CHEONG, H.-K., LEE, M.-K., KANG, J.-W., LEE, Y.-C., OH, I.-H. & EUN, Y.-G. 2022. Metabolic Diseases and Risk of Head and Neck Cancer: A Cohort Study Analyzing Nationwide Population-Based Data. *Cancers,* 14**,** 3277.

CHUANG, S.-C., JENAB, M., HECK, J. E., BOSETTI, C., TALAMINI, R., MATSUO, K., CASTELLSAGUE, X., FRANCESCHI, S., HERRERO, R., WINN, D. M., LA VECCHIA, C., MORGENSTERN, H., ZHANG, Z.-F., LEVI, F., DAL MASO, L., KELSEY, K., MCCLEAN, M. D., VAUGHAN, T., LAZARUS, P., MUSCAT, J., RAMROTH, H., CHEN, C., SCHWARTZ, S. M., ELUF-NETO, J., HAYES, R. B., PURDUE, M., BOCCIA, S., CADONI, G., ZARIDZE, D., KOIFMAN, S., CURADO, M. P., AHRENS, W., BENHAMOU, S., MATOS, E., LAGIOU, P., SZESZENIA-DABROWSKA, N., OLSHAN, A. F.,

FERNANDEZ, L., MENEZES, A., AGUDO, A., DAUDT, A. W., MERLETTI, F., MACFARLANE, G. J., KJAERHEIM, K., MATES, D., HOLCATOVA, I., SCHANTZ, S., YU, G.-P., SIMONATO, L., BRENNER, H., MUELLER, H., CONWAY, D. I., THOMSON, P., FABIANOVA, E., ZNAOR, A., RUDNAI, P., HEALY, C. M., FERRO, G., BRENNAN, P., BOFFETTA, P. & HASHIBE, M. 2012a. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer causes & control : CCC,* 23**,** 69-88.

CHUANG, S. C., JENAB, M., HECK, J. E., BOSETTI, C., TALAMINI, R., MATSUO, K., CASTELLSAGUE, X., FRANCESCHI, S., HERRERO, R., WINN, D. M., LA VECCHIA, C., MORGENSTERN, H., ZHANG, Z. F., LEVI, F., DAL MASO, L., KELSEY, K., MCCLEAN, M. D., VAUGHAN, T., LAZARUS, P., MUSCAT, J., RAMROTH, H., CHEN, C., SCHWARTZ, S. M., ELUF-NETO, J., HAYES, R. B., PURDUE, M., BOCCIA, S., CADONI, G., ZARIDZE, D., KOIFMAN, S., CURADO, M. P., AHRENS, W., BENHAMOU, S., MATOS, E., LAGIOU, P., SZESZENIA-DABROWSKA, N., OLSHAN, A. F., FERNANDEZ, L., MENEZES, A., AGUDO, A., DAUDT, A. W., MERLETTI, F., MACFARLANE, G. J., KJAERHEIM, K., MATES, D., HOLCATOVA, I., SCHANTZ, S., YU, G. P., SIMONATO, L., BRENNER, H., MUELLER, H., CONWAY, D. I., THOMSON, P., FABIANOVA, E., ZNAOR, A., RUDNAI, P., HEALY, C. M., FERRO, G., BRENNAN, P., BOFFETTA, P. & HASHIBE, M. 2012b. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer Causes Control,* 23**,** 69-88.

CHUNG, Y. K., KIM, M. K., KIM, J., LEE, D. O., LEE, C. W., LIM, M. C., OH, H. Y., PARK, S. & SEO, S. 2015. Alcohol consumption and persistent infection of high-risk human papillomavirus. *Epidemiology and Infection,* 143**,** 1442-1450.

CLARKSON, J. E., PITTS, N. B., FEE, P. A., GOULAO, B., BOYERS, D., RAMSAY, C. R., FLOATE, R., BRAID, H. J., ORD, F. S., WORTHINGTON, H. V., VAN DER POL, M., YOUNG, L., FREEMAN, R., GOUICK, J., HUMPHRIS, G. M., MITCHELL, F. E., MCDONALD, A. M., NORRIE, J. D. T., SIM, K., DOUGLAS, G. & RICKETTS, D. 2021. Examining the effectiveness of different dental recall strategies on maintenance of optimum oral health: the INTERVAL dental recalls randomised controlled trial. *British Dental Journal,* 230**,** 236-243.

CLIN RISK QCancer

CLIN RISK 2013. QCancer

CLINTON, S. K., GIOVANNUCCI, E. L. & HURSTING, S. D. 2020. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *The Journal of Nutrition,* 150**,** 663-671.

COGLIANO, V. J., BAAN, R., STRAIF, K., GROSSE, Y., LAUBY-SECRETAN, B., EL GHISSASSI, F., BOUVARD, V., BENBRAHIM-TALLAA, L., GUHA, N., FREEMAN, C., GALICHET, L. & WILD, C. P. 2011. Preventable exposures associated with human cancers. *J Natl Cancer Inst,* 103**,** 1827-39.

COLLINS, G. S. & ALTMAN, D. G. 2013. Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer(®) (Pancreas). *Br J Gen Pract,* 63**,** e636-42.

COLLINS, G. S., DE GROOT, J. A., DUTTON, S., OMAR, O., SHANYINDE, M., TAJAR, A., VOYSEY, M., WHARTON, R., YU, L.-M., MOONS, K. G. & ALTMAN, D. G. 2014. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Medical Research Methodology,* 14**,** 40.

COLLINS, G. S., REITSMA, J. B., ALTMAN, D. G. & MOONS, K. G. 2015a. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Cancer,* 112**,** 251-9.

COLLINS, G. S., REITSMA, J. B., ALTMAN, D. G. & MOONS, K. G. M. 2015b. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine,* 13**,** 1.

CONROY, M. C., LACEY, B., BEŠEVIĆ, J., OMIYALE, W., FENG, Q., EFFINGHAM, M., SELLERS, J., SHEARD, S., PANCHOLI, M., GREGORY, G., BUSBY, J., COLLINS, R. & ALLEN, N. E. 2022. UK Biobank: a globally important resource for cancer research. *British Journal of Cancer*.

CONWAY, D. I., ALEX D. MCMAHON, DENISE BROWN & LEYLAND., A. H. 2019a. *Reducing social inequalities in cancer: evidence and priorities for research. Chapter 4: Reducing social inequalities in cancer: evidence and priorities for research.*, IARC Scientific Publications,.

CONWAY, D. I., BRENNER, D. R., MCMAHON, A. D., MACPHERSON, L. M., AGUDO, A., AHRENS, W., BOSETTI, C., BRENNER, H., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., CURIONI, O. A., DAL MASO, L., DAUDT, A. W., DE GOIS FILHO, J. F., D'SOUZA, G., EDEFONTI, V., FABIANOVA, E., FERNANDEZ, L., FRANCESCHI, S., GILLISON, M., HAYES, R. B., HEALY, C. M., HERRERO, R., HOLCATOVA, I., JAYAPRAKASH, V., KELSEY, K., KJAERHEIM, K., KOIFMAN, S., LA VECCHIA, C., LAGIOU, P., LAZARUS, P., LEVI, F., LISSOWSKA, J., LUCE, D., MACFARLANE, T. V., MATES, D., MATOS, E., MCCLEAN, M., MENEZES, A. M., MENVIELLE, G., MERLETTI, F., MORGENSTERN, H., MOYSICH, K., MÜLLER, H., MUSCAT, J., OLSHAN, A. F., PURDUE, M. P., RAMROTH, H., RICHIARDI, L., RUDNAI, P., SCHANTZ, S., SCHWARTZ, S. M., SHANGINA, O., SIMONATO, L., SMITH, E., STUCKER, I., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., THOMSON, P., VAUGHAN, T. L., WEI, Q., WINN, D. M., WUNSCH-FILHO, V., YU, G. P., ZHANG, Z. F., ZHENG, T., ZNAOR, A., BOFFETTA, P., CHUANG, S. C., GHODRAT, M., AMY LEE, Y. C., HASHIBE, M. & BRENNAN, P. 2015. Estimating and explaining the effect of education and income on head and neck cancer risk: INHANCE consortium pooled analysis of 31 case-control studies from 27 countries. *Int J Cancer,* 136**,** 1125-39.

CONWAY, D. I., CULSHAW, S., EDWARDS, M., CLARK, C., WATLING, C., ROBERTSON, C., BRAID, R., O'KEEFE, E., MCGOLDRICK, N., BURNS, J., PROVAN, S., VANSTEENHOUSE, H., HAY, J. & GUNSON, R. 2021a. SARS-CoV-2 Positivity in Asymptomatic-Screened Dental Patients. *Journal of Dental Research,* 100**,** 583-590.

CONWAY, D. I., HOVANEC, J., AHRENS, W., ROSS, A., HOLCATOVA, I., LAGIOU, P., SERRAINO, D., CANOVA, C., RICHIARDI, L., HEALY, C., KJAERHEIM, K., MACFARLANE, G. J., THOMSON, P., AGUDO, A., ZNAOR, A., BRENNAN, P., LUCE, D., MENVIELLE, G., STUCKER, I., BENHAMOU, S., RAMROTH, H., BOFFETTA, P., VILENSKY, M., FERNANDEZ, L., CURADO, M. P., MENEZES, A., DAUDT, A., KOIFMAN, R., WUNSCH-FILHO, V., YUAN-CHIN, A. L., HASHIBE, M., BEHRENS, T. & MCMAHON, A. D. 2021b. Occupational socioeconomic risk associations for head and neck cancer in Europe and South America: individual participant data analysis of pooled case–control studies within the INHANCE Consortium. *Journal of Epidemiology and Community Health,* 75**,** 779-787.

CONWAY, D. I., MCMAHON, A. D., BROWN, D. & LEYLAND, A. H. 2019b. *Measuring socioeconomic status and inequalities*, IARC-WHO.

CONWAY, D. I., MCMAHON, A. D., SMITH, K., BLACK, R., ROBERTSON, G., DEVINE, J. & MCKINNEY, P. A. 2010. Components of socioeconomic risk associated with head and neck cancer: a population-based case-control study in Scotland. *Br J Oral Maxillofac Surg,* 48**,** 11-7.

CONWAY, D. I., PETTICREW, M., MARLBOROUGH, H., BERTHILLER, J., HASHIBE, M. & MACPHERSON, L. M. D. 2008. Socioeconomic inequalities and oral cancer risk: A systematic review and meta-analysis of case-control studies. *International Journal of Cancer,* 122**,** 2811-2819.

CONWAY, D. I., PURKAYASTHA, M. & CHESTNUTT, I. G. 2018a. The changing epidemiology of oral cancer: definitions, trends, and risk factors. *British Dental Journal,* 225**,** 867-873.

CONWAY, D. I., PURKAYASTHA, M. & CHESTNUTT, I. G. 2018b. The changing epidemiology of oral cancer: definitions, trends, and risk factors. *Br Dent J,* 225**,** 867-873.

CONWAY, D. I., ROBERTSON, C., GRAY, H., YOUNG, L., MCDAID, L. M., WINTER, A. J., CAMPBELL, C., PAN, J., KAVANAGH, K., KEAN, S., BHATIA, R., CUBIE, H., CLARKSON, J. E., BAGG, J., POLLOCK, K. G. & CUSCHIERI, K. 2016. Human Papilloma Virus (HPV) Oral Prevalence in Scotland (HOPSCOTCH): A Feasibility Study in Dental Settings. *PLOS ONE,* 11**,** e0165847.

CONWAY, D. I., STOCKTON, D. L., WARNAKULASURIYA, K. A. A. S., OGDEN, G. & MACPHERSON, L. M. D. 2006. Incidence of oral and oropharyngeal cancer in United Kingdom (1990–1999)—recent trends and regional variation. *Oral Oncology,* 42**,** 586-592.

CREANEY, G., MCMAHON, A. D., ROSS, A. J., BHATTI, L. A., PATERSON, C. & CONWAY, D. I. 2022. Head and neck cancer in the UK: what was the stage before COVID-19? UK cancer registries analysis (2011-2018). *British Dental Journal,* 233**,** 787-793.

CUSCHIERI, K. S., CUBIE, H. A., WHITLEY, M. W., GILKISON, G., ARENDS, M. J., GRAHAM, C. & MCGOOGAN, E. 2005. Persistent high risk HPV infection associated with development of cervical neoplasia in a prospective population study. *Journal of Clinical Pathology,* 58**,** 946-950.

CUYPERS, M., CAIRNS, D. & ROBB, K. A. 2024. Identifying the deficits in cancer care for people with intellectual disabilities. *BMJ Oncology,* 3**,** e000171.

D'SOUZA, G., KREIMER, A. R., VISCIDI, R., PAWLITA, M., FAKHRY, C., KOCH, W. M., WESTRA, W. H. & GILLISON, M. L. 2007. Case–Control Study of Human Papillomavirus and Oropharyngeal Cancer. *New England Journal of Medicine,* 356**,** 1944-1956.

DAHLSTROM, K. R., BELL, D., HANBY, D., LI, G., WANG, L.-E., WEI, Q., WILLIAMS, M. D. & STURGIS, E. M. 2015a. Socioeconomic characteristics of patients with oropharyngeal carcinoma according to tumor HPV status, patient smoking status, and sexual behavior. *Oral Oncology,* 51**,** 832-838.

DAHLSTROM, K. R., BELL, D., HANBY, D., LI, G., WANG, L. E., WEI, Q., WILLIAMS, M. D. & STURGIS, E. M. 2015b. Socioeconomic characteristics of patients with oropharyngeal carcinoma according to tumor HPV status, patient smoking status, and sexual behavior. *Oral Oncol,* 51**,** 832-8.

DAL MASO, L., TORELLI, N., BIANCOTTO, E., DI MASO, M., GINI, A., FRANCHIN, G., LEVI, F., LA VECCHIA, C., SERRAINO, D. & POLESEL, J. 2016. Combined effect of tobacco smoking and alcohol drinking in the risk of head and neck cancers: a re-analysis of case–control studies using bi-dimensional spline models. *European Journal of Epidemiology,* 31**,** 385-393.

DARAEI, P. & MOORE, C. E. 2015. Racial Disparity Among the Head and Neck Cancer Population. *J Cancer Educ,* 30**,** 546-51.

DAVEY SMITH, G., HOFMAN, A. & BRENNAN, P. 2023. Chance, ignorance, and the paradoxes of cancer: Richard Peto on developing preventative strategies under uncertainty. *European Journal of Epidemiology,* 38**,** 1227-1237.

DAVID W. HOSMER JR., S. L., RODNEY X. STURDIVANT 2013. *Applied Logistic Regression, Third Edition*, Wiley.

DE SANJOSE, S., QUINT, W. G. V., ALEMANY, L., GERAETS, D. T., KLAUSTERMEIER, J. E., LLOVERAS, B., TOUS, S., FELIX, A., BRAVO, L. E., SHIN, H.-R., VALLEJOS, C. S., DE RUIZ, P. A., LIMA, M. A., GUIMERA, N., CLAVERO, O., ALEJO, M., LLOMBART-BOSCH, A., CHENG-YANG, C., TATTI, S. A., KASAMATSU, E., ILJAZOVIC, E., ODIDA, M., PRADO, R., SEOUD, M., GRCE, M., USUBUTUN, A., JAIN, A., SUAREZ, G. A. H., LOMBARDI, L. E., BANJO, A., MENÉNDEZ, C., DOMINGO, E. J., VELASCO, J., NESSA, A., CHICHAREON, S. C. B., QIAO, Y. L., LERMA, E., GARLAND, S. M., SASAGAWA, T., FERRERA, A., HAMMOUDA, D., MARIANI, L., PELAYO, A., STEINER, I., OLIVA, E., MEIJER, C. J. L. M., AL-JASSAR, W. F., CRUZ, E., WRIGHT, T. C., PURAS, A., LLAVE, C. L., TZARDI, M., AGORASTOS, T., GARCIA-BARRIOLA, V., CLAVEL, C., ORDI, J., ANDÚJAR, M., CASTELLSAGUÉ, X., SÁNCHEZ, G. I., NOWAKOWSKI, A. M., BORNSTEIN, J., MUÑOZ, N. & BOSCH, F. X. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *The Lancet Oncology,* 11**,** 1048-1056.

DE VITO, R., LEE, Y. C. A., PARPINEL, M., SERRAINO, D., OLSHAN, A. F., ZEVALLOS, J. P., LEVI, F., ZHANG, Z. F., MORGENSTERN, H., GARAVELLO, W., KELSEY, K., MCCLEAN, M., SCHANTZ, S., YU, G. P., BOFFETTA, P., CHUANG, S. C., HASHIBE, M., LA VECCHIA, C., PARMIGIANI, G. & EDEFONTI, V. 2019. Shared and Study-specific Dietary Patterns and Head and Neck Cancer Risk in an International Consortium. *Epidemiology (Cambridge, Mass.),* 30**,** 93-102.

DEPARTMENT OF HEALTH AND SOCIAL CARE 2021a. Delivering better oral health: an evidence-based toolkit for prevention. *In:* CARE, D. O. H. A. S. (ed.).

DEPARTMENT OF HEALTH AND SOCIAL CARE 2021b. Delivering better oral health: an evidence based toolkit for prevention - Chapter 11: smoking and tobacco use

DEPARTMENT OF HEALTH AND SOCIAL CARE 2021c. Delivering better oral health: an evidence based toolkit for prevention - Chapter 12: Alcohol

DEPARTMENT OF HEALTH AND SOCIAL CARE 2023. Green Book Chapter 18a Human Pappiloma Virus (HPV). *In:* CARE, D. O. H. A. S. (ed.). UK Government.

DESCHLER, D. G., RICHMON, J. D., KHARIWALA, S. S., FERRIS, R. L. & WANG, M. B. 2014a. The "new" head and neck cancer patient-young, nonsmoker, nondrinker, and HPV positive: evaluation. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery,* 151**,** 375-380.

DESCHLER, D. G., RICHMON, J. D., KHARIWALA, S. S., FERRIS, R. L. & WANG, M. B. 2014b. The "New" Head and Neck Cancer Patient—Young, Nonsmoker, Nondrinker, and HPV Positive:Evaluation. *Otolaryngology–Head and Neck Surgery,* 151**,** 375-380.

DHULL, A. K., ATRI, R., DHANKHAR, R., CHAUHAN, A. K. & KAUSHAL, V. 2018. Major Risk Factors in Head and Neck Cancer: A Retrospective Analysis of 12-Year Experiences. *World journal of oncology,* 9**,** 80-84.

DI CREDICO, G., POLESEL, J., DAL MASO, L., PAULI, F., TORELLI, N., LUCE, D., RADOÏ, L., MATSUO, K., SERRAINO, D., BRENNAN, P., HOLCATOVA, I., AHRENS, W., LAGIOU, P., CANOVA, C., RICHIARDI, L., HEALY, C. M., KJAERHEIM, K., CONWAY, D. I., MACFARLANE, G. J., THOMSON, P., AGUDO, A., ZNAOR, A., FRANCESCHI, S., HERRERO, R., TOPORCOV, T. N., MOYSES, R. A., MUSCAT, J., NEGRI, E., VILENSKY, M., FERNANDEZ, L., CURADO, M. P., MENEZES, A., DAUDT, A. W., KOIFMAN, R., WUNSCH-FILHO, V., OLSHAN, A. F., ZEVALLOS, J. P., STURGIS, E. M., LI, G., LEVI, F., ZHANG, Z. F., MORGENSTERN, H., SMITH, E., LAZARUS, P., LA VECCHIA, C., GARAVELLO, W., CHEN, C., SCHWARTZ, S. M., ZHENG, T., VAUGHAN, T. L., KELSEY, K., MCCLEAN, M., BENHAMOU, S., HAYES, R. B., PURDUE, M. P., GILLISON, M., SCHANTZ, S., YU, G. P., CHUANG, S. C., BOFFETTA, P., HASHIBE, M., YUAN-CHIN, A. L. & EDEFONTI, V. 2020. Alcohol drinking and head and neck cancer risk: the joint effect of intensity and duration. *Br J Cancer,* 123**,** 1456-1463.

DONÀ, M. G., ROLLO, F., PICHI, B., SPRIANO, G., MORETTO, S., COVELLO, R., PELLINI, R. & BENEVOLO, M. 2020. Evolving Profile of HPV-Driven Oropharyngeal Squamous Cell Carcinoma in a National Cancer Institute in Italy: A 10-Year Retrospective Study. *Microorganisms,* 8.

DOROBISZ, K., DOROBISZ, T. & ZATOŃSKI, T. 2023. The Microbiome's Influence on Head and Neck Cancers. *Curr Oncol Rep,* 25**,** 163-171.

DURRANT, F. G., GUTIERREZ III, J. A., NGUYEN, S. A., NATHAN, C.-A. O. & NEWMAN, J. G. 2024. Sexual history of patients with human papillomavirus positive and negative oropharyngeal cancer: A systematic review and meta-analysis. *Head & Neck,* n/a.

EDEFONTI, V., HASHIBE, M., AMBROGI, F., PARPINEL, M., BRAVI, F., TALAMINI, R., LEVI, F., YU, G., MORGENSTERN, H., KELSEY, K., MCCLEAN, M., SCHANTZ, S., ZHANG, Z., CHUANG, S., BOFFETTA, P., LA VECCHIA, C. & DECARLI, A. 2012. Nutrient-based dietary patterns and the risk of head and neck cancer: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. *Annals of Oncology,* 23**,** 1869-1880.

ELREFAEY, S., MASSARO, M. A., CHIOCCA, S., CHIESA, F. & ANSARIN, M. 2014. HPV in oropharyngeal cancer: the basics to know in clinical practice. *Acta Otorhinolaryngol Ital,* 34**,** 299-309.

ESPRESSIVO, A., PAN, Z. S., USHER-SMITH, J. A. & HARRISON, H. 2024. Risk Prediction Models for Oral Cancer: A Systematic Review. *Cancers,* 16**,** 617.

EVANS, D., MILLS, I., BURNS, L., BRYCE, M. & HANKS, S. 2023. The dental workforce recruitment and retention crisis in the UK. *Br Dent J,* 234**,** 573-577.

FAKHRY, C., ZHANG, Q., NGUYEN-TAN, P. F., ROSENTHAL, D., EL-NAGGAR, A., GARDEN, A. S., SOULIERES, D., TROTTI, A., AVIZONIS, V., RIDGE, J. A., HARRIS, J., LE, Q. T. & GILLISON, M. 2014. Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma. *J Clin Oncol,* 32**,** 3365-73.

FALCARO, M., CASTAÑON, A., NDLELA, B., CHECCHI, M., SOLDAN, K., LOPEZ-BERNAL, J., ELLISS-BROOKES, L. & SASIENI, P. 2021. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. *The Lancet,* 398**,** 2084-2092.

FANG, Z., ROSSATO, S. L., HANG, D., KHANDPUR, N., WANG, K., LO, C.-H., WILLETT, W. C., GIOVANNUCCI, E. L. & SONG, M. 2024. Association of ultra-processed food consumption

with all cause and cause specific mortality: population based cohort study. *BMJ,* 385**,** e078476.

FITZMAURICE, C., ALLEN, C., BARBER, R. M., BARREGARD, L., BHUTTA, Z. A., BRENNER, H., DICKER, D. J., CHIMED-ORCHIR, O., DANDONA, R., DANDONA, L., FLEMING, T., FOROUZANFAR, M. H., HANCOCK, J., HAY, R. J., HUNTER-MERRILL, R., HUYNH, C., HOSGOOD, H. D., JOHNSON, C. O., JONAS, J. B., KHUBCHANDANI, J., KUMAR, G. A., KUTZ, M., LAN, Q., LARSON, H. J., LIANG, X., LIM, S. S., LOPEZ, A. D., MACINTYRE, M. F., MARCZAK, L., MARQUEZ, N., MOKDAD, A. H., PINHO, C., POURMALEK, F., SALOMON, J. A., SANABRIA, J. R., SANDAR, L., SARTORIUS, B., SCHWARTZ, S. M., SHACKELFORD, K. A., SHIBUYA, K., STANAWAY, J., STEINER, C., SUN, J., TAKAHASHI, K., VOLLSET, S. E., VOS, T., WAGNER, J. A., WANG, H., WESTERMAN, R., ZEEB, H., ZOECKLER, L., ABD-ALLAH, F., AHMED, M. B., ALABED, S., ALAM, N. K., ALDHAHRI, S. F., ALEM, G., ALEMAYOHU, M. A., ALI, R., AL-RADDADI, R., AMARE, A., AMOAKO, Y., ARTAMAN, A., ASAYESH, H., ATNAFU, N., AWASTHI, A., SALEEM, H. B., BARAC, A., BEDI, N., BENSENOR, I., BERHANE, A., BERNABÉ, E., BETSU, B., BINAGWAHO, A., BONEYA, D., CAMPOS-NONATO, I., CASTAÑEDA-ORJUELA, C., CATALÁ-LÓPEZ, F., CHIANG, P., CHIBUEZE, C., CHITHEER, A., CHOI, J. Y., COWIE, B., DAMTEW, S., DAS NEVES, J., DEY, S., DHARMARATNE, S., DHILLON, P., DING, E., DRISCOLL, T., EKWUEME, D., ENDRIES, A. Y., FARVID, M., FARZADFAR, F., FERNANDES, J., FISCHER, F., TT, G. H., GEBRU, A., GOPALANI, S., HAILU, A., et al. 2017. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol,* 3**,** 524-548.

FLOR, L. S., REITSMA, M. B., GUPTA, V., NG, M. & GAKIDOU, E. 2021. The effects of tobacco control policies on global smoking prevalence. *Nature Medicine,* 27**,** 239-243.

FRAZER, K., CALLINAN, J. E., MCHUGH, J., VAN BAARSEL, S., CLARKE, A., DOHERTY, K. & KELLEHER, C. 2016. Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database of Systematic Reviews*.

FREEDMAN, A. N., SEMINARA, D., GAIL, M. H., HARTGE, P., COLDITZ, G. A., BALLARD-BARBASH, R. & PFEIFFER, R. M. 2005. Cancer Risk Prediction Models: A Workshop on Development, Evaluation, and Application. *JNCI: Journal of the National Cancer Institute,* 97**,** 715-723.

FREEDMAN, N. D., PARK, Y., SUBAR, A. F., HOLLENBECK, A. R., LEITZMANN, M. F., SCHATZKIN, A. & ABNET, C. C. 2008. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. *Int J Cancer,* 122**,** 2330-6.

FRY, A., LITTLEJOHNS, T. J., SUDLOW, C., DOHERTY, N., ADAMSKA, L., SPROSEN, T., COLLINS, R. & ALLEN, N. E. 2017. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology,* 186**,** 1026-1034.

FUNDAKOWSKI, C. E. 2020. *Head and neck cancer: psychological and psychosocial effects*, Springer.

GAPSTUR, S. M., BOUVARD, V., NETHAN, S. T., FREUDENHEIM, J. L., ABNET, C. C., ENGLISH, D. R., REHM, J., BALBO, S., BUYKX, P., CRABB, D., CONWAY, D. I., ISLAMI, F., LACHENMEIER, D. W., MCGLYNN, K. A., SALASPURO, M., SAWADA, N., TERRY, M. B., TOPORCOV, T. & LAUBY-SECRETAN, B. 2023. The IARC Perspective on Alcohol Reduction or Cessation and Cancer Risk. *New England Journal of Medicine,* 389**,** 2486-2494.

GAUDET, M. M., OLSHAN, A. F., CHUANG, S.-C., BERTHILLER, J., ZHANG, Z.-F., LISSOWSKA, J., ZARIDZE, D., WINN, D. M., WEI, Q., TALAMINI, R., SZESZENIA-DABROWSKA, N., STURGIS, E. M., SCHWARTZ, S. M., RUDNAI, P., ELUF-NETO, J., MUSCAT, J., MORGENSTERN, H., MENEZES, A., MATOS, E., BUCUR, A., LEVI, F., LAZARUS, P., LA VECCHIA, C., KOIFMAN, S., KELSEY, K., HERRERO, R., HAYES, R. B., FRANCESCHI, S., WUNSCH-FILHO, V., FERNANDEZ, L., FABIANOVA, E., DAUDT, A. W., DAL MASO, L., CURADO, M. P., CHEN, C., CASTELLSAGUE, X., BENHAMOU, S., BOFFETTA, P., BRENNAN, P. & HASHIBE, M. 2010a. Body mass index and risk of head and neck cancer in a pooled analysis of case-control

studies in the International Head and Neck Cancer Epidemiology (INHANCE) Consortium. *International journal of epidemiology,* 39**,** 1091-1102.

GAUDET, M. M., OLSHAN, A. F., CHUANG, S. C., BERTHILLER, J., ZHANG, Z. F., LISSOWSKA, J., ZARIDZE, D., WINN, D. M., WEI, Q., TALAMINI, R., SZESZENIA-DABROWSKA, N., STURGIS, E. M., SCHWARTZ, S. M., RUDNAI, P., ELUF-NETO, J., MUSCAT, J., MORGENSTERN, H., MENEZES, A., MATOS, E., BUCUR, A., LEVI, F., LAZARUS, P., LA VECCHIA, C., KOIFMAN, S., KELSEY, K., HERRERO, R., HAYES, R. B., FRANCESCHI, S., WUNSCH-FILHO, V., FERNANDEZ, L., FABIANOVA, E., DAUDT, A. W., DAL MASO, L., CURADO, M. P., CHEN, C., CASTELLSAGUE, X., BENHAMOU, S., BOFFETTA, P., BRENNAN, P. & HASHIBE, M. 2010b. Body mass index and risk of head and neck cancer in a pooled analysis of case-control studies in the International Head and Neck Cancer Epidemiology (INHANCE) Consortium. *Int J Epidemiol,* 39**,** 1091-102.

GAUDET, M. M., PATEL, A. V., SUN, J., HILDEBRAND, J. S., MCCULLOUGH, M. L., CHEN, A. Y. & GAPSTUR, S. M. 2012. Prospective studies of body mass index with head and neck cancer incidence and mortality. *Cancer Epidemiol Biomarkers Prev,* 21**,** 497-503.

GILLISON, M. L., CHATURVEDI, A. K., ANDERSON, W. F. & FAKHRY, C. 2015. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol,* 33**,** 3235-42.

GILLISON, M. L., D'SOUZA, G., WESTRA, W., SUGAR, E., XIAO, W., BEGUM, S. & VISCIDI, R. 2008. Distinct Risk Factor Profiles for Human Papillomavirus Type 16–Positive and Human Papillomavirus Type 16–Negative Head and Neck Cancers. *JNCI: Journal of the National Cancer Institute,* 100**,** 407-420.

GINSBURG, G. S. & PHILLIPS, K. A. 2018. Precision Medicine: From Science To Value. *Health Aff (Millwood),* 37**,** 694-701.

GLOBAL BURDEN OF DISEASE CANCER COLLABORATION 2022. Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncology,* 8**,** 420-444.

GLOBAL CANCER OBSERVATORY, G. 2020. Estimated number of new cases in 2020, lip, oral cavity, oropharynx, nasopharynx, hypopharynx, both sexes, all ages. International Agency for Research on Cancer, (IARC).

GLOBOCAN. 2022a. *Cancer Today: Incidence, Both sexes, in 2022 (UK)*

Lip, oral cavity + Oropharynx + Nasopharynx + Hypopharynx + Larynx [Online]. IARC-WHO. Available: https://gco.iarc.fr/today/en/dataviz/tables?mode=population&cancers=3_14_1_5_4&multiple_cancers=1&sexes=0&populations=100_112_191_196_203_208_233_246_250_276_300_348_352_372_380_40_428_440_442_470_498_499_528_56_578_616_620_642_643_688_70_703_705_724_752_756_8_804_807_826&multiple_populations=1 [Accessed].

GLOBOCAN. 2022b. *Cancer Today: Incidence, Both sexes, in 2022*

Lip, oral cavity + Oropharynx + Nasopharynx + Hypopharynx + Larynx [Online]. IARC-WHO. Available: https://gco.iarc.fr/today/en/dataviz/tables?mode=population&cancers=3_14_1_5_4&multiple_cancers=1 [Accessed].

GLOBOCAN 2022 -a. Cancer Today - Incidence, Both sexes, in 2022.

GLOBOCAN. 2022 -b. *Cancer Today Age-Standardized Rate (World) per 100 000, Mortality, Both sexes, in 2022* [Online]. Available: https://gco.iarc.fr/today/en/dataviz/bars?mode=population&cancers=1_3_14_5_4&group_cancers=1&multiple_cancers=1&types=1&sort_by=value0 [Accessed].

GOLDSTONE, S., PALEFSKY, J. M., GIULIANO, A. R., MOREIRA, E. D., JR., ARANDA, C., JESSEN, H., HILLMAN, R. J., FERRIS, D. G., COUTLEE, F., LIAW, K. L., MARSHALL, J. B., ZHANG, X., VUOCOLO, S., BARR, E., HAUPT, R. M., GURIS, D. & GARNER, E. I. 2011. Prevalence of and

risk factors for human papillomavirus (HPV) infection among HIV-seronegative men who have sex with men. *J Infect Dis,* 203**,** 66-74.

GOMES, E. P. A. D. A., ARANHA, A. M. F., BORGES, A. H. & VOLPATO, L. E. R. 2020. Head and Neck Cancer Patients' Quality of Life: Analysis of Three Instruments. *Journal of dentistry (Shiraz, Iran),* 21**,** 31-41.

GOPINATH, D., KUNNATH MENON, R., K VEETTIL, S., GEORGE BOTELHO, M. & JOHNSON, N. W. 2020. Periodontal Diseases as Putative Risk Factors for Head and Neck Cancer: Systematic Review and Meta-Analysis. *Cancers,* 12**,** 1893.

GORMLEY, M., CREANEY, G., SCHACHE, A., INGARFIELD, K. & CONWAY, D. I. 2022a. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *British Dental Journal,* 233**,** 780-786.

GORMLEY, M., CREANEY, G., SCHACHE, A., INGARFIELD, K. & CONWAY, D. I. 2022b. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *Br Dent J,* 233**,** 780-786.

GORMLEY, M., DUDDING, T., SANDERSON, E., MARTIN, R. M., THOMAS, S., TYRRELL, J., NESS, A. R., BRENNAN, P., MUNAFÒ, M., PRING, M., BOCCIA, S., OLSHAN, A. F., DIERGAARDE, B., HUNG, R. J., LIU, G., DAVEY SMITH, G. & RICHMOND, R. C. 2020. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. *Nat Commun,* 11**,** 6071.

GORMLEY, M., DUDDING, T., THOMAS, S. J., TYRRELL, J., NESS, A. R., PRING, M., LEGGE, D., DAVEY SMITH, G., RICHMOND, R. C., VINCENT, E. E. & BULL, C. 2023. Evaluating the effect of metabolic traits on oral and oropharyngeal cancer risk using Mendelian randomization. *eLife,* 12**,** e82674.

GOYAL, N., HENNESSY, M., LEHMAN, E., LIN, W., AGUDO, A., AHRENS, W., BOCCIA, S., BRENNAN, P., BRENNER, H., CADONI, G., CANOVA, C., CHEN, C., CONWAY, D., CURADO, MARIA P., DAL MASO, L., DAUDT, A. W., EDEFONTI, V., FABIANOVA, E., FERNANDEZ, L., FRANCESCHI, S., GARAVELLO, W., GILLISON, M., HAYES, R. B., HEALY, C., HERRERO, R., HOLCATOVA, I., KANDA, J. L., KELSEY, K., HANSEN, BO T., KOIFMAN, R., LAGIOU, P., LA VECCHIA, C., LEVI, F., LI, G., LISSOWSKA, J., MENDOZA LÓPEZ, R., LUCE, D., MACFARLANE, G., MATES, D., MATSUO, K., MCCLEAN, M., MENEZES, A., MENVIELLE, G., MORGENSTERN, H., MOYSICH, K., NEGRI, E., OLSHAN, A. F., PANDICS, T., POLESEL, J., PURDUE, M., RADOI, L., RAMROTH, H., RICHIARDI, L., SCHANTZ, S., SCHWARTZ, S. M., SERRAINO, D., SHANGINA, O., SMITH, E., STURGIS, E. M., ŚWIĄTKOWSKA, B., THOMSON, P., VAUGHAN, T. L., VILENSKY, M., WINN, D. M., WUNSCH-FILHO, V., YU, G.-P., ZEVALLOS, J. P., ZHANG, Z.-F., ZHENG, T., ZNAOR, A., BOFFETTA, P., HASHIBE, M., LEE, Y.-C. A. & MUSCAT, J. E. 2023. Risk factors for head and neck cancer in more and less developed countries: Analysis from the INHANCE consortium. *Oral Diseases,* 29**,** 1565-1578.

GRAY, E. P., TEARE, M. D., STEVENS, J. & ARCHER, R. 2016. Risk Prediction Models for Lung Cancer: A Systematic Review. *Clin Lung Cancer,* 17**,** 95-106.

GREER, C. E., WHEELER, C. M., LADNER, M. B., BEUTNER, K., COYNE, M. Y., LIANG, H., LANGENBERG, A., YEN, T. S. & RALSTON, R. 1995. Human papillomavirus (HPV) type distribution and serological response to HPV type 6 virus-like particles in patients with genital warts. *J Clin Microbiol,* 33**,** 2058-63.

GRIBB, J. P., WHEELOCK, J. H. & PARK, E. S. 2023. Human Papilloma Virus (HPV) and the Current State of Oropharyngeal Cancer Prevention and Treatment. *Dela J Public Health,* 9**,** 26-28.

GUHA, N., WARNAKULASURIYA, S., VLAANDEREN, J. & STRAIF, K. 2014. Betel quid chewing and the risk of oral and oropharyngeal cancers: A meta-analysis with implications for cancer control. *International Journal of Cancer,* 135**,** 1433-1443.

GUO, K., XIAO, W., CHEN, X., ZHAO, Z., LIN, Y. & CHEN, G. 2021. Epidemiological Trends of Head and Neck Cancer: A Population-Based Study. *BioMed research international,* 2021**,** 1738932-1738932.

GUPTA, B., KUMAR, N. & JOHNSON, N. W. 2019. Evidence of past dental visits and incidence of head and neck cancers: a systematic review and meta-analysis. *Systematic Reviews,* 8**,** 43.

HARDMAN, J. C., TIKKA, T. & PALERI, V. 2021a. Remote triage incorporating symptom-based risk stratification for suspected head and neck cancer referrals: A prospective population-based study. *Cancer,* 127**,** 4177-4189.

HARDMAN, J. C., TIKKA, T., PALERI, V., ENT UK, B. & INTEGRATE 2021b. Remote triage incorporating symptom-based risk stratification for suspected head and neck cancer referrals: A prospective population-based study. *Cancer,* 127**,** 4177-4189.

HARRISON, H., PENNELLS, L., WOOD, A., ROSSI, S. H., STEWART, G. D., GRIFFIN, S. J. & USHER-SMITH, J. A. 2022. Validation and public health modelling of risk prediction models for kidney cancer using the UK Biobank. *BJU International,* 129**,** 498-511.

HASHIBE, M., BRENNAN, P., BENHAMOU, S., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., MASO, L. D., DAUDT, A. W., FABIANOVA, E., WÜNSCH-FILHO, V., FRANCESCHI, S., HAYES, R. B., HERRERO, R., KOIFMAN, S., LA VECCHIA, C., LAZARUS, P., LEVI, F., MATES, D., MATOS, E., MENEZES, A., MUSCAT, J., ELUF-NETO, J., OLSHAN, A. F., RUDNAI, P., SCHWARTZ, S. M., SMITH, E., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., WEI, Q., WINN, D. M., ZARIDZE, D., ZATONSKI, W., ZHANG, Z.-F., BERTHILLER, J. & BOFFETTA, P. 2007. Alcohol Drinking in Never Users of Tobacco, Cigarette Smoking in Never Drinkers, and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *JNCI: Journal of the National Cancer Institute,* 99**,** 777-789.

HASHIBE, M., BRENNAN, P., CHUANG, S. C., BOCCIA, S., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., DAL MASO, L., DAUDT, A. W., FABIANOVA, E., FERNANDEZ, L., WÜNSCH-FILHO, V., FRANCESCHI, S., HAYES, R. B., HERRERO, R., KELSEY, K., KOIFMAN, S., LA VECCHIA, C., LAZARUS, P., LEVI, F., LENCE, J. J., MATES, D., MATOS, E., MENEZES, A., MCCLEAN, M. D., MUSCAT, J., ELUF-NETO, J., OLSHAN, A. F., PURDUE, M., RUDNAI, P., SCHWARTZ, S. M., SMITH, E., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., WEI, Q., WINN, D. M., SHANGINA, O., PILARSKA, A., ZHANG, Z. F., FERRO, G., BERTHILLER, J. & BOFFETTA, P. 2009. 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,* 18**,** 541-50.

HASHIBE, M., HUNT, J., WEI, M., BUYS, S., GREN, L. & LEE, Y.-C. A. 2013. Tobacco, alcohol, body mass index, physical activity, and the risk of head and neck cancer in the prostate, lung, colorectal, and ovarian (PLCO) cohort. *Head & Neck,* 35**,** 914-922.

HASHIM, D., GENDEN, E., POSNER, M., HASHIBE, M. & BOFFETTA, P. 2019a. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. *Annals of Oncology,* 30**,** 744-756.

HASHIM, D., GENDEN, E., POSNER, M., HASHIBE, M. & BOFFETTA, P. 2019b. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. *Ann Oncol,* 30**,** 744-756.

HASHIM, D., SARTORI, S., BRENNAN, P., CURADO, M. P., WÜNSCH-FILHO, V., DIVARIS, K., OLSHAN, A. F., ZEVALLOS, J. P., WINN, D. M., FRANCESCHI, S., CASTELLSAGUÉ, X., LISSOWSKA, J., RUDNAI, P., MATSUO, K., MORGENSTERN, H., CHEN, C., VAUGHAN, T. L., HOFMANN, J. N., D'SOUZA, G., HADDAD, R. I., WU, H., LEE, Y. C., HASHIBE, M., VECCHIA, C. L. & BOFFETTA, P. 2016a. The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Annals of Oncology,* 27**,** 1619-1625.

HASHIM, D., SARTORI, S., BRENNAN, P., CURADO, M. P., WÜNSCH-FILHO, V., DIVARIS, K., OLSHAN, A. F., ZEVALLOS, J. P., WINN, D. M., FRANCESCHI, S., CASTELLSAGUÉ, X., LISSOWSKA, J., RUDNAI, P., MATSUO, K., MORGENSTERN, H., CHEN, C., VAUGHAN, T. L., HOFMANN, J. N., D'SOUZA, G., HADDAD, R. I., WU, H., LEE, Y. C., HASHIBE, M., VECCHIA, C. L. & BOFFETTA, P. 2016b. The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Ann Oncol,* 27**,** 1619-25.

HAWKINS, D. M. 2004. The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences,* 44**,** 1-12.

HAYES, R. B., AHN, J., FAN, X., PETERS, B. A., MA, Y., YANG, L., AGALLIU, I., BURK, R. D., GANLY, I., PURDUE, M. P., FREEDMAN, N. D., GAPSTUR, S. M. & PEI, Z. 2018. Association of Oral

Microbiome With Risk for Incident Head and Neck Squamous Cell Cancer. *JAMA Oncol,* 4**,** 358-365.

HIPPISLEY-COX, J. & COUPLAND, C. 2015. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. *BMJ Open,* 5**,** e007825.

HIPPISLEY-COX, J., COUPLAND, C. & BRINDLE, P. 2017. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ,* 357**,** j2099.

HIYAMA, T., YOSHIHARA, M., TANAKA, S. & CHAYAMA, K. 2008. Genetic polymorphisms and head and neck cancer risk (Review). *Int J Oncol,* 32**,** 945-973.

HOFFMAN, C. & PARADISE, J. 2008. Health Insurance and Access to Health Care in the United States. *Annals of the New York Academy of Sciences,* 1136**,** 149-160.

HOFFMAN, H. T., KARNELL, L. H., FUNK, G. F., ROBINSON, R. A. & MENCK, H. R. 1998. The National Cancer Data Base report on cancer of the head and neck. *Arch Otolaryngol Head Neck Surg,* 124**,** 951-62.

HOLZINGER, D., WICHMANN, G., BABOCI, L., MICHEL, A., HÖFLER, D., WIESENFARTH, M., SCHROEDER, L., BOSCOLO-RIZZO, P., HEROLD-MENDE, C., DYCKHOFF, G., BOEHM, A., DEL MISTRO, A., BOSCH, F. X., DIETZ, A., PAWLITA, M. & WATERBOER, T. 2017a. Sensitivity and specificity of antibodies against HPV16 E6 and other early proteins for the detection of HPV16-driven oropharyngeal squamous cell carcinoma. *International Journal of Cancer,* 140**,** 2748-2757.

HOLZINGER, D., WICHMANN, G., BABOCI, L., MICHEL, A., HÖFLER, D., WIESENFARTH, M., SCHROEDER, L., BOSCOLO-RIZZO, P., HEROLD-MENDE, C., DYCKHOFF, G., BOEHM, A., DEL MISTRO, A., BOSCH, F. X., DIETZ, A., PAWLITA, M. & WATERBOER, T. 2017b. Sensitivity and specificity of antibodies against HPV16 E6 and other early proteins for the detection of HPV16-driven oropharyngeal squamous cell carcinoma. *Int J Cancer,* 140**,** 2748-2757.

HOMER, J. J. & WINTER, S. C. 2024. Head and Neck Cancer: United Kingdom National Multidisciplinary Guidelines, Sixth Edition. *The Journal of Laryngology & Otology,* 138**,** S1-S224.

HOSMER JR, D. W., LEMESHOW, S. & STURDIVANT, R. X. 2013. *Applied logistic regression*, John Wiley & Sons.

HUGHES, R. A., HERON, J., STERNE, J. A. C. & TILLING, K. 2019. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology,* 48**,** 1294-1304.

IARC-WHO 2004. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum,* 83**,** 1-1438.

IARC-WHO 2023. Oral Cancer Prevention Volume 19 - IARC Handbooks

IARC-WHO 2024. Global cancer burden growing, amidst mounting need for services.

INGARFIELD, K., MCMAHON, A. D., DOUGLAS, C. M., SAVAGE, S.-A., MACKENZIE, K. & CONWAY, D. I. 2019. Inequality in the Survival of Patients With Head and Neck Cancer in Scotland. *Frontiers in Oncology,* 8.

INGARFIELD, K., MCMAHON, A. D., DOUGLAS, C. M., SAVAGE, S. A., MACKENZIE, K. & CONWAY, D. I. 2018. Inequality in the Survival of Patients With Head and Neck Cancer in Scotland. *Front Oncol,* 8**,** 673.

INGARFIELD, K., MCMAHON, A. D., HURLEY, K., TOMS, S., PRING, M., THOMAS, S. J., WAYLEN, A., PAWLITA, M., WATERBOER, T., NESS, A. R. & CONWAY, D. I. 2021. Inequality in survival of people with head and neck cancer: Head and Neck 5000 cohort study. *Head & Neck,* 43**,** 1252-1270.

IOCCA, O., SOLLECITO, T. P., ALAWI, F., WEINSTEIN, G. S., NEWMAN, J. G., DE VIRGILIO, A., DI MAIO, P., SPRIANO, G., PARDIÑAS LÓPEZ, S. & SHANTI, R. M. 2020. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. *Head & Neck,* 42**,** 539-555.

JACKSON, C., SWEETING, H. & HAW, S. 2012. Clustering of substance use and sexual risk behaviour in adolescence: analysis of two cohort studies. *BMJ Open,* 2**,** e000661.

JACKSON, S. E. & CHESTER, J. D. 2015. Personalised cancer medicine. *International Journal of Cancer,* 137**,** 262-266.

JANES, H., PEPE, M. S. & GU, W. 2008. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med,* 149**,** 751-60.

JETHWA, A. R. & KHARIWALA, S. S. 2017. Tobacco-related carcinogenesis in head and neck cancer. *Cancer metastasis reviews,* 36**,** 411-423.

JIT, M., VYSE, A., BORROW, R., PEBODY, R., SOLDAN, K. & MILLER, E. 2007. Prevalence of human papillomavirus antibodies in young female subjects in England. *British Journal of Cancer,* 97**,** 989-991.

JOHNSON, D. E., BURTNESS, B., LEEMANS, C. R., LUI, V. W. Y., BAUMAN, J. E. & GRANDIS, J. R. 2020a. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers,* 6**,** 92.

JOHNSON, D. E., BURTNESS, B., LEEMANS, C. R., LUI, V. W. Y., BAUMAN, J. E. & GRANDIS, J. R. 2020b. Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers,* 6**,** 92.

JOHNSON, S., MCDONALD, J. T. & CORSTEN, M. J. 2008. Socioeconomic factors in head and neck cancer. *J Otolaryngol Head Neck Surg,* 37**,** 597-601.

JOSHI, P., DUTTA, S., CHATURVEDI, P. & NAIR, S. 2014. Head and neck cancers in developing countries. *Rambam Maimonides medical journal,* 5**,** e0009-e0009.

KASRADZE, D., JUODZBALYS, G., GUOBIS, Z., GERVICKAS, A. & CICCIÙ, M. 2020. Genetic and proteomic biomarkers of head-and-neck cancer: A systematic review. *Journal of Cancer Research and Therapeutics,* 16.

KAWAKITA, D. & MATSUO, K. 2017. Alcohol and head and neck cancer. *Cancer and Metastasis Reviews,* 36**,** 425-434.

KHETAN, P., BOFFETTA, P., LUCE, D., STUCKER, I., CURADO, M. P., MENEZES, A., WUNSCH-FILHO, V., AHRENS, W., LAGIOU, P., SERRAINO, D., RICHIARDI, L., KJAERHEIM, K., CONWAY, D., THOMSON, P., MUSCAT, J., MATES, D., RAMROTH, H., MENVIELLE, G., VAUGHAN, T. L., BRENNER, H., LEE, Y.-C. A., LA VECCHIA, C., HASHIBE, M. & HASHIM, D. 2019. Occupations and the Risk of Head and Neck Cancer: A Pooled Analysis of the International Head and Neck Cancer Epidemiology (INHANCE) Consortium. *Journal of occupational and environmental medicine,* 61**,** 397-404.

KISLING LA, D. J. 2023. *Prevention Strategies*, StatPearls Publishing.

KLEIN, E. A., RICHARDS, D., COHN, A., TUMMALA, M., LAPHAM, R., COSGROVE, D., CHUNG, G., CLEMENT, J., GAO, J., HUNKAPILLER, N., JAMSHIDI, A., KURTZMAN, K. N., SEIDEN, M. V., SWANTON, C. & LIU, M. C. 2021. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Annals of Oncology,* 32**,** 1167-1177.

KOBAYASHI, K., HISAMATSU, K., SUZUI, N., HARA, A., TOMITA, H. & MIYAZAKI, T. 2018. A Review of HPV-Related Head and Neck Cancer. *Journal of clinical medicine,* 7**,** 241.

KOSTOPOULOU, O., ARORA, K. & PÁLFI, B. 2022. Using cancer risk algorithms to improve risk estimates and referral decisions. *Commun Med (Lond),* 2**,** 2.

KOTZUR, M., MACDONALD, S., O'CARROLL, R. E., O'CONNOR, R. C., IRVINE, A., STEELE, R. J. C. & ROBB, K. A. 2022. What are common barriers and helpful solutions to colorectal cancer screening? A cross-sectional survey to develop intervention content for a planning support tool. *BMJ Open,* 12**,** e062738.

KOYANAGI, Y. N., ITO, H., OZE, I., HOSONO, S., TANAKA, H., ABE, T., SHIMIZU, Y., HASEGAWA, Y. & MATSUO, K. 2017. 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,* 26**,** 38-47.

KREIMER, A. R. 2014. Prospects for prevention of HPV-driven oropharynx cancer. *Oral Oncol,* 50**,** 555-9.

KREIMER, A. R., BHATIA, R. K., MESSEGUER, A. L., GONZÁLEZ, P., HERRERO, R. & GIULIANO, A. R. 2010. Oral human papillomavirus in healthy individuals: a systematic review of the literature. *Sex Transm Dis,* 37**,** 386-91.

KRISHNA RAO, S., MEJIA, G. C., LOGAN, R. M., KULKARNI, M., KAMATH, V., FERNANDES, D. J., RAY, S. & ROBERTS-THOMSON, K. 2016. A screening model for oral cancer using risk scores: development and validation. *Community Dentistry & Oral Epidemiology,* 44**,** 76-84.

KURKI, M. I., KARJALAINEN, J., PALTA, P., SIPILÄ, T. P., KRISTIANSSON, K., DONNER, K. M., REEVE, M. P., LAIVUORI, H., AAVIKKO, M., KAUNISTO, M. A., LOUKOLA, A., LAHTELA, E., MATTSSON, H., LAIHO, P., DELLA BRIOTTA PAROLO, P., LEHISTO, A. A., KANAI, M., MARS, N., RÄMÖ, J., KIISKINEN, T., HEYNE, H. O., VEERAPEN, K., RÜEGER, S., LEMMELÄ, S., ZHOU, W., RUOTSALAINEN, S., PÄRN, K., HIEKKALINNA, T., KOSKELAINEN, S., PAAJANEN, T., LLORENS, V., GRACIA-TABUENCA, J., SIIRTOLA, H., REIS, K., ELNAHAS, A. G., SUN, B., FOLEY, C. N., AALTO-SETÄLÄ, K., ALASOO, K., ARVAS, M., AURO, K., BISWAS, S., BIZAKI-VALLASKANGAS, A., CARPEN, O., CHEN, C.-Y., DADA, O. A., DING, Z., EHM, M. G., EKLUND, K., FÄRKKILÄ, M., FINUCANE, H., GANNA, A., GHAZAL, A., GRAHAM, R. R., GREEN, E. M., HAKANEN, A., HAUTALAHTI, M., HEDMAN, Å. K., HILTUNEN, M., HINTTALA, R., HOVATTA, I., HU, X., HUERTAS-VAZQUEZ, A., HUILAJA, L., HUNKAPILLER, J., JACOB, H., JENSEN, J.-N., JOENSUU, H., JOHN, S., JULKUNEN, V., JUNG, M., JUNTTILA, J., KAARNIRANTA, K., KÄHÖNEN, M., KAJANNE, R., KALLIO, L., KÄLVIÄINEN, R., KAPRIO, J., KERIMOV, N., KETTUNEN, J., KILPELÄINEN, E., KILPI, T., KLINGER, K., KOSMA, V.-M., KUOPIO, T., KURRA, V., LAISK, T., LAUKKANEN, J., LAWLESS, N., LIU, A., LONGERICH, S., MÄGI, R., MÄKELÄ, J., MÄKITIE, A., MALARSTIG, A., MANNERMAA, A., MARANVILLE, J., MATAKIDOU, A., MERETOJA, T., MOZAFFARI, S. V., et al. 2023. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature,* 613**,** 508-518.

LAGIOU, P., GEORGILA, C., MINAKI, P., AHRENS, W., POHLABELN, H., BENHAMOU, S., BOUCHARDY, C., SLAMOVA, A., SCHEJBALOVA, M., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., AGUDO, A., CASTELLSAGUE, X., MACFARLANE, T. V., MACFARLANE, G. J., TALAMINI, R., BARZAN, L., CANOVA, C., SIMONATO, L., LOWRY, R., CONWAY, D. I., MCKINNEY, P. A., ZNAOR, A., MCCARTAN, B. E., HEALY, C., NELIS, M., METSPALU, A., MARRON, M., HASHIBE, M. & BRENNAN, P. J. 2009. Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. *Eur J Cancer Prev,* 18**,** 76-84.

LANDY, R., PESOLA, F., CASTAÑÓN, A. & SASIENI, P. 2016. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case-control study. *Br J Cancer,* 115**,** 1140-1146.

LAU, K., WILKINSON, J. & MOORTHY, R. 2018. A web-based prediction score for head and neck cancer referrals. *Clin Otolaryngol,* 43**,** 1043-1049.

LAUBY-SECRETAN, B., SCOCCIANTI, C., LOOMIS, D., GROSSE, Y., BIANCHINI, F. & STRAIF, K. 2016. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine,* 375**,** 794-798.

LAWSON, P. J. & FLOCKE, S. A. 2009. Teachable moments for health behavior change: a concept analysis. *Patient Educ Couns,* 76**,** 25-30.

LECHNER, M., LIU, J., MASTERSON, L. & FENTON, T. R. 2022. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nature Reviews Clinical Oncology,* 19**,** 306-327.

LEE, Y.-C. A., BOFFETTA, P., STURGIS, E. M., WEI, Q., ZHANG, Z.-F., MUSCAT, J., LAZARUS, P., MATOS, E., HAYES, R. B., WINN, D. M., ZARIDZE, D., WÜNSCH-FILHO, V., ELUF-NETO, J., KOIFMAN, S., MATES, D., CURADO, M. P., MENEZES, A., FERNANDEZ, L., DAUDT, A. W., SZESZENIA-DABROWSKA, N., FABIANOVA, E., RUDNAI, P., FERRO, G., BERTHILLER, J., BRENNAN, P. & HASHIBE, M. 2008. Involuntary Smoking and Head and Neck Cancer Risk: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology, Biomarkers & Prevention,* 17**,** 1974-1981.

LEE, Y. A., AL-TEMIMI, M., YING, J., MUSCAT, J., OLSHAN, A. F., ZEVALLOS, J. P., WINN, D. M., LI, G., STURGIS, E. M., MORGENSTERN, H., ZHANG, Z. F., SMITH, E., KELSEY, K., MCCLEAN, M., VAUGHAN, T. L., LAZARUS, P., CHEN, C., SCHWARTZ, S. M., GILLISON, M., SCHANTZ, S., YU, G. P., D'SOUZA, G., GROSS, N., MONROE, M., KIM, J., BOFFETTA, P. & HASHIBE, M. 2020.

Risk Prediction Models for Head and Neck Cancer in the US Population From the INHANCE Consortium. *American Journal of Epidemiology,* 189**,** 330-342.

LEE, Y. C., JUNG, S. H., SHIVAKUMAR, M., CHA, S., PARK, W. Y., WON, H. H., EUN, Y. G., BIOBANK, P. M. & KIM, D. 2024. Polygenic risk score-based phenome-wide association study of head and neck cancer across two large biobanks. *BMC Med,* 22**,** 120.

LEEMANS, C. R., SNIJDERS, P. J. F. & BRAKENHOFF, R. H. 2018. The molecular landscape of head and neck cancer. *Nature Reviews Cancer,* 18**,** 269-282.

LESSEUR, C., DIERGAARDE, B., OLSHAN, A. F., WÜNSCH-FILHO, V., NESS, A. R., LIU, G., LACKO, M., ELUF-NETO, J., FRANCESCHI, S., LAGIOU, P., MACFARLANE, G. J., RICHIARDI, L., BOCCIA, S., POLESEL, J., KJAERHEIM, K., ZARIDZE, D., JOHANSSON, M., MENEZES, A. M., CURADO, M. P., ROBINSON, M., AHRENS, W., CANOVA, C., ZNAOR, A., CASTELLSAGUÉ, X., CONWAY, D. I., HOLCÁTOVÁ, I., MATES, D., VILENSKY, M., HEALY, C. M., SZESZENIA-DĄBROWSKA, N., FABIÁNOVÁ, E., LISSOWSKA, J., GRANDIS, J. R., WEISSLER, M. C., TAJARA, E. H., NUNES, F. D., DE CARVALHO, M. B., THOMAS, S., HUNG, R. J., PETERS, W. H., HERRERO, R., CADONI, G., BUENO-DE-MESQUITA, H. B., STEFFEN, A., AGUDO, A., SHANGINA, O., XIAO, X., GABORIEAU, V., CHABRIER, A., ANANTHARAMAN, D., BOFFETTA, P., AMOS, C. I., MCKAY, J. D. & BRENNAN, P. 2016. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet,* 48**,** 1544-1550.

LI, P., STUART, E. A. & ALLISON, D. B. 2015. Multiple Imputation: A Flexible Tool for Handling Missing Data. *Jama,* 314**,** 1966-7.

LIEDERBACH, E., KYRILLOS, A., WANG, C.-H., LIU, J. C., STURGIS, E. M. & BHAYANI, M. K. 2017. The national landscape of human papillomavirus-associated oropharynx squamous cell carcinoma. *International Journal of Cancer,* 140**,** 504-512.

LIN, L., BENARD, V. B., GREEK, A., HAWKINS, N. A., ROLAND, K. B. & SARAIYA, M. 2015. Racial and ethnic differences in human papillomavirus positivity and risk factors among low-income women in Federally Qualified Health Centers in the United States. *Prev Med,* 81**,** 258-61.

LINDSON, N., BUTLER, A. R., MCROBBIE, H., BULLEN, C., HAJEK, P., BEGH, R., THEODOULOU, A., NOTLEY, C., RIGOTTI, N. A., TURNER, T. & ET AL. 2024. Electronic cigarettes for smoking cessation. *Cochrane Database of Systematic Reviews*.

LIU, H., WANG, D., YANG, Z., LI, S., WU, H., XIANG, J., KAN, S., HAO, M. & LIU, W. 2022. Regulation of epigenetic modifications in the head and neck tumour microenvironment. *Front Immunol,* 13**,** 1050982.

LIU, L. 2018. Chapter 1 - Introduction. *In:* LIU, L. (ed.) *Heart Failure: Epidemiology and Research Methods.* Elsevier.

LIU, Y., LI, Y., FU, Y., LIU, T., LIU, X., ZHANG, X., FU, J., GUAN, X., CHEN, T., CHEN, X. & SUN, Z. 2017. Quantitative prediction of oral cancer risk in patients with oral leukoplakia. *Oncotarget,* 8**,** 46057-46064.

LOOMIS, D., GUHA, N., HALL, A. L. & STRAIF, K. 2018. Identifying occupational carcinogens: an update from the IARC Monographs. *Occupational and Environmental Medicine,* 75**,** 593-603.

LORENZONI, V., CHATURVEDI, A. K., VIGNAT, J., LAVERSANNE, M., BRAY, F. & VACCARELLA, S. 2022. The Current Burden of Oropharyngeal Cancer: A Global Assessment Based on GLOBOCAN 2020. *Cancer Epidemiol Biomarkers Prev,* 31**,** 2054-2062.

LOUIE, K. S., MEHANNA, H. & SASIENI, P. 2015. Trends in head and neck cancers in England from 1995 to 2011 and projections up to 2025. *Oral Oncology,* 51**,** 341-348.

LOURO, J., POSSO, M., HILTON BOON, M., ROMÁN, M., DOMINGO, L., CASTELLS, X. & SALA, M. 2019. A systematic review and quality assessment of individualised breast cancer risk prediction models. *British Journal of Cancer,* 121**,** 76-85.

LUBIN, J. H., MUSCAT, J., GAUDET, M. M., OLSHAN, A. F., CURADO, M. P., DAL MASO, L., WÜNSCH-FILHO, V., STURGIS, E. M., SZESZENIA-DABROWSKA, N., CASTELLSAGUE, X., ZHANG, Z. F., SMITH, E., FERNANDEZ, L., MATOS, E., FRANCESCHI, S., FABIANOVA, E., RUDNAI, P., PURDUE, M. P., MATES, D., WEI, Q., HERRERO, R., KELSEY, K., MORGENSTERN, H., SHANGINA, O., KOIFMAN, S., LISSOWSKA, J., LEVI, F., DAUDT, A. W., NETO, J. E., CHEN, C., LAZARUS, P., WINN, D. M., SCHWARTZ, S. M., BOFFETTA, P., BRENNAN, P., MENEZES, A.,

LA VECCHIA, C., MCCLEAN, M., TALAMINI, R., RAJKUMAR, T., HAYES, R. B. & HASHIBE, M. 2011. An examination of male and female odds ratios by BMI, cigarette smoking, and alcohol consumption for cancers of the oral cavity, pharynx, and larynx in pooled data from 15 case-control studies. *Cancer Causes Control,* 22**,** 1217-31.

LUBIN, J. H., PURDUE, M., KELSEY, K., ZHANG, Z.-F., WINN, D., WEI, Q., TALAMINI, R., SZESZENIA-DABROWSKA, N., STURGIS, E. M., SMITH, E., SHANGINA, O., SCHWARTZ, S. M., RUDNAI, P., NETO, J. E., MUSCAT, J., MORGENSTERN, H., MENEZES, A., MATOS, E., MATES, I. N., LISSOWSKA, J., LEVI, F., LAZARUS, P., VECCHIA, C. L., KOIFMAN, S., HERRERO, R., FRANCESCHI, S., WÜNSCH-FILHO, V., FERNANDEZ, L., FABIANOVA, E., DAUDT, A. W., MASO, L. D., CURADO, M. P., CHEN, C., CASTELLSAGUE, X., BRENNAN, P., BOFFETTA, P., HASHIBE, M. & HAYES, R. B. 2009. Total Exposure and Exposure Rate Effects for Alcohol and Smoking and Risk of Head and Neck Cancer: A Pooled Analysis of Case-Control Studies. *American Journal of Epidemiology,* 170**,** 937-947.

LYALL, D. M., QUINN, T., LYALL, L. M., WARD, J., ANDERSON, J. J., SMITH, D. J., STEWART, W., STRAWBRIDGE, R. J., BAILEY, M. E. S. & CULLEN, B. 2022. Quantifying bias in psychological and physical health in the UK Biobank imaging sub-sample. *Brain Communications,* 4.

LYDIATT, W., O'SULLIVAN, B. & PATEL, S. 2018. Major Changes in Head and Neck Staging for 2018. *American Society of Clinical Oncology Educational Book***,** 505-514.

LYONS, M., PHILPOTT, J., HORE, I. & WATTERS, G. 2004. Audit of referrals for head and neck cancer - the effect of the 2-week, fast track referral system. *Clin Otolaryngol Allied Sci,* 29**,** 143-5.

MAASLAND, D. H., VAN DEN BRANDT, P. A., KREMER, B. & SCHOUTEN, L. J. 2015a. Body mass index and risk of subtypes of head-neck cancer: the Netherlands Cohort Study. *Sci Rep,* 5**,** 17744.

MAASLAND, D. H. E., BRANDT, P. A. V. D., KREMER, B. & SCHOUTEN, L. J. 2015b. Body mass index and risk of subtypes of head-neck cancer: the Netherlands Cohort Study. *Scientific Reports,* 5**,** 17744.

MACFARLANE, T. V., MACFARLANE, G. J., OLIVER, R. J., BENHAMOU, S., BOUCHARDY, C., AHRENS, W., POHLABELN, H., LAGIOU, P., LAGIOU, A., CASTELLSAGUE, X., AGUDO, A., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., SLAMOVA, A., SCHEJBALOVA, M., CANOVA, C., SIMONATO, L., TALAMINI, R., BARZAN, L., CONWAY, D. I., MCKINNEY, P. A., ZNAOR, A., LOWRY, R. J., THOMSON, P., HEALY, C. M., MCCARTAN, B. E., MARRON, M., HASHIBE, M. & BRENNAN, P. 2010. The aetiology of upper aerodigestive tract cancers among young adults in Europe: the ARCAGE study. *Cancer Causes Control,* 21**,** 2213-21.

MACFARLANE, T. V., MACFARLANE, G. J., THAKKER, N. S., BENHAMOU, S., BOUCHARDY, C., AHRENS, W., POHLABELN, H., LAGIOU, P., LAGIOU, A., CASTELLSAGUE, X., AGUDO, A., SLAMOVA, A., PLZAK, J., MERLETTI, F., RICHIARDI, L., TALAMINI, R., BARZAN, L., KJAERHEIM, K., CANOVA, C., SIMONATO, L., CONWAY, D. I., MCKINNEY, P. A., THOMSON, P., SLOAN, P., ZNAOR, A., HEALY, C. M., MCCARTAN, B. E., MARRON, M. & BRENNAN, P. 2012. Role of medical history and medication use in the aetiology of upper aerodigestive tract cancers in Europe: the ARCAGE study. *Ann Oncol,* 23**,** 1053-60.

MACILWRAITH, P., MALSEM, E. & DUSHYANTHEN, S. 2023. The effectiveness of HPV vaccination on the incidence of oropharyngeal cancers in men: a review. *Infect Agent Cancer,* 18**,** 24.

MAHAL, B. A., CATALANO, P. J., HADDAD, R. I., HANNA, G. J., KASS, J. I., SCHOENFELD, J. D., TISHLER, R. B. & MARGALIT, D. N. 2019. Incidence and Demographic Burden of HPV-Associated Oropharyngeal Head and Neck Cancers in the United States. *Cancer Epidemiology, Biomarkers & Prevention,* 28**,** 1660-1667.

MALCOMSON, F. C., WIGGINS, C., PARRA-SOTO, S., HO, F. K., CELIS-MORALES, C., SHARP, L. & MATHERS, J. C. 2023. Adherence to the 2018 World Cancer Research Fund/American Institute for Cancer Research Cancer Prevention Recommendations and cancer risk: A systematic review and meta-analysis. *Cancer,* 129**,** 2655-2670.

MANDREKAR, J. N. 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology,* 5**,** 1315-1316.

MARIZ, B., KOWALSKI, L. P., WILLIAM, W. N., JR., DE CASTRO, G., JR., CHAVES, A. L. F., SANTOS, M., DE OLIVEIRA, T. B., ARAÚJO, A. L. D., NORMANDO, A. G. C., RIBEIRO, A. C. P., BRANDÃO, T. B., VARGAS, P. A., LOPES, M. A. & SANTOS-SILVA, A. R. 2020. Global prevalence of human papillomavirus-driven oropharyngeal squamous cell carcinoma following the ASCO guidelines: A systematic review and meta-analysis. *Crit Rev Oncol Hematol,* 156**,** 103116.

MARKOWETZ, F. 2024. All models are wrong and yours are useless: making clinical prediction models impactful for patients. *npj Precision Oncology,* 8**,** 54.

MARKS, J. A., SWITCHENKO, J. M., STEUER, C. E., RYAN, M., PATEL, M. R., MCDONALD, M. W., HIGGINS, K., BEITLER, J. J., SHIN, D. M., GILLESPIE, T. W. & SABA, N. F. 2021. Socioeconomic Factors Influence the Impact of Tumor HPV Status on Outcome of Patients With Oropharyngeal Squamous Cell Carcinoma. *JCO Oncol Pract,* 17**,** e313-e322.

MARRON, M., BOFFETTA, P., MØLLER, H., AHRENS, W., POHLABELN, H., BENHAMOU, S., BOUCHARDY, C., LAGIOU, P., LAGIOU, A., SLÁMOVÁ, A., SCHEJBALOVÁ, M., MERLETTI, F., RICHIARDI, L., KJAERHEIM, K., AGUDO, A., CASTELLSAGUE, X., MACFARLANE, T. V., MACFARLANE, G. J., TALAMINI, R., BARZAN, L., CANOVA, C., SIMONATO, L., BIGGS, A. M., THOMSON, P., CONWAY, D. I., MCKINNEY, P. A., ZNAOR, A., HEALY, C. M., MCCARTAN, B. E., BRENNAN, P. & HASHIBE, M. 2012. Risk of upper aerodigestive tract cancer and type of alcoholic beverage: a European multicenter case-control study. *Eur J Epidemiol,* 27**,** 499-517.

MARRON, M., BOFFETTA, P., ZHANG, Z.-F., ZARIDZE, D., WÜNSCH-FILHO, V., WINN, D. M., WEI, Q., TALAMINI, R., SZESZENIA-DABROWSKA, N., STURGIS, E. M., SMITH, E., SCHWARTZ, S. M., RUDNAI, P., PURDUE, M. P., OLSHAN, A. F., ELUF-NETO, J., MUSCAT, J., MORGENSTERN, H., MENEZES, A., MCCLEAN, M., MATOS, E., MATES, I. N., LISSOWSKA, J., LEVI, F., LAZARUS, P., VECCHIA, C. L., KOIFMAN, S., KELSEY, K., HERRERO, R., HAYES, R. B., FRANCESCHI, S., FERNANDEZ, L., FABIANOVA, E., DAUDT, A. W., MASO, L. D., CURADO, M. P., CADONI, G., CHEN, C., CASTELLSAGUE, X., BOCCIA, S., BENHAMOU, S., FERRO, G., BERTHILLER, J., BRENNAN, P., MØLLER, H. & HASHIBE, M. 2009. Cessation of alcohol drinking, tobacco smoking and the reversal of head and neck cancer risk. *International Journal of Epidemiology,* 39**,** 182-196.

MASSAT, N. J., DIBDEN, A., PARMAR, D., CUZICK, J., SASIENI, P. D. & DUFFY, S. W. 2016. Impact of Screening on Breast Cancer Mortality: The UK Program 20 Years On. *Cancer Epidemiology, Biomarkers & Prevention,* 25**,** 455-462.

MATHUR, S., CONWAY, D. I., WORLLEDGE-ANDREW, H., MACPHERSON, L. M. D. & ROSS, A. J. 2015. Assessment and prevention of behavioural and social risk factors associated with oral cancer: protocol for a systematic review of clinical guidelines and systematic reviews to inform Primary Care dental professionals. *Systematic Reviews,* 4**,** 184.

MCCARTHY, C. E., BONNET, L. J., MARCUS, M. W. & FIELD, J. K. 2020. Development and validation of a multivariable risk prediction model for head and neck cancer using the UK Biobank. *International Journal of Oncology,* 57(5)**,** 1192-1202.

MCCARTHY, C. E., FIELD, J. K., RAJLAWAT, B. P., FIELD, A. E. & MARCUS, M. W. 2015. Trends and regional variation in the incidence of head and neck cancers in England: 2002 to 2011. *Int J Oncol,* 47**,** 204-210.

MCDERMOTT, J. D. & BOWLES, D. W. 2019. Epidemiology of Head and Neck Squamous Cell Carcinomas: Impact on Staging and Prevention Strategies. *Current Treatment Options in Oncology,* 20**,** 43.

MCKIE, C., AHMAD, U. A., FELLOWS, S., MEIKLE, D., STAFFORD, F. W., THOMSON, P. J., WELCH, A. R. & PALERI, V. 2008. The 2-week rule for suspected head and neck cancer in the United Kingdom: referral patterns, diagnostic efficacy of the guidelines and compliance. *Oral Oncol,* 44**,** 851-6.

MEDINA-LARA, A., GRIGORE, B., LEWIS, R., PETERS, J., PRICE, S., LANDA, P., ROBINSON, S., NEAL, R., HAMILTON, W. & SPENCER, A. E. 2020. Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis. *Health Technol Assess,* 24**,** 1-332.

MEHANNA, H., PALERI, V., WEST, C. M. L. & NUTTING, C. 2010. Head and neck cancer—Part 1: Epidemiology, presentation, and prevention. *BMJ,* 341**,** c4684.

MEHTA, R. & CHINTHAPALLI, K. 2019. Glasgow coma scale explained. *BMJ,* 365**,** l1296.

MELO FILHO, M. R., ROCHA, B. A., PIRES, M. B., FONSECA, E. S., FREITAS, E. M., MARTELLI JUNIOR, H. & SANTOS, F. B. 2013. Quality of life of patients with head and neck cancer. *Braz J Otorhinolaryngol,* 79**,** 82-8.

MENEZES, F. D. S., FERNANDES, G. A., ANTUNES, J. L. F., VILLA, L. L. & TOPORCOV, T. N. 2021. Global incidence trends in head and neck cancer for HPV-related and -unrelated subsites: A systematic review of population-based studies. *Oral Oncology,* 115**,** 105177.

MENTZER, A. J., BRENNER, N., ALLEN, N., LITTLEJOHNS, T. J., CHONG, A. Y., CORTES, A., ALMOND, R., HILL, M., SHEARD, S., MCVEAN, G., AIELLO, A., BANGHAM, C., BORROW, R., BREUER, J., BROOKS, T., FRANCESCHI, S., GKRANIA-KLOTSAS, E., GREENWOOD, B., GRIFFITHS, P., GUY, E., JEFFERY, K., KELLY, D., KLENERMAN, P., VAN DER KLIS, F., KNIGHT, J., MCMICHAEL, A., NARANBHAI, V., PEBODY, R., PETO, T., POLLARD, A. J., SCHULZ, T., SOLDAN, K., TAYLOR, G., TOWERS, G., TOMMASINO, M., WEISS, R., WHITBY, D., WILD, C., WYLLIE, D., COLLINS, R., HILL, A. V. S., WATERBOER, T. & BOARD, U. K. B. I. A. 2022. Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. *Nature Communications,* 13**,** 1818.

MESA-EGUIAGARAY, I., WILD, S. H., ROSENBERG, P. S., BIRD, S. M., BREWSTER, D. H., HALL, P. S., CAMERON, D. A., MORRISON, D. & FIGUEROA, J. D. 2020. Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: a population-based study of Scottish cancer registry data. *British Journal of Cancer,* 123**,** 852-859.

METTIAS, B., CHARLTON, A. & ASHOKKUMAR, S. 2021. Outcome of two-week head and neck cancer pathway for the otolaryngology department in a tertiary centre. *The Journal of Laryngology & Otology,* 135**,** 869-873.

MIRANDA-FILHO, A. & BRAY, F. 2020. Global patterns and trends in cancers of the lip, tongue and mouth. *Oral Oncol,* 102**,** 104551.

MIRZA, A. H., AYLIN, P., MIDDLETON, S., KING, E. V., NOURAEI, R. A. R. & REPANOS, C. 2019. Impact of social deprivation on the outcome of major head and neck cancer surgery in England: A national analysis. *Head Neck,* 41**,** 692-700.

MOHEBBI, E., ALIMORADI, Z., ROZEK, L. S., MOHEBBI, A. & JAFARI, E. 2024. Risk of head and neck cancer among opium users. *Archives of Oral Biology,* 157**,** 105846.

MOON, P. K., MA, Y. & MEGWALU, U. C. 2022. Head and Neck Cancer Stage at Presentation and Survival Outcomes Among Native Hawaiian and Other Pacific Islander Patients Compared With Asian and White Patients. *JAMA Otolaryngology–Head & Neck Surgery,* 148**,** 636-645.

MOONS, K. G., DE GROOT, J. A., BOUWMEESTER, W., VERGOUWE, Y., MALLETT, S., ALTMAN, D. G., REITSMA, J. B. & COLLINS, G. S. 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med,* 11**,** e1001744.

MOURAD, M., JETMORE, T., JATEGAONKAR, A. A., MOUBAYED, S., MOSHIER, E. & URKEN, M. L. 2017. Epidemiological Trends of Head and Neck Cancer in the United States: A SEER Population Study. *Journal of Oral and Maxillofacial Surgery,* 75**,** 2562-2572.

NARISAWA-SAITO, M. & KIYONO, T. 2007. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: Roles of E6 and E7 proteins. *Cancer Science,* 98**,** 1505-1511.

NASSIF, S. J., STEINWALD, P. & TRACY, J. C. 2022. Human Papillomavirus Vaccine and Head and Neck Cancer. *JAMA Otolaryngology–Head & Neck Surgery,* 148**,** 204-204.

NATIONAL CANCER INSTITUE (NCI) 2022 Cancer Stat Facts: Female Breast Cancer.

NATIONAL CANCER INSTITUE (NCI) 2024. Cancer Stat Facts: Colorectal Cancer.

NATIONAL CANCER INSTITUTE, N. 2021 *Head and Neck Cancers* [Online]. Available: https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet#what-are-cancers-of-the-head-and-neck [Accessed].

NATIONAL RECORDS OF SCOTLAND (NRS). 2022. *Population Estimates by Scottish Index of Multiple Deprivation (SIMD)* [Online]. Available: https://www.nrscotland.gov.uk/statistics-

and-data/statistics/statistics-by-theme/population/population-estimates/2011-based-special-area-population-estimates/population-estimates-by-simd-2016 [Accessed].
NEGRI, E., BOFFETTA, P., BERTHILLER, J., CASTELLSAGUE, X., CURADO, M. P., DAL MASO, L., DAUDT, A. W., FABIANOVA, E., FERNANDEZ, L., WÜNSCH-FILHO, V., FRANCESCHI, S., HAYES, R. B., HERRERO, R., KOIFMAN, S., LAZARUS, P., LENCE, J. J., LEVI, F., MATES, D., MATOS, E., MENEZES, A., MUSCAT, J., ELUF-NETO, J., OLSHAN, A. F., RUDNAI, P., SHANGINA, O., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., WEI, Q., WINN, D. M., ZARIDZE, D., LISSOWSKA, J., ZHANG, Z. F., FERRO, G., BRENNAN, P., LA VECCHIA, C. & HASHIBE, M. 2009. Family history of cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Int J Cancer,* 124**,** 394-401.
NHS. 2021. *Galleri Trial* [Online]. Available: https://www.nhs-galleri.org/about-the-trial [Accessed].
NHS ONLINE. 2021. *Conditions - Head and neck cancer* [Online]. Available: https://www.nhs.uk/conditions/head-and-neck-cancer/ [Accessed 8/3/22].
NHS ONLINE. 2022. *NHS screening* [Online]. Available: https://www.nhs.uk/conditions/nhs-screening/ [Accessed].
NHS SCOTLAND 2019. Scottish Referral Guidelines for Suspected Cancer.
NHS SCOTLAND 2023. Optimal Head and Neck Cancer Diagnostic Pathway.
NHS SCOTLAND 2024. NHS Dental Reform: Section I - Oral Health Examination and Diagnosis.
NHS UK 2022. NHS Long Term Plan - Chapter 3: Further progress on care quality and outcomes - Cancer
NICE 2020. Dental checks: intervals between oral health reviews.
NICE 2021a. NICE Guidelines: Head and neck cancers - recognition and referral.
NICE, T. N. I. F. H. A. C. E. 2015. Suspected cancer: recognition

and referral
NICE, T. N. I. F. H. A. C. E. 2021b. *Symptoms suggestive of head and neck cancers* [Online]. Available: https://cks.nice.org.uk/topics/head-neck-cancers-recognition-referral/diagnosis/symptoms-suggestive-of-head-neck-cancers/ [Accessed 8/3/22].
NIH NATIONAL CANCER INSTITUTE 2020. BRCA Gene Mutations: Cancer Risk and Genetic Testing.
NORTHERN IRELAND CANCER REGISTRY 2019. Head and neck cancer Report.
OFFICE FOR NATIONAL STATISTICS 2023. Approximated Social Grade, England and Wales: Census 2021.
OKUNADE, K. S. 2020. Human papillomavirus and cervical cancer. *J Obstet Gynaecol,* 40**,** 602-608.
OMAÑA-CEPEDA, C., JANÉ-SALAS, E., ESTRUGO-DEVESA, A., CHIMENOS-KÜSTNER, E. & LÓPEZ-LÓPEZ, J. 2016. Effectiveness of dentist's intervention in smoking cessation: A review. *J Clin Exp Dent,* 8**,** e78-83.
PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., BOUTRON, I., HOFFMANN, T. C., MULROW, C. D., SHAMSEER, L., TETZLAFF, J. M., AKL, E. A., BRENNAN, S. E., CHOU, R., GLANVILLE, J., GRIMSHAW, J. M., HRÓBJARTSSON, A., LALU, M. M., LI, T., LODER, E. W., MAYO-WILSON, E., MCDONALD, S., MCGUINNESS, L. A., STEWART, L. A., THOMAS, J., TRICCO, A. C., WELCH, V. A., WHITING, P. & MOHER, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* 372**,** n71.
PALERI, V., HARDMAN, J., TIKKA, T., BRADLEY, P., PRACY, P. & KERAWALA, C. 2020. Rapid implementation of an evidence-based remote triaging system for assessment of suspected referrals and patients with head and neck cancer on follow-up after treatment during the COVID-19 pandemic: Model for international collaboration. *Head & neck,* 42**,** 1674-1680.
PALMER, T., WALLACE, L., POLLOCK, K. G., CUSCHIERI, K., ROBERTSON, C., KAVANAGH, K. & CRUICKSHANK, M. 2019. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. *BMJ,* 365**,** l1161.
PALMER, T. J., KAVANAGH, K., CUSCHIERI, K., CAMERON, R., GRAHAM, C., WILSON, A. & ROY, K. 2024a. Invasive cervical cancer incidence following bivalent human papillomavirus

vaccination: a population-based observational study of age at immunization, dose, and deprivation. *JNCI: Journal of the National Cancer Institute*.

PALMER, T. J., KAVANAGH, K., CUSCHIERI, K., CAMERON, R., GRAHAM, C., WILSON, A. & ROY, K. 2024b. Invasive cervical cancer incidence following bivalent human papillomavirus vaccination: a population-based observational study of age at immunization, dose, and deprivation. *JNCI: Journal of the National Cancer Institute,* 116**,** 857-865.

PARIKH, R., MATHAI, A., PARIKH, S., CHANDRA SEKHAR, G. & THOMAS, R. 2008. Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology,* 56**,** 45-50.

PARK, J. O., NAM, I. C., KIM, C. S., PARK, S. J., LEE, D. H., KIM, H. B., HAN, K. D. & JOO, Y. H. 2022. Sex Differences in the Prevalence of Head and Neck Cancers: A 10-Year Follow-Up Study of 10 Million Healthy People. *Cancers (Basel),* 14.

PARK, S. L., LEE, Y. C., MARRON, M., AGUDO, A., AHRENS, W., BARZAN, L., BENCKO, V., BENHAMOU, S., BOUCHARDY, C., CANOVA, C., CASTELLSAGUE, X., CONWAY, D. I., HEALY, C. M., HOLCÁTOVÁ, I., KJAERHEIM, K., LAGIOU, P., LOWRY, R. J., MACFARLANE, T. V., MACFARLANE, G. J., MCCARTAN, B. E., MCKINNEY, P. A., MERLETTI, F., POHLABELN, H., RICHIARDI, L., SIMONATO, L., SNEDDON, L., TALAMINI, R., TRICHOPOULOS, D., ZNAOR, A., BRENNAN, P. & HASHIBE, M. 2011. The association between change in body mass index and upper aerodigestive tract cancers in the ARCAGE project: multicenter case-control study. *Int J Cancer,* 128**,** 1449-61.

PARKIN, D. M. & BRAY, F. 2006. Chapter 2: The burden of HPV-related cancers. *Vaccine,* 24**,** S11-S25.

PATHIRANA, T. I. & JACKSON, C. A. 2018. Socioeconomic status and multimorbidity: a systematic review and meta-analysis. *Australian and New Zealand Journal of Public Health,* 42**,** 186-194.

PETO, R. 1986. Influence of dose and duration of smoking on lung cancer rates. *IARC Sci Publ***,** 23-33.

PICCIRILLO, J. F. 2000. Importance of Comorbidity in Head and Neck Cancer. *The Laryngoscope,* 110**,** 593-602.

POSNER, M. R., LORCH, J. H., GOLOUBEVA, O., TAN, M., SCHUMAKER, L. M., SARLIS, N. J., HADDAD, R. I. & CULLEN, K. J. 2011. Survival and human papillomavirus in oropharynx cancer in TAX 324: a subset analysis from an international phase III trial. *Annals of Oncology,* 22**,** 1071-1077.

PRICE, S., SPENCER, A., MEDINA-LARA, A. & HAMILTON, W. 2019. Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care. *British Journal of General Practice,* 69**,** e437-e443.

PUBLIC HEALTH SCOTLAND 2021. Cancer incidence in Scotland; Cancer incidence and prevalence in Scotland to December 2019; Cancer Statistics - Head and Neck Cancer.

PUBLIC HEALTH SCOTLAND 2022. Head and neck cancer Quality Performance Indicators.

PUBLIC HEALTH SCOTLAND 2023a. Dental statistics - registration and participation

NHS registration and participation.

PUBLIC HEALTH SCOTLAND 2023b. HPV immunisation statistics Scotland

School year 2022/23.

PUBLIC HEALTH SCOTLAND. 2023c. *HPV vaccine - NHSInform.scot* [Online]. Available: https://www.nhsinform.scot/healthy-living/immunisation/vaccines/hpv-vaccine#overview [Accessed].

PUBLIC HEATH SCOTLAND 2022. Scottish Cancer Registry

PULTE, D. & BRENNER, H. 2010a. Changes in Survival in Head and Neck Cancers in the Late 20th and Early 21st Century: A Period Analysis. *The Oncologist,* 15**,** 994-1001.

PULTE, D. & BRENNER, H. 2010b. Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis. *Oncologist,* 15**,** 994-1001.

PURDUE, M. P., HASHIBE, M., BERTHILLER, J., LA VECCHIA, C., DAL MASO, L., HERRERO, R., FRANCESCHI, S., CASTELLSAGUE, X., WEI, Q., STURGIS, E. M., MORGENSTERN, H., ZHANG, Z. F., LEVI, F., TALAMINI, R., SMITH, E., MUSCAT, J., LAZARUS, P., SCHWARTZ, S. M., CHEN, C., NETO, J. E., WÜNSCH-FILHO, V., ZARIDZE, D., KOIFMAN, S., CURADO, M. P., BENHAMOU, S., MATOS, E., SZESZENIA-DABROWSKA, N., OLSHAN, A. F., LENCE, J., MENEZES, A., DAUDT, A. W., MATES, I. N., PILARSKA, A., FABIANOVA, E., RUDNAI, P., WINN, D., FERRO, G., BRENNAN, P., BOFFETTA, P. & HAYES, R. B. 2009. Type of alcoholic beverage and risk of head and neck cancer--a pooled analysis within the INHANCE Consortium. *Am J Epidemiol,* 169**,** 132-42.

PURKAYASTHA, M., MCMAHON, A. D., GIBSON, J. & CONWAY, D. I. 2016. Trends of oral cavity, oropharyngeal and laryngeal cancer incidence in Scotland (1975-2012) - A socioeconomic perspective. *Oral Oncol,* 61**,** 70-5.

Q-CANCER. 2017. *Q-Cancer* [Online]. Available: https://www.qcancer.org/ [Accessed].

RAJ, A. T., SUJATHA, G., MURUGANANDHAN, J., KUMAR, S. S., BHARKAVI, S. I., VARADARAJAN, S., PATIL, S. & AWAN, K. H. 2020. Reviewing the oral carcinogenic potential of E-cigarettes using the Bradford Hill criteria of causation. *Transl Cancer Res,* 9**,** 3142-3152.

RAMSPEK, C. L., JAGER, K. J., DEKKER, F. W., ZOCCALI, C. & VAN DIEPEN, M. 2020. External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal,* 14**,** 49-58.

RECALDE, M., PISTILLO, A., DAVILA-BATISTA, V., LEITZMANN, M., ROMIEU, I., VIALLON, V., FREISLING, H. & DUARTE-SALLES, T. 2023. Longitudinal body mass index and cancer risk: a cohort study of 2.6 million Catalan adults. *Nature Communications,* 14**,** 3816.

REKVE, D., BANATVALA, N., KARPATI, A., TARLTON, D., WESTERMAN, L., SPERKOVA, K., CASSWELL, S., DUENNBIER, M., ROJHANI, A., BAKKE, Ø., MONTEIRO, M., LINOU, N., KULIKOV, A. & POZNYAK, V. B. 2019. Prioritising action on alcohol for health and development. *BMJ,* 367**,** l6162.

RENOU, A., GUIZARD, A. V., CHABRILLAC, E., DEFOSSEZ, G., GROSCLAUDE, P., DENEUVE, S., VERGEZ, S., LAPOTRE-LEDOUX, B., PLOUVIER, S. D., DUPRET-BORIES, A. & FRANCIM, N. 2023. Evolution of the Incidence of Oral Cavity Cancers in the Elderly from 1990 to 2018. *J Clin Med,* 12.

REYNOLDS, R. M., WEIR, J., STOCKTON, D. L., BREWSTER, D. H., SANDEEP, T. C. & STRACHAN, M. W. J. 2005. Changing trends in incidence and mortality of thyroid cancer in Scotland. *Clinical Endocrinology,* 62**,** 156-162.

RIAZ, N., MORRIS, L. G., LEE, W. & CHAN, T. A. 2014. Unraveling the molecular genetics of head and neck cancer through genome-wide approaches. *Genes & diseases,* 1**,** 75-86.

RICHARD D RILEY, P. O. B., JOIE ENSOR, RESEARCH ASSOCIATE IN BIOSTATISTICS1, KYM I E SNELL, RESEARCH FELLOW IN BIOSTATISTICS2, THOMAS P A DEBRAY, ASSISTANT PROFESSOR34, DOUG G ALTMAN, PROFESSOR OF STATISTICS IN MEDICINE5, KAREL G M MOONS, PROFESSOR OF CLINICAL EPIDEMIOLOGY34, GARY S COLLINS, ASSOCIATE PROFESSOR5 2019. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ,* 365**,** l4379.

RILEY, R. D., ENSOR, J., SNELL, K. I. E., DEBRAY, T. P. A., ALTMAN, D. G., MOONS, K. G. M. & COLLINS, G. S. 2016. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ,* 353**,** i3140.

RISK, C. 2013. QCancer Julia Hippisley-Cox, Carol Coupland.

ROBINSON, K. L. & MACFARLANE, G. J. 2003. Oropharyngeal cancer incidence and mortality in Scotland: are rates still increasing? *Oral Oncology,* 39**,** 31-36.

RODERICK J. A. LITTLE, D. B. R. 2002. *Statistical Analysis with Missing Data*, Wiley

RODRÍGUEZ, A. C., SCHIFFMAN, M., HERRERO, R., WACHOLDER, S., HILDESHEIM, A., CASTLE, P. E., SOLOMON, D. & BURK, R. 2008. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst,* 100**,** 513-7.

RORY, C. 2007. UK Biobank: Protocol for a large-scale prospective epidemiological resource. United Kingdom: UK Biobank.

RUBIN, D. B. 2018. Multiple imputation. *Flexible Imputation of Missing Data, Second Edition.* Chapman and Hall/CRC.

RUSSO, D., MARIANI, P., CAPONIO, V. C. A., LO RUSSO, L., FIORILLO, L., ZHURAKIVSKA, K., LO MUZIO, L., LAINO, L. & TROIANO, G. 2021. Development and Validation of Prognostic Models for Oral Squamous Cell Carcinoma: A Systematic Review and Appraisal of the Literature. *Cancers,* 13**,** 5755.

RUSSO, D. P., THAM, T., BARDASH, Y. & KRAUS, D. 2020. The effect of race in head and neck cancer: A meta-analysis controlling for socioeconomic status. *American Journal of Otolaryngology,* 41**,** 102624.

SABATINI, M. E. & CHIOCCA, S. 2020a. Human papillomavirus as a driver of head and neck cancers. *British Journal of Cancer,* 122**,** 306-314.

SABATINI, M. E. & CHIOCCA, S. 2020b. Human papillomavirus as a driver of head and neck cancers. *Br J Cancer,* 122**,** 306-314.

SADETZKI, S., BENSAL, D., NOVIKOV, I. & MODAN, B. 2003. The limitations of using hospital controls in cancer etiology--one more example for Berkson's bias. *Eur J Epidemiol,* 18**,** 1127-31.

SAKA HERRÁN, C., JANÉ-SALAS, E., ESTRUGO DEVESA, A. & LÓPEZ-LÓPEZ, J. 2018. Protective effects of metformin, statins and anti-inflammatory drugs on head and neck cancer: A systematic review. *Oral Oncology,* 85**,** 68-81.

SANDERSON, R. J. & IRONSIDE, J. A. 2002a. Squamous cell carcinomas of the head and neck. *Bmj,* 325**,** 822-7.

SANDERSON, R. J. & IRONSIDE, J. A. D. 2002b. Squamous cell carcinomas of the head and neck. *BMJ (Clinical research ed.),* 325**,** 822-827.

SANKARANARAYANAN, R., RAMADAS, K., THOMAS, G., MUWONGE, R., THARA, S., MATHEW, B. & RAJAN, B. 2005. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. *The Lancet,* 365**,** 1927-1933.

SASSANO, M., TABORELLI, M., BOCCIA, S., CADONI, G., LA VECCHIA, C., GARAVELLO, W., LAZARUS, P., LEE, Y.-C. A., HASHIBE, M. & BOFFETTA, P. 2024. Aspirin intake and head and neck cancer: A pooled analysis within the INHANCE consortium. *Head & Neck,* 46**,** 926-935.

SCHABATH, M. B., THOMPSON, Z. J., EGAN, K. M., TORRES, B. N., NGUYEN, A., PAPENFUSS, M. R., ABRAHAMSEN, M. E. & GIULIANO, A. R. 2015. Alcohol consumption and prevalence of human papillomavirus (HPV) infection among US men in the HPV in Men (HIM) study. *Sex Transm Infect,* 91**,** 61-7.

SCHABATH, M. B., VILLA, L. L., LAZCANO-PONCE, E., SALMERÓN, J., QUITERIO, M., GIULIANO, A. R. & STUDY, F. T. H. 2012. Smoking and Human Papillomavirus (HPV) Infection in the HPV in Men (HIM) Study. *Cancer Epidemiology, Biomarkers & Prevention,* 21**,** 102-110.

SCHACHE, A. G., POWELL, N. G., CUSCHIERI, K. S., ROBINSON, M., LEARY, S., MEHANNA, H., RAPOZO, D., LONG, A., CUBIE, H., JUNOR, E., MONAGHAN, H., HARRINGTON, K. J., NUTTING, C. M., SCHICK, U., LAU, A. S., UPILE, N., SHEARD, J., BROUGHAM, K., WEST, C. M. L., OGUEJIOFOR, K., THOMAS, S., NESS, A. R., PRING, M., THOMAS, G. J., KING, E. V., MCCANCE, D. J., JAMES, J. A., MORAN, M., SLOAN, P., SHAW, R. J., EVANS, M. & JONES, T. M. 2016. HPV-Related Oropharynx Cancer in the United Kingdom: An Evolution in the Understanding of Disease Etiology. *Cancer Research,* 76**,** 6598-6606.

SCHROEDER, L., PRING, M., INGARFIELD, K., PAWLITA, M., LEARY, S. D., THOMAS, S. J., WAYLEN, A., WATERBOER, T. & NESS, A. R. 2020. HPV driven squamous cell head and neck cancer of unknown primary is likely to be HPV driven squamous cell oropharyngeal cancer. *Oral Oncology,* 107**,** 104721.

SCOTLAND'S CENSUS 2022. Scotland's Census 2022 - Rounded population estimates.

SCOTLAND, P. H. 2023. Dental statistics - registration and participation.

SCOTTISH GOVERNMENT 2018. Alcohol Framework 2018.

SCOTTISH GOVERNMENT 2020. Scottish Index of Multiple Deprivation.

SCOTTISH GOVERNMENT 2021. The Scottish Health Survey Main Report - Cabinet Secretary for NHS Recovery, Health and Social Care.

SCOTTISH GOVERNMENT 2023a. Cancer strategy 2023 to 2033.

SCOTTISH GOVERNMENT 2023b. The Scottish Health Survey 2022 – volume 1: main report.
SEITZ, H. K. & STICKEL, F. 2010. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. *Genes & nutrition,* 5**,** 121-128.
SEMPRINI, J. & WILLIAMS, J. C. 2023. Community socioeconomic status and rural/racial disparities in HPV−/+ head and neck cancer. *Technical Innovations & Patient Support in Radiation Oncology,* 26**,** 100205.
SHANG, C., FENG, L., GU, Y., HONG, H., HONG, L. & HOU, J. 2021. Impact of Multidisciplinary Team Management on the Survival Rate of Head and Neck Cancer Patients: A Cohort Study Meta-analysis. *Frontiers in Oncology,* 11.
SHARMA, N. & OM, H. 2015. Usage of Probabilistic and General Regression Neural Network for Early Detection and Prevention of Oral Cancer. *Scientific World Journal,* 2015 (no pagination).
SHAW, R. & BEASLEY, N. 2016. Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *J Laryngol Otol,* 130**,** S9-s12.
SHETE, S., LIU, H., WANG, J., YU, R., STURGIS, E. M., LI, G., DAHLSTROM, K. R., LIU, Z., AMOS, C. I. & WEI, Q. 2020. A Genome-Wide Association Study Identifies Two Novel Susceptible Regions for Squamous Cell Carcinoma of the Head and Neck. *Cancer Res,* 80**,** 2451-2460.
SHIELD, K. D., FERLAY, J., JEMAL, A., SANKARANARAYANAN, R., CHATURVEDI, A. K., BRAY, F. & SOERJOMATARAM, I. 2017. The global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2012. *CA Cancer J Clin,* 67**,** 51-64.
SIEGEL, R. L., MILLER, K. D., FUCHS, H. E. & JEMAL, A. 2021. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians,* 71**,** 7-33.
SIM, J. & LEWIS, M. 2012. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *Journal of Clinical Epidemiology,* 65**,** 301-308.
SIMARD, E. P., TORRE, L. A. & JEMAL, A. 2014. International trends in head and neck cancer incidence rates: Differences by country, sex and anatomic site. *Oral Oncology,* 50**,** 387-403.
SIMPSON, C. D., HENNIGAN, M. & MORRISON, J. 2023. Accuracy of the Head and Neck Cancer Risk Calculator in identifying maxillofacial malignancies. *Br J Oral Maxillofac Surg,* 61**,** 356-361.
SMĘDRA, A. & BERENT, J. 2023. The Influence of the Oral Microbiome on Oral Cancer: A Literature Review and a New Approach. *Biomolecules,* 13**,** 815.
SMITH, C. D. L., MCMAHON, A. D., ROSS, A., INMAN, G. J. & CONWAY, D. I. 2022a. Risk prediction models for head and neck cancer: A rapid review. *Laryngoscope Investigative Otolaryngology,* 7**,** 1893-1908.
SMITH, C. D. L., MCMAHON, A. D., ROSS, A., INMAN, G. J. & CONWAY, D. I. 2022b. Risk prediction models for head and neck cancer: A rapid review. *Laryngoscope Investig Otolaryngol,* 7**,** 1893-1908.
SMITH, C. D. L., MCMAHON, A. D., ROSS, A., INMAN, G. J. & CONWAY, D. I. 2022c. Risk prediction models for head and neck cancer: A rapid review. *Laryngoscope Investigative Otolaryngology,* n/a.
SONAWANE, K., SUK, R., CHIAO, E. Y., CHHATWAL, J., QIU, P., WILKIN, T., NYITRAY, A. G., SIKORA, A. G. & DESHMUKH, A. A. 2017. Oral Human Papillomavirus Infection: Differences in Prevalence Between Sexes and Concordance With Genital Human Papillomavirus Infection, NHANES 2011 to 2014. *Ann Intern Med,* 167**,** 714-724.
SPEIGHT, P. M., EPSTEIN, J., KUJAN, O., LINGEN, M. W., NAGAO, T., RANGANATHAN, K. & VARGAS, P. 2017. Screening for oral cancer—a perspective from the Global Oral Cancer Forum. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology,* 123**,** 680-687.
SPENCE, T., BRUCE, J., YIP, K. W. & LIU, F. F. 2016. HPV Associated Head and Neck Cancer. *Cancers (Basel),* 8.
STANG, A. & JÖCKEL, K.-H. 2018. The Impact of Cancer Screening on All-Cause Mortality. *Deutsches Arzteblatt international,* 115**,** 481-486.

STINGONE, J. A., FUNKHOUSER, W. K., WEISSLER, M. C., BELL, M. E. & OLSHAN, A. F. 2013. Racial differences in the relationship between tobacco, alcohol, and squamous cell carcinoma of the head and neck. *Cancer Causes Control,* 24**,** 649-64.

SUBRAMANIAN, S., SANKARANARAYANAN, R., BAPAT, B., SOMANATHAN, T., THOMAS, G., MATHEW, B., VINODA, J. & RAMADAS, K. 2009. Cost-effectiveness of oral cancer screening: results from a cluster randomized controlled trial in India. SciELO Public Health.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J., LANDRAY, M., LIU, B., MATTHEWS, P., ONG, G., PELL, J., SILMAN, A., YOUNG, A., SPROSEN, T., PEAKMAN, T. & COLLINS, R. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med,* 12**,** e1001779.

SUN, Z., LIU, L., WANG, C., QIAN, L., YANG, J., ZHAO, Z., FAN, Y. & PENG, Z. 2019. Assessment of the cancerization risk for oral potentially malignant disorders by clinical risk model combined with autofluorescence and brush biopsy with DNA-image cytometry. *European Archives of Oto-Rhino-Laryngology,* 276(9)**,** 2549-2557.

SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians,* 71**,** 209-249.

SUTTON, R. T., PINCOCK, D., BAUMGART, D. C., SADOWSKI, D. C., FEDORAK, R. N. & KROEKER, K. I. 2020. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine,* 3**,** 17.

SZUKALSKA, M., SZYFTER, K., FLOREK, E., RODRIGO, J. P., RINALDO, A., MÄKITIE, A. A., STROJAN, P., TAKES, R. P., SUÁREZ, C., SABA, N. F., BRAAKHUIS, B. J. M. & FERLITO, A. 2020. Electronic Cigarettes and Head and Neck Cancer Risk-Current State of Art. *Cancers (Basel),* 12.

TABERNA, M., GIL MONCAYO, F., JANÉ-SALAS, E., ANTONIO, M., ARRIBAS, L., VILAJOSANA, E., PERALVEZ TORRES, E. & MESÍA, R. 2020. The Multidisciplinary Team (MDT) Approach and Quality of Care. *Front Oncol,* 10**,** 85.

TAN, J. Y., CALLAGHAN, C. J., LEWTHWAITE, A. W., CHAN, C., WEE, C. T., TO, E. Y. H., SUMMERS, I., NELSON, J. W., SMITH, M. B., LI, L. Q., MORTON, C., PORTEOUS, L., EVANS, A. S. & NIXON, I. J. 2024. Impact of Primary Care Triage using Head and Neck Cancer Risk Calculator Version 2 (HaNC-RCv2) on Tertiary Head and Neck Service in Post-COVID period. *J Laryngol Otol***,** 1-29.

TANTON, C., MESHER, D., BEDDOWS, S., SOLDAN, K., CLIFTON, S., PANWAR, K., FIELD, N., MERCER, C. H., JOHNSON, A. M. & SONNENBERG, P. 2017. Human papillomavirus (HPV) in young women in Britain: Population-based evidence of the effectiveness of the bivalent immunisation programme and burden of quadrivalent and 9-valent vaccine types. *Papillomavirus Res,* 3**,** 36-41.

TAYLOR, L. C., HUTCHINSON, A., LAW, K., SHAH, V., USHER-SMITH, J. A. & DENNISON, R. A. 2023. Acceptability of risk stratification within population-based cancer screening from the perspective of the general public: A mixed-methods systematic review. *Health Expectations,* 26**,** 989-1008.

TENNY, S., KERNDT, C. C. & HOFFMAN, M. R. 2022. Case Control Studies. *StatPearls.* Treasure Island (FL): StatPearls Publishing

THAM, T., MACHADO, R., HERMAN, S. W., KRAUS, D., COSTANTINO, P. & ROCHE, A. 2019. Personalized prognostication in head and neck cancer: A systematic review of nomograms according to the AJCC precision medicine core (PMC) criteria. *Head & Neck,* 41**,** 2811-2822.

THOMPSON-HARVEY, A., YETUKURI, M., HANSEN, A. R., SIMPSON, M. C., ADJEI BOAKYE, E., VARVARES, M. A. & OSAZUWA-PETERS, N. 2020. Rising incidence of late-stage head and neck cancer in the United States. *Cancer,* 126**,** 1090-1101.

THUN MJ, J. A. 2003. Analytic Epidemiology.

Holland-Frei Cancer Medicine. 6th edition.

TIKKA, T., KAVANAGH, K., LOWIT, A., JIAFENG, P., BURNS, H., NIXON, I. J., PALERI, V. & MACKENZIE, K. 2020. Head and neck cancer risk calculator (HaNC-RC)-V.2. Adjustments and addition of symptoms and social history factors. *Clinical Otolaryngology,* 45**,** 380-388.

TOMAR, S. L. 2001. Dentistry's role in tobacco control. *The Journal of the American Dental Association,* 132**,** 30S-35S.

TOPORCOV, T. N., ZNAOR, A., ZHANG, Z.-F., YU, G.-P., WINN, D. M., WEI, Q., VILENSKY, M., VAUGHAN, T., THOMSON, P., TALAMINI, R., SZESZENIA-DABROWSKA, N., STURGIS, E. M., SMITH, E., SHANGINA, O., SCHWARTZ, S. M., SCHANTZ, S., RUDNAI, P., RICHIARDI, L., RAMROTH, H., PURDUE, M. P., OLSHAN, A. F., ELUF-NETO, J., MUSCAT, J., MOYSES, R. A., MORGENSTERN, H., MENEZES, A., MCCLEAN, M., MATSUO, K., MATES, D., MACFARLANE, T. V., LISSOWSKA, J., LEVI, F., LAZARUS, P., VECCHIA, C. L., LAGIOU, P., KOIFMAN, S., KJAERHEIM, K., KELSEY, K., HOLCATOVA, I., HERRERO, R., HEALY, C., HAYES, R. B., FRANCESCHI, S., FERNANDEZ, L., FABIANOVA, E., DAUDT, A. W., CURIONI, O. A., MASO, L. D., CURADO, M. P., CONWAY, D. I., CHEN, C., CASTELLSAGUE, X., CANOVA, C., CADONI, G., BRENNAN, P., BOCCIA, S., ANTUNES, J. L. F., AHRENS, W., AGUDO, A., BOFFETTA, P., HASHIBE, M., LEE, Y.-C. A. & FILHO, V. W. 2015. Risk factors for head and neck cancer in young adults: a pooled analysis in the INHANCE consortium. *International Journal of Epidemiology,* 44**,** 169-185.

TOTA, J. E., BEST, A. F., ZUMSTEG, Z. S., GILLISON, M. L., ROSENBERG, P. S. & CHATURVEDI, A. K. 2019a. Evolution of the Oropharynx Cancer Epidemic in the United States: Moderation of Increasing Incidence in Younger Individuals and Shift in the Burden to Older Individuals. *J Clin Oncol,* 37**,** 1538-1546.

TOTA, J. E., GILLISON, M. L., KATKI, H. A., KAHLE, L., PICKARD, R. K., XIAO, W., JIANG, B., GRAUBARD, B. I. & CHATURVEDI, A. K. 2019b. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer,* 125**,** 4407-4416.

TOTA, J. E., GILLISON, M. L., KATKI, H. A., KAHLE, L., PICKARD, R. K., XIAO, W., JIANG, B., GRAUBARD, B. I. & CHATURVEDI, A. K. 2019c. Development and validation of an individualized risk prediction model for oropharynx cancer in the US population. *Cancer,* 125(24)**,** 4407-4416.

TSENG, K.-S., LIN, C., LIN, Y.-S. & WENG, S.-F. 2014. Risk of Head and Neck Cancer in Patients With Diabetes Mellitus: A Retrospective Cohort Study in Taiwan. *JAMA Otolaryngology–Head & Neck Surgery,* 140**,** 746-753.

TSENTEMEIDOU, A., FYRMPAS, G., STAVRAKAS, M., VLACHTSIS, K., SOTIRIOU, E., POUTOGLIDIS, A. & TSETSOS, N. 2021. Human Papillomavirus Vaccine to End Oropharyngeal Cancer. A Systematic Review and Meta-Analysis. *Sex Transm Dis,* 48**,** 700-707.

TURNER, S., BREWSTER, L., KIDD, J., GNICH, W., BALL, G. E., MILBURN, K., PITTS, N. B., GOOLD, S., CONWAY, D. I. & MACPHERSON, L. M. 2010. Childsmile: the national child oral health improvement programme in Scotland. Part 2: Monitoring and delivery. *Br Dent J,* 209**,** 79-83.

TWEED, E. J., ALLARDICE, G. M., MCLOONE, P. & MORRISON, D. S. 2018. Socio-economic inequalities in the incidence of four common cancers: a population-based registry study. *Public Health,* 154**,** 1-10.

UK BIOBANK. 2019. *Serological measurement of infectious agents in UK Biobank: a pilot study in 10,000 samples* [Online]. Available: https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/infdisease.pdf [Accessed].

UK GOVERNMENT 2023. Modelling for the smokefree generation policy. *In:* CARE, D. O. H. A. S. (ed.).

UK NATIONAL SCREENING COMMITTEE 2003. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme.

USHER-SMITH, J., EMERY, J., HAMILTON, W., GRIFFIN, S. J. & WALTER, F. M. 2015. Risk prediction tools for cancer in primary care. *British Journal of Cancer,* 113**,** 1645-1650.

USHER-SMITH, J. A., EMERY, J., KASSIANOS, A. P. & WALTER, F. M. 2014. Risk Prediction Models for Melanoma: A Systematic Review. *Cancer Epidemiology, Biomarkers & Prevention,* 23**,** 1450-1463.

USHER-SMITH, J. A., HARSHFIELD, A., SAUNDERS, C. L., SHARP, S. J., EMERY, J., WALTER, F. M., MUIR, K. & GRIFFIN, S. J. 2018. External validation of risk prediction models for incident colorectal cancer using UK Biobank. *British Journal of Cancer,* 118**,** 750-759.

USHER-SMITH, J. A., WALTER, F. M., EMERY, J. D., WIN, A. K. & GRIFFIN, S. J. 2016. Risk Prediction Models for Colorectal Cancer: A Systematic Review. *Cancer Prev Res (Phila),* 9**,** 13-26.

VALERO, C. & SHAH, J. P. 2021. Staging of human papilloma virus related cancers of the oropharynx. *J Surg Oncol,* 124**,** 931-934.

VAN CALSTER, B., MCLERNON, D. J., VAN SMEDEN, M., WYNANTS, L., STEYERBERG, E. W., BOSSUYT, P., COLLINS, G. S., MACASKILL, P., MCLERNON, D. J., MOONS, K. G. M., STEYERBERG, E. W., VAN CALSTER, B., VAN SMEDEN, M., VICKERS, ANDREW J., ON BEHALF OF TOPIC GROUP 'EVALUATING DIAGNOSTIC, T. & PREDICTION MODELS' OF THE, S. I. 2019. Calibration: the Achilles heel of predictive analytics. *BMC Medicine,* 17**,** 230.

VAN CALSTER, B., NIEBOER, D., VERGOUWE, Y., DE COCK, B., PENCINA, M. J. & STEYERBERG, E. W. 2016. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology,* 74**,** 167-176.

VAN DYNE, E. A., HENLEY, S. J., SARAIYA, M., THOMAS, C. C., MARKOWITZ, L. E. & BENARD, V. B. 2018. Trends in Human Papillomavirus-Associated Cancers - United States, 1999-2015. *MMWR Morb Mortal Wkly Rep,* 67**,** 918-924.

VAN MONSJOU, H. S., WREESMANN, V. B., VAN DEN BREKEL, M. W. M. & BALM, A. J. M. 2013. Head and neck squamous cell carcinoma in young patients. *Oral Oncology,* 49**,** 1097-1102.

VINODHINI, K., SHANMUGHAPRIYA, S., DAS, B. C. & NATARAJASEENIVASAN, K. 2012. Prevalence and risk factors of HPV infection among women from various provinces of the world. *Archives of Gynecology and Obstetrics,* 285**,** 771-777.

VOLTZKE, K. J., LEE, Y. A., ZHANG, Z. F., ZEVALLOS, J. P., YU, G. P., WINN, D. M., VAUGHAN, T. L., STURGIS, E. M., SMITH, E., SCHWARTZ, S. M., SCHANTZ, S., MUSCAT, J., MORGENSTERN, H., MCCLEAN, M., LI, G., LAZARUS, P., KELSEY, K., GILLISON, M., CHEN, C., BOFFETTA, P., HASHIBE, M. & OLSHAN, A. F. 2018. Racial differences in the relationship between tobacco, alcohol, and the risk of head and neck cancer: pooled analysis of US studies in the INHANCE Consortium. *Cancer Causes Control,* 29**,** 619-630.

WAKEHAM, K., PAN, J., POLLOCK, K. G., MILLAN, D., BELL, S., MCLELLAN, D., MCPHADEN, A., CONWAY, D. I., GRAHAM, S. V., KAVANAGH, K. & CUSCHIERI, K. 2019a. A Prospective Cohort Study of Human Papillomavirus-Driven Oropharyngeal Cancers: Implications for Prognosis and Immunisation. *Clinical Oncology,* 31**,** e132-e142.

WAKEHAM, K., PAN, J., POLLOCK, K. G., MILLAN, D., BELL, S., MCLELLAN, D., MCPHADEN, A., CONWAY, D. I., GRAHAM, S. V., KAVANAGH, K. & CUSCHIERI, K. 2019b. A Prospective Cohort Study of Human Papillomavirus-Driven Oropharyngeal Cancers: Implications for Prognosis and Immunisation. *Clin Oncol (R Coll Radiol),* 31**,** e132-e142.

WALKER, J. G., LICQURISH, S., CHIANG, P. P., PIROTTA, M. & EMERY, J. D. 2015a. Cancer risk assessment tools in primary care: a systematic review of randomized controlled trials. *Ann Fam Med,* 13**,** 480-9.

WALKER, J. G., LICQURISH, S., CHIANG, P. P. C., PIROTTA, M. & EMERY, J. D. 2015b. Cancer Risk Assessment Tools in Primary Care: A Systematic Review of Randomized Controlled Trials. *The Annals of Family Medicine,* 13**,** 480-489.

WALSER, T., CUI, X., YANAGAWA, J., LEE, J. M., HEINRICH, E., LEE, G., SHARMA, S. & DUBINETT, S. M. 2008. Smoking and lung cancer: the role of inflammation. *Proc Am Thorac Soc,* 5**,** 811-5.

WALSH, C. G., SHARMAN, K. & HRIPCSAK, G. 2017. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. *J Biomed Inform,* 76**,** 9-18.

WALSH, T., MACEY, R., KERR, A. R., LINGEN, M. W., OGDEN, G. R. & WARNAKULASURIYA, S. 2021. Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. *Cochrane Database of Systematic Reviews*.

WANG, T., LI, Y. & ZHENG, X. 2024. Association of socioeconomic status with cardiovascular disease and cardiovascular risk factors: a systematic review and meta-analysis. *Journal of Public Health,* 32**,** 385-399.

WARNAKULASURIYA, S. 2020. Oral potentially malignant disorders: A comprehensive review on clinical aspects and management. *Oral Oncology,* 102**,** 104550.

WARNAKULASURIYA, S., SUTHERLAND, G. & SCULLY, C. 2005. Tobacco, oral cancer, and treatment of dependence. *Oral Oncology,* 41**,** 244-260.

WATERBOER, T., SEHR, P., MICHAEL, K. M., FRANCESCHI, S., NIELAND, J. D., JOOS, T. O., TEMPLIN, M. F. & PAWLITA, M. 2005a. Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. *Clin Chem,* 51**,** 1845-53.

WATERBOER, T., SEHR, P., MICHAEL, K. M., FRANCESCHI, S., NIELAND, J. D., JOOS, T. O., TEMPLIN, M. F. & PAWLITA, M. 2005b. Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-Transferase Fusion Proteins. *Clinical Chemistry,* 51**,** 1845-1853.

WATERBOER, T., SEHR, P. & PAWLITA, M. 2006. Suppression of non-specific binding in serological Luminex assays. *J Immunol Methods,* 309**,** 200-4.

WELLINGS, K., NANCHAHAL, K., MACDOWALL, W., MCMANUS, S., ERENS, B., MERCER, C. H., JOHNSON, A. M., COPAS, A. J., KOROVESSIS, C., FENTON, K. A. & FIELD, J. 2001. Sexual behaviour in Britain: early heterosexual experience. *The Lancet,* 358**,** 1843-1850.

WHITE, C. 2008. Scotland starts HPV vaccination programme. *BMJ,* 337**,** a1543.

WHO 2017 'Best buys' and other recommended interventions for the

prevention and control of noncommunicable diseases. *In:* DEPARTMENT FOR MANAGEMENT OF NCDS & DISABILITY, V. A. I. P. N. (eds.). (Online) 920 Avenue Appia

1211 Geneva 27

Switzerland.

WHO. 2022. *List of consortia led by IARC* [Online]. Available: https://www.iarc.who.int/list-of-consortia-led-by-iarc/ [Accessed].

WHO 2023. Agents classified by the IARC Monographs, Volumes 1–133.

WHO 2023 Oral Cancer Prevention IARC Handbooks of Cacner Prevention

WILLETT, W. 1987. Nutritional epidemiology: issues and challenges. *Int J Epidemiol,* 16**,** 312-7.

WILLIAMS, T. G., CUBIELLA, J., GRIFFIN, S. J., WALTER, F. M. & USHER-SMITH, J. A. 2016. Risk prediction models for colorectal cancer in people with symptoms: a systematic review. *BMC Gastroenterol,* 16**,** 63.

WILSON, J. C., ANDERSON, L. A., MURRAY, L. J. & HUGHES, C. M. 2011. Non-steroidal anti-inflammatory drug and aspirin use and the risk of head and neck cancer: a systematic review. *Cancer Causes & Control,* 22**,** 803-810.

WILSON, J. M. G., JUNGNER, G. & ORGANIZATION, W. H. 1968. Principles and practice of screening for disease.

WILSON, L. M., AVILA TANG, E., CHANDER, G., HUTTON, H. E., ODELOLA, O. A., ELF, J. L., HECKMAN-STODDARD, B. M., BASS, E. B., LITTLE, E. A., HABERL, E. B. & APELBERG, B. J. 2012. Impact of Tobacco Control Interventions on Smoking Initiation, Cessation, and Prevalence: A Systematic Review. *Journal of Environmental and Public Health,* 2012**,** 961724.

WINN, D., LEE, Y.-C., HASHIBE, M., BOFFETTA, P. & CONSORTIUM, T. I. 2015a. The INHANCE consortium: toward a better understanding of the causes and mechanisms of head and neck cancer. *Oral Diseases,* 21**,** 685-693.

WINN, D. M., LEE, Y. C., HASHIBE, M. & BOFFETTA, P. 2015b. The INHANCE consortium: toward a better understanding of the causes and mechanisms of head and neck cancer. *Oral Dis,* 21**,** 685-93.

WOLFF, R. F., MOONS, K. G. M., RILEY, R. D., WHITING, P. F., WESTWOOD, M., COLLINS, G. S., REITSMA, J. B., KLEIJNEN, J. & MALLETT, S. 2019. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med,* 170**,** 51-58.

WOODLEY, N., SLIM, M. A. M., TON, T., MONTGOMERY, J. & DOUGLAS, C. 2022. Does recreational drug use influence survival and morbidity associated with laryngeal cancer. *Addict Health,* 14**,** 115-126.

WORLD HEALTH, O. 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision. 2nd ed ed. Geneva: World Health Organization.

WORLD HEALTH ORGANISATION (WHO) 2024a. The HEAD and neck cancer in South America and Europe (HEADSpAcE) Study.

WORLD HEALTH ORGANISATION (WHO) 2024b. VOYAGER (Human Papillomavirus, Oral and Oropharyngeal Cancer Genomic Research).

WORLD HEALTH ORGANISATION (WHO), I. A. F. R. O. C. I. 2023. Agents classified by the IARC Monographs, Volumes 1–133.

WORLD HEALTH ORGANIZATION 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision. 2nd ed ed. Geneva: World Health Organization.

WORLD HEALTH ORGANIZATION 2014. Human papillomavirus vaccines: WHO position paper, October 2014. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire,* 89**,** 465-491.

WORLD HEALTH ORGANIZATION 2022. Human papillomavirus vaccines: WHO position paper, December 2022.

WORLD HEALTH ORGANIZATION (WHO) 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision. 2nd ed ed. Geneva: World Health Organization.

WYPER, G. M. A., MACKAY, D. F., FRASER, C., LEWSEY, J., ROBINSON, M., BEESTON, C. & GILES, L. 2023. Evaluating the impact of alcohol minimum unit pricing on deaths and hospitalisations in Scotland: a controlled interrupted time series study. *The Lancet,* 401**,** 1361-1370.

WYSS, A. B., HASHIBE, M., LEE, Y. A., CHUANG, S. C., MUSCAT, J., CHEN, C., SCHWARTZ, S. M., SMITH, E., ZHANG, Z. F., MORGENSTERN, H., WEI, Q., LI, G., KELSEY, K. T., MCCLEAN, M., WINN, D. M., SCHANTZ, S., YU, G. P., GILLISON, M. L., ZEVALLOS, J. P., BOFFETTA, P. & OLSHAN, A. F. 2016. Smokeless Tobacco Use and the Risk of Head and Neck Cancer: Pooled Analysis of US Studies in the INHANCE Consortium. *Am J Epidemiol,* 184**,** 703-716.

YOUNG, D., XIAO, C. C., MURPHY, B., MOORE, M., FAKHRY, C. & DAY, T. A. 2015. Increase in head and neck cancer in younger patients due to human papillomavirus (HPV). *Oral Oncology,* 51**,** 727-730.

ZENG, X.-T., DENG, A.-P., LI, C., XIA, L.-Y., NIU, Y.-M. & LENG, W.-D. 2013. Periodontal disease and risk of head and neck cancer: a meta-analysis of observational studies. *PloS one,* 8**,** e79017-e79017.

ZHANG, M., CHEN, C., LI, G., KORIC, A., LEE, Y.-C. A., MORGENSTERN, H., SCHWARTZ, S. M., STURGIS, E. M., BOFFETTA, P., HASHIBE, M. & ZHANG, Z.-F. 2024. Cocaine use and head and neck cancer risk: A pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Medicine,* 13**,** e7019.

ZHANG, Y., FAKHRY, C. & D'SOUZA, G. 2021. Projected Association of Human Papillomavirus Vaccination With Oropharynx Cancer Incidence in the US, 2020-2045. *JAMA Oncol,* 7**,** e212907.