University of Strathclyde

Faculty of Humanities and Social Sciences (HASS)

School of Social Work and Social Policy

Development and validation of a head and neck cancer risk calculator

A thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy in Public Health and Health Policy

by

Theofano Tikka

January 2023

Declaration of Authenticity and Author's Rights

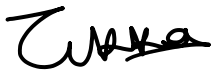
This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: 

Date: 14/01/2023

Published work

Chapter 4 of this thesis includes previously published work (Tikka *et al.*, 2020) for which I have been responsible. I was the first author of this work, taking the lead in the planning, data collection, analysis and writing of the manuscript under the guidance of my supervisors.

Signed: 

Date: 14/01/2023

COVID-19 Impact Statement

I would like to make a note here of how the COVID-19 pandemic affected the output of my thesis. The aim of my thesis was to further increase the predictive power of the HaNC-RC (Tikka, Pracy and Paleri, 2016) with the initial research question being:
1. Are there any new variables that can be added to the existing HNC risk calculator (Tikka, Pracy and Paleri, 2016) and/or could its current variables be refined based on a new large dataset to improve the risk calculator diagnostic efficacy?

Following the publication of the results of this work in January 2020 (Tikka *et al.*, 2020) and while I was still a PhD student, the refined version of the calculator (HaNC-RC v.2) was used in a prospective service evaluation study designed and led by the British Association of Otorhinolaryngology (ENT UK) (ENTUK, 2020) and UK ENT Trainee Research Network (INTEGRATE) (INTEGRATE, 2020) organisations. The study was registered by INTEGRATE as a service evaluation audit based on the output of the NHS Health Research authority online tool (www.hra.nhs.uk, 2021) and no further ethics approval was sought. The HaNC-RC v.2 tool was used as an aid to facilitate telephone consultations in head and neck virtual clinics during the first wave of the COVID-19 pandemic (Hardman *et al.*, 2021).

Following my request, the anonymised prospective database generated by this study was shared with me by the INTEGRATE committe. This was an opportunity for an external validation of the HaNC-RC v.2 in a separate prospective cohort of patients. Hence, an additional research question was added to my PhD thesis output that was:
2. How does the new, updated version of the HNC risk calculator perform in triaging a new cohort of patients referred to HaN clinics across the UK?

Acknowledgements

I would like to immensely thank my supervisors, Prof Anja Lowit, Prof Kenneth MacKenzie and Dr Kimberely Kavanagh, for their continuous help, encouragement, guidance and support throughout my PhD years. I am ever so grateful to them for helping me achieve my academic goal

Table of Contents

List of Figures

List of Tables

Abstract

Background: Most new head and neck cancer (HNC) cases in the UK are diagnosed in an advanced disease stage. This is despite the availability of the 2-week wait (2ww), urgent suspected cancer referral pathway from primary to secondary care. Most HNCs are diagnosed from routes other than the 2ww, despite an increasing number of 2ww referrals. A symptom-based risk calculator had been previously designed to identify patients at high risk of HNC (AUC: 77%) but has not been widely adopted to date, having a lower AUC compared to other common cancer risk calculators (AUC >80%).

Aim and Objectives: The aim of this study was to develop and validate a refined version of the HNC symptom-based calculator with the objective of increasing its prediction potential to be more in line with other cancer risk calculators.

Design, Setting and Participants: The study was performed in two stages. The calculator development phase was based on a prospective cohort of new head and neck referrals to a secondary care centre in Glasgow (n=3,531, following sample size calculation). The validation phase was performed in a new prospective cohort of patients referred via the 2ww pathway in 41 secondary care centres across the UK (n=4,569) during the first wave of the COVID-19 pandemic.

Main outcome measures: The main outcome measure was the area under the curve (AUC) and sensitivity and specificity combination of the final selected model at internal and external validation. Data collected included demographics, social history, presenting symptoms and signs and HNC diagnosis. Binary logistic regression analysis and random forest modelling with internal validation were performed to identify the best-performing model, followed by logistic regression external validation of the updated (HaNC-RC v.2) model.

Results: The HaNC-RC v.2 had an improved AUC of 88.6% at internal validation. The model included age, gender, unintentional weight loss, smoking and alcohol history and a refined list of positive and negative symptoms of HNC. Two recommended referral thresholds were introduced based on sensitivity and specificity combinations for a 2ww referral (cut-off: 7.1%; sensitivity: 85%, specificity: 78.3%) and urgent referral (cut-off: 2.2%; sensitivity: 97.1%; specificity: 52.9%). The AUC remained high at external validation (AUC: 83.96%; sensitivity:70%; specificity: 81%). The use of the HaNC-RC v.2 resulted in a reduction of the 2ww appointments by 70% during the first wave of the COVID-19 pandemic. Of the total of 256 cancers, 73.2% were seen in the high-risk group (2ww referral) and 16.5% in the moderate-risk group (urgent referral). These figures were much improved compared to those based on GP triaging using the national referral guidelines (59.9% and 25.4%, respectively) in the Glasgow region, without affecting the total numbers seen in each clinical setting.

Conclusions: This study achieved its aim and objectives of developing and validating an updated version of a previously designed HNC risk calculator. The HaNC-RC v.2 has a much-improved AUC that remained high at external validation, and it could be used as a triaging aid for head and neck referrals in secondary or primary care pathways.

1 Introduction

This thesis will present the results of designing an early detection algorithm for head and neck cancer (HNC) in the UK. This research idea sprung from an earlier work of mine, having developed and published the first symptom-based head and neck cancer risk calculator in 2016. Its design was based on a retrospective database from a cohort of patients seen in urgent suspected cancer head and neck clinics in the UK (Tikka, Pracy and Paleri, 2016). The calculator had a satisfactory predictive power and was validated in a further retrospective cohort of patients maintaining a good prediction level (Tikka, Paleri and MacKenzie, 2018). Nevertheless, there were limitations in the design and performance of the tool compared to other available risk calculators for common cancers (Steyerberg *et al.*, 2004;Steyerberg, 2019). The retrospective methodology, lack of a priori sample size analysis, and less than 80% predictive power prevented it from gaining popularity and being considered for implementation within the early HNC diagnosis pathways in the UK. A summary of this calculator's performance and drawbacks will be covered in section 1.6 of this chapter and discussed further in section 2.3.3.12 of the literature review chapter. This thesis work aimed to develop an updated version of the previously designed head and neck cancer risk calculator, using a robust methodology to eliminate the limitations noted in the first version of the tool.

This introductory chapter begins with a brief presentation of cancer statistics within Europe. Then it focuses on the UK cancer figures and current early diagnosis strategies. This information is essential to understanding the cancer burden and early cancer detection strategies within the wider part of the world and the country within which an early cancer detection solution is to be introduced. HNC statistics, diagnostic challenges, and the rationale behind developing a HNC triage tool for early HNC diagnosis are then introduced, followed by this thesis's aim and research questions.

1.1 Cancer statistics in Europe

In Europe, approximately 4 million new cancer cases are diagnosed every year. Cancer is the second most common cause of death after cardiovascular disease, with more than one in four deaths being due to cancer, with an estimated 1.9 million deaths per year. Studies have shown a larger share of cancer-related deaths in wealthier European countries. Furthermore, in the UK, Denmark, France and the Netherlands, cancer deaths were higher than those due to cardiovascular disease (Hofmarcher *et al.*, 2019). Nevertheless, the 5-year relative cancer survival has steadily increased for all European countries according to the latest EUROCARE study, but variations were noted, with Eastern European figures being lower than the rest of Europe. Overall, for all countries, the 1-year relative survival rate varied from 58.2% to 81.1% for all cancers. The cancer survival figures for UK and Ireland were intermediate for most cancers (rectal, breast, prostate, skin melanoma, non-Hodgkin's lymphoma) and low for kidney, stomach, ovarian, colon and lung cancer, as seen in Figure 1-1 (De Angelis *et al.*, 2014).

Similarly, low survival rates have been noted in the UK compared to the rest of Europe in previous EUROCARE studies spanning over 20 years (Coleman *et al.*, 2003;Berrino *et al.*, 1995;Berrino *et al.*, 2007;Berrino *et al.*, 1998). The following section will cover this phenomenon and more details about UK cancer statistics. This information is essential in understanding why there is a need for improvement in the cancer detection pathways in the UK and the rationale of this thesis work, that is, the development of an assessment tool for the early identification of symptomatic patients with HNC.

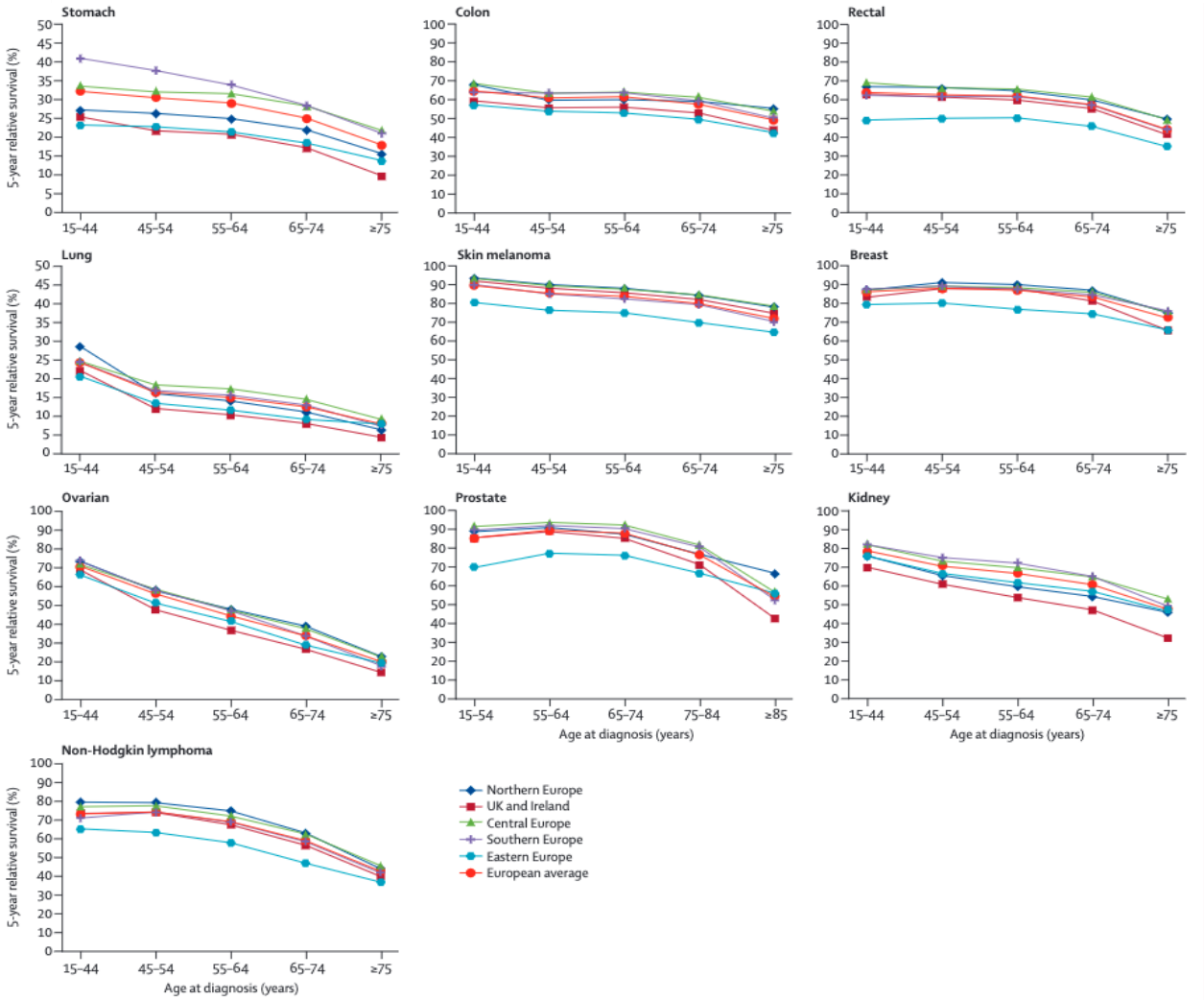


Figure 1-1. Age-specific 5-year relative survival of cancer patients in Europe in 2000-2007. Source: De Angelis et al., 2014

1.2 Cancer statistics in the UK

Cancer incidence in the UK is among the highest worldwide, and within Europe, being ranked 16th and 11^{th,} respectively, that means that the UK cancer incidence rate for all cancers is higher than 90% of the world rankings. An estimated 2.5 million people live with cancer in the UK, which is predicted to rise to 5 million by 2040 (Maddams, Utley and Møller, 2012). Every day an average of 1,200 new patients are diagnosed with cancer in the UK, with around 450,000 new cases per year, according to the most recent statistics from 2020. Over half of these are due to breast, prostate, lung and bowel malignancy (Figure 1-2) (IARC, 2020).

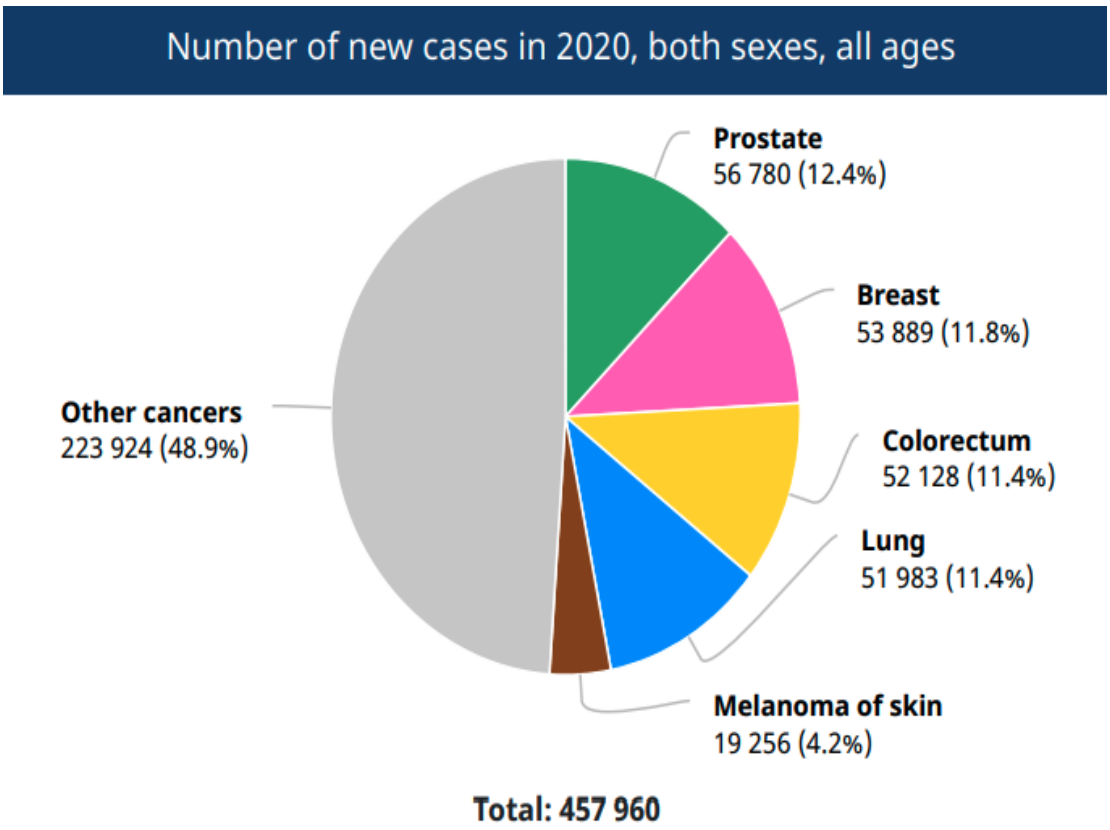


Figure 1-2. UK cancer incidence per most common cancer sites. Source: The Global Cancer Observatory.

Cancer mainly affects the older population, with a third of cases being over 75 years of age. Over 166,000 patients with cancer die every year in the UK due to the four most common cancers mentioned above. Over 54% of mortality is in patients above 75 years of age, with the highest rates in patients over 90 years old. It is estimated that one in two people will be diagnosed with cancer during their lifetime in the UK (Ahmad, Ormiston-Smith and Sasieni, 2015).

Cancer survival is improving, with doubled survival rates over the past 4 decades, but it still remains low, with an overall 50% 10-year survival reported. The cancer-related death rate per 100,000 population has slightly decreased since an initial steep upward trend and a peak in the 1990s and has now stabilised for females with a slight increase noticeable for males in the last decade (Figure 1-3). In accordance with the worldwide data, cancer incidence varies with socioeconomic status, with some cancers having a higher incidence in the more deprived UK regions. These include cancers of the lung, larynx, oesophagus, stomach, bladder, kidney, oral cavity, pancreas, and cervix. On

the other hand, the incidence of breast, prostate and skin cancer is higher in the less deprived groups (NCRI, 2010).

The reasons behind the high cancer incidence in the UK compared to the worldwide statistics and the high reported mortality compared to other Western European countries will be discussed in the following section. This review will help better understand potential gaps in cancer management strategies and the importance of conducting research focused on optimising current pathways.

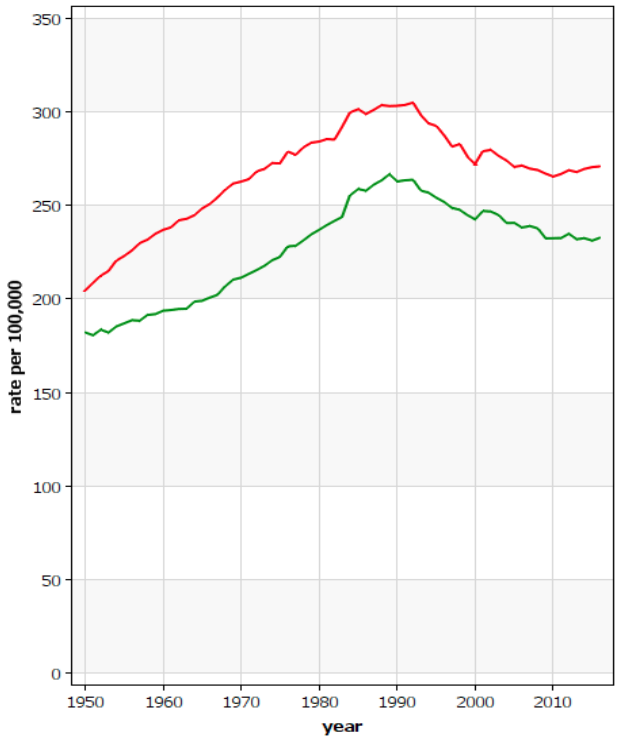


Figure 1-3. All cancer mortality rates (per 100.000) in the UK for males and females. Source: International Agency for Research on Cancer (IARC)

1.3 Factors related to the high cancer burden in the UK

The high mortality rates in the UK, compared to the rest of Western Europe, have been extensively investigated and debated over the past few years. It was initially hypothesised that the difference could be the result of mistakes in cancer registration (Beral and Peto, 2010), but subsequent studies have shown that this is not the case as it would have required missing data on cancer survival of over 60% of cancer cases (Woods *et al.*, 2011).

A review of the literature on the potential reasons behind the high mortality rates for the most common cancers highlights issues with delayed cancer presentation with patients waiting longer to be seen by a specialist; low uptake of screening programs for some cancers; reduced provision of more aggressive treatment regimes, perhaps because of the advanced disease stage at diagnosis precluding curative intent treatments (Coleman *et al.*, 2011). Other issues highlighted were the increased number of co-morbidities and unhealthy lifestyles of the UK population with a high rate of obesity, smoking and alcoholism. Differences in cancer biology could also be part of the problem, with more aggressive cancer behaviour seen in those diagnosed with cancer in the UK (Richards, 2009). Delays in early cancer detection and lack of cancer symptom awareness have been highlighted as part of the problem (Møller *et al.*, 2010), as well as non-standardised treatment and management pathways (Thomson and Forman, 2009). Moreover, it is noted that the use of cancer medications, including immunotherapy, across 7 common cancer types has been consistently lower than in other developed countries, which highlights room for improvement in treatment provision through policies and evidence-based, cost-effective cancer care (Hofmarcher *et al.*, 2019).

The results of the previous EUROCARE studies have informed changes in UK primary care referring systems to ensure timely referral of a patient with suspected cancer symptoms and generated national awareness and early diagnosis initiatives and international cancer benchmarking partnerships (DH, 2011;DH, 2007;Butler *et al.*, 2013). People worried about cancer due to the onset of relevant signs and symptoms

are usually first seen in the primary care setting by their general practitioners (GP). In the UK, guidelines are now in place by NICE in England (NICE, 2021) and SIGN in Scotland (NHSScotland, 2019), helping GPs to refer patients urgently to the hospital for specialised assessment when red flags symptoms for cancer are identified following patients' history and examination. All patients with red flag symptoms are referred urgently to secondary care and should be seen within 2 weeks from referral (NICE, 2021). Similar primary care referral pathways and targeted times from referral to first hospital appointment are present worldwide. However, significant variation exists as to how the referral is made, associated costs per referral, primary care clinic set-up, and availability of investigations outside the hospital setting. These could cause variations in the staging of cancer diagnosis, which may explain some of the survival variations seen in the EUROCARE studies (Harris *et al.*, 2018;Brown *et al.*, 2014).

Nevertheless, even when studies have focused on referral outcomes in countries with similar primary care set-ups, poorer outcomes were seen for UK regions, with significant delays between the primary care referral date and first hospital appointment as well as the delay from first hospital appointment and diagnosis (Murchie *et al.*, 2012). Despite actions being in place to promote early cancer diagnosis and increase public awareness, almost half of the patients with cancer are diagnosed at a late stage in the UK. Although cancer pick-up via the urgent 2-week pathway has improved over the past decade, increasing from 41% to 52% within the last decade, it is still considered low (Round *et al.*, 2021). Screening programs are available for breast, prostate and bowel cancer with variable uptake over the past years, with breast and cervical screening coverage slowly but steadily falling (Round *et al.*, 2021;Hamilton, 2010). A detailed presentation of the different cancer referral pathways across the world, cancer symptoms and diagnostic tools will be presented in detail in the next chapter of this thesis.

When cancer suspicion is established as a possible cause of a patient's symptoms, the 31-day target needs to be met in the UK. This time target is set as the UK government's standard for completing all necessary investigations to reach a final diagnosis and, with a 62-day target for initiation of treatment. Recent cancer statistics show that even

though all patients are seen within 2 weeks from GP referral, the 31-day-target is currently not met in Northern Ireland and Wales, and the 62-day-target is currently not met by any of the UK countries (*International data analysis of Cancer Incidence statistics for Egland in 2015*, 2015;NICR, 2016). This failure may reflect why the UK has worse cancer mortality outcomes than other nations, as the earlier treatment is initiated, the more likely it is to have a positive clinical response on overall outcomes, including mortality (Neal, 2009). National cancer patient experience surveys are reporting good experience with hospital cancer teams, but many patients state that they are dissatisfied with the care received at the start of the journey from their GP doctors and other GP practice staff they have seen (*Cancer patient experience survey, England. 2016,* 2016).

Moreover, studies have shown that almost half of the patients with cancer are not picked up by GPs as potential cancer cases; hence they are not referred via the urgent suspected cancer pathway. This delay can add up to 12 weeks of additional waiting time for a cancer patient to be seen in the hospital before any investigation is initiated (Lewis, Le Jeune and Baldwin, 2005). Additionally, it is noted that since the introduction of the guidelines, the number of urgent suspicion of cancer referrals has increased. However, the cancer conversion rate of urgent cancer appointments continues to drop, despite a steady rise in overall cancer incidence. Hence, more patients are referred with potential cancer, but fewer patients are diagnosed with cancer via the urgent pathway, with an increasing number of cancer cases diagnosed via other routes (Round *et al.*, 2021;Lewis, Le Jeune and Baldwin, 2005). This mismatch could result from a wrong interpretation of the guidelines or abuse of the urgent cancer referral slots being offered to the 'worried-well' patients without significant cancer symptoms at the expense of cancer patients who may not stress the severity of their symptoms during the primary care consultation enough.

This thesis will explore a potential solution to the currently problematic triaging of cancer referrals, focusing on the suspected HNC referrals in the UK. Similar attempts have been made in other cancers, which will be discussed in-depth in the literature review part of the thesis. In the following section, the presented HNC statistics will

show the currently very low cancer detection rates at an early disease stage and the reasons behind the need for a tool to help triage the suspected HNC cases more effectively.

1.4 Head and neck cancer definition and statistics

Head and neck cancers (HNCs) include malignancies of the larynx, pharynx, thyroid, lip and oral cavity, nasal cavities and paranasal sinuses, also including skin malignancies of the head and neck (HaN) region and cancers of the salivary glands and ear (Deschler, Moore and Smith, 2014). The most common cancer subsite is the larynx, followed by the oropharynx. The majority of cancers are squamous cell carcinomas originating from the squamous cell lining of the upper aerodigestive tract. Rarer cancer types include lymphomas, sarcomas, adenocarcinoma, melanomas and other rare cell types usually arising from the salivary gland cells (Mody *et al.*, 2021). It is the 7th most common cancer worldwide and accounts for 1.1 million new cases per year and 4.1 million prevalent cases, resulting annually in 500.000 deaths (Vos *et al.*, 2017). Putting these into context, approximately 19 million people are diagnosed every year with cancer worldwide, with as many as 10 million deaths being recorded annually worldwide as a result of cancer (IARC, 2020).

HNC is part of the rare cancers group that includes all cancers with an incidence of less than 6 per 100.000 population (Figure 1-4). 5-year survival has been found to be worse for these cancers compared to more common cancers, being 55% for rare cancer versus 75% for common cancer in the USA between 2009 and 2013 and 49% vs 63% during 2007 in Europe (Gatta *et al.*, 2017). It was also found that rare cancers are diagnosed in a more advanced stage (Mathoulin-Pélissier and Pritchard-Jones, 2019). This trend is indeed the case for HNC as well. Large multicentre studies from the USA and Europe note that the majority of HNC cases are diagnosed in disease stages 3 and 4 (Abrahão *et al.*, 2020;Guizard *et al.*, 2017;Gatta *et al.*, 2015).

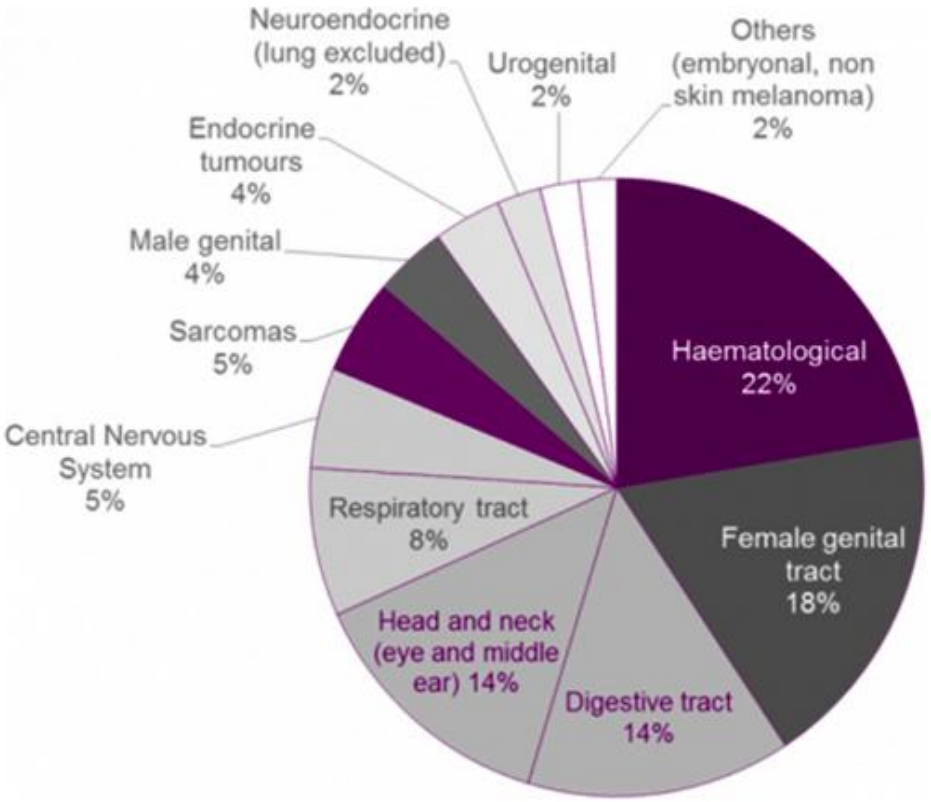


Figure 1-4. Rare cancer groups in the adult population. Source: www.rarecancereurope.or

HNC is the 8th most common cancer in the UK, with its incidence continuing to rise. An average of 12,000 patients are diagnosed every year with a HaN malignancy, accounting for 3% of all cancers. It is most common in people aged 70 to 75, with a fifth of the total cases being over 75 years of age. It is more common in males, being the 4th most common cancer. On the other hand, it is a relatively rare cancer in the female population, ranked 13th among all cancers in females. Nevertheless, an increase in HNC incidence has been noted over the last 3 decades, with an overall rise of 33%, with a more pronounced increase for females when compared to males (43% vs 23%) (*Internal data analysis of cancer incidence statistics for England in 2015.*, 2015; *Internal data analysis of Cancer Incidence for Scotland in 2015*, 2015; *Internal data analysis of Cancer Incidence Statistics for Wales in 2015*, 2015; *Internal data analysis of Cancer Incidence Statistics for Northern Ireland in 2015*, 2015). A projected 50% increase in overall cancer incidence is also expected worldwide, which is based on the expected population growth and ageing and an increase in risk factors prevalence (Sung *et al.*, 2021).

HNC is a cancer strongly linked with deprivation, with its incidence being 64% higher in females from the most deprived UK regions and 101% higher for the most deprived male population. Socioeconomic deprivation is also linked with common cancers, but this relationship is not uniform. This has been explored in studies exploring variations in cancer incidence and mortality for different social deprivation indices worldwide, such as the human development index (HDI) (Bray *et al.*, 2012;Thun *et al.*, 2017) and the education and income index (EDI) (Cao *et al.*, 2017). Lip and oral cavity malignancy is commoner in low HDI countries due to the high incidence in Asia of betel nut chewing (Thun *et al.*, 2017), but no clear association is found between EDI and incidence or mortality for these cancers (Lortet-Tieulent *et al.*, 2020). For cancers of the larynx, there is a decline in incidence and mortality as EDI increases, whereas the opposite is seen for the incidence of thyroid cancers without much effect on their mortality. HNC is more common in males in older age groups, and it is linked to smoking and alcohol consumption, human papillomavirus for cancers of the oropharynx, Epstein Barr virus for nasopharyngeal cancers, poor diet and poor oral hygiene, exposure to chemicals and ultraviolet radiation secondary to long periods of sun exposure (Mehanna *et al.*, 2010).

The latest EUROCARE study found higher 5-year relative survival figures for the UK compared to the rest of Europe for all HNC subsites with small variations and improved overall survival over time compared to previous iterations of the EUROCARE studies (Gatta *et al.*, 2015). This is in contrast to the survival statistics for other cancers, showing worse survival rates for cancer in the UK compared to the rest of Europe (Coleman *et al.*, 2003;Berrino *et al.*, 1995;Berrino *et al.*, 2007;Berrino *et al.*, 1998;De Angelis *et al.*, 2014). More specifically, the average European figure for laryngeal cancer is 59% for men, compared to 75% in Northern Ireland, 63% in England and Scotland and 59% in Wales. Figures for females are all very close to the European average. For oropharyngeal cancer, the UK's 5-year relative survival is 44%, the same as the European average for men. The figures are the same for females as the mean Europe figure of 50%, with Scotland slightly lower at 44% (De Angelis *et al.*, 2014;Gatta *et al.*, 2015).

HNC mortality is ranked 16th among all cancer deaths in the UK. An average of 4,100 HNC patients die every year, attributing 2% to the total number of cancer deaths (www.ons.gov.uk, 2016). Over a third is in people over 75 years of age. Even though statistics from 3-4 decades back had shown a steady decrease in HNC mortality, data from the last decade has shown an overall 17% increase in mortality, being equal for males and females but greater for people living in deprived UK regions (www.ons.gov.uk, 2016). The rise in mortality can be partially explained by the rising death rates related to human papillomavirus (HPV) associated oropharyngeal cancers (Siegel, Miller and Jemal, 2020). The percentage of oropharyngeal cancer associated with HPV rose from 10% in the 1980s to over 70% in recent epidemiological studies, with the incidence exceeding HPV-related cervical cancers (Pan, Issaeva and Yarbrough, 2018). Primary prevention with HPV vaccination programmes has reduced HPV-related cancer incidence. However, it is estimated that the HPV-related oropharyngeal cancer epidemic will continue until 2060 due to the initially low vaccine uptake with significant geographical variations, the long latency of cancer presentation following initial exposure, as well as variation in the vaccine program policies that only recently included male vaccinations (Pan, Issaeva and Yarbrough, 2018). The rise in HNC mortality has also been linked with advanced-stage of cancer at the time of diagnosis. The 5-year survival is 84% for early-stage HNC, dropping to 39% for late-stage HNC (Siegel et al., 2019). In the last decade, approximately less than a third of HNCs have been diagnosed at an early stage, with the incidence of late-stage disease rising and affecting more the socio-economically deprived population (Siegel, Miller and Jemal, 2020;Thompson-Harvey *et al.*, 2020).

The survival rates vary across the different HNC subsites, with the 10-year survival ranging from 19% to 59% in reports from the last 10 years. Hypopharyngeal cancer has the worst 5-year survival, ranging from 20-34% worse for those over 70 years of age. On the contrary, the 5-year survival for oropharyngeal cancer is as high as 82% for the 15 to 49 age group (www.ons.gov.uk, 2016). A more recent Scottish study has shown 71% 5-year survival for laryngeal cancer and 32% for hypopharyngeal cancer (Douglas *et al.*, 2018). Aside from the cancer site, the survival was affected by the number and type of symptoms at the initial presentation following adjustments for age,

stage of disease and cancer site. Median survival was 5.3 years when only 1 symptom was present, reduced to 1.1 years in the presence of 3 symptoms. The weight loss symptom had the worst prognosis, with a median survival of 0.8 years (Douglas *et al.*, 2018). In the absence of benefit in mortality by asymptomatic population screening (Moyer, 2014) these results highlight the importance of patients' and GPs' education in symptoms related to HNC, as well as the development of guidelines to enable prompt identification of patients with potential HNC based on the symptom history profile at an early disease stage (Luryi *et al.*, 2014). Current strategies to account for these are presented in the following section, alongside their outcomes in early HNC detection.

1.5 Current pathways to HNC diagnosis in the UK and treatment strategies

To date, there is no available established screening test for HNC as opposed to other common malignancies such as cervical and breast cancer (Vineis and Wild, 2014). All patients presenting with red flags symptoms for HNC in the UK, as assessed by the primary care doctors, are referred within 2 weeks to the hospital for specialised assessment and further investigations to rule out malignancy. This pathway is known as the 2-week-wait pathway (2ww) in England and Wales, and the urgent suspicion of cancer (USOC) pathway in Scotland, with slight variation in the included symptoms, which will be covered in the literature review (NICE, 2021;NHSScotland, 2019). Similar primary care referral pathways and targeted times from referral to first hospital appointment are present worldwide. However, significant variation exists as to how the referral is made, associated costs per referral, primary care clinic set-up, and availability of investigations outside the hospital setting. These could cause variations in the staging of cancer diagnosis, which may explain some of the survival variations seen in the EUROCARE studies (Harris *et al.*, 2018;Brown *et al.*, 2014)

Due to the HNC rarity, an average GP would only see 1 new case of HNC every six years, which can make diagnosis difficult and result in a delay in referring urgently as a potential malignant case (NICE, 2004). The most common presenting symptoms are difficulty swallowing (dysphagia), pain in swallowing (odynophagia), otalgia with normal otoscopy due to referred ear pain, hoarseness, mucosal ulceration and growths

in the mouth pharynx, oral and neck pain, weight loss and neck lumps (Mody *et al.*, 2021). Patients can also present with non-specific symptoms and present late to their GP due to a lack of awareness of possible symptoms associated with HNC. Previous audits from England and Ireland showed that the majority of cancers (60%) are diagnosed at a disease stage III or IV, with 21% of HNC patients visiting their GP more than twice before being diagnosed with cancer (*Cancer Patient Experience survey, England.*, 2016), with these numbers being in accordance with other worldwide multicentre studies (Gatta *et al.*, 2015;Abrahão *et al.*, 2020).

A systematic review of the literature on the efficacy of the HNC 2ww pathway is in agreement with studies from other cancer sites, showing that although the number of urgent cancer referrals has increased, the cancer yield from these referrals is low (Round *et al.*, 2021;Lewis, Le Jeune and Baldwin, 2005). For HNC, this has dropped further over the years from 8% to 6% in more recent studies. Furthermore, 60% of HNC are diagnosed by routes other than the 2ww pathway (Langton, Siau and Bankhead, 2016;NCRAS, 2016). With an average of 100 000 HNC urgent suspicion of cancer (USOC) referrals annually in the UK (*Delivering cancer waiting times: a good practical guide.*, 2015) and an annual HNC incidence of 12 000 (*Internal data analysis of cancer incidence statistics for England in 2015.*, 2015), a UK-wide 2ww HNC conversion rate of 4.3% can be extrapolated.

Treatment recommendations depend on the cancer stage and subsite location when the cancer diagnosis is established. Single-modality treatment, with either surgery or radical radiotherapy, is preferred as there is increased morbidity associated with multimodality treatment (Simon *et al.*, 2020). This can be achieved for early-stage disease, but multimodality treatment with either a combination of radiotherapy and chemotherapy or initial surgery by means of total laryngectomy, followed by postoperative adjuvant chemoradiotherapy, is usually required for advanced-stage laryngeal HNC with adverse histopathological features following surgical resection (Pignon *et al.*, 2009). The latter treatments significantly affect the short-term and long-term quality of life of the HNC survivors. There are limited options for salvage treatments if recurrence occurs, with studies showing a significant difference in

survival rates between early and late-stage HNC of all subsites (Bernier *et al.*, 2004;Weber *et al.*, 2003).

For early-stage HNC, new surgical technologies are also now available. Transoral robotic surgical resection is an alternative to radiotherapy for early-stage cancer of the tongue base, supraglottis and pharynx (Byrd and Ferris, 2016;White, 2013). Studies show equivalent local control and long-term survival rates, with improved short-term and long-term morbidity (de Almeida *et al.*, 2014;Morisod and Simon, 2016). Moreover, early evidence shows promising results from a de-escalation of the intensity of adjuvant treatment if this is required post-robotic-assisted resection. More morbid salvage surgical and systemic treatment options remain available if recurrence occurs. The former, less anatomically destructive and morbid options, will not suffice for the treatment of advanced HNC at the time of initial presentation, where a more complex multimodality treatment approach is usually required from the outset (Ferris *et al.*, 2020;Swisher-McClure *et al.*, 2020).

Hence, to summarise, despite the advances in the treatment of HNC and the development of referral pathways and day-targets to encourage prompt patient referral, diagnosis and initiation of investigations and treatment, the cancer statistics for HNC remain suboptimal when compared to commoner cancers. The suboptimal outcomes include the cancer stage at presentation and the survival rates, with the two being directly linked. Over half of the HNC cases are diagnosed in an advanced disease stage which reduces the chances of treatment with curative intent, with many patients' treatment being of palliative intent by the time they are seen in the hospital setting (Goy *et al.*, 2009). Cancer survival is directly linked to the stage of disease; hence the earliest the cancer is found and treated, the least the associated patient morbidity is, and the greater the chance of improved mortality rates (Siegel et al., 2019). The 2ww referral guidelines for HNC have a low cancer conversion rate, despite the number of yearly referrals steadily increasing over the past 10 years, with only a third of HNC being diagnosed via the 2ww pathway. Therefore, there is an urgent need for improvement in the current pathway to identify and refer patients with a high risk of having HNC and, in turn, improve HNC survival outcomes.

The focus of this thesis is to design a HNC risk calculator informed by the literature and using a robust methodology which will be described in the methodology chapter. The design of a HNC risk calculator will be based on currently recognised HNC risk factors and knowledge available from other common cancers. In the following section, currently available risk calculators for common cancers will be introduced, their design process and current implementation, followed by any current knowledge of HNC risk models and how the results of this thesis will add to the available literature.

1.6 Cancer risk calculators

Early referral of patients with potential cancer is of paramount importance for prompt diagnosis and treatment and improved long-term outcomes, including cancer survival (Neal, 2009). Most patients present initially in the primary care setting with symptoms suggestive of cancer where, following history and examination taken by general practitioners, a referral to the hospital is made for those deemed high risk for malignancy (Emery *et al.*, 2014). Multiple appointments prior to a referral to a hospital can delay diagnosis, especially for rare cancers or those with non-specific symptoms (Lyratzopoulos *et al.*, 2012). A review of the literature on potential interventions for reducing diagnostic errors in primary care suggests using technology-based interventions such as computer-assistive diagnostic aids, decision-support algorithms and text message alerting (McDonald *et al.*, 2013). In recent years, there has been a drive to develop risk calculators designed to identify cancer at early stages. Risk calculators not only have the potential to contribute to the earlier diagnosis of cancers but could also lead to service delivery improvements (Usher-Smith *et al.*, 2015). This is a potential area of great opportunity for improvement in patients' cancer journeys from initial presentation in primary care, to diagnosis in the hospital setting and initiation of treatment (Niederhuber, 2006).

The first attempts to quantify the risk of malignancy in patients presenting with a suspicion of cancer date back to the late 80s - early 90s. Gail et al. in 1989 established a prediction tool for breast cancer based on age, past medical history and family history (Gail *et al.*, 1989). Logistic modelling was used to establish the likelihood of ovarian

cancer based on the patient's age and ultrasonographic tumour characteristics. The collection of this information required the patient with suspected cancer to be seen by a specialist for imaging before all the necessary information was available to calculate the cancer probability (Minaretzis *et al.*, 1994). Many years after these first attempts at cancer prediction, a landmark study on prostate cancer established a prostate cancer risk calculator based on patients' age and race, the value of the prostate-specific antigen (PSA), family history of prostate cancer, digital examination findings and previous prostate biopsies. It is known as the Prostate Cancer Prevention Trial (PCPT) (Thompson *et al.*, 2006).

At present, several risk calculators are available for common cancers, such as prostate, lung or ovarian cancer, which have been externally validated and are recommended to aid prompt referral of high-risk individuals to specialist clinics for further assessment. Most of these cancer risk calculators have achieved high discrimination performance with predictive power measured as the area under the receiver operating characteristic curve (AUC) of over 0.8 (Steyerberg, 2019). However, most of these nomograms require blood tests and radiological findings—in addition to patients' symptoms and demographics—to calculate cancer probability, which potentially limits their widespread adoption in primary care settings (Usher-Smith *et al.*, 2015). On the other hand, there are also examples where risk can be established solely based on symptoms and demographics, such as for lung and colorectal cancer (Gray *et al.*, 2016;Williams *et al.*, 2016;Hippisley-Cox and Coupland, 2013a;Hippisley-Cox and Coupland, 2013b).

Although early diagnosis cancer risk calculators have been available for the last 10-20 years for common cancers, prediction models for HNC have only recently emerged. The first symptom-based head and neck cancer risk calculator (HaNC-RC) was published in 2016 (being the research output of my dissertation for the award of a Master's degree in Medical Statistics). It was based on patients' symptoms, signs and demographics using data from patients referred to the hospital via the urgent suspicion of cancer route in England (Tikka, Pracy and Paleri, 2016). It was subsequently externally validated with a cohort from a different UK region with high overall

prediction power and sensitivity and specificity combination. Its prediction power was 0.77 and 0.81 in the development and validation cohorts (Tikka, Paleri and MacKenzie, 2018). Nevertheless, the retrospective design of both the development and external validation studies of the tool limited the generalisability of the results, with information on important risk factors associated with HNC missing from the database (such as smoking and alcohol status). Moreover, the lack of a standardised uniform method for data recording could introduce collection bias in the recorded symptoms included in the calculator. Finally, the AUC, despite being satisfactory, was lower than the level expected in the cancer risk calculators literature, which is expected to be over 0.8 (Steyerberg, 2019). Therefore, it was evident that there was scope for an update in the previously designed HNC risk calculator using a robust methodology and having the first version of the tool as groundwork to achieve outcomes more in line with other common cancer risk calculators.

A few years later, another symptom-based HNC risk calculator was proposed, applying a different symptom combination and also including demographics and smoking and alcohol information (Lau, Wilkinson and Moorthy, 2018). It was also based on a cohort of urgent suspicion of HNC referrals similar to the previously mentioned tool (Tikka, Pracy and Paleri, 2016). Although the prediction power was high, the sensitivity was low at 31%, with high false-negative figures in their internal validation cohort (Lau, Wilkinson and Moorthy, 2018). My previously designed HNC risk calculator (Tikka, Pracy and Paleri, 2016) and the calculator by Lau et al. (2018) are the only two reported HNC risk calculators for symptomatic patients, both being based on multivariate logistic regression analysis. Artificial intelligence methods have also been attempted in the development of HNC risk calculators, with logistic regression being found again to be the method having the higher predictive power using variational inference that approximates probability densities for each variable through optimisation (Moor, Paleri and Edwards, 2019).

Current trends in the development and refinement of risk calculators learned from reviews of nomograms for common cancers lean towards validation of existing risk calculators, combined with continuous improvement through further iterations for

increased predictive power instead of continuous generation of new prediction models (Louie *et al.*, 2015). It has been noted that many models are developed for the same outcome of interest when previous models do not perform well on external validation. This not only leads to confusion as to which calculator to use in clinical practice but also causes the loss of all the previous information captured from the older tools (Steyerberg *et al.*, 2004). Any new model will again require external validation, which will likely lead to another replacement of the existing model with a newer one. Therefore, it is suggested that an alternative solution is the redevelopment of the existing calculators, assessing for adjustment of the intercepts or some covariates using the external validation dataset (Louie *et al.*, 2015;Janssen *et al.*, 2008;Steyerberg *et al.*, 2004). This method has been employed in prostate cancer risk calculators, such as the addition of the prostate cancer antigen 3 parameter in the prostate cancer prevention trial risk calculator (Ankerst *et al.*, 2019). No symptom-based only cancer risk calculator was identified to have followed this process. More extensive model revision with recalculation of the model intercept, addition of new variables, and adjustment of already available covariates with recalculation of the regression coefficients should be attempted when large cohorts are available (Steyerberg *et al.*, 2004). This is the approach most commonly used in the available cancer risk calculators, which are based on a combination of symptoms, signs, family history variables, as well as the results of specialised blood tests, radiological examinations, and biopsy results. Prime examples are prostate cancer risk calculators (Ankerst et al., 2014; Ankerst et al., 2018) and breast cancer risk tools (Berry et al., 1997; Fischer *et al.*, 2013).

1.7 Aim of project and research questions

Taking all the above information into consideration, this research work aimed to further increase the predictive power of the HaNC-RC (Tikka, Pracy and Paleri, 2016) by updating this previously designed HNC risk calculator, assessing the potential for inclusion of other significant HNC symptoms and the addition of any relevant social history factors based on a large prospective patient cohort. The development phase included internal validation of the model and assessment of performance statistics, and

external validation in a separate prospective cohort of patients. The research questions were as follows:

1. Are there any new variables that can be added to the existing HNC risk calculator (Tikka, Pracy and Paleri, 2016) and/or could its current variables be refined based on a new large dataset to improve the risk calculator diagnostic efficacy?
2. How does the new, updated version of the HNC risk calculator perform in triaging a new cohort of patients referred to HaN clinics across the UK?

As mentioned in the COVID-19 Impact Statement at the beginning of my thesis, my PhD thesis initially had one research question (question 1, as seen above), assessing for any new variables or refined current variables that can be included in an updated version of the HNC risk calculator. Therefore, only the development phase of the calculator was initially planned for my PhD research. Nevertheless, an opportunity arose to also externally validate the tool using a prospectively collected database of patients being triaged with the new version of the calculator. This became possible as the tool was used as a triaging aid for HNC referrals during the difficult times of the first wave of the COVID-19 pandemic. The ENTUK and INTEGRATE organisations led the design and delivery of this UK-wide collaborative service evaluation work. The anonymised database later became available to me for analysis hence resulting in the addition of the second research question in my thesis (as listed in the paragraph above).

1.8 Thesis outline

The first introductory chapter of this thesis presented a general outline of important cancer statistics, followed by a focused presentation of HaN malignancies statistics. The incidence, mortality, and referral pathways for suspected HNC malignancy were discussed, and a brief overview of the current use of available cancer risk calculators was provided. The rationale for the need for change in the current referral pathway and the potential use of a risk calculator for this purpose was covered, followed by stating the aim and objectives of this research work.

In the chapters to follow, a detailed review of the literature in chapter 2 will cover the referral pathways for all cancers, their red flag symptoms and how the pathways vary across the UK and compare to international cancer networks. A focused review of the above topics specifically for HNC will follow. The literature review will then focus on an overview of the use of risk calculators for triaging and diagnosis of common cancers.

In the third chapter, the methodology around the development of cancer risk calculators will be covered, and the different statistical models and artificial intelligence networks will be explained, discussed, and compared. This will be followed by a detailed presentation of the methodology used in this research work.

The fourth chapter includes results of the statistical analysis covering the development and internal validation of the updated version of the HNC risk calculator using logistic regression and random forest analysis, as well as its external pan-UK validation during the COVID-19 pandemic.

The fifth chapter of this thesis covers a critical discussion of the results and a comparison with the available literature. It will also present how the output of this research work has been used already and reported in other research works, including a critical discussion of the challenges faced by its use for patients' triaging using

telephone clinics during the first wave of the COVID-19 pandemic. Future work based on the output of this research is outlined, as well as its limitations.

The thesis concludes with a summary of the findings, results and salient discussion points and remarks.

2 Literature review

Reviewing the literature carefully prior to starting the design process of the HNC risk assessment tool was of paramount importance as it is crucial to understand the context within which the tool could be potentially used in the future to aid early HNC detection. Even though the risk calculator was designed and validated internally and externally using UK patient cohorts with the scope of being used within the UK health care system, the long-term vision is that it can also be incorporated into the diagnostic pathways of countries other than the UK. Nevertheless, this can be challenging as not all healthcare systems share similar models and cancer prevention and detection strategies to allow such tools to be incorporated into their protocols. The initial triaging of patients with potential cancer is performed by GPs acting as gatekeepers to secondary care in many, but not all, countries, with the responsibility of cancer screening and early diagnosis resting in other levels of health care, that is secondary and tertiary care including private health care provision.

Even within the countries where GPs act as gatekeepers to secondary care, it is important to understand how they currently perform cancer triaging for common cancers, what are the main problems, if any, potential limitations and clinicians' views of the current pathways, and how this relates to HNC. With that knowledge, the HNC risk calculator can then be designed, taking into consideration not only the primary care set up within the UK but also the worldwide primary care links to secondary care. This can lead to the design of a tool that has the flexibility of being used in different clinical settings according to the clinical needs and circumstances, such as being used either in primary care for triaging or being implemented directly in secondary care clinics, being used not only by doctors but also by appropriately trained allied health care professionals.

In the first section of the literature review, current healthcare models across the globe will be presented, including a focused review of the role of GPs in each of the healthcare systems, where the responsibility for cancer screening, early diagnosis, and

any problems resulting in diagnostic delays lies. This will be followed by international and then UK-specific cancer early detection guidelines for common cancers and any practice variation and barriers to early diagnosis prior to focusing on the HNC related literature.

2.1 Context

Early cancer detection is important for successful treatment, and this is dependent on effective processes being in place to ensure patients are seen promptly by the appropriate secondary care specialists (Sung *et al.*, 2021). This section will describe the various approaches that exist to identify patients in primary care with a possible cancer diagnosis across different healthcare systems for onward secondary care referral. Barriers to early cancer detection will also be explored. This information is important for gaining an understanding of how the introduction of a cancer triaging intervention could be incorporated into the available healthcare systems and also to what extent it could help reduce some of the barriers to early cancer diagnosis.

2.1.1 Worldwide health care models and the role of primary care in cancer detection

Referral to secondary care for patients with symptoms suggestive of cancer can be filtered by primary care doctors acting as "gatekeepers" to secondary care or can be patient-initiated depending on the referral system of each country. Often, the cancer referral gatekeepers are general practitioners (GPs) who assess patients in the community, perform the initial examination, history, and baseline bloodwork investigation, and then refer onwards to secondary care patients deemed high risk for malignancy (Vedsted and Olesen, 2011).

Countries with GPs acting as gatekeepers report better outcomes of health care in terms of costs, patient satisfaction, and overall health status. Moreover, the overall mortality from diseases amenable to healthcare interventions is much reduced in countries with gatekeeper facilities compared to the US (Nolte and McKee, 2008). On the other hand, looking specifically at cancer-related 1-year mortality based on cancer mortality figures from 19 European countries, it has been found that European countries without a gatekeeper policy are performing better when compared to countries with GP-

generated referrals to secondary care (Vedsted and Olesen, 2011). The authors suggested many theories for this finding. It could be because of longer waiting times for the initial diagnosis-focused investigations attributed to system-related delays that are seen in countries with GPs acting as gatekeepers, making the primary care doctors reluctant to refer patients to a long waiting list, adopting instead a "wait and watch" behaviour. This is also part of the GPs' role in cost-containment of healthcare resources as GPs cannot request specialist investigations, which is termed "double gatekeeping". GPs may also delay the referral until symptoms are more prominent due to possible fear of negative judgment by secondary care if a referral for suspected cancer is incorrect. It is also noted that patients may feel ashamed of asking their trusted-doctor for a referral if this is not initiated by their doctor, which is also adding to the overall delay to diagnosis. Hence, there is a need for improvement of the cancer diagnostic phase in the countries with healthcare gatekeepers, possibly by giving GPs access to specialised diagnostic work-up (Vedsted and Olesen, 2011).

Most European countries have implemented the GP service as gatekeepers to secondary care. The GP services can be operated solely by public tax funds but private primary care doctors can also exist that can be funded from premiums allocated to healthcare from the financial contributions of working individuals (Saltman, Rico and Boerma, 2006). Tax-funded primary care use the GPs as a gatekeeper to secondary care with the exception of emergency care that does not require GP referral. For the rest of the services that are privately funded, there are no clear boundaries between primary and secondary care, and patients can seek direct consultation with secondary care providers with an “open access” policy that varies depending on the level of private funds available (Boerma, 2003). Just over half of the European countries (n=18, 52%) have GPs as gatekeepers to secondary care. In most of these countries, there are also private speciality doctors and institutions that can see patients directly without the need for primary care referral. Due to the associated expenses, these options are approachable only to the minority of the population that can cover the associated costs (Boerma, 2003). Australia, Canada and New Zealand have adopted a similar healthcare system to the UK, with GPs acting as gatekeepers to secondary care (Groenewegen, Schellevis and Boerma, 2016;Cheng *et al.*, 2018).

On the other hand, the American healthcare system is mainly market-driven, with minimum state involvement and all healthcare provision is privately operated. This results in a healthcare system that is not accessible to the majority of the population; health insurance coverage is not comprehensive and varies significantly in the levels of provision (Boerma, 2003). Public health insurance is available for individuals over 65 years of age, the disabled population, and the socioeconomically deprived groups, but despite this, 16% of the population remains uninsured. Access to care is limited for this group of individuals, mainly via limited public clinics and hospitals (Ridic, Gleason and Ridic, 2012). Patients usually initiate the referral to the hospital service (Allen *et al.*, 2002;Nolte and McKee, 2008). and primary care referral pathways are not in place due to the predominantly private health care model, as individuals can directly see a specialist and attend secondary care if they are concerned about cancer. Early cancer detection is focused on cancer screening, with screening programs currently available for breast, cervix, colorectal and lung cancer, but there is little significant effort put into early cancer detection of symptomatic cancer patients via referral guidelines and pathways (Sarma, Kobrin and Thompson, 2020).

Looking overall at Asia, significant variations are seen in the provision of primary health care services and referral processes. This reflects the differences in economic development, investment, and provision of health care policies. With the exclusion of the prosperous economies of Japan, Hong Kong, Saudi Arabia, Singapore, The Republic of Korea, Qatar, Bahrain and Kuwait, the rest of the Asian countries are still developing. The latter can only allocate limited resources to cancer screening, early diagnosis and treatment (Sankaranarayanan, Ramadas and Qiao, 2014). In poorer countries, very limited facilities are available for cancer prevention and diagnosis, with weak referral systems. Even though access to primary care is generally good, the quality of care is poor and inefficient due to a lack of facilities and outdated protocols and guidelines for referral to secondary care. Hence most patients initially seen in primary care are subsequently referred to the hospital for specialist assessment without GPs acting in their expected gatekeeper role (McKee, Healy and Falkingham, 2002). Awareness of cancer symptoms and risk factors requiring an urgent suspected cancer

referral to secondary care is lacking not only amongst the general population but also the primary care doctors (WHO, 2002). There are insufficient human resources to respond to the high caseload demand despite the recruitment of other healthcare professionals, such as nurses, to work alongside doctors, and inadequate financial support from state funds to support early cancer detection. The majority of provision comes from private resources, mainly for out-of-pocket payments that lead to unbearable family debts (Sloan and Gelband, 2007). Patients that can afford to pay usually self-refer to secondary care when they have persistent symptoms and are worried about cancer. Unfortunately, this is usually at a late presentation, due to a lack of awareness of cancer symptoms, with over 70% of the cancer cases in low and mid-income Asian countries being diagnosed in an advanced stage with an overall 5-year survival of less than 50% (Sankaranarayanan *et al.*, 2010).

Similar issues, with poor access to health care and lack of a structured way to refer patients to hospital for suspected cancer, is the case for sub-Saharan Africa. The cancer burden may be underestimated due to poor access to care, shortage of medical workforce and limitations in case reporting and cancer database maintenance, with a reported 4.3% of deaths due to cancer compared to the worldwide figure of 12.6% according to WHO (WHO, 2016). Cancer incidence follows mortality figures given the lack of referral infrastructures and the low numbers of cancer specialists and treatment availability. As there is a high burden of death from communicable, maternal, perinatal and nutritional diseases (68.3%) as well as infectious diseases (43.1%), funding received from worldwide resources targets these areas for improvement and reduction in the death burden, as such improvement in cancer diagnosis remains a low priority (Morhason-Bello *et al.*, 2013). Nevertheless, cancer incidence is projected to increase by over 85% by 2030 based on population change, improvement in socioeconomic levels, and the effects of westernisation (Bray *et al.*, 2012). Cancer awareness amongst the general population is deficient, as well as amongst the healthcare providers causing delays in referral to secondary care and late diagnosis exacerbated by the lack of established national cancer prevention and control programmes (Lingwood *et al.*, 2008;Morhason-Bello *et al.*, 2013).

The cancer burden in Latin America and the Caribbean is much higher. The health care system is not integrated in these countries, with each operating a different health care plan. Even within each country, there are discrepancies and significant differences in health care provision, with most hospitals and specialist health care providers concentrated in large urban centres. This makes it difficult to set up cancer control strategies and primary to secondary care referral pathways (Curado and Bezerra de Souza, 2014). National health plans and public policies lack cancer control, early prevention, and referral for specialist care from primary care providers in most countries. A few are currently working on their development or have started operating with newly established policies, such as in Brazil, Bolivia, Costa Rica, Cuba, and Mexico. Similar issues exist as those discussed for Asia and Africa, with a small proportion of the financial expenses spent on cancer care due to other diseases mainly contributing to health care burdens such as infectious diseases and nontransmissible chronic diseases. Even then, the budget is mainly from private healthcare organisations rather than the state (Curado and Bezerra de Souza, 2014).

In summary, cancer statistics across the world show that the type of health system and gatekeeper practice can have a significant impact on outcomes. The primary care jurisdiction in triaging patients for onward referral to secondary care for suspected cancer varies significantly across the world. It depends on the individual healthcare model for every country and the distribution and availability of public and private funding sources. In most developed countries, a mixed public/private health care model exists where patients can be referred to secondary care either after a GP consultation or a private specialist clinic appointment. Predominantly privately funded healthcare appointments are more common in the USA, whereas a mostly public primary care system with a gatekeeping role to secondary care is found in the UK and some European countries that follow a similar healthcare model. In the developing world, access to care is overall difficult, with scarce primary care resources and cancer detection being primarily performed in secondary care. Therefore, any developed cancer risk tool should be available for use in all of the above-described settings, being free and easily accessible in both primary and secondary care clinics but also potentially directly by patients, using a language that can be understood by a specialist

in the specific cancer type but also by GPs and potentially other health care professionals and the public. Prior to presenting the evidence relating to the current suspected cancer referral guidelines and pathways with any country-level variations, it is also important to have an appreciation of potential barriers to early cancer diagnosis. This knowledge helps when it comes to designing any early cancer diagnosis intervention to ensure that its implementation will not be significantly affected by these factors.

2.1.2 Barriers to early cancer detection

Having presented how the provision of primary health care for cancer triaging and referral varies worldwide, the focus of the discussion will now shift to the barriers to early diagnosis aside from the country-level differences mentioned above. This section will incorporate issues associated with GP-level and policy-level attributes as well as patient-level characteristics affecting the cancer diagnosis journey. Three main steps have been described to achieve early cancer diagnosis and treatment: symptoms awareness and access to care; clinical evaluation, diagnosis and staging and referral for treatment; access to treatment. WHO has published recommendations to inform and enable primary care improvements towards early cancer diagnosis, which is a global health priority (WHO, 2020).

Diagnosing cancer at an early stage is associated with improved morbidity and mortality and reduced healthcare costs and cancer burden (WHO, 2007). Early diagnosis means identifying cancer in individuals who have developed symptoms and signs of the disease; hence it is different to cancer screening, which aims to identify cancer in the asymptomatic population. The cost of treatment is less for cancer at its early stages, but also patients can either continue to work or return to work soon after the successful completion of treatment and hence keep supporting their families (WHO, 2007). The majority of cancers are amenable to early detection, with many studies showing that every year millions of cases of breast, cervical, colorectal and oral cancers could have been detected earlier (Loud and Murphy, 2017). The primary care doctors' attributes needed for early cancer detection, as well as cancer diagnosis

policy-related characteristics and associated barriers, will be discussed first, followed by patient-level characteristics and related barriers.

Education of doctors in the prompt identification of potential cancer cases covers an important aspect of early cancer detection (Weller *et al.*, 2012). Awareness of cancer symptoms by primary health care professionals requires a high index of suspicion, especially for patients that are at higher risk for malignancy due to social history, previous related medical history, and family history factors (WHO, 2013). Primary health care professionals should receive appropriate training and have the relevant background knowledge and clinical examination skills to be able to identify symptoms and signs that can be attributed to cancer and act upon them by a timely referral to secondary care. Variations in the education level of the GPs could affect this stage of the early detection pathway significantly (Macleod *et al.*, 2009). Identifying early symptomatic patients with cancer is very important as only a few cancers can be diagnosed at the asymptomatic stage via national screening programs; hence the majority of cancer will be diagnosed after the initial review of a symptomatic patient (Elliss-Brookes *et al.*, 2012) and protocols should be available to ensure sufficient time is allocated for primary care appointments.

As will be described in the following sections, the use of referral guidelines and risk assessment tools, where available, can help to further reduce primary care cancer referral delays (WHO, 2013). Appropriate diagnostic tools should be available to aid cancer risk assessment, relevant training being provided, and a clearly established referral mechanism should be in place. Targets should also be in place to ensure any change has achieved significant improvement in referral pathways. WHO has set a >80% target of patients to be diagnosed within 1 month from initial presentation to primary care, using all available resources (WHO, 2017).

A 2001 WHO survey assessing national cancer control programmes found that only half of the 167 surveyed countries had cancer control policies or cancer management guidelines to include prevention, screening, early diagnosis, and referral of suspected cancer patients. Further assessment of the available cancer referral guidelines for each

country was challenging as only a third of the countries provided documents to support the presence of national cancer prevention and management guidelines. (WHO, 2002). Even when management guidelines existed, these were mainly consisting of guidance on cancer prevention through screening and public awareness of cancer symptoms and guidelines of treatment following initial diagnosis but lacked a plan of action for early detection of cancer in patients that have already developed cancer and are symptomatic (WHO, 2001). This reflects that over the past years, the focus of cancer control strategies shifted to prevention and screening rather than on early detection of symptomatic patients. Prevention strategies and screening have been proven to be an effective approach for cancer of the cervix, breast, colon and stomach, whereas they have been found to be largely ineffective and very expensive for cancers of the mouth/pharynx, larynx, oesophagus, liver and lung cancer (Stjernswärd, 1985;WHO, 2002). Hence the latter cancers, including HNC, which is studied in this thesis, are being neglected by most national cancer control programs, resulting in a diagnosis at a later cancer stage.

Aside from doctor-level and policy-level attributes and barriers to early cancer detection, patient characteristics can also affect the diagnostic pathway. Physical, psychological, and other socioeconomic barriers to early cancer detection have been described, which focus on age and gender issues, limited awareness of red flag symptoms for cancer and the psychological burden of fear and other negative emotions associated with seeking medical assessment for potential cancer symptoms (Chojnacka-Szawłowska *et al.*, 2017). To reduce the time to a patient's first attendance at primary care, patients should have an awareness of the symptoms possibly related to cancer, realise that they need to urgently see a doctor if they develop such symptoms as well as have the support that is needed to overcome any associated fears and stigma associated with cancer and medical assessment for its exclusion (WHO, 2013;McCutchan *et al.*, 2016).

Older people can find it difficult to realise that they have worrisome symptoms and to explain and communicate this effectively. Realising that a symptom may be related to cancer has also been found to be more challenging for the male population. This is due

to long-established assumptions about male strength that make some males challenging to overcome the beliefs of female-only vulnerability and susceptibility to illness. There is also a tendency in the media to mainly promote female health issues, which can prevent males from seeking health advice early due to a lack of education and awareness of cancer symptoms. Studies have shown that males usually seek medical advice following encouragement from the family and spouses (Walter *et al.*, 2012).

Another barrier relates to the emotional burden in potential cancer consultation seeking. Studies show increased fear and anxiety in patients seeking medical help for their symptoms when they are worried the symptoms may be related to cancer. There is fear of the possible need for extensive investigations as well as a belief that they waste the time of the doctor if they seek help (Whitaker *et al.*, 2015). Another reason for delayed patient presentation is the worry about the financial burden of a cancer diagnosis, not only because of the direct costs related to treatment but also indirect costs due to missed wages or unemployment (Azzani, Roslani and Su, 2015). Feelings of shame, in addition to cultural red lines and misbeliefs, can delay first medical contact. This can also include women seeking a female medical professional for breast and cervix examination for potential cancer, as was noted in a study performed in Indonesia (Iskandarsyah *et al.*, 2014). In addition to the above, a review of the literature has also shown that patients may also believe that all cancers are incurable or that any proposed treatment will be painful or ultimately result in death or keep them away from family and friends (Macleod *et al.*, 2009).

Socio-economic barriers to an early patient presentation also pose a significant problem in timely cancer diagnosis (Weller *et al.*, 2012). This is particularly evident in the low- and middle-income countries' cancer statistics, showing the majority of the cancer cases presenting at a late stage, being an effect of scarce healthcare resources, as discussed in section 2.1. Additionally, in each country, a late cancer diagnosis is more common in socio-economically deprived regions. This is because of lack of education of the socio-economically deprived groups of the cancer red-flag symptoms despite the availability of primary care resources (Smits *et al.*, 2018).

Nevertheless, despite all the above-mentioned characteristics and barriers, it has been found that most patients will present to the primary care doctors for the first time within a year before the cancer diagnosis, and the diagnostic interval is reduced for countries with guidelines in place to help identify patients with high risk for malignancy (Neal *et al.*, 2014). Better coordination between health facilities with timely referral to secondary care can further reduce diagnostic delays (Richards *et al.*, 1999). Reducing delays in cancer diagnosis, even by a few months (from 3-6 months to less than 3 months), has been linked to improved survival (Richards *et al.*, 1999). The magnitude differs depending on the cancer type, but overall, an increased 5 and 10-year mortality has been found for delays in treatment initiation. This was more pronounced for colon cancer and lymphoma, whilst prostate cancer was less affected, followed by breast cancer (Cone *et al.*, 2020).

To summarise, patient related barriers to early diagnosis include: limited access to primary care for the socioeconomically deprived population, lack of awareness of symptoms associated with cancer that is more evident in the elderly and socio-economically deprived population, negative feelings associated with a potential cancer diagnosis, cultural and gender-related misbeliefs of health values as well as worries related to loss of income. Additionally, poor cancer symptom awareness amongst some primary health care professionals and variations in cancer control policies also affect early cancer detection. The barriers related to GP consultations and primary to secondary care interlink services are found to affect cancer outcomes more compared to the other early diagnosis barriers. Hence, introducing a triaging aid to help structure and streamline GP consultations and referrals can help improve early detection outcomes, although other barriers will remain to be addressed. Cancer risk calculators have the potential to be incorporated in national and worldwide early cancer detection strategies that are currently based on guidelines for suspected cancer referrals comprised of lists of red-flag cancer symptoms that patients can complain of at the point of first contact with a healthcare professional. These guidelines will be presented in the following section of the literature review to cover worldwide referral protocols and then move to more in-depth information on the UK cancer referral pathway.

2.2 Suspected cancer referral guidelines

This section will discuss the guidelines that are in place across the world to enable structured suspected cancer referral triaging. This knowledge will facilitate an understanding of variations in guidelines, their pros and cons and how the referral processes can be potentially improved with the introduction of a triaging aid. International guidelines will be mentioned in this introductory subsection, followed by a more in-depth presentation of the UK cancer referral guidelines, as the HNC risk calculator was designed for initial implementation within the UK.

Only a few countries have guidelines relating to the referral of symptomatic patients with suspected cancer. Urgent suspected cancer (USOC) pathways for referral to the hospital following a primary care consultation with suspected cancer symptoms are currently established in the UK, Denmark, New Zealand, Australia and Spain (Koo *et al.*, 2021). The nature of presenting symptoms for the various cancers is not expected to vary by country, and this is reflected in the similarities found in the referral guidelines for suspected cancer based on patient symptomatology, currently available in the countries with such policies available (Koo *et al.*, 2021). What does vary is the perception of symptoms by the patient that is based, as discussed earlier, on the level of health literacy of the country, or regions within the country, cancer symptoms awareness and the fear of the stigma associated with health-seeking assessment for potential cancer. These also depend on population health education, socioeconomic level, gender, and age, as was already discussed in 2.1.2 (Moffat *et al.*, 2015;McCutchan *et al.*, 2015;Humphrys *et al.*, 2019).

The USOC referral pathways focus on the patient's presenting symptoms. Understanding symptoms epidemiology, that is, the frequency and type of symptoms in the population seeking medical attention, can help identify those patients in need of urgent specialist medical review. An innovative systematic review of the literature on cancer symptoms has introduced the term "symptom signature" to describe the nature and relative frequency of symptoms leading to a particular cancer diagnosis. A cancer taxonomy system was then described based on the "symptom signature" - being narrow (most patients have one particular symptom) or broad (large range of symptoms) - and

the symptoms' predictive value to achieve diagnosis (Koo *et al.*, 2018). Cancers that can be identified based on a single or only a few alarm symptoms that all have high positive predictive value (PPV) are linked to shorter time to diagnosis and better outcomes; that is the case for breast, testicular and thyroid cancer. At the same time, cancers such as oropharyngeal, oesophageal, colorectal and lung cancer can present with many different symptoms, not all being typical alarm symptoms. These atypical symptoms have low PPV for cancer hence being associated with delays in diagnosis (Koo *et al.*, 2018). Symptom prevalence, positive predictive value for cancer, their association with cancer stage at presentation and time from symptoms awareness to seeking medical assessment are the main features that can help symptom prioritisation for early cancer detection targeting. Symptoms with low awareness amongst the population and a long time for symptom onset to health review seeking should be targeted first for early detection strategies, as well as those with positive predictive value for cancer and high frequency of presentation (Koo *et al.*, 2018). Data on presenting symptoms can be collected retrospectively from the patients or get them extracted from health records that were prospectively recorded. Asking the patients to retrospectively recall their symptoms can introduce recall bias but also excludes a group of patients that might be too unwell to take part in such studies. It is preferable to use prospectively collected data available in health records for adequate capturing of patient symptomatology, frequency and duration, but the main issue in this endeavour is that often data are missing as potential symptoms were not assessed for their presence at the time of patient consultation (Verheij *et al.*, 2018).

As mentioned above, symptoms-based country-specific referral guidelines are in place in the UK, Denmark, Australia, New Zealand, and Spain. Similar guidelines have also been published by the American and WHO Cancer committees (WHO, 2007). The New Zealand guidance is based on the UK NICE guidance, which will be discussed in detail in the next section of this chapter, with only minor changes, mainly to wording, made when publishing their referral guidance (NZGG, 2009). Guidelines in Australia are based again on the best available literature that was used to develop the UK NICE guidelines. A detailed review of the most relevant literature with a recommendation summary for symptoms associated with malignancy is available on the Cancer Council

Australia website for cancers of the bowel, endometrium, cervix, lung, oesophagus, prostate, sarcoma and skin (CCA, 2021). Spain has also implemented an urgent cancer referral pathway since 2005 (Prades *et al.*, 2011) and Denmark since 2007 (Probst, Hussain and Andersen, 2012). In both countries, the guidelines are consensus-based, agreed upon by clinical working groups commissioned by their national board of health (Probst, Hussain and Andersen, 2012) (Prades *et al.*, 2011). The American Cancer Society has published guidelines for early cancer detection, which are available to both doctors and the general population, with specific symptoms and signs for the common cancers (breast, colon/rectum, cervical cancer, endometrial cancer, lung cancer and prostate cancer) (cancer.org, 2021) but also a list of cancer red-flag symptoms to incorporate the rarer cancer types (cancer.org, 2020). The WHO has also produced a list of red-flag symptoms for each cancer site which are also in agreement with the NICE recommendation but without adjustments made to take into consideration age and other co-morbidities (WHO, 2007).

As the focus of this thesis is the design of a cancer risk calculator for implementation within the UK healthcare system, the following section will present in more detail the current cancer referral pathways in the UK, regional variations, outcomes, and shortcomings that can then be accounted for when a new early detection intervention is designed.

2.2.1 Suspected cancer referral guidelines in the UK

In this section, the cancer referral guidelines across the UK will be discussed, from initial development to the current framework for use across the different UK regions. Differences in their implementation across GPs will be covered, and how this reflects in the cancer detection and conversion rates for different cancer sites. This is important to understand as similar variations may be noticed if a HNC triaging aid is introduced, which is the research outcome of this thesis.

In the UK, in 2000, the department of health issued the UK National Guidelines for urgent cancer referrals for suspected cancers seen by primary care professionals. The target was to ensure that these high-risk patients would be seen within 2 weeks of

referral to secondary care for further assessment. These guidelines were further updated in 2005 by NICE and again in 2015, with a further iteration since then with minor adjustments taking into consideration blood results or other tests that can be done in primary care to aid diagnosis prior to referral to the hospital (NICE, 2021). The guidance includes symptoms or a combination of symptoms that are found to be associated with a high risk of malignancy based on the most recent review of the literature performed by the NICE guidance working group.

In the early iterations of the guidance, symptoms were included when there was available literature associated with a cancer diagnosis based on papers published by primary care institutions with no explicit guidelines to suggest criteria for specific symptom inclusion (NICE, 2015). Since 2005, the concept of positive predictive value (PPV) threshold was introduced to allow filtering of the symptoms. The NICE panel of experts, taking into consideration the available literature, had initially adopted the 5% PPV threshold for a symptom to be included in the guidance. This felt that it needed revision in the 2015 update, which brought the threshold down to 3% as it was accepted that this is a reasonable threshold to boost early detection without overburdening the pathway with a large number of referrals with the associated financial implication, but also the patient-related implication of increased anxiety associated with a suspected cancer referral (NICE, 2015). In young adults and children, the PPV threshold was decided to be lower than 3% across all cancers to ensure early detection in this special population group with a long-life expectancy after early identification and successful cancer treatment. Aside from symptoms, other risk factors were also considered for inclusion in the guidance that could potentially change the weighting of symptoms to be associated with cancer. From all the factors tested, including the history of cancer, occupation, social history factors, age, gender, exposure to chemicals, and others, only age and smoking (for lung cancers) were significantly altering the risk of cancer when a specific symptom was present hence these two factors were incorporated in the guidance (NICE, 2015).

In England and Wales, all patients with suspected cancer are referred to secondary care following an initial review by a GP using the above-mentioned 2ww referral pathway (NICE, 2021). Northern Ireland’s referral guidelines were updated in 2021 but are still

based on the 2005 NICE guidelines (NICaN, 2021). Scotland has its own cancer referral symptoms checklist, and the pathway is called the urgent suspected cancer (USOC) pathway (NHSScotland, 2019). The list of symptoms is similar to those published by NICE, but differences exist. For example, for HNC, no age limit for referrals with suspicious symptoms is set, whereas it is over 45 years of age for suspected laryngeal cancer referrals for NICE (NICE, 2021). Dysphagia was part of the Scottish guidelines but was removed in the latest iteration, with no reason being mentioned in the guidance for this deletion (NHSScotland, 2019). Odynophagia and sore throat is part of the Scottish guidelines, but these symptoms are not included in the NICE guidance (NICE, 2021) (NHSScotland, 2019). Thus, there are currently no unified cancer referral guidelines across the UK. As the pool of evidence remains the same, there is no explanation as to the reason for the differences in the referral guidelines between Scotland and the rest of the UK beyond the fact that the advisory boards that provide the expert opinion in the two regions have different members. Variations in the symptoms included in the referral guidelines are also found within England, for example London has separate pan-London cancer referral guidelines (myhealth.london.nhs.uk, 2017). The lack of uniformity in the referral criteria can introduce variation in the 2ww referral numbers and outcomes (Dodds *et al.*, 2004) (Blank *et al.*, 2014) as will be also discussed in detail in section 2.2.2.2 of this thesis, in relation to the HNC referral guidelines. The above differences in the referral guidelines should be taken into consideration in the evidence that will be presented below about referral rates and cancer pick up rates across the UK nations.

2ww pathway referral rates

A recent large-scale study of 14 million referrals has shown that overall the 2ww referrals in England have doubled over the past 10 years, being over 2.2 million in 2018-2019, with a 10% increase each year (Round *et al.*, 2021). The evidence shows that there are differences in how often GPs use the guidelines for referring suspected cancer cases. A review of pan-Scottish GP compliance with the suspected cancers guidelines showed a large variation in the referral rates ranging from 3.7% to 24% per 1.000 per annum for 512 practices (median: 11%). A similar audit from England,

including 8049 practices, again highlighted a significant variation in the referral ratio from -50% to over 50% from the mean (Meechan *et al.*, 2012).

Although the referral guidelines are neither very sensitive nor specific for cancer diagnosis and significant variations exist, the use of a structured way of referring suspected cancer cases is linked with better mortality outcomes and patient satisfaction. Moller et al. (2015) study based on 200,000 cancer patients across English GP practices showed that GP practices that tend to refer many patients via the 2ww pathway are linked with better mortality outcomes for their registered cancer patients, with a 4% lower hazard ratio compared to a 7% higher hazard for practices with the low use of the pathway. Higher practice detection rates (the proportion of cancers being identified via the urgent suspected cancer referral route out of the total cancers) were also linked with better mortality outcomes, whereas no association was found with practice conversion rates (the proportion of urgent suspected cancer referrals that result in a cancer diagnosis) (Møller *et al.*, 2015). In agreement with the above findings, a recent study from England has shown better mortality outcomes in GP practices that had a high referral ratio via the 2ww pathway. This was a large-scale study including over 9,000 English General Practices, analysing 6.9 million urgent cancer referrals. It was also found that a 2ww high referral ratio was also linked to an earlier stage at the time of diagnosis (Round *et al.*, 2020). Also, interestingly, GP practices that have scored high in patient satisfaction for doctor-patient communication have higher rates of USOC referrals and less proportion of cancer patients diagnosed after emergency hospital admission (Lyratzopoulos *et al.*, 2018). Therefore, it is important to explore potential reasons for the 2ww referral variations as they affect patient outcomes.

An English audit, including 8,049 practices, showed a higher referral ratio for practices of over 6000 patients (Meechan *et al.*, 2012). There was no significant variation depending on the deprivation index of the practice population, and small variations were seen for the different health authority regions (Meechan *et al.*, 2012). Practices with doctors in training, with younger GPs, and with many partners have a higher number of referrals, but no difference has been seen in relation to the urban or rural

location of the practice or the country of medical qualification of the GPs (Mendonca *et al.*, 2019). GP gender was speculated in the past to be a factor affecting the referral rates, having been seen by female GPs being linked with longer diagnostic intervals and more advanced cancer stage at presentation in some small sample size studies (Maclean *et al.*, 2015;Hansen *et al.*, 2011). However, the results of the large-scale pan-UK audit by Mendonca et al. (2019), which included over 7000 GP practices, did not identify any GP gender-related variations to affect the cancer referral rates (Mendonca *et al.*, 2019). In a Scottish study including 500 GPs, younger patients were more commonly referred via the pathway despite the fact that a cancer diagnosis was higher in the older population (Baughan, Keatings and O'Neill, 2011). On the other hand, Mendonca et al.'s 2019 study of all English GPs, showed a high rate of 2ww referrals for older patients and those from the most deprived areas in the country. It may be that variations in GPs' referral attitudes exist between Scotland and England that can explain this variation. Additionally, the difference between the Scottish and English lists of referral symptoms could also play a factor. Furthermore, the study by Mendonca et al. (2019) is likely to show a more accurate reflection of the current situation as it included data from over 7000 GP practices, whereas 500 GP practices were included in the Scottish study (Baughan, Keatings and O'Neill, 2011), therefore it is less likely for the former study results to be affected by selection bias. In the English study, male patients and those with Asian or "other" ethnicity were less likely to be referred (Mendonca *et al.*, 2019). Adding to the gender difference in referral rates, another study also found minor variations for gender, with females more likely to be referred via the urgent route (56% vs 47%). Negligible variations were found per deprivation quintile (Zhou *et al.*, 2018). In terms of cancer sites, considerable variations are seen in the proportion of urgent cancer referrals per cancer site in England. Breast and testicular cancer have the highest proportions (73% and 71%) compared to only 6% for brain cancers, with over 50-fold variation (Zhou *et al.*, 2018). So, there are differences in the referral rates across GP practices, but no explicit reason for these differences has been identified apart from an association with the size of the practice, the presence of doctors in training and the cancer site and minor variations depending on patients' demographics. It has been proposed that the rest of the variation might be related to the referral threshold at which each clinician feels that a referral is

needed and justified. Despite the fact that all clinicians use the same referral guidelines, they can interpret them differently depending on how risk-averse they are (Djulbegovic *et al.*, 2015). It has also been suggested that the variation in referral rates might also be affected by GPs taking into consideration other features aside from patients' symptoms, including patients 'wishes and preferences, health-seeking behaviours and demographics, but this has yet to be proved (Møller *et al.*, 2015).

2ww pathway detection rates

Apart from the above-mentioned referral rate variations, the overall increase in the referral rates via the urgent cancer route has led to an increased number of cancers detected via this pathway (detection rate) from 41% to 52% over a ten-year period (2009/10 to 2018/19). This still remains low as half of the cancers are diagnosed via routes other than the urgent cancer pathway, as demonstrated in a study of 14 million referrals (Round *et al.*, 2021). Apart from the overall increase in urgent cancer referrals, there are other factors that have been found to affect cancer detection rates. Practices with a large number of registered patients and younger GPs have higher cancer detection rates via the 2ww pathway, whereas the detection rate is lower in GPs in deprived areas (Round *et al.*, 2021). This agrees with an earlier study by Meechan et al. (2012) analysing over 800.000 referrals, also showing that the detection rate was lower in the most deprived area and for GPs with a small number of registered patients in the practice. The study also found that detection rates were lower for younger patients (<65 years of age) and for GPs with a low referral ratio (Meechan *et al.*, 2012). So, it appears that differences in the population demographics of each GP practice affect the detection rates. Nevertheless, statistical analysis performed in two studies by (Murchie *et al.*, 2015;Sullivan *et al.*, 2005) showed that after adjusting for these differences, the variation in the detection rates was much smaller; hence these metrics alone should not be used as a way to assess GPs performance but could also be attributed to chance. In a more recent study, the variation in cancer detection rate was found to be significantly affected by chance after adjusting for the GP population metric; therefore, it was again evident that these metrics should not be used to assess GP's performance in relation to early cancer detection interventions (Abel *et al.*, 2018). Nevertheless, the significant variations in the detection rates should not only be

attributed to GP – level variations, regional demographics and cancer proportions attributed to chance. A recent study has shown that these variations can also be partly explained by healthcare infrastructures above the GP – level, at the clinical commissioning group (CCG) and at the acute hospital trust level. One-third of the variation in the detection rate was found to be attributed to CCG clusters (Burton *et al.*, 2020). Moreover, the acute care hospital accounts for two-thirds of the CCG – level variation within the CCG clusters. Hence, the variations in cancer detection via the urgent cancer route go beyond the GP level or patient-related traits but also depend on the CCG and acute hospital infrastructures (Burton *et al.*, 2020). These could be differences in access to secondary hospitals, waiting times for hospital appointments depending on the level of urgency and other issues related to access to investigations and treatment (Blank *et al.*, 2014).

2ww pathway conversion rates

Focusing now on conversion rates (how many cancer found from the total referral being made), the increased number of referrals has led to an overall reduction in the cancer conversion rates across England, from 10.8% to 7.3% as more patients without cancer are being referred (Round *et al.*, 2021). However, national variations of the mean conversation rates also exist. The cancer conversion rate has been found to be lower for small GP practices as well as for GPs with a high referral rate (Meechan *et al.*, 2012). Nevertheless, after adjusting for differences in the population demographics of each GP practice, the variation in the conversion rates was much smaller, similar to what was seen above for detection rates (Murchie *et al.*, 2015;Sullivan *et al.*, 2005) and also affected by chance (Abel *et al.*, 2018). Conversion rate has also been found to vary for different cancer sites, being higher for haematological, prostate and lung cancers but low for HNCs, melanoma, gastro-oesophageal and brain cancer, with laryngeal cancer having the lowest conversion rate at 7.8% (Figure 2-1). Additionally, GP compliance with the guidelines varies significantly per cancer group, as seen in Figure 2-2, which can also affect the total referrals being made and with a subsequent effect on the cancer conversion rates (Baughan, Keatings and O'Neill, 2011).

Figure 2-1. Proportion of urgent cancer referrals that lead to a cancer diagnosis per cancer site. Source: Baughan et al., 2011

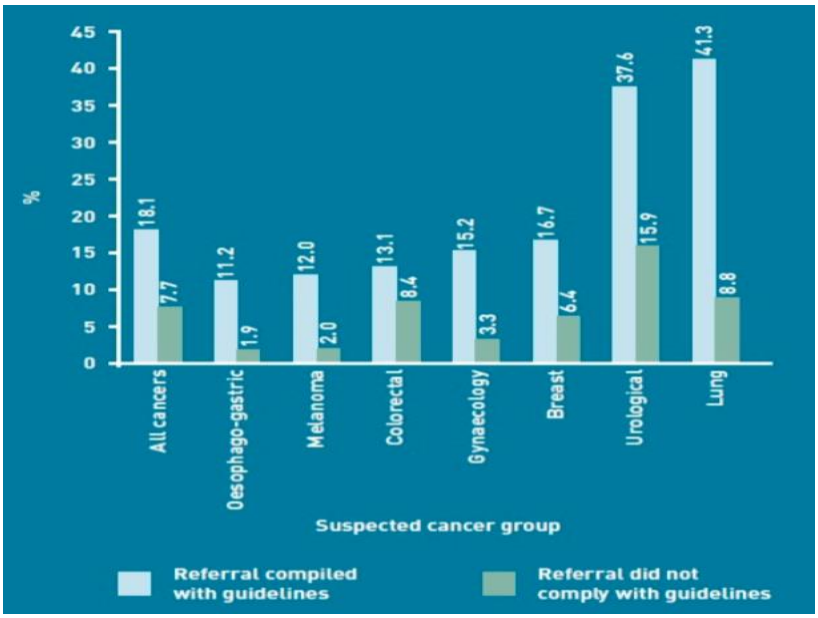
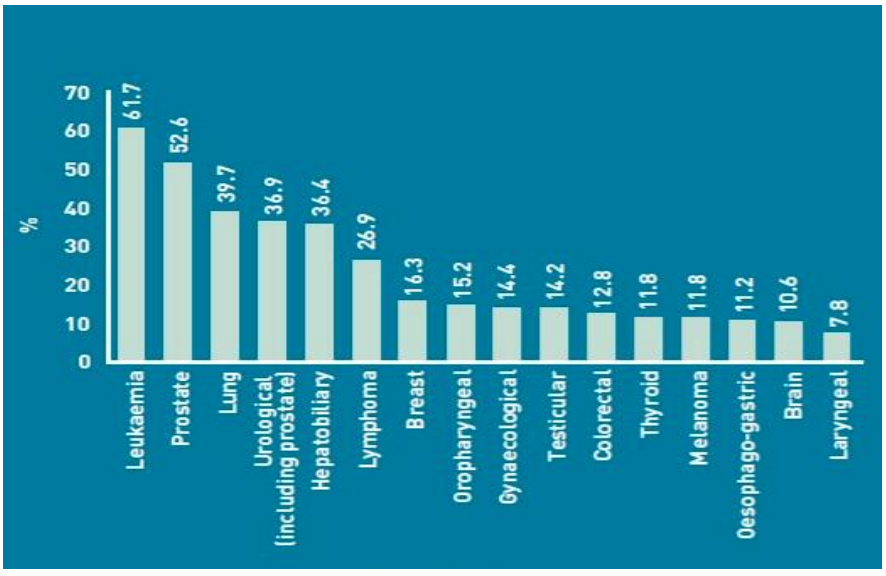


Figure 2-2. Proportion of referrals in compliance with the guidelines per cancer site. Source: Baughan et al., 2011



To summarise the above, there is currently significant variation in the use of urgent cancer referral appointments across the UK, but overall, an increasing trend in the number of total referrals is noted. The high rate of referrals has resulted in low conversion rates, especially for rarer cancers, but the overall detection rate has increased. Despite that, it still remains low overall, with the majority of cancers still being diagnosed via non-urgent cancer referrals. Many studies have been conducted to

assess traits that can influence referral rates. It appears that those that included robust, large sample, nationwide data have detected higher referrals rates in elderly patients, female patients, non-ethnic minorities, the socioeconomically deprived and in practices with younger doctors and a large number of registered patients; without this being translated necessarily into higher detection rates in all these subgroups. Favourable outcomes have been noted for those practices referring more patients despite the conversion rate being low. Appreciating that there are differences in the use of the urgent cancer pathway and exploring the likely aetiology of such variation is important when considering the design of cancer triaging tools that could act as an alternative or addition to the referral criteria for a cancer referral. This is because similar variations are likely to be observed in the uptake and use of triage tools that could potentially affect its nationwide rollout and performance. Any identified factors should be anticipated, and solutions designed and implemented to manage them appropriately.

2.2.2 Head and Neck Cancer referral guidelines and pathways to referral

As per the WHO recommendations for early cancer detection strategies, many nations have implemented specific referral guidelines for HNC based on patients' symptoms in an attempt to facilitate early cancer detection in the primary care setting. For countries where GPs act as gatekeepers, these guidelines are used to aid referral to secondary care. For countries with mixed or mainly privately based primary care, including private GPs or private practising ENT specialists, the guidelines are used as a filtering function to refer high-risk patients to secondary care (WHO, 2007). This sometimes follows initial investigations performed in an out-of-hospital setting, such as the flexible nasal endoscopy investigation, office-based biopsies and scans in privately operated primary care facilities (Lee *et al.*, 2018)

International and National Societies and Government-led health organisation bodies have published guidelines for the management of patients presenting with HaN symptoms that raise a possibility of malignancy. The symptoms included in these recommendations are summarised in Table 2-1. As the table demonstrates, the included symptoms are not uniform across the different regulatory bodies. Their

similarities, differences and potential strengths and weaknesses are presented and discussed in the following paragraphs.

2.2.2.1 Referral guidelines of International HNC societies

Several guidelines have been published providing recommendations on red flag symptoms for HNC requiring prompt referral to secondary care for further investigations. Despite the fact that there are many papers available in the literature reporting presenting symptoms for HNC, as summarised in section 2.4.1 of this thesis, most of the guidelines are based on expert opinion collation of lists of symptoms per HNC subsite. A review of the international medical organisation societies performed as part of this thesis identified three medical societies that have published their recommendations on symptoms that should prompt urgent referral for suspected HNC. These organisations are the Society of Surgical Oncology (SSO), the American Cancer Society and the European Head and Neck Society (www.surgonc.org, 2021), (cancer.org, 2021), (European Head & Neck Society, 2020).

Societies-Led referral guidelines

The SSO is a global community of cancer surgeons that publishes guidelines and recommendations for the diagnosis and management of cancers (www.surgonc.org, 2021). They have published guidelines for the identification of oral cavity and oropharyngeal cancer based on presenting symptoms and signs, and they recommend being used alongside good medical judgement in referring patients to secondary care. The SSO guidelines are based on expert opinion from clinician members of the SSO aiming to produce guidelines that "were not likely to result in significant controversy" (Shaha, Byers and Terz, 1997b), (Shaha, Byers and Terz, 1997a).

The American Cancer Society has also published on its website general information on cancer symptoms and signs for all cancers accessible to the public, advising them to seek medical attention if they have such symptoms (cancer.org, 2021). The list of symptoms included in this guideline is based on the head and neck chapter of Abeloff's Clinical Oncology book (Leeman JE, 2020).

The European Head and Neck Society has also published referral guidelines as part of the Make Sense Annual Campaign. As with the SSO guidelines, the list of symptoms here is again based on leading experts in head and neck across Europe recommendations (European Head & Neck Society, 2020). Therefore, it is evident that the available international guidelines are based on low-level evidence from clinical experts rather than an in-depth review of the literature. This limits the clinical value and strength of the recommendations and their generalisability outside the geographical region from where they have been derived. The list of symptoms included in these guidelines is summarised in Table 2-1. Symptoms that were present in all guidelines were: mouth ulcer, mouth/tongue pain, hoarseness, throat pain and neck mass. The rest of the symptoms are present in only one or two of the three guidelines. When compared to the results of the HNC symptoms literature review that is presented in the later section 2.4.1, neck mass, hoarseness, sore throat, and dysphagia were the most common presenting symptoms linked to a HNC diagnosis. Only SSO included dysphagia in the list of symptoms, with the rest of the common symptoms being included in the guidelines.

No specific recommendations about the process of secondary care referral are made as part of the international societies' guidelines. This is expected as each country or region has its own pathways from primary to secondary care referral, as discussed early on in the literature review chapter (section 2.1) of this thesis. The specific HNC pathways for each region will be covered in the section below, alongside the government-led HNC referral guidelines. Such information is available and will be discussed for Australia, New Zealand, and some European countries, with a specific focus on the UK. Very limited literature is available on the pathway of referral to secondary care in the USA, Canada, or other world regions, with no state-wide records of possible diagnostic delays prior to the secondary care clinic review. This is the case across all cancer, but specifically relating to HNC, the example of the USA and Canada will be briefly discussed. The issues relating to the lack of a uniform HNC referral process in the USA were recognised in a study by Ohlstein et al. (2015). In their research, they have described a pathway of actions from the first secondary care review to initiation of treatment, introducing a 2-week timeframe. Nevertheless, this move does not

address the referral pathway shortcomings of the USA model of care prior to a patient being referred to a HaN clinic (Ohlstein *et al.*, 2015). Similarly, in Canada, no referral guidelines exist, with the majority of patients with suspected HNC being referred to secondary care following a review in private or state-funded ENT offices or primary care doctors. Usually, a biopsy with confirmation of malignancy is required for patients to be urgently referred to the hospital, with specific referral proformas being available for each Canadian region after a diagnosis of HNC has been made in primary care (Toronto Central Regional Cancer Program, 2016).

Table 2-1. Symptoms included in the HNC referral guidelines per regulatory body

Referral guidelines regulators	Symptoms																						
	Oral							Neck		Nose		Throat						Other					
	mouth ulcer	oral lesion	mouth/ tongue pain	red/white patch	trismus	mouth numbness	tooth mobility	neck pain	neck mass	unilateral nasal blockage/mass	unilateral epistaxis	odynophagia	dysphagia	hoarseness	throat pain	haemoptysis	breathing difficulty/ Stridor	cough	otalgia	slurred speech	unilateral facial paralysis	(unilateral) vision, hearing,	skin changes	weight loss
Societies																								
International Society of surgical oncology	✓	✓	✓		✓				✓				✓	✓	✓	✓	✓		✓	✓				
American cancer society	✓	✓	✓			✓		✓	✓					✓	✓	✓		✓				✓	✓	✓
European Head and Neck Society	✓		✓	✓					✓	✓	✓	✓		✓	✓									
Government-led																								
Australia	✓	✓		✓	✓	✓	✓		✓	✓				✓	✓	✓			✓		✓	✓		✓
New Zealand	✓	✓		✓			✓		✓					✓										
Denmark, Norway, Sweden	✓	✓							✓	✓	✓			✓					✓		✓	✓		
Spain	✓			✓					✓					✓										✓
England and Wales (NICE)	✓	✓		✓					✓					✓										
Scotland	✓	✓		✓					✓			✓		✓	✓		✓							

Government-led HNC referral guidelines

Aside from the above-mentioned international organisations' published guidelines on referral symptoms, government-led national bodies have also published similar guidelines also, incorporating the pathway process from primary referral to secondary care review in each nation. A summary of the included symptoms in these guidelines is seen in Table 2-1. Of the total of 24 symptoms recorded across the guidelines, the symptoms most commonly reported are mouth ulcer, oral lesion, neck mass and hoarseness, being broadly in agreement with the symptoms included in the international recommendations. The presence of red and white patch lesion symptoms is also included in the majority of the government-led guidelines but was only mentioned in the European Head and Neck Society recommendation. It is a symptom that was reported in 10 out of the 37 papers included in the literature review that is presented later in section 2.4.1 of this chapter. The mean percentage of patients presenting with this symptom is 13.7%, while commoner symptoms such as hoarseness and neck lump have a mean presentation value of over 25%. On the other hand, the symptom of mouth/tongue pain that was part of all 3 international societies' guidelines is not included in the government-led guidelines, and it has a low frequency of 2.9% in the HNC symptoms literature review (section 2.4.1).

The evidence used in producing the guidelines is again primarily level 4 - expert opinion recommendations - as was the international societies guidelines, with the addition of a review of the limited primary care HNC literature influencing the UK NICE guidelines and the New Zealand guidelines (Cancer Council Victoria and Department of Health Victoria, 2021) (NZGG, 2009), (NICE, 2021), (cancerreferral.scot.nhs.uk, 2019). The primary care paper quoted in the guidelines is the paper by Alho (2006), which presented the frequency of symptoms in a relatively small number of patients, n=221, with HNC seen in the primary care setting. Five more papers performed in primary care were identified during the literature review presented earlier in this chapter. Of these, three have been published since 2019 and, as such, were unlikely to be available for inclusion in the most updated review of literature that informed the guidelines (Shephard, Parkinson and Hamilton, 2019)**,** (Nieminen *et al.*, 2021), (Talwar *et al.*, 2020). It is unlikely, however, that the latter two papers would

have influenced the guidelines significantly as they have a very small number of included HNC patients (n=6 and n=40, respectively), but the paper by Shephard et al. (2019) adds significantly to the available literature with a total of 813 HNC patients seen in primary care. Finally, it appears that the studies by Merletti et al. (1990) of 279 HNC patients and (Koivunen *et al.*, 2001) of 80 patients were missed from the review of the primary care literature during the development of the guidelines. Hoarseness and dysphagia were the most common symptoms reported by Merletti et al. (1990), and sore throat and otalgia by Koivenum et al. (2001). Interestingly, none of the government-led guidelines includes dysphagia in their list of symptoms, and sore throat and otalgia are only part of two guidelines (Table 2-1). The list of symptoms in the guidelines and the referral pathways for each region will be now discussed in detail in the paragraphs below.

In Australia, nationwide government-led guidelines are followed by referral of patients with red flag symptoms for HNC to be seen within 2 weeks by a HNC specialist (Cancer Council Victoria and Department of Health Victoria, 2021). In their guidelines, there is also a particular recommendation that the first appointment in secondary care should be a specialist who regularly participates in HaN multidisciplinary meetings. These guidelines are available on the Australian Cancer Council website and include the following signs and symptoms being persistent for 3 weeks or more: hoarseness, dysphagia, persistent sore throat, particularly in association with otalgia, unexplained neck/parotid lump, oral ulcer, or mass, oral white or red patch, unexplained tooth mobility, haemoptysis, unilateral nasal blockage especially if associated with double vision or eye swelling, lip ulcers or patches, weight loss and trismus. The guidelines also include unilateral otology symptoms (unilateral pain/pressure/ringing in the ear or hearing loss) and cranial nerve neuropathies (unilateral paralysis of facial muscles, unilateral tingling feeling in the face) as red flags for an urgent referral (Cancer Council Victoria and Department of Health Victoria, 2021). Nevertheless, as it will be described later in section 2.4.1, the literature review on symptoms associated with HNC diagnosis showed that these symptoms are very non-specific for a HNC diagnosis. Neuropathies as a presenting symptom were found in 1.4% of newly diagnosed HNC patients, and unilateral ear symptoms in 2.8%

of patients. These numbers are very low compared to the incidence of the cardinal red flag presenting symptoms, such as neck lump and hoarseness, being noted in 29.9% and 28.4% respectively of HNC patients.

In New Zealand, uniform country-wide guidelines exist for urgent referral of suspected HNC patients available on the government website (NZGG, 2009). They are based on the current and previous versions of the NICE UK urgent 2ww guidelines and include recommendations for referral subdivided for oral cavity, laryngeal, and thyroid cancers. Despite their origin from the NICE recommendations, minor but important differences are noted, having a more detailed description of the nature of main symptoms, duration, and association with other symptoms. The New Zealand guidelines are as follows: urgent referral for consideration of oral or larynx cancer should be made within 2 weeks for persistent unexplained symptoms that have not disappeared within 6 weeks from being noticed from the first time in primary care, hence eluding that a primary care review will be put in place following initial patient review with a symptom suggestive of possible HaN pathology. This is a significant difference compared to the NICE guidelines that advise referral of all patients to secondary care after initial consultation without a period of controlled, watchful waiting to allow for the resolution of symptoms possibly associated with a benign condition. An urgent referral is required for all the below symptoms: persistent hoarseness for more than 3 weeks (particularly in smokers over 50 years old or heavy drinkers), ulceration or mass in the oral mucosa of 3 weeks or more, unexplained tooth mobility of more than 3 weeks, presence of red and white patch being also painful, swollen or with associated bleeding. In the absence of the latter associated symptoms, when a patient has a red/white oral patch, a non-urgent referral should be made (NZGG, 2009). A new neck lump should also be referred urgently with particular attention to the lump being painless and new or pre-existing lumps that have recently changed over a period of 3-6 weeks as well as persistent sore/painful throat or mouth, particularly unilateral for over 4 weeks or unilateral HaN pain/paraesthesia/dysesthesia of 4 weeks or more (NZGG, 2009). The number of weeks duration of the main symptoms and their association with other features, as well as the smoking and alcohol history in association with some symptoms, makes the New Zealand recommendations

more specific, allowing differentiation from benign pathologies, in contrast to the generic symptoms description of the NICE guidelines (NICE, 2021). Similar differences are seen for the thyroid lump symptom, with an urgent referral being recommended in the New Zealand guidelines only if the thyroid swelling is associated with one or more of the following symptoms: increase in size, neck irradiation history, family history of endocrine tumour, hoarseness, lymphadenopathy, pre-pubertal age or over 65 years old. All other cases should first have thyroid function checked with onward endocrinology referral (NZGG, 2009).

Some European countries have their own referral guidelines published on their respective government websites, such as Denmark, which introduced a fast-track clinical pathway solution for cancer, including HNC in collaboration with the Danish National Board of health and the national multidisciplinary cancer groups in 2007. Their HNC referral guidelines include a list of red-flag cancer symptoms per HaN subsite and are summarised in Table 2-2 (Roennegaard *et al.*, 2018). In Denmark, the majority of patients are first seen by privately practising ENT doctors, and then a referral is made to the hospital based on the Danish red-flag fast-track referral pathway. Audits of their pathway have found a high detection rate of cancer in the hospital urgent cancer clinics following a priority GP referral, being 40.6% in a large cohort study of 3.165 patients, of which 71.9% were of HNC origin and the remainder metastatic cancer disease presenting with HaN symptomatology (Roennegaard *et al.*, 2018). Allocated cancer slots are available in the hospital for those referred urgently based on the Danish National Board of health referral guidelines, which in collaboration with the Danish Head and Neck Cancer Group (DAHANCA), have streamlined that urgent HNC cases are seen in the hospital within 1 day of referral. Hospitals have implemented telephone referrals rather than paper/online referrals in an attempt to achieve the 1-day target for hospital appointment following primary care/private ENT referral, with very promising results. Reported cancer pick up rates have been noted to be high ranging from 41% to 52%, with a HNC in 17% to 21% of referrals (Sorensen *et al.*, 2014;Toustrup *et al.*, 2011).

Table 2-2. Danish health board HNC red-flag symptoms for urgent referral

Head and Neck cancer subsite	Symptoms and Signs
Sino-nasal cancer	Unilateral nasal stenosis Bloody secretion Recurrent nasal haemorrhage Nasal wounds tumour in the nasal cavity
Nasopharyngeal cancer	Unilateral, secretory otitis media Affection of cranial nerves Special attention on high-risk ethnicity
Oral cavity and Oropharynx cancer	Wounds in the oral cavity or oropharynx tumour in the oral cavity of the oropharynx Pain radiation to ear Enlarged submandibular lymph node
Larynx and Hypopharynx cancer	Hoarseness Difficulty swallowing/globulus Pain radiation to ear
Salivary gland cancer	tumour in the salivary gland Growth of known salivary gland tumour tumour in the salivary gland with simultaneous affection of the facial nerve
Thyroid cancer	tumour in the thyroid gland with simultaneous hoarseness The rapid growth of known thyroid tumours Hard, immobile tumour in the thyroid gland
Unknown primary (neck metastasis)	Enlarged lymph nodes with no infectious or benign cause A lateral neck cyst in patients more than 40 years of age

Influenced by the Danish fast tract programme and based on their red-flag symptoms list, similar HNC fast track programs were later developed in Norway, with the referral guidelines available on their government website (Norwegian Directorate of health, 2015), followed by Finland (Nieminen *et al.*, 2021) and Sweden (the Swedish Standardised Care Pathway program) with the implementations to include HNC spanning from 2009 to 2015 (Wilkens *et al.*, 2016). Similar fast tract referral guidelines for HNC are also available in Spain. The Spanish HNC referral guidelines include oral cavity/lip ulcer >3 weeks, presence of erythroplakia/leukoplakia, hoarseness >6 weeks , dysphagia >3 weeks as well as cervical lymph nodes over >2 cm for >4 weeks with

a note of any generalised symptoms (weight loss, sweating, fever) (Martínez *et al.*, 2021;Martínez *et al.*, 2015). Since their implementation in 2009, the time for primary care visits to secondary care appointments has been steadily reducing, being an average of 7.5 days for HNC. Overall, the number of referrals is also increasing, having doubled compared to the first 5 versus the later 5 years of the implementation of the recommendations (Martínez *et al.*, 2021).

2.2.2.2 The NICE HNC referral guidelines

In the UK, the department of health (DoH) issued the UK National Guidelines for urgent cancer referral of a suspected HNC by primary care professionals with a target of 2 weeks waiting between the referral and the patient's first visit to a HaN or Oral and maxillofacial surgeon. The UK was one of the first countries to introduce a standardised referral pathway for HNC in 2000 (DoH, 2000). This move seemed to be an important step forward in the early detection and management of HNCs. The guidelines included symptoms suggestive of cancer for each HaN subsite as well as a list of symptoms to prompt urgent within 2 weeks review in the hospital. Laryngeal cancer symptoms were hoarseness, pain on swallowing and dysphagia; For the nasopharynx: neck lump, nose obstruction, hearing loss, postnasal discharge; oral cavity symptoms comprised of ulcer, pain, and neck lump; for the oropharynx: persistent sore throat, neck lump, otalgia; hypopharynx symptoms included: otalgia, pain on swallowing, hoarseness; for the nasal cavity: blocked nose and bleeding. For the thyroid, the symptoms were a thyroid lump and discomfort in the lower neck, whereas for the submandibular glands: a lump in the glands, pain, lump in the neck. Of these symptoms, an urgent 2ww referral was recommended for the following list: Hoarseness > 6 weeks, ulcer in mouth >3 weeks, red/white patch in the mouth, dysphagia > 3 weeks, unilateral nasal obstruction, unexplained tooth mobility, persistent neck mass, cranial neuropathies, orbit mass (DoH, 2000). Further updates of the guidance have been made since their initial inception, with the most crucial updates done in 2005 and 2015. Minor further changes were made in 2019. The current guidelines include persistent hoarseness, neck lump over 45 years old for suspected laryngeal cancer referral; ulcer in the mouth for 3 weeks or more, neck lump, oral lump, red/white patch in the mouth for oral cancer possibility and an unexplained thyroid lump for possible thyroid malignancy (NICE, 2021). The otalgia, sore throat

and dysphagia symptoms were removed in the 2015/2021 update and age limits were introduced for laryngeal cancer patients prompting hoarseness referral for those over 45 years of age (NICE, 2021). Another change to the guidance is that oral cavity signs of a lump, swelling and red and white patch should be referred to secondary care following a dental referral rather than after a GP visit. This change was made in an attempt to reduce inappropriate GP referrals by ensuring dentists with significant experience in oral lesions have seen the patients first hence filtering the referrals to those that actually have suspicious lesions. Nevertheless, the dental review approach has raised concerns as to potential delays in such referrals as it is known that just over 55% of the adult English population is recorded to have seen an NHS dentist at least once within a 2-year period and waiting times to see an NHS dentist are not well known (Grimes, Patel and Avery, 2017). Many individuals are not registered with a dentist, and there is no clear guidance as to the pathway of referral when such an occasion arises, which is common, especially for populations of low socioeconomic backgrounds (Grimes, Patel and Avery, 2017). Presumably, patients without a registered dentist still attend their GPs when they have oral cavity-related problems, but no figures are available to further explore this. Lack of access to NHS dentistry, with difficulty seeing but also registering with an NHS dentist, has been an issue prior to the COVID-19 pandemic, but it has been exacerbated during and after the pandemic, with 7 in 10 currently finding it hard to access support (healthwatch, 2022). The reduced access to NHS dental appointments has led to a 65% reduction in oral cancer referrals 6 months after the first lockdown, and it has raised concerns about the impact of this on early diagnosis of oral cavity cancer (Carter, 2021).

During the many iterations of the NICE guidelines, audits were conducted to assess their performance. The first audit of the guidance was published in 2004 from the ENT departments in Essex, showing a cancer pick-up rate of 15%, and in the same period, most cancer (71%) were diagnosed via other routes (Lyons *et al.*, 2004) with similar findings reported by other research groups (Patel, Khan and Thiruchelvam, 2011;White *et al.*, 2004). A similar audit this time from the maxillofacial department of Southmead Hospital in Bristol (Shah, Williams and Irvine, 2006) showed that from a total of 150 referrals for potential oral cancer, only 6% were diagnosed with cancer

during the study period, with most of the cancers seen in the department referred from other routes (67%). The majority of patients were referred with an oral ulcer, and the remainder with white/red/speckled lesions, with the latter not being a good cancer discriminator (Shah, Williams and Irvine, 2006). Further audits across the country showed a low cancer detection rate for the 2ww referral system, which was not significantly different from the non-urgent referrals detection rates with many inappropriately "urgent" referrals by general practitioners and an overwhelmed system (East, Stocker and Avery, 2005;Duvvi *et al.*, 2006;Singh and Warnakulasuriya, 2006). In the 2005 NICE guideline update, dysphagia > 3 weeks, unilateral nasal obstruction, unexplained tooth mobility, cranial neuropathies and orbit mass were removed from the guidance, whereas the requirement of a CXR was added when the symptom of hoarseness was present. A further addition was an unexplained persistent sore throat and unexplained unilateral pain in HaN of 4 weeks or more (The National Collaborating Centre for Primary Care, 2005). Further audits and research papers, following the change in guidance, showed that the cancer detection rate remains low and even lower when compared to earlier years (reduction from 9% to 5% in some audits), with poor compliance to guidelines, whilst at the same, there is a steep increase in the referral rates of over 450% within a 10 year period (Williams *et al.*, 2014;Joshi and McPartlin, 2012). A 2012 systematic review of the 2ww audits confirmed a low conversion rate of 11% via the 2ww route (average number from 6 studies) as well as very low cancer detection via this route when compared to total cancer identified from other referrals, with a mean of 26.3% (Kumar *et al.*, 2012). More recent audits, following the latest update, continued to show overall poor compliance with the referral guidelines (55%) from GP referrals, with an inappropriate referral of patients complaining of globus symptoms (38%) or inappropriate interpretation of the periodicity and persistency of symptoms. High positive predictive values were found for symptoms removed from the guidelines: otalgia: 9.5%; sore throat: 4%; dysphagia:8%; oral bleeding 8%. Overall a 6% cancer detection rate was noted using the 2015 referral guidelines (Mettias, Charlton and Ashokkumar, 2021). These numbers are confirmed by a primary care audit that identified a low conversion rate of 5.5% from 3 primary care practices in the Merseyside region with highlighted issues of limited ENT/Maxfax training within the GPs, making them more likely to over-

refer patients (Talwar *et al.*, 2020). A study looking particularly at the persistent unilateral sore throat symptom for over 4 weeks with or without the presence of otalgia that was present in the previous NICE iteration and subsequently removed was found to have a 9.5% positive predictive value for HNC. The authors have raised concerns as to the NICE decision to remove these symptoms for the HNC referral guidance given the high PPV for HNC (Allam and Nijim, 2019).

A more recent meta-analysis of 17 studies showed, yet again, a pooled conversion rate of 8.8% and a detection rate of 40.8%, including studies from 2000 to 2014 (Langton, Siau and Bankhead, 2016). Over two-thirds of the 2ww patients are referred by GPs, the rest of the referrals coming from dental practices. Of the total referral, 30% - 40% will require a biopsy to establish a diagnosis (Patel *et al.*, 2020;Metcalfe *et al.*, 2019;Piggott, 2015). It has been found that referrals from dental practices have a higher conversion rate, but unfortunately, the numbers of such referrals are low (Metcalfe *et al.*, 2019). This is speculated to be because the majority of patients are from low socioeconomic classes and do not have easy access to dental care due to the perceived high costs associated with dental visits. The dental referral conversion rate is two to over three times higher than the GP referrals (11.9% - 20.8% vs 6.2 % - 6.3%) for oral and oropharyngeal cancers with presenting symptoms of mouth ulcer, lesion, mass, or white patch (Metcalfe *et al.*, 2019).

Despite the NICE national guidelines existing in the UK for GP referrals of patients with suspected HNC for two decades now, as described above (NICE, 2021) (DoH, 2000), some UK regions also have their own separate guidelines, using a combination of symptoms from the current and older NICE guidance (myhealth.london.nhs.uk, 2017). This can cause confusion within the GP community as to when a patient should be referred urgently. The Pan-London Suspected Cancer Referral Guide for HNC is a good example of this. It includes symptoms that were present in 2000, 2005 and 2015 updates of the NICE guidelines as well as a list of risk factors for HNC that include: smoking, oral tobacco use, alcohol, previous HPV or HIV infection, prior irradiation to the HaN region and family history of thyroid cancer (myhealth.london.nhs.uk, 2017). A recent study looking specifically at the detection rate of the pan-London

referral guidelines showed that none of the patients diagnosed with cancer had the following symptoms that are still included in the guidance but have been excluded from NICE since 2005: unilateral tinnitus, nasal discharge, cranial neuropathies, presence of orbital masses, facial pain, unilateral nasal obstruction, tooth mobility, non-healing gums after teeth extraction (Gao *et al.*, 2019). Confusion can arise when, in addition to the national referral guidelines, local cancer networks ask for separate referral forms to be filled alongside the 2ww referrals. Especially when different sets of symptoms are present in the local cancer network forms. This has been raised as a concern in a GP survey of the use of the 2ww referral pathway (Dodds *et al.*, 2004). Ensuring uniform services in secondary care could help manage variations in cancer referral rates (Blank *et al.*, 2014) that cannot be otherwise explained after GP and patient-related demographics have been taken into consideration, as described earlier in section 2.2.1.

To summarise, despite the multiple iterations of the NICE referral guidelines, audits and systematic reviews have failed to show an improvement in the HNC detection rate, with an increase in the referral numbers that are not reflected in a higher number of cancer cases being identified. Many patients are referred without having symptoms included in the guidance, which burdens the referral pathway. Locally designed referral criteria also exist, contradicting the symptoms included in the NICE guidelines, which also introduces confusion to the GPs as to which referral form should be used.

2.2.2.3 The Scottish HNC Referral guidelines

In Scotland, the referral pathway is called urgent suspicion of cancer (USOC), with the same target outcomes as the NICE guidelines but with a slightly different list of included symptoms which has changed in recent years to align more with NICE. The referral symptoms had multiple iterations, with the most recent update in 2019 that include the following symptoms: >3 weeks of persistent unexplained hoarseness, ulcer or swelling of oral cavity >3 weeks, red/white patches in oral cavity >3 weeks, persistent neck lump >3 weeks, persistent pain in throat > 3 weeks, persistent odynophagia >3 weeks, stridor. Previously the dysphagia symptom was also part of the guidance but was removed in the latest review with no reason given for this change

(cancerreferral.scot.nhs.uk, 2019). Audits performed over the years, looking particularly at the Scottish guidelines conversion rates, have found similar outcomes as for the English cohorts. Kennedy et al. (2012) reported an 8% HNC conversion rate from a total of 190 referrals over a 1-year audit period, with a 15% rate when cancers of non-HaN origin with HaN manifestations were included. The detection rate was only 14%, as the majority of HNCs were diagnosed via other routes. Highlighted issues were the incomplete referral letters with a high proportion of missing information about patients' alcohol consumption (78% missing information), smoking history (36% missing information), lack of mention of any red flag symptoms in 12% of cases with no details of symptoms duration in 24% of referrals. Similar to the English cohort, 74% of patients were diagnosed with an advanced disease stage (Kennedy *et al.*, 2012).

Ongoing concerns regarding the quality of information of the urgent suspicion of cancer referral letters have also been expressed and described in the literature from another Scottish group (Moloney and Stassen, 2010). More recent studies of the Scottish referral guidelines (Douglas, Carswell and Montgomery, 2019) have shown persistent problems with non-compliance of the USOC referrals with the most recent guidelines (55% compliance) as GPs were referring patients urgently but without them meeting the set referral criteria. Looking at the overall conversion rate, it was much improved, with an 18.8% cancer detection for the referrals being in accordance with the guidelines. Nevertheless, the total conversion rate, taking into account the inappropriate referrals, was still low at 8.6% for HNC and 11.6% for all types of malignancies identified (Fingland *et al.*, 2018). A previous requirement of the Scottish HNC referral guidelines was for a chest radiography (CXR) to be instructed at the time of referral when the patient has persistent hoarseness of more than 3 weeks. Nevertheless, the study by Fingland et al. identified only 1 X-ray with a positive cancer finding (0.8% conversion rate); hence this requirement was removed from the latest version of the guidelines (cancerreferral.scot.nhs.uk, 2019). A large study from Scotland with thousands of patients (n=2,116) reviewing the outcomes of their USOC clinics found that 42% of patients are reassured and discharged following the first clinic consultation, 27% were followed up for benign disease, and another 22%

followed with suspicion of cancer, with a final cancer diagnosis made in 12% of referrals (Douglas, Carswell and Montgomery, 2019).

To summarise, the Scottish head and neck cancer referral guidelines have also undergone multiple iterations over the years, similar to the NICE guidelines. The HNC conversion and detection rates are still low ranging from 8% to 14%, with a large number of inappropriately filled referrals.

2.2.2.4 Head and neck cancer referral guidelines - section summary

Many different guidelines exist worldwide advising patients and primary care clinicians on red-flag symptoms of HNC. The list of symptoms differs from world region to region and even within the countries of the same geographical area, for example, within Europe. What is more, the list of recommended symptoms can differ even within the same country, as the UK example has demonstrated with a different list of symptoms for England and Scotland. Even within England, different HNC referral guidelines exist in different cities. The evidence behind selecting the recommended list of symptoms is currently based on expert panel recommendations from international or national societies, the government and very limited primary care literature. This is a significant drawback of the design of the current guidelines as they are mainly based on 'expert opinion', which is the lowest level of evidence. It highlights the need for the development of an evidence-based HNC risk stratification guidance for early diagnosis at the point of first contact with a healthcare professional. The literature review performed is targeted at guidelines published to aid the decision-making of healthcare professionals on referring patients with suspected HaN symptoms to rule out cancer. Some of these guidelines, such as the American Cancer Society recommendations and the Make Sense campaign on behalf of the European Head and Neck Society (European Head & Neck Society, 2020;cancer.org, 2021), are also addressed to the general public to inform them about suspicious symptoms for HNC, advising them to see a doctor for further assessment. Many other resources exist that are designed primarily to give information to the general public about HNC symptoms and how to seek medical advice, such as the MacMillan Cancer Support HNC booklet(MacMillan Cancer Support, 2018). A review of HNC symptoms resources addressed directly to patients is vast and is outside the scope of this thesis

work that looks at the understanding of the guidelines addressed to the healthcare professionals to enable the prompt referral of high-risk patients for specialist assessment.

The development and utilisation of a HNC risk calculator informed by the literature and utilising a robust design process will be explored in this thesis as a potential solution to the currently poor HNC early diagnosis statistics and the lack of a strong evidence base in the current referral guidelines for suspected HNC. This endeavour will be based on knowledge of the design and use of similar risk calculators for other common cancers and any current attempts at designing HNC risk assessment tools. The final section of this literature review will explore the evidence on all available risk calculators for common cancers and how these are presently used in medical practice, as well as their outcomes. In developing a cancer risk calculator for HNC, the knowledge of implementation of such tools in other cancers is important, appreciating how cancer risk calculators have been developed and adopted for other cancers, their prediction outcomes and revision process throughout the years. After a review of the available risk assessment tools for all cancers, the remainder of the section will focus on HNC-specific risk calculators setting the scene for this thesis.

2.3 Cancer risk calculators

One of the main future goals of the NHS long-term plan is the diagnosis of over two-thirds of cancers at an early stage (stages 1 and 2). Currently, only 50% of cancers are diagnosed early, and in HNC, this is only a third (NHS, 2019). Improvement in the referral system for suspected cancer and the use of clinical cancer decision tools is one of the focus areas for improvement of early cancer detection. Numerous publications are available that developed cancer risk calculators for common and uncommon cancers. Nevertheless, the majority of these tools have not been widely adopted, used and externally validated; hence their potential applications in primary care for cancer triaging are limited (Steyerberg *et al.*, 2013).

2.3.1 Primary care designed cancer risk calculators

In the UK, a limited number of these cancer decision tools have been developed based on primary care data and have been made available in the electronic system used by GPs when accessing patients' data and are also available online. The three most well-known cancer risk calculators in UK primary care are the Risk Assessment Tool (RAT) (Hamilton, 2009), the Qcancer (Qcancer.org, 2017), and the 7-point checklist (7PCL) for assessment of pigmented skin lesion (MacKie, 1990) and are available in the GP software systems.

The RAT generates positive predictive values for symptoms of 14 cancers; lung, colorectal, prostate, brain, ovarian, oesophageal, kidney, bladder, pancreas, breast, uterine, Hodgkin's and non-Hodgkin's lymphoma, multiple myeloma and metastatic cancer, and is based on symptoms in isolation, a combination of symptoms or relevant blood work (Hamilton, 2010). Early RATs were generated using case-control data from Devon, UK, in the CAPER studies (Hamilton, 2009). Later studies were based on UK-wide data from primary care practices (the Health Improvement Network database – THIN, and the Clinical Practice Research Datalink – CPRD database) (Hamilton *et al.*, 2009;Shephard *et al.*, 2013). Univariate and multivariate logistic regression analysis was performed to identify independent risk factors for cancer. The final model was based on the positive predictive value of one or a combination of two symptoms for each cancer, with a PPV of over 3% being considered a significant threshold, in line with the NICE guidelines recommendation (Hamilton, 2009). The gold standard for conducting studies looking at the association between cancer outcomes and symptoms is a prospective cohort study. Nevertheless, the RATs were generated from the CAPER dataset, which is a pre-populated GP registry with symptoms prior to a diagnosis of cancer being made (Hamilton, 2010). It is unclear if the GP registries were prospectively or retrospectively collected and by whom the data collection and entry in the database were performed. Bias could be an issue here if the data were incorrectly coded by staff who lack understanding of clinical language used for symptoms description. Moreover, a case-control methodology was used in these studies, as the authors felt that it would have been impossible to conduct a cohort study in primary care due to the relative frequency of symptoms and the rarity of the cancer

diagnosis (Hamilton, 2010). However, the case-control format does not give information about the true incidence of cancer and symptoms but rather a conventional estimation of the likelihood ratio of symptoms association with a cancer diagnosis (Cole *et al.*, 2011).

Qcancer calculates the probability of having a common undiagnosed cancer based on an individual's age, gender, postcode, smoking and alcohol history, family history of some cancers and other chronic conditions, symptoms and signs and body mass index (Hippisley-Cox and Coupland, 2013a). The cancer probabilities included in the calculator are for prostate, pancreas, gastro-oesophageal, haematological, lung, colorectal, renal, and testicular cancer. No Qcancer tool is available for HNC, but a percentage calculation for the likelihood of "other" cancers is given, which should include HNC. Within the list of symptoms, there are some HaN symptoms (weight loss, difficulty swallowing, neck lump) and risk factors for HNC (smoking, alcohol), so it could be that in future iterations of the Qcancer tool, separate probabilities could be given for HNC (Qcancer.org, 2017). A score is also given based only on the patient's age and gender alone. A cohort study design was used for the development of the Qcancer algorithm using multivariate logistic regression analysis using a large database of over 12 million primary care patients' records from the QResearch database (Qcancer.org, 2017;Hippisley-Cox and Coupland, 2013a;Hippisley-Cox and Coupland, 2013b). This design eliminates the previously mentioned issues of the case-control format used in the RAT studies, but the problem of using already populated pre-recorded databases remains. In cohort studies, ideally, the list of possible symptoms should have been already decided at the beginning of the study based on an in-depth review of the literature to ensure all symptoms, even rare ones, are considered during data collection (Cole *et al.*, 2011). If a symptom is not being asked about, the patient will not necessarily mention it during the consultation, hence introducing reporting bias.

7PCL is a scoring system for the diagnosis of melanoma, with a score of 3 raising the possibility of melanoma. This was based on a case-control study with an initial recommendation based on univariate proportion analysis (MacKie, 1990). The same

methodological issues discussed above relating to the case-control format also exist for this study. The data were recorded by interviews with the patients that were scheduled after the diagnosis of skin cancer was made, therefore introducing recollection bias that can compromise the quality of the collected data. The cut-off point has been more recently adjusted to 4, with a higher prediction power based on the area under the receiver operating curve analysis (Walter *et al.*, 2013).

Despite the availability of these risk assessment tools since 2013 in GPs' online software, a recent GP survey has shown that only a third of the participating practices have downloaded and used these resources (Price *et al.*, 2019). A qualitative synthesis of UK-based GPs' views on using the above cancer risk tools suggests that GPs understand and are aware of the need for knowledge of cancer red flag symptoms and signs and that these tools can help them towards this (Bradley *et al.*, 2021). The tools have been used for re-assuring anxious patients elaborating on their low risk of malignancy, and also have helped educate patients in altering lifestyle behaviour that can reduce their lifelong risk of developing preventable cancers. Issues identified were lack of training in using such tools, IT integration problems, as well as fear that these tools may not be endorsed by the referred secondary care speciality, resulting in a dismissal of the high-risk probability and the referral returned to the primary care doctor (Bradley *et al.*, 2021).

A common theme is a misconception that the risk assessment tools will replace the GP's clinical judgement and gravitas as clinicians. The realisation of the aiding nature of these tools is essential in order to be effectively used, as current views include annoyance of the alert messages generated by the risk tools that GPs feel can disrupt their consultation. Another point of concern is that the risk tools are not always in accordance with the NICE guidance recommendations for referral of suspected cancer of a given subsite, causing confusion and collaboration between primary and secondary care, and NICE is required for uniform policies incorporating the available risk calculators to be achieved (Bradley *et al.*, 2021). Issues that other studies have highlighted have been the potential increase in the number of possible cancer referrals by using these tools, and that could potentially cause extra burden to secondary care

(Hamilton *et al.*, 2013;Green *et al.*, 2014). It has also been noted that there has been variation in the uptake of the cancer diagnostic tools in different GPs evaluated, with some regions more in favour of their use than others (Dikomitis, Green and Macleod, 2015).

A recent review of the literature identified only 9 tools that have data available evaluating any improvement in cancer diagnosis decision-making from their use in primary care (Chima *et al.*, 2019). Only one study was found showing that they can help reduce the time to cancer diagnosis, three improved decision making and three helped in the improvement of the content of the referral letters, prescribing and cost-effectiveness (Chima *et al.*, 2019). A negative issue of the use of diagnostic tools that was highlighted in the review was GPs' mistrust of the result of the tools when there was disagreement between their decision to refer or not and the tool's recommendation. A lack of understanding of the evidence base behind the design of the tools made the trust issue worse. Moreover, GPs felt that the use of the tools would increase the referral to secondary care by over-referring low-risk patients hence compromising their role as gatekeepers. Finally, challenges were found in using the tool during the consultation as it was noted to increase the time needed to complete each consultation (Chima *et al.*, 2019). In an English study, the use of RAT and Qcancer has resulted in an average of 19% increase in the cases investigated and/or referred to hospital for potential cancer, but significant variations were noted across the country. On 54% of occasions, the GPs' perception of cancer risk was the same as the one indicated by the risk tool. The calculated risk was lower at 15% and higher at 31%. No results were available in the final cancer diagnosis; hence assessment of diagnostic outcomes was not possible (Moffat, Ironmonger and Green, 2014). Again, another highlighted negative issue was the concern that using the tools takes over 10 minutes, which is the usually allocated consultation time per patient. Another expressed concern was that the use of technology could impact the GP-patient interaction as well the reliance of the tool in coded data from the GPs registries that can vary for each practice. Finally, the selected threshold to prompt a warning for referral was challenged by GPs that wanted to know more about how this is decided (Moffat, Ironmonger and Green, 2014). Despite the negative points being highlighted, it appears to be increased

awareness of the commonly used cancer diagnostic tools and appreciation of the potential benefit of their integration into primary care consultations (Dikomitis, Green and Macleod, 2015).

Aside from these three risk calculators that have been recently added to the UK GP software, many others exist looking at one specific cancer site or assessing multiple sites. Two more models were found looking at predicting multiple cancers similarly to the RAT and Qcancer tools, developed by Holtedahl et al., and by Muris et al., for abdominal cancers (Holtedahl *et al.*, 2018) (Muris *et al.*, 1995) with only the latter being externally validated. The external validation cohort of 810 patients from the Netherlands was prospectively collected in the past for the different purpose of assessing abdominal symptoms in relation to cost and resource use in primary care. This approach can introduce potential bias in the reported symptoms in the database as they were not collected primarily to investigate cancer occurrence in the studied population (Elias *et al.*, 2017). The development phase of the two models had a solid study design methodology. The Holtedahl et al. (2018) had a very robust methodology, using a pre-designed form for symptoms list collection using a prospective methodology over a period of 10 consecutive days resulting in the collection of data from 61,802 consultations from GPs in Norway, Denmark, Sweden, Scotland, Belgium and the Netherlands (Holtedahl *et al.*, 2018). Muris et al. (1995) performed a 1-year prospective cohort study across 80 GPs in the Netherlands, collecting data for a total of 933 patients with abdominal symptoms. No prior sample size calculation was performed in these studies. Overall, despite a plethora of developed risk calculators for common cancer, only a minority has been externally validated. The evidence about externally validated risk calculators targeted for primary care use will be presented in the paragraphs below.

2.3.2 Externally validated cancer risk calculators for use in primary care

The Q cancer tool has been externally validated using primary care cohorts for pancreatic and colorectal cancer, with cohort size ranging from 618 cancer cases to 4798, preserving a good prediction power (AUC 0.89 – 0.92) (Collins and Altman,

2013;Collins and Altman, 2012;Hippisley-Cox and Coupland, 2012b;Hippisley-Cox and Coupland, 2012a) as well as for gastroesophageal cancer, validated in a cohort of 2527 cancer cases with AUC 0.89 - 0.92 (Hippisley-Cox and Coupland, 2011), ovarian cancer, including 976 cases and an AUC of 0.84 (Hippisley-Cox and Coupland, 2012d) and renal tract cancers with 2878 cases achieving AUC of 0.91 - 0.94 (Hippisley-Cox and Coupland, 2012c). The Qcancer and THIN longitudinal databases were used to perform the validation studies for this model based on millions of primary care consultations. Similarly, the RAT has been externally validated in large prospective cohort studies for lung and colorectal cancer in datasets ranging from 810 up to 1433, following appropriate sample size calculations (Hamilton *et al.*, 2013;Elias *et al.*, 2017;Hamilton *et al.*, 2005). Elias et al. (2017) identified a similar AUC at external validation, of 0.79, for colorectal cancer. A rise in the identification of lung cancers by 37% (from 127 to 174) was found in a pilot study of implementing the lung RAT over a 6 month period in practices across the UK, alongside a 19% increase in identifying early-stage cancer. In the same study, new colorectal cancer diagnoses increased by 7% (from 134 to 144) with no change in the cancer stage at the time of diagnosis (Hamilton *et al.*, 2013).

Three studies have been identified assessing the use of skin cancer support tools (English *et al.*, 2003;Del Mar and Green, 1995;Gulati *et al.*, 2015). Both validation studies were performed in Australia. Del Mar and Green (1995) performed a prospective randomised control trial based on two GPs assessing a total of 5,823 cases of melanotic skin lesions. No benefit was found in using the decision aid algorithm, with no reduction in the proportion of benign lesions compared to malignant lesions excised, whereas English et al., 2003 randomised control trial including a much larger number of GPs (223 practices and 468 practitioners), with a prior sample size calculation, including a total of 8,563 cases, using a slightly updated version of the skin cancer tool algorithm, found a reduction in the number of benign skin lesions removed from 93.8% to 88.8% but no difference in the number of melanomas excised. A survey performed by Gulati et al. (2015) assessed the impact of the GP Skin cancer toolkit that was created by the British Association of Dermatologists and the Cancer Research UK (CancerResearchUK, 2020). The study included 20% of GPs in England

that used the tool for a period of 6 months and identified an increase in the number of appropriate GP referrals when using the decision aid without increasing the overall number of urgent referrals (Gulati *et al.*, 2015). The evaluation of the use of red-flag symptom flashcards and the local cancer referral pathways by GPs in Australia did not improve the mean time to cancer diagnosis (Emery *et al.*, 2017).

A 2016 systematic review assessed all symptom-based colorectal cancer risk calculators designed for primary care use to identify those that had been externally validated (Williams *et al.*, 2016). Fifteen colorectal cancer models were identified that met the tight inclusion criteria ensuring that all models had information on performance, accuracy, and discrimination. Of those, only six had been validated in external populations. Only those being either developed or validated using primary care data will be presented here. The Qcancer and RAT colorectal cancer tools, as previously described, and the Bristol-Birmingham (BB) equation model were the only ones being developed and validated in a primary care population (Marshall *et al.*, 2011). The BB equation model used data from the THIN longitudinal database employing a case and matched control study setting, including 5,477 cancer cases and 38,314 controls, and the validation was done using the CAPER dataset (Marshall *et al.*, 2011). The Fijten model was developed in primary care but employed a cross-sectional retrospective data collection study setting that increased bias relating to data recording (Fijten *et al.*, 1995). The study cohort was also small, including 290 patients and only 9 cancer cases, with no sample size assessment being performed. Despite the fact that the recorded AUC was very high at 0.97, the results are questionable due to the population and study setting drawbacks (Fijten *et al.*, 1995). The Fijten model was validated in two large prospective secondary care databases from the UK database (n=3,302) and the Netherlands (n=933), with its AUC dropping to below 0.775 (Hodder *et al.*, 2005;Elias *et al.*, 2017). An additional Danish colorectal cancer model (Nørrelund and Nørrelund, 1996) has also been validated using the Elias et al. (2017) primary care database. Giving a total of 5 colorectal cancer risk calculators, which are either developed or validated with primary care data. Nevertheless, only the RAT tool has been assessed in a prospective cohort study setting with no control group, hence having the most robust methodology (Hamilton *et al.*, 2013). Assessing for future

development of colorectal cancer in a currently asymptomatic population, 52 models have been identified in a literature review by Usher-Smith *et al.* (2016). Of these, 37 mentioned accuracy and discrimination values, and external validation has shown adequate discrimination (AUC >0.7) in only 10 models (Usher-Smith *et al.*, 2016).

A similar systematic review for lung cancer has identified seven cancer risk tools that can be used in primary care, but none of these has been externally validated (Schmidt-Hansen *et al.*, 2017). A further tool that had been excluded by Schmidt-Hansen et al.'s (2017) tight exclusion criteria is a lung cancer risk tool by Iyer-Omofoman et al. (2013). It was excluded from the review as the same database was used to extract the development and validation cohort for the risk calculator, which was considered of high risk for bias (Schmidt-Hansen *et al.*, 2017). They developed a calculator using logistic regression analysis and validated it on an independent dataset from the same large primary care database used to develop the model, both deriving from the THIN database (Iyen-Omofoman *et al.*, 2013). The variables included by Iyen-Omofoman et al. (2013) were age, sex, smoking history, socioeconomic status, symptoms of cough, haemoptysis, dyspnoea, weight loss, chest infections, chest pain, hoarseness, upper respiratory tract infections and COPD. The model performed well at internal validation with an AUC of 0.88 (Iyen-Omofoman *et al.*, 2013).

There is overall limited good-quality evidence of the use of cancer risk calculators outside their development population. Few randomised controlled trials have been conducted looking at the effectiveness of such tools in screening healthy individuals for future cancer development in primary care, but none looking at currently symptomatic individuals (Walker *et al.*, 2015). McCowan et al. have developed a breast cancer risk calculator based on a Scottish dataset which was validated using a further Scottish dataset from 11 GPs. Both datasets were prospectively collected, but no power calculation was performed. Both derivation and validation cohorts are probably underpowered, with 807 patients and 59 cancers in the development and 97 patients with 5 having breast cancer in the validation study. The predictive variables included age, a discrete breast lump, skin thickening, lymphadenopathy, and a breast

lump of over 2cm. No sensitivity, specificity or AUC measures were recorded (McCowan *et al.*, 2011).

To summarise, many risk calculators exist for common cancer based on symptoms that can be potentially targeted for use in primary care. Of these, external validation evidence of use by GPs with improved cancer prediction outcomes exists for the RATs for common cancers, Qcancer tool for common cancers, the 7PCL; the English et al.(2003) model and GP Skin cancer toolkit for skin cancer and finally the Bristol-Birmingham equation and Fijten et al.,(1995) model for colorectal cancer. It is evident that the main issue of the methodology of the currently available calculators designed for primary care use is that they are based and/or validated on datasets that are not collected for the purpose of cancer risk model development, and sample size calculations are rarely reported in the studies even in those with prospective data collection with databases designed for the purpose of cancer assessment tools development. Aside from these risk calculators that have been externally validated for use in primary care, numerous others exist that have been developed and validated using secondary care cohorts. These will be presented per cancer site in the following sections of this chapter. These mainly focus on screening asymptomatic individuals, but any available risk calculators for symptomatic patients will also be discussed. It is important to pool this evidence to understand how the concept of model prediction for cancer diagnosis was first derived and then developed, progressed, and changed over the years for different cancer sites, incorporating diagnostic advances in medicine but also technological developments in data science analytics. This knowledge of the methodology used, data collected and analysed, and outcomes achieved for other cancer calculators is fundamental when new risk calculators are to be designed. This is the aim of this thesis for HNC. The evidence for each cancer will be presented separately in the sections below, followed by a summary section at the end of the chapter where their outcomes will be summarised, helping to make informed decisions for the methodology that will be followed for the development of the HNC risk calculator.

2.3.3 Secondary care designed risk calculators

2.3.3.1 Prostate cancer

Prostate cancer research has been at the forefront of the development and use of cancer risk prediction models. The models were initially based on the prostate-specific antigen (PSA) levels, risk stratifying patients requiring prostate biopsy (Caras and Sterbis, 2014). PSA testing alone has been found to be a weak discriminator of prostate cancer with a pooled AUC of 0.66 (0.59 – 0.73) after a meta-analysis of 16 studies discriminating men at risk of being diagnosed with any prostate cancer (Louie *et al.*, 2015).

In 2006, a landmark study on prostate cancer established a prostate cancer risk calculator based on patients' age and race, the value of the PSA, family history of prostate cancer, digital examination findings and previous prostate biopsies. It is known as the Prostate Cancer Prevention Trial (PCPT) risk assessment tool and was developed based on data from 5,519 men that had previously participated in the PCPT trial finasteride vs placebo for men 55 years old or more, with normal rectal examination and a PSA equal of less than 3ng/ml (Thompson *et al.*, 2006). No sample size calculation specific to this study design was mentioned. The predictive power of the tool was 0.702 for cancer diagnosis and 0.698 for high-grade cancer diagnosis (Thompson *et al.*, 2006). Since then, it has had several iterations in an attempt to further improve its diagnostic efficacy with the addition of further significant factors (Ankerst *et al.*, 2008;Ankerst *et al.*, 2014;Ankerst *et al.*, 2013). External validation of the prostate cancer risk calculator was successful in different populations using either retrospective datasets (n=1,280), contemporary longitudinal cohorts (n=3,482), or datasets initially collected for other purposes (i.e. the SABOR dataset n=3,379, PRAP dataset n=624, Early Detection Research Network Cohort n=5.519), achieving similar discriminatory power ranging from 0.60 to 0.80 (Parekh *et al.*, 2006;Hernandez *et al.*, 2009;Kaplan *et al.*, 2010;Eyre *et al.*, 2009;Nguyen *et al.*, 2010;Auffenberg *et al.*, 2017). It was the first risk calculator to become available online, with a further update in 2012 to incorporate the feature of differentiating between low and high-grade (Gleason >7) prostate cancer and the addition of further biomarkers (free PSA, PCA3 and T2:ERG) using again patients from the PCPT database (n=6,664). Power analysis

was performed in this study. The addition of the biomarkers increased the AUC to 79.8% (compared to 0.702 for PCPT v.1) for differentiating cancer from no cancer, but no difference was found in differentiated low-grade cancer from high-grade cancer or low-grade cancer from no cancer (Ankerst *et al.*, 2014). Its performance has been assessed in another large longitudinal data cohort of 4,289 patients, showing a worse AUC of 0.62 (Auffenberg *et al.*, 2017). It has been recently replaced by the newest PBCG (prostate biopsy collaborative group) risk calculator developed by the same research team using a newer cohort of 15,611 patients prospectively followed up from 2006 to 2017, showing a higher AUC (75.5% compared to the PCPT at 72.3% on cross-validation (Ankerst *et al.*, 2018). It is now recommended to be used instead of the PCPT as an online tool. It uses similar variables for risk stratification: race, age, PSA, findings of the digital rectal examination, prior biopsy findings and family history of prostate cancer (PCPT, 2006-2018).

The European Randomised study of screening for prostate cancer (ERSPC) tool was also introduced in 2007 (Steyerberg *et al.*, 2007) based on the validation and refinement of a previously designed risk tool by Kattan et al. (2003) designed to predict indolent prostate cancer. The model AUC at the development phase ranged from 0.64 - 0.79 and 0.61 - 0.76 at validation, with the prediction power of the full model (base model plus millimetres noncancerous tissue, millimetres of cancer from the biopsies and medium model (base model plus ultrasound volume, % of positive cores) being higher compared to the base model (PSA, biopsy Gleason grade, clinical stage). The model development was based on a prospectively collected sample size of 409 patients, whereas the validation and refinement were performed using 279 patients from the ERSCP database (Kattan *et al.*, 2003;Steyerberg *et al.*, 2007). This database was developed as part of a multicentre randomised control trial with the aim to compare prostate cancer mortality between an intervention arm of screening for cancer versus a control group including 182,000 patients (Roobol and Schröder, 2003). No sample size analysis was performed as part of the development and validation studies, and it is mentioned as one of the limitations of the methodology. Six versions of the ERSPC calculator are currently available. The first two are designed for use by the public (Risk calculator 1 variable: family history, age, urinary symptoms, Risk

Calculator 2: PSA value), whereas the rest are targeted for use by clinicians. The latter versions are based on patients' age and race, the value of the prostate-specific antigen (PSA), family history of prostate cancer, digital examination findings, prostate volume data from prostate ultrasound, the prostate index blood test (phi), and results from previous prostate biopsies with the prediction power of the models increasing the more complex the calculators' variables are (SWOP, 2022). The ERSPC has been found to outperform PCPT at external validation with AUC values between 0.65-0.8 across various prospective and retrospective validation cohorts with sample sizes ranging from 390 - 2000 cases (Cavadas *et al.*, 2010;Oliveira *et al.*, 2011;van Vugt *et al.*, 2011;Trottier *et al.*, 2011;Poyet *et al.*, 2016;Foley *et al.*, 2016). It has been recommended for use by general practitioners, and it is available online, and recently a smartphone version of the calculator was developed, which was received well by primary care doctors and allied professionals in the decision-making process for urology referrals for suspected prostate cancer (Pereira-Azevedo *et al.*, 2017). In 2014, The Dutch government introduced new guidance for GP referral of patients with possible prostate cancer to the hospital for urology review. Under the new guidance, a GP referral can be made with a PSA of 3mg/ml (previously being 4 or more) with the condition that the hospital urologists will use the ERSPC tool (calculator 3 and 4) to decide if a prostate biopsy is required (Federatie Medisch Specialisten, 2016). Its use resulted in a 50% reduction in the number of GP referrals for biopsy, having a positive predictive value of 79%. Follow-up of patients that did not have biopsy based on the risk calculator guidance showed a 100% negative predictive value for clinically significant prostate cancer (Gleason >7) in subsequent biopsies and 96% for any prostate cancer. Additionally, GPs were 94% compliant with the advice on referral received by the online risk calculator. The results of this study are very promising as a groundwork for establishing individualised risk stratification for prostate cancer in the primary care setting, which has not been assessed before (Osses *et al.*, 2018)

In 2014, a systematic review of the literature identified 127 unique prostate cancer prediction models (up to June 2012), with the 6 most commonly used models having an AUC ranging from 0.66-0.79 following meta-analysis (Louie *et al.*, 2015). Since then, many research groups have attempted to introduce further similar prostate cancer

risk calculators. The c-statistic remains in the range of 0.55-0.938, with variations depending on the tested population (Hernandez *et al.*, 2009;Foley *et al.*, 2015;Bandala-Jacques *et al.*, 2021). The majority of these calculators are based on a combination of clinical and biochemical markers (blood and urinary markers, including genetic tests) to predict prostate cancers in the pre-biopsy setting. Several updates of the two main calculators have been proposed, incorporating these new biomarkers, and internally validated. Nevertheless, there is currently no sufficient number of external validation studies to allow a pooled analysis of AUC and head-to-head comparison with the PCPT and ERCPC (Loeb and Dani, 2017;Bandala-Jacques *et al.*, 2021) .

2.3.3.2 Breast cancer

Several breast cancer risk calculators exist, but a previous review of the literature identified only six of 17 being externally validated and also demonstrating relatively low prediction power at validation (Meads, Ahmed and Riley, 2012). The Breast Cancer Risk Assessment Tool (BCRAT), also known as the Gail model, was first developed in 1989 (Gail *et al.*, 1989). The initial model included 5 factors: age, menarche age, age at first live birth, number of first-degree female relatives with breast cancer and number of previous breast biopsies. It estimates future breast cancer development in a currently asymptomatic population, reporting on hazard ratio values rather than AUC. The results were based on a case-control study from the Breast Cancer Detection Demonstration Project in the USA, including 5,998 patients (Gail *et al.*, 1989). Since its first inception, it has undergone several modifications over the years with the addition and adjustment of included risk factors, and it is currently available on the National Cancer Institute website (Costantino *et al.*, 1999;National Cancer Intitute). Race, ethnicity and history of atypia were included in the most recent version of the calculator (National Cancer Intitute;Costantino *et al.*, 1999;Gail *et al.*, 1989). The BCRAT is currently used to evaluate the 5-year risk of invasive breast cancer of at least 1.67% that would benefit from chemoprevention (National Comprehensive Cancer Network, 2020). Numerous studies have externally validated the BCRAT with an average AUC value of 0.6 (National Comprehensive Cancer Network, 2020;Terry *et al.*, 2019;Rockhill *et al.*, 2001).

The later developed Breast Cancer Surveillance Consortium Model (BCSC) is also based on the BCRAT model but also includes breast density as one of the risk factors in the model (Breast Cancer Surveillance Consortium Risk Calculator, 2017), achieving slightly higher AUCs compared to the base BCRAT model of 0.63 to 0.66 at large external validation studies (n=4,000- 250,000) (Vachon *et al.*, 2015a;Tice *et al.*, 2019). The Rosner-Colditz model includes age, age at menarche, age at first birth and each subsequent birth and menopause age (Rosner and Colditz, 1996). Further large sample size studies (n>5,000 cases) updated the risk calculator to include more detailed variables, such as results requiring investigations from breast tissue biopsy and ultrasonographic breast assessment, but the AUC of these models was only increased by a modest 0.031 (Colditz and Rosner, 2000;Rice *et al.*, 2017). It has also been externally validated, having a moderate discrimination ability (AUC 0.59 – 0.60 (Rosner *et al.*, 2013). More recently, a risk model, retrospectively analysing data from 2,283 patients, had incorporated artificial intelligence achieving higher AUCs (AUC 0.65 and 0.60 vs 0.57 for standard density-based calculations) (Dembrower *et al.*, 2020;Yala *et al.*, 2019). Similar results of improved AUC for deep-learning breast density evaluation have also been reported in a much larger retrospective study of 39,571 patients (Yala *et al.*, 2019).

Aside from the models including only demographic and lifestyle risk factors as well as biopsy and ultrasonographic data, more recent calculators also include genetic data in the evaluation of breast cancer prediction. An example is the Tyrer-Cuzick model, also known as the IBIS model, which incorporated a segregation analysis that assesses the probability of the presence of genetic mutation based on pedigree family data (Brentnall and Cuzick, 2020). Later model iterations included more risk factors and adjustment of the polygenic risk scores based on the BRCA1 and BRCA2 gene locus and other lower penetrance genes (Brentnall *et al.*, 2020;Brentnall *et al.*, 2015;Brentnall *et al.*, 2018;Brentnall *et al.*, 2019;Vilmun *et al.*, 2020). It performs well on external validation even in women with a low risk of developing breast cancer (Terry *et al.*, 2019;Cintolo-Gonzalez *et al.*, 2017). The Claus model is also based on genetic risk factors but does not include any other demographic or lifestyle factors apart from the previous family history of breast cancer and/or ovarian cancer, which

is a prerequisite for its use, reporting on cumulative lifetime risk of breast cancer development, found to be 92% for women carrying the susceptibility allele (Claus, Risch and Thompson, 1991;Claus, Risch and Thompson, 1993;Claus, Risch and Thompson, 1994). The model did not perform that well on external validation with a very low specificity of 9.6%, despite a very high sensitivity of 98% and an AUC of 0.745, likely due to the limited risk factors included in the model (Fischer *et al.*, 2013;Amir *et al.*, 2003).

Similarly, the BRCAPRO model also includes genetic risk factors (Berry *et al.*, 1997;Berry *et al.*, 2002;Mazzola *et al.*, 2015), with the updated version also incorporating other risk factors being ethnicity and tumour markers with good discrimination on external validation studies (AUC: 0.76 for the base model; AUC: 0.81 for the model with additional biomarkers) (Fischer *et al.*, 2013;Biswas *et al.*, 2012;Mazzola *et al.*, 2015). The later developed Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) is also based on genetic risk factors (Antoniou *et al.*, 2002) but also includes other risk factors such as breast density and hormonal factors in more recent updates (Antoniou *et al.*, 2008;Antoniou *et al.*, 2003). It performs better than the previously mentioned genetic models on external validation as it is highlighted that family cancer history extends beyond first- and second-degree relatives (Fischer *et al.*, 2013;Cintolo-Gonzalez *et al.*, 2017). Finally, the Myriad model based on genetic as well as family and demographic factors was first developed in 1997 (Shattuck-Eidens *et al.*, 1997) with later iterations in 1998 and 2002 with reported sensitivity and specificity combinations of 0.71 and 0.63, respectively (Frank *et al.*, 2002;Frank *et al.*, 1998). Head to Head comparison for models including genetic test variables compared to the simpler models mentioned earlier, such as the BRCA, shows an average increase in AUC by 0.03 from 0.66 to 0.69 (Vachon *et al.*, 2015b). Looking at performance within the risk calculators that include genetic tests, a large-scale external validation, including 7,354 patients, found the BOADICEA to have an AUC of 0.791, being 0.796 for the BRCAPRO with the sensitivities of 82.1% and 84.3% respectively combined with slightly lower specificities of 56.8% and 55.5% respectively. These two being the best performing models (Fischer *et al.*, 2013).

2.3.3.3 Lung cancer

Prediction models for malignant pulmonary nodules have also been introduced and externally validated. A systematic review of the literature in 2016 identified 25 distinct prediction models for lung cancer screening. Of these, 11 included only epidemiological factors hence requiring no specialist input for the cancer probability to be calculated (Gray *et al.*, 2016). The variables included in these models were: personal risk factors: age, sex, ethnicity, body mass index, history of previous X-rays, socioeconomic status (education level), history of previous malignancy; smoking history: smoking status, start age, cessation age, smoking duration, cigarettes per day, pack-years, quit duration, passive smoking; family history of cancer; exposure to asbestos, dust, a past medical history of asthma, hay fever, emphysema, COPD, pneumonia. These models have performed reasonably well on internal and external validation, with reported AUC ranging from 0.57 – 0.859. The two best-performing models were the Hoggart (AUC: 0.843) and PLCO models (0.859), but none of these models has been externally validated in more than 5 new cohorts (Gray *et al.*, 2016). The Hoggart model includes age, smoking status, smoking start age, smoking duration, and cigarettes per day as the model variables, and it is restricted to patients over 35 years of age (Hoggart *et al.*, 2012). The PLCO model includes more variables that are: age, BMI, history of previous chest X-rays, education level, smoking status, smoking duration, pack-years, family history of cancer and past medical history of COPD without age restrictions (Tammemägi *et al.*, 2013). Newer models tested the addition of genetic tests in the above epidemiological factors but did not perform better when compared to the above-mentioned epidemiological models, with AUC values ranging from 0.639 – 0.773) (Gray *et al.*, 2016).

More recent models include patient demographics and past medical history (age, sex, family history of lung cancer, history of emphysema) but also lung nodule characteristics based on CT scanning. The size of the nodule being the most important predictor in the multivariate models. These models have very high discriminatory power, some of them with AUC values over 0.9 in both the design and the validation cohorts (McWilliams *et al.*, 2013;White *et al.*, 2017). In contrast, other models are based on lung nodule scan characteristics alone for lung cancer prediction

(Balagurunathan *et al.*, 2019;Soardi *et al.*, 2017;Soardi *et al.*, 2015;Perandini *et al.*, 2017). Their drawback is that a detailed assessment of scans is required by specialists. Despite their high AUC, which is at a range of 0.822 - 0.893, they cannot be used as an aid to triage patients in nurse-led clinics or in the primary care setting(Gray *et al.*, 2016).

In the UK, the Liverpool Lung Project Model (LLP) is used in selective screening trials (McRonald *et al.*, 2014). Despite the fact that it has a lower AUC than the PLCO and Hoggart models, it has been externally validated in 4 large UK cohorts with AUC ranging from 0.67 to 0.8, which is considered satisfactory and applicable to the UK population. Its other advantage is that it can be used in a primary care setting with variables included in the model being: age, sex, previous malignant tumour, smoking duration, family history of lung cancer (Cases and age of onset), history of asbestos exposure and COPD (Cassidy *et al.*, 2008).

2.3.3.4 Colorectal cancer

Several colorectal cancer prediction models exist. The majority of them are based on non-invasive variables based on patients' demographics, medical history, and social and dietary habits. A recent systematic review of the literature identified 16 such models, with all demonstrating similar discrimination power on external validation with the same validation cohorts (Smith *et al.*, 2019). This was an update of an older review that identified six out of nine models being validated with highlighted low prediction power capabilities at external validation studies (Win *et al.*, 2012). The best-performing models were the colorectal cancer predicted risk online (CRC-PRO) (Wells *et al.*, 2014), The Steffen et al. (2014) and Shin et al. (2014) models with AUCs ranging from 0.68 – 0.71 (Smith *et al.*, 2019). In the Well's model, the included variables are age, ethnicity, smoking pack-years, alcoholic drinks per day, BMI, years of education, aspirin use, family history of colon cancer, use of multivitamins, red meat intake per day, diabetes, hours of physical activity per day, regular use of NSAIDS and oestrogen (for women) (Wells *et al.*, 2014). Steffen's and Shin's model include fewer variables but directly comparable AUCs to Well's model (Steffen et al. variables: age, sex, BMI, diabetes, previous endoscopy findings, smoking and alcohol

history; Shin et al. variables: age, BMI, family history of cancer, alcohol, and red meat consumption) (Steffen *et al.*, 2014;Shin *et al.*, 2014).

These models are a good example of screening tools that can be used in the primary care setting as they are based on patients' demographics and past medical history. Hence, no specialist input is required. What is not taken into account in these models is the potential additive effect of patients' symptoms in the cancer diagnosis. A recent review of the literature identified 15 risk calculators that include symptoms as part of their algorithms, nine of them being developed in primary care and the remainder in a secondary care setting (Williams *et al.*, 2016). The symptoms included in these models are rectal bleeding, change in bowel habits, diarrhoea, constipation, abdominal pain, weight loss, loss of appetite, mucous secretion, and rectal examination findings. A few of the models also included results from blood tests easily obtainable in the primary care setting: haemoglobin count and MCV. The models developed in primary care setting had very high AUCs (all above 0.89). The models developed in the secondary care had a lower AUC (0.8 – 0.9) but were still better than those that did not include symptoms. These symptom-based models have a great potential for triaging patients requiring referral to secondary care for exclusion of cancer and have been found to perform better when compared to the NICE guidance criteria for a 2ww referral for suspected cancer (Williams *et al.*, 2016).

Several prediction models also exist that have incorporated genetic variables in the risk calculation of colorectal cancer, focusing on asymptomatic populations. A review of the literature identified 29 such models, including at least one single nucleotide polymorphism. Models including only genetic variables had very low discriminatory power (AUC: 0-56 – 0.57) (McGeoch *et al.*, 2019). The addition of other past medical history risk factors increased the AUC ranges to 0.61 – 0.63. Models with genetic variables and age had AUC ranges of 0.56 to 0.71. It is, therefore, questionable if screening using prediction models should include genetic variables that require expensive genetic testing. Moreover, if predictive models are included in cancer detection, it needs to be carefully considered if they should focus on the general

population or on the symptomatic cohorts that seek review by their general practitioners as in the latter, the efficacy of predictive models is much higher.

Logistic regression analysis was used in the development of most of the predictive models. One study that assessed the use of machine learning approaches found that the random forest analysis performed better than logistic regression in terms of overall predictive power, with the rest of the artificial intelligence approaches giving a lower prediction (Kop *et al.*, 2015). Nevertheless, during the qualitative analysis of the models' output, the random forest had worse performance when compared to logistic regression, with many of the predictors included in the random forest model being not meaningful or attributed to an incidental finding (Kop *et al.*, 2015).

2.3.3.5 Uterine cancer

Only a few attempts have been made to develop risk calculators for uterine cancers. Husing et al. (2016) published a prediction tool for the development of endometrial cancer in an asymptomatic population. The factors included in the model were BMI, menopausal status, menarche age, age at menopause, use of the contraceptive pill, parity, age of first live birth, smoking status, and duration of menopausal hormone therapy. The AUC was high at internal validation (77%), but no external validation was performed (Hüsing *et al.*, 2016). The latter model was based on a European cohort study, whereas Pfeiffer et al. developed a similar calculator based on a USA database (Pfeiffer *et al.*, 2013). The AUC of this model was lower at 68%, with the included variables being: BMI, menopause status, menopause age, BMI parity, smoking status, contraceptive pill use and hormone replacement therapy use. An update of the Husing et al. calculator, with the addition of significant biomarkers (IL1Ra and estrone), only marginally increased the prediction power of the calculator by 1.7% (Fortner *et al.*, 2017).

A review of the literature up to 2011 identified nine risk calculators of endometrial cancer in women with postmenopausal bleeding, but none of the tools has been validated (van Hanegem *et al.*, 2012). One risk calculator was identified since that review that is based on symptomatic individuals for assessment of endometrial cancer in postmenopausal patients with vaginal bleeding called the RHEA risk model, which

is internally validated (Giannella *et al.*, 2014). The variables included in the logistic regression model were recurrent vaginal bleeding, hypertension, an endometrial thickness of over 8mm and age over 65 years old. The AUC at internal validation was 87.8% at the best cut-off point of a score of 4 or more (Giannella *et al.*, 2014). External validation of two earlier developed calculators, the DEFAB (risk factors: diabetes age, BMI, frequency of bleeding endometrial thickness) and DFAB (risk factors included in the DEFAB bar the endometrial thickness was performed by Musonda et al., 2011 found both calculators to have 95% discriminatory power for endometrial cancer (Musonda *et al.*, 2011). These results were promising as they show that a risk calculator based on symptoms and demographics alone can be used for endometrial cancer identification without the need for specialist assessment of endometrial thickness information at the point of initial triaging. One risk calculator was also identified, focusing this time on pre-menopausal symptomatic patients with good AUC (84%), but it has not yet been validated (Bagepalli Srinivas *et al.*, 2020). More recently, another risk tool has been developed, triaging patients to low, medium and high risk based on a scoring system from -6 to 8 or more (Kitson, Evans and Crosbie, 2017). This was based on risk scoring from four components: insulin, reproduction, obesity, and genetics. A generated risk prediction model including all factors in each of the four components is still awaited from that research group. The lack of external validation in most endometrial cancer risk models has also been highlighted in a recent review of the literature, as well as issues related to missing information with regard to handling missing data and the clarity of methods used for identification of risk factors (Alblas *et al.*, 2018).

Two recent publications have used artificial intelligence to develop an endometrial cancer risk prediction model based on a neural network algorithm for population screening. In the study by Hutt et al., high BMI, diabetes, contraceptive pill, null parity, and noncontinuous hormone replacement therapy were factors associated with increased risk of malignancy, whereas other contraceptive methods and continuous hormone replacement therapy decreased cancer risk. The AUC was high at 98.6% but was validated in a small cohort of only 40 patients, so further studies are needed to assess its generalisability (Hutt *et al.*, 2021). Hart et al. performed an assessment of

seven machine learning algorithms to identify that random forest analysis achieved the highest AUC of 96%, followed by the neural network algorithm at 91% (Hart *et al.*, 2020).

2.3.3.6 Ovarian cancer

The first attempts to quantify the risk of malignancy in patients presenting with a suspicion of ovarian cancer dates back to the early 90s (Minaretzis *et al.*, 1994). Logistic modelling was used to establish the likelihood of ovarian cancer based on the patient's age and ultrasonographic tumour characteristics (tumour size, consistency, laterality). The collection of this information required the patient with suspected cancer to be seen by a specialist for imaging before all the necessary information was available to calculate the cancer probability (Minaretzis *et al.*, 1994).

Since then, many predictive models have been proposed. A systematic review of the literature included studies up to March 2008 that identified 109 studies with 83 predictive models for ovarian cancer (Geomini *et al.*, 2009). The model created by Sassone (Sassone *et al.*, 1991) has been more extensively externally validated. It is based on tumour ultrasound characteristics and had a pooled sensitivity and specificity of 84% and 80% based on 18 studies (Geomini *et al.*, 2009). Some studies incorporated the Sassone variables and added further ultrasonographic and blood test features (Leeners *et al.*, 1996;Alcázar and Jurado, 1998;Sengoku *et al.*, 1994). Further models based on ultrasonographic characteristics have been proposed and validated. The De Priest et al. model (DePriest *et al.*, 1993) has been validated in 10 studies with a pooled sensitivity of 91% and specificity of 69%.

The Risk of Malignancy Index (RMI) I (Jacobs *et al.*, 1990) and Risk of Malignancy Index II (Tingulstad *et al.*, 1996) were the best predictors of cancer following meta-analysis. The variables included in these models include serum Ca125, USS findings and menopause state. The first model has been validated in 16 studies and the latter in 7, both having high pooled sensitivity (ranging from 78 – 91% for different cut-offs) and specificity (74% - 87%). Many other prediction models exist but have not been validated in many studies to allow for meaningful comparison for the above-described externally validated models (Geomini *et al.*, 2009). The RMI was also, at that point,

endorsed for use by NICE and the Royal College of Obstetrics and Gynaecologists as a diagnostic tool in secondary care (NICE, 2011).

Many studies published since this review led to a further systematic review and meta-analysis in 2014, identifying 19 models and 96 studies for evaluation (Kaijser *et al.*, 2014). New models that emerged from this review had higher predictive power compared to the RMI, that was the IOTA LR2 model (sensitivity 0.92: 95% CI 0.88 – 0.95; specificity 0.83: 95% CI 0.77 – 0.88) and the SR model (sensitivity 0.93: 95% CI 0.89 – 0.95; specificity 0.81: 95% CI 0.76 – 0.85) that achieved the highest diagnostic accuracy. The LR2 model includes sonographic and blood test variables, incorporating neural network analysis, having similar prediction power to a logistic regression methodology (Timmerman *et al.*, 1999). In contrast, the Simple Rules (SR) score is based on sonographic characteristics alone (Timmerman *et al.*, 2008), requiring a radiologist with expertise in ovarian USS. A later meta-analysis by Meys et al. (2016) confirmed the above findings (Meys *et al.*, 2016). The SR had a pooled sensitivity of 93% and specificity of 80%, with similar values found for the LR2 (sensitivity: 93%; specificity: 84%), outperforming the RMI (sensitivity 75%; specificity: 92%). Attempts to simplify the complexity of the ultrasonographic characteristics have been made by newer models. Stukan et al., 2019 developed a model based on two ultrasonographic characteristics only: vascularisation and solid areas, in addition to plasma d-dimer levels. The model reported a similar AUC to the more complex models described above (0.977, 95% CI 0.955 – 0.999) (Stukan, Badocha and Ratajczak, 2019).

Aside from the above-mentioned risk calculators for ovarian cancer that require the input of a radiologist, a recent systematic review (Funston et al., 2020) identified 14 models that included only symptoms and seven models that included a combination of symptoms, risk factors and blood tests. Only four of these are externally validated, having moderate accuracy at validation compared to the above-described tools based on USS characteristics. These models are the Goff-Symptom Index (sensitivity 56.9%, specificity 83.3%), the modified Goff symptoms index (71.6%, specificity 88.5%), the Society of Gynaecologic Oncologists (SGO) Consensus criteria (sensitivity: 65.3-

71.5%, specificity: 82.9 - 93.9%) and the Qcancer ovarian model (sensitivity: 64.1%, specificity: 90.1%, at the 10% threshold) (Funston et al., 2020). The symptoms included in the models are abdominal pain, pelvic pain, increase in abdominal size/distention, bloating, feeling full, and difficulty eating for the Goff et al. (2007) model. The modified Goff model added urinary frequency and urinary urgency symptoms and achieved a 15% increase in sensitivity (Kim *et al.*, 2009); abdominal pain, pelvic pain, bloating, feeling full, urinary frequency and urgency were included in the SGO model (Rossing *et al.*, 2010) whereas the Qcancer model aside of the pain in the abdomen, and abdomen size increase has also included appetite loss, postmenopausal bleeding, rectal bleeding and weight loss, with no inclusion of urological symptoms. Family history, age and haemoglobin count are also included in the Qcancer calculator (Hippisley-Cox, 2012). AUC values are only reported for the Qcancer model being 0.86 at external validation, but only sensitivity and specificity combinations are available for the other three externally validated tools, making difficult comparisons with the previously described models. The Goff et al. (2007) model is the most widely validated in 6 studies, with sensitivity ranging from 56.9 % to 83.3% and specificity from 48.3% to 94.9% (Goff *et al.*, 2007) (Funston et al., 2020). These symptom-based models can be incorporated into primary care assessment, but their prediction is not as accurate as the model that includes radiological findings from USS investigations. Hence it is likely that more specialised assessments will be needed to rule out ovarian cancer in the high-risk patients being identified via the symptoms risk tools, that is currently the Ca125 cancer marker blood test (Andersen *et al.*, 2008)

2.3.3.7 Oesophageal cancer

Only a few studies have been found exploring oesophageal cancer risk prediction models. The model by Xie and Lagergen (2016) includes symptoms and demographics as variables in the model, making it a good candidate for use in the non-hospital setting (Xie and Lagergren, 2016). The variables included in the model are reflux symptoms/use of anti-reflux medication, BMI, smoking, duration of living with a partner, previous diagnosis of oesophagitis, diaphragmatic hernia, previous surgery for oesophagitis/hernia/reflux/gastritis/ulcer with an AUC of 0.84. This model is yet to be externally validated. A specific prediction model for oesophageal adenocarcinoma

(AUC: 0.84) and squamous cell carcinoma (AUC: 0.681 – 0.795) have also been proposed based on symptoms and demographics but again lacks external validation (Xie *et al.*, 2018;Liu *et al.*, 2017a). A recent review identified only 2 models out of 13 that have been externally validated for oesophageal cancer, with an AUC >0.7 (Li *et al.*, 2021). The addition of genetic factors was found to increase the AUC by 7% in a study combining common genetic variants and lifestyle variables in the prediction of cancer in the Chinese population (Chang *et al.*, 2013).

2.3.3.8 Hepatocellular and Pancreatic cancer

Keane et al. developed a risk calculator for biliary tract cancers and one for pancreatic ductal adenocarcinoma based on the THIN database (Keane *et al.*, 2014). The variables were weight loss, abdominal pain, nausea and vomiting, bloating, dyspepsia, diabetes, change in bowel habits, pruritus, lethargy, back pain, jaundice, and shoulder pain that overlap in the two types of cancer. However, back pain, lethargy and new-onset diabetes were specific to pancreatic ductal cancer only (Keane *et al.*, 2014).

Six models were found assessing hepatocellular cancer for patients with a previous diagnosis of hepatitis B that are externally validated, showing poor discrimination power in patients with cirrhosis (AUC <0.7), being slightly better in the cohort subgroup that was on antiviral treatment (AUC max= 0.778) (Yang *et al.*, 2021).

2.3.3.9 Skin Cancer

Many risk prediction tools exist for aiding melanoma diagnosis, with a review of literature having identified 28 risk prediction tools with AUCs ranging from 0.62 to 0.86. The most common predictive factors included in the models were the number of naevi, type of skin, freckle density, hair colour and history of sunburn and age (Vuong *et al.*, 2014). Another review published the same year identified only two calculators being validated in different cohorts (Usher-Smith *et al.*, 2014). Some of these are designed as self-assessment tools by the patients prior to presentation in primary care. One of the first calculators was developed in Scotland in 1989, with the included risk factors being the total number of pigmented naevi above 2mm, the tendency to freckle, the number of atypical naevi over 5 mm and a history of sunburn at any time in life (MacKie, Freudenberger and Aitchison, 1989). It was validated in an English

population in 1998 and recommended for use as a self-assessment questionnaire (Jackson *et al.*, 1998). Williams et al. developed a self-assessment melanoma prediction tool achieving a 70% AUC. The final logistic regression model included gender, age, hair colour, freckles' density, number of raised moles on the arms, number of severe sunburns in childhood/adolescence and history of non-melanoma skin cancer as the predictive factors (Williams *et al.*, 2011), which has also been externally validated in a UK population (Usher-Smith *et al.*, 2017). Further studies have validated previously reported calculators with variations seen in accuracy between males and females but also failed to attain both high sensitivity and specificity (Olsen *et al.*, 2015a). Wide variation has been noticed in the available calculators that lack consistency in terms of methodology and validation strategies that make direct comparison difficult (Kaiser *et al.*, 2020). More recently published melanoma calculators have achieved slightly lower AUC values (0.69-0.72) with similar risk factors to the previously mentioned models, such as age, gender, tanning ability, number of moles at the age of 21, number of skin lesions previously requiring removal (Olsen *et al.*, 2018). Others have also included genetic factors in addition to demographics and examination findings. The addition of polygenic risk scores has shown to increase the AUC by 2.3 up to 7 % (Gu *et al.*, 2018;Cust *et al.*, 2018).

Whiteman et al. (2016) developed and validated the first risk calculator based on logistic regression analysis for the identification of basal cell and squamous cell carcinomas in the general population (Whiteman *et al.*, 2016). The included risk factors in the final selected model being age, gender, smoking status, race, skin colour, tanning ability, freckling tendency, times of a sunburn prior to 10 years of age, the number of previously excised skin cancers and the number of any previous skin lesions removed. The AUC was high at 0.80, with subgroup analysis also performed for separate calculators for those with or without previous history of excised cancerous skin lesions (Whiteman *et al.*, 2016), being subsequently externally validated (Shetty *et al.*, 2021). Several calculators have been published since then, also having high prediction power of over 80% (Wang *et al.*, 2018), with a Ukraine study reporting an AUC of 97% in their model (Oshyvalova, Ziukov and Gurianov, 2019). The latter predictive power is impressively high, but the model requires external validation. The

risk factors included in the model are carefully tailored to account for previous medical history related to skin cancer and include sunburns, use of sunscreens, family history of skin cancer, recent sun exposure, exposure to radiological material, drug consumption including cardiac, antihypertensive, contraceptives and antibiotics (Oshyvalova, Ziukov and Gurianov, 2019). More specialised non-melanoma cancer calculators are also available such as for the prediction of the risk of a second basal cell carcinoma (Verkouteren *et al.*, 2015) or specifically for individuals with previous actinic keratosis (Tokez *et al.*, 2020) with moderate observed discrimination (0.6 – 0.65). The calculator developed by Tokez et al. included coffee consumption as a significant risk factor, a variable that has not been previously included in other cancer calculators (Tokez *et al.*, 2020), with polygenic risk factors also available in other tools (Stapleton *et al.*, 2019;Fontanillas *et al.*, 2021). The majority of skin calculators have been developed using logistic regression analysis, a recent study used an artificial intelligence deep learning approach for the development of a non-melanoma skin cancer prediction tool, also achieving a high AUC of 89%, but a complex dataset was used with numerous factors included in the model, including extensive drug and past medical history that could perhaps be difficult to use in a non-research setting (Wang *et al.*, 2019).

2.3.3.10 Kidney and Bladder cancer

Many risk prediction models are available for kidney cancer; nevertheless, very few have information available to draw conclusions on model performance in the development cohort or after external validation (Harrison *et al.*, 2021). The model by Frantzi et al. is based on urinary results screening, looking at the presence of renal cancer-specific peptides with a sensitivity of 80% and a specificity of 87%. The results of this model have been externally validated using a history cohort (Frantzi *et al.*, 2014). Urine biomarkers, this time urine AQP1 and PL1N2 were assessed by Morrissey et al. in the development of a model with an AUC of 99%. These biomarkers are suggested to be used for population screening in asymptomatic patients (Morrissey *et al.*, 2015). Another model also based on biomarkers analysis for early kidney cancer detection has reported a high AUC of 0.932 using a combination of 3 plasma biomarkers: NNMT, LCP1 and NM23A. The prediction power of the NNMT marker alone was also very high at 91.3% (Su Kim *et al.*, 2013). The addition of KIM-1 plasma

biomarkers had been found to increase by 10% the prediction power of a model that was previously based only on demographic risk factors. The base model included the following risk factors: smoking, gender, BMI, diabetes, hyperlipidaemia and hypertension (Scelo *et al.*, 2018). Other biomarkers, such as the serum-circulating long noncoding RNA signature, has shown promising results in renal cancer identification with AUCs of 0.82 – 0.9 (Wu *et al.*, 2016a). Genetic scores have more recently been investigated for population screening, with their discrimination power currently being moderate with AUC 0.62 – 0.65 (Wu *et al.*, 2016b). Out of a total of eleven models assessed in a review of the literature, six models were validated, with only 2 in external populations. The AUC was >0.7 for most of them, but the validation was performed in small case-control study settings, and the sensitivity was low in the majority of the models despite an acceptable discrimination power (Harrison *et al.*, 2021).

Many risk calculators are also available for bladder cancer, with a recent review of the literature having identified 28 models, including those generating the RAT and Qcancer algorithms (Shephard *et al.*, 2013;Harrison *et al.*, 2022;Price *et al.*, 2014). The majority of models were developed using logistic regression analysis, but only 8 models have been externally validated (Harrison *et al.*, 2022). The model by Loo et al. identified age over 50 years, gross haematuria, and male sex to be significant indicators of bladder or renal cancer. These factors comprise the Haematuria Risk Index, with an AUC of 0.809, with preserved high prediction in the validation cohort (Loo *et al.*, 2013). A similar model developed by Matulewicz et al. identified older age, haematuria, as well as smoking status or current or previous smoking history as significant factors of bladder malignancy. The 1% was used for best discrimination with an AUC of 0.79 (Matulewicz, Rademaker and Meeks, 2020). The haematuria cancer risk score was developed based on the DETECT trial in the UK, based on a large prospective cohort with external validation in a Swiss population (Tan *et al.*, 2019). Age, gender, haematuria type and smoking history were included in the model with a high discriminatory power of 83.5% AUC, which had much-improved cancer detection compared with the UK and American urological societies guidelines (Tan *et al.*, 2019). Male gender, smoking history and gross haematuria have also been included in older nomograms with good AUC of over 80% but developed from smaller cohorts

without external validation(Hee *et al.*, 2013), whereas other models also included voided cytology results (Cha *et al.*, 2012b) or the addition of immunocytology results achieving even higher AUC of 88% to over 90% (Cha *et al.*, 2012a;Beukers *et al.*, 2013).

Further studies attempted to add biomarkers to models with demographics and medical history factors. The NMP22 protein assay results in combination with age, race, gender, smoking status, haematuria presence and its extent resulted in a model with an AUC of 82.4% at external validation with no further improvement in bladder cancer detection when the urinary cytology result factor was added to the model(Lotan *et al.*, 2009). The addition of NMP22 protein results in the common risk factors model has been confirmed by later studies, with a net improvement of 8.2%, showing this factor to be a strong independent indicator of bladder cancer(Barbieri *et al.*, 2012).

2.3.3.11 Paediatric cancers

Only one model was identified looking at the prediction of paediatric malignancy (Dommett *et al.*, 2013). This was developed in a case-control study designed from UK primary care health records. They have identified twelve alarm symptoms being pallor, HaN lumps, masses elsewhere, bruising, abnormal movements, lymphadenopathy, fatigue, pain, bleeding, visual symptoms, and musculoskeletal symptoms. No AUCs were reported, but only PPVs per symptom were identified as significant following logistic regression analysis (Dommett *et al.*, 2013).

2.3.3.12 Head and neck Cancer

The first head and neck cancer risk calculator (HNC-RC) based on symptoms and demographics was developed in 2016 in England using a large cohort of over 5,000 patients and is available online (Sensitivity:74.8%; specificity:65.9%; overall predictive power (AUC):0.77) (Tikka, Pracy and Paleri, 2016). It has been externally validated in a Scottish cohort (n=2,000), maintaining its discriminatory ability with even higher AUC in this cohort (sensitivity: 79.3%; specificity: 68.6%; AUC:0.81) (Tikka, Paleri and MacKenzie, 2018). The variables currently included in this predictive model are age, gender, and symptoms of dysphagia, odynophagia, neck lump, hoarseness (persistent or intermittent), feeling of something in the throat,

otalgia, oral swelling, oral ulcer, and haemoptysis. The principle of a symptom and demographics-based prediction model is suitable for use in the primary care setting as it does not require any specialised input, as discussed earlier for the epidemiologically based lung cancer predictive models and symptom-based colorectal cancer risk calculators.

Using the same cohort of 5,000 patients used to generate the HNC-RC, Moor et al. (2019) attempted to employ artificial intelligence methods to generate a HNC risk calculator. Eleven machine learning options were used, with the variational inference logistic regression having the highest discriminatory ability. Nevertheless, despite the very low false-negative proportions (0.1% vs 7% for conventional logistic regression), the proportion of false positives was very high at 66.4% compared to 0.3% for logistic regression. This questions the use of this model for triaging referrals. It would have resulted in a large number of patients needing to be seen urgently, compared to the more balanced conventional logistic regression output. Moreover, the study did not present the combination of variables used in the resulting model, which may imply the difficulty in interpreting the results of artificial intelligence algorithms (Moor, Paleri and Edwards, 2019).

Another research group (Lau, Wilkinson and Moorthy, 2018) published their proposed HNC prediction score two years following the Tikka et al. (2016) publication based on a cohort of 1,075 retrospectively collected patients. They have accepted a sensitivity of 31% as this yielded the best discriminatory combination using the F-statistics (92% specificity, AUC: 0.79). Clinically, this translates into 2 out of 3 patients with cancer being misdiagnosed. Their false-negative figures were very high in the external validation cohort (Lau, Wilkinson and Moorthy, 2018). A systematic review of the efficacy of the 2ww HNC clinics in the UK showed a pooled sensitivity of 40.8% (Langton, Siau and Bankhead, 2016). Hence, proposing a scoring tool that gives a lower sensitivity than the current standard is controversial. Looking at the significant factors included in the Lau et al. model, well-known red flags such as dysphagia, odynophagia and oral swellings were not included. Their model included as significant variables: smoking pack-years, alcohol units over the recommended weekly intake,

oral ulcer, neck lump, ear lesion, facial lesion, age, tongue ulcer, weight loss, unilateral hearing loss and thyroid swelling (Lau, Wilkinson and Moorthy, 2018). The ear and facial lesions could have potentially been grouped together, as well as the tongue and oral ulcer variables and the thyroid swelling included within the neck lump symptoms, making it easier to use by GPs and reducing the complexity of the model. Unilateral hearing loss was one of the statistically significant risk factors in their model, but its positive predictive value (PPV) was not stated (Lau, Wilkinson and Moorthy, 2018). The PPV for HNC of unilateral hearing loss has been found to be very low in previous studies (Tikka, Paleri and MacKenzie, 2018;Tikka, Pracy and Paleri, 2016).

Attempts have also been made to predict thyroid cancer based on ultrasonographic characteristics of thyroid nodules (Choi *et al.*, 2015). A web-based risk estimation has been created based on thyroid nodule characteristics: solid content, taller-than-wider shape, spiculated margin, ill-defined margin, hypoechogenicity, marked hypoechogenicity and rim calcifications. The model performed very well on internal and external validation with AUC of 0.903 and 0.897, respectively (Choi *et al.*, 2015). Similar ultrasonographic characteristics were used by a model developed by the Korean Society of Thyroid Radiology (Kwak-TIRADS), having an AUC of 0.872, performing similarly to the prediction model devised by the American College of Radiology (ACR-TIRADS) (AUC 0.867) (Kwak *et al.*, 2013). In further validation studies, their AUCs were directly comparable (0.884 – Choi et al. model; 0.891 – Korean model; 0.875 - American model) with the Choi et al. model having the best agreement in calibration analysis (Ha *et al.*, 2017). European guidelines also exist (EU-TIRADS), as well as a classification by the American thyroid association (ATA) and the British Thyroid Association. On a head-to-head comparison of the first four models, the Kwak-TIRADS has performed marginally better with an AUC of 0.896 compared to a range of 0.869 – 0.879 for the rest of the models (Shen *et al.*, 2019). More recently, demographic characteristics (age) have been incorporated into the predictive modelling alongside ultrasonographic characteristics – irregular shape, microcalcification, absent halo, homogeneous echotexture, and solid content (Girardi, Silva and Flores, 2019). This model still awaits external validation, but in the development cohort, the AUC values were not reported to allow comparison with other

predictive models. Despite the usefulness of the thyroid nodule prediction models for malignancy, referral to secondary care centres is required as a radiologist with sub-specialisation in thyroid ultrasonography is required to assess the thyroid nodules.

Similar efforts have been made in the development of prediction tools for the diagnosis of oral cancer. A model was identified that could predict the malignant potential of leukoplakia lesions but was designed from a small cohort of 22 cancer cases and 138 controls. This model was based on clinical biomarkers following excision of the lesion: p53, CA9 combined with age and degree of dysplasia, with an AUC of 0.88 (Zhang *et al.*, 2017), with further models available using cytology variables of dysplastic lesions achieving 100% discrimination (AUC:1) but was based on a small development cohort of 87 patients and a validation phase with 277 patients only (Liu *et al.*, 2017b) or salivary CD44 and protein measurements (AUC: 0.763) based on a case-control study of 300 patients (Pereira *et al.*, 2016). Sun et al. (2019) developed a more clinically orientated prediction model for the cancerous progression of a leukoplakic lesion, with the variables included being patients' age, gender, site of lesion, history of local stimulus (severe periodontitis, sharp/broken tooth, bad prosthesis) and alcohol consumption. The AUC was high at 0.83, and cut-off values for high-risk cases were identified based on the coefficient values of the model variables. The sensitivity was 67%, and the specificity was 81% (Sun *et al.*, 2019). The model has not yet been validated, and it was developed based on a small sample size of 77 cancer cases in a total of 269 patients. Sharma et al. (2015) also developed a model of prediction of malignancy in patients presenting with any oral lesion - not only confirmed leukoplakia – which is based on demographic factors and the site of the lesion rather than other cytological characteristics (Sharma and Om, 2015). It was based on a retrospective dataset of 1,025 patients. A Probabilistic neural network and general regression neural network was used for the analysis. Significant factors for progression to malignancy were stated to be: socioeconomic status, clinical symptoms, history of addiction, comorbid conditions, clinical examination findings, site of lesion, presence of neck nodes, and tumour size. Nevertheless, no further details are given in the paper regarding the specific data included to characterise each of the variables, which makes the interpretation and applicability of the model not practical. Moreover, there is no

mention of how many cancer cases their cohort had. The reported AUC of 0.9974 shows perfect discrimination, which would be useful to be related to the number of cancer cases they included in their dataset. If the numbers are too small, then generalisability will be difficult to be assumed (Sharma and Om, 2015). Recent studies have also been working on early HNC cancer detection based on non-invasive exhaled breath tests and circulating tumour DNA on blood sample tests. Dharmawardana et al. (2020) showed promising results in the identification of early and advanced HNC in patients presenting with symptoms suggestive of HNC, with a sensitivity of 80% and specificity of over 86% (AUC: 82.1%) based on breath analysis for volatile organic compounds using a selected ion flow-tube mass spectrometer (Dharmawardana *et al.*, 2020). Early identification of HNC based on circulating tumour DNA, circulating tumour cells and exosome miRNA from serum, plasma, and saliva, all of which are currently being tested in trials and results from large sample studies are awaited (Arantes *et al.*, 2018;Hudečková *et al.*, 2021). These biomarkers could be incorporated into future early diagnosis risk assessment tools when these tests are widely available and cost-effective.

Aside from the above models designed to be used to predict cancer probability when symptoms or signs are already developed, other calculators are available focusing on HNC prediction of asymptomatic population. A screening tool for identifying existing oral cancer in high-risk but asymptomatic individuals has been developed by researchers in Sri Lanka (Amarasinghe *et al.*, 2010). It was developed based on multivariate logistic regression modelling and included age, socioeconomic status, smoking and alcohol history, as well as a history of betel-quid chewing in the resulting calculator. A cut-off for identifying high-risk populations was calculated based on a scoring system generated by the sum of the odds ratios of the variables included in the model. The cut-off point that maximised the area under the ROC was 12, yielding a sensitivity of 93.7% and a specificity of 67.7% with an AUC of 84% (Amarasinghe *et al.*, 2010). It was developed based on a cohort of 1,029 patients in a prospective case-control set-up and was externally validated in a sample of 410 patients. Some of the limitations of this model are that it will be difficult to be used outside its development region as betel nut chewing is uncommon in other parts of the world. Additionally, no

sample size calculation was performed prior to the development of the calculator which can limit its generalisability. Moreover, the scoring system used to develop the cut-off points has not been used before in risk cancer calculators. Using the odds ratio calculation to assess a model performance has been criticised as a poor discriminator, that does not provide meaningful classification information unless employed in large epidemiological studies of large sample size magnitude (Pepe *et al.*, 2004). A very similar model has also been developed for screening Indian population for oral and oropharyngeal cancer, achieving an AUC of 0.866 (Krishna Rao *et al.*, 2016). It includes similar variables to the model mentioned above by Amarasinghe et al. (2010), being smoking, chewing tobacco, chewing quid with tobacco, alcohol, family history of HNC, diet (spiciness of the food, fruit consumption) and oral hygiene (rinsing mouth with water after eating/chewing). This model has not been validated yet. Aside from the fact that it will be difficult to be validated in populations other than Indian due to the addition of quid chewing, no sample size calculation was performed prior to the development of the model that is based on a relatively small number of 180 cases and 272 controls (Krishna Rao *et al.*, 2016). Another research group in India (Cheung *et al.*, 2021), using data from the Kerala oral cancer screening trial, developed an oral cancer calculator predicting the 7-year likelihood of oral cancer incidence. The calculator was based on prospective data collected from over 90,000 cases and controls that participated in the Kerala trial. COX proportional hazard modelling was used for the analysis taking into consideration a follow-up period of 7 years, with significant variables included in the model being sex, age, education level, BMI, tobacco chewing, tobacco smoking, chewing-smoking tobacco interaction and alcohol use. Internal validation was performed, achieving an AUC of 0.75. Looking at the variables included in the model, it appears that a backward elimination process was not performed, as BMI was included in the model despite having a non-statistically significant p-value of 0.44 (Cheung *et al.*, 2021). This may have affected the odds ratios and p-values of the other variables included in the model, making generalisability of the results difficult to appreciate. It remains to be assessed in this calculator can be validated in population outside the development cohort. Its applicability in other than Indian cohorts will be difficult given it included tobacco chewing which is not common in other regions.

Koyanagi et al. (2017) looked at the estimation of future oral cancer incidence incorporating genetic results for the aldehyde dehydrogenase 2 (ALDH2) gene polymorphism, which has been found to have a risk of HNC development when associated with alcohol consumption in the Japanese population. The model has been externally validated in a different Japanese cohort achieving an AUC of 0.82. Predictors included in this model were: age, sex, smoking, drinking and the ALDH2 genotype (Koyanagi *et al.*, 2017). Another model identifying high-risk Taiwanese population was recently developed by Yu et al. (2022). Patient information was prospectively collected from the Taiwan biobank, including 11,462 controls and 3,313 patients with HNC. The calculator included age, sex, education level, marital status, mother ethnicity and father ethnicity (Taiwanese, Hakka, other), occupation, alcohol and smoking status, betel nut chewing, coffee consumption, BMI, and family history of oral cancer. Separate models were run for males, females, and cancer subsites with reported AUCs ranging from 0.93 to 0.98 at internal validation. No external validation was performed. Moreover, there was no mention of the sensitivity and specificity combinations that achieved the AUCs (Yu *et al.*, 2022). It could be that despite a high AUC, the sensitivity is low to achieve high specificity and vice versa, making it an inappropriate tool for population screening. Finally, due to the nature of the data included in the study, this screening tool is only applicable to the Taiwanese population, making difficult a more widespread adaptation. A normogram predicting oral cancer occurrence in asymptomatic Chinese population has been developed by Chen et al. (2018), employing a similar statistical analysis to the model by Yu et al. (2022). Separate models were created for males and females, based on a prospective cohort of 978 oral cancer patients and 2,646 controls achieving AUCs of 0.768 and 0.7, respectively. The risk factors included in the model for males were: smoking status (packyears), alcohol consumption (g/day), repetitive dental ulcer and teeth loss (more than five). Moreover, factors negatively correlated with cancer were tea consumption), fish intake (more than once/week), seafood intake (more than once/week) and regular dental visits (more than once/year). In females, the risk factors were: passive smoking, cooking oil fume exposure, teeth loss (more than five), repetitive dental ulcer and day of first intercourse (before 22 years of age). Negatively correlated factors for cancers

were: tea drinking, vegetable consumption (more than once/week), beans consumption (more than once per week), and fruit intake (more than three/week). Other demographic characteristics relating to marital status, residence type, family history of cancer, educational level and age were assessed in univariate analysis but not included in the multivariate regression model despite the first two being significantly associated with an oral cancer diagnosis (Chen *et al.*, 2018). These risk factors were also assessed for inclusion in a model built using UK population data from the UK biobank study (McCarthy *et al.*, 2020). Laryngeal cancer was excluded from this screening-focused model as it was felt that the identification of laryngeal cancer is not possible in primary care screening due to the location of the disease that requires specialised equipment for examination. The resulting risk prediction model, again based on asymptomatic population similar to the Chen et al. (2018) model, predicted future development of HNC with an AUC of 0.69 at internal validation (of 232 HNC cases and 396.947 controls) and 0.64 during external validation (of 157 cases and 78,895 controls), with the model slightly underpredicting HNC compared to the development phase. The factors included in the McCarthy model that were positively associated with a cancer diagnosis were age (increasing age), gender (male), smoking and alcohol consumption, and level of deprivation (higher level of material deprivation). Significantly but negatively associated with HNC was found to be the consumption of at least five portions of fruits and vegetables per day, exercise at least once per week and higher BMI (McCarthy *et al.*, 2020). The AUC of this model was lower compared to the previously described calculators, also designed to predict the absolute risk of developing HNC, which achieved AUCs of 0.98 - 0.768 (Chen *et al.*, 2018;Yu *et al.*, 2022;Amarasinghe *et al.*, 2010;Krishna Rao *et al.*, 2016). The better performance may be attributed to the larger HNC sample size in these models as well as the included information on family history of cancer, which was not available in the UK Biobank dataset. Moreover, the external validation was performed using a sub-cohort of the UK biobank database. Ideally, the model should also be validated using a different dataset to allow an assessment of its transportability (McCarthy *et al.*, 2020).

Apart from the screening-based risk calculators mentioned above-covering populations in India, Japan, China, Taiwan and Sri-Lanka and UK, Recently, a US-

based HNC risk calculator has been developed for estimating the hazard rate for the future development of HNC in an attempt to identify high-risk populations for focused cancer prevention strategies (Lee *et al.*, 2020). It was developed using a large US database of The International Head and Neck Cancer Epidemiology (INHANCE) Consortium, which included 7.299 HNC cases and over 10.000 controls. The developed model can predict the future risk for HNC development for the US population, and it has been internally validated as part of the same study. Separate models were also developed for each HNC subsite. The risk factors included were age, sex, race, education level, smoking status and intensity, alcohol drinking intensity, as well as family history of HNC, with the latter not being a significant predictor for oropharyngeal cancer or for laryngeal cancer in males. A model without the cancer family history was also designed as it was noted that not all patients could know if such family history existed. Tobacco and alcohol interaction was only significant for the oral cavity subsite. The AUC was over 70% for most of the cancer subsite models (Lee *et al.*, 2020). Tota et al. (2019) also developed a similarly designed calculator based on a US population, with prospectively collected data including 241 cases and 9,327 controls. The focus of this calculator was future oropharynx cancer absolute risk. The variables included in the final model were: age, sex, race, smoking and alcohol history, lifetime sexual partners and oral HPV status. Internal validation was performed achieving an excellent discrimination with an AUC of 0.94. External validation was also performed with also a very good model performance of 0.87, nevertheless a drop of 0.07 units is noted (Tota *et al.*, 2019). The very high AUC could be attributed to focus of this calculator to capture only oropharyngeal cancer cases that makes the cohort more homogeneous.

To conclude, the review of the literature has identified eight screening-based calculators for the estimation of future HNC development and further thirteen calculators assessing HNC prediction using specialised investigations in patients presenting with HNC symptoms. Only two models based on currently symptomatic population are available that include a list of demographics and symptoms without the need for investigations that could be potentially used for triaging suspected HNC patients in primary or secondary care (Lau, Wilkinson and Moorthy, 2018;Tikka,

Pracy and Paleri, 2016). Of these, only one achieved high discrimination at external validation, but its development was based on a retrospective cohort of patients and achieved moderate performance at its development phase (AUC: 0.77) and validation phase (AUC: 0.8) (Tikka, Paleri and MacKenzie, 2018;Tikka, Pracy and Paleri, 2016). In view of this, there is the potential for the development of better-performing models to approach the high AUCs of symptom-based cancer calculators being reported for other common cancers (AUC>0.8), as mentioned in the previous sections. The fact that only one externally validated symptom-based calculator exists for HNC is a unique opportunity to avoid developing multiple new HNC calculators, introducing confusion as to the most appropriate tool selection, but instead focusing on the further development of the existing tool (Steyerberg *et al.*, 2004).

2.4 Literature Review of Head and neck cancer symptoms

In addition to the common themes highlighted in the first part of the literature review chapter, each cancer has specific patterns of symptom presentation, cancer stage at the time of diagnosis and barriers to early detection. A review of the literature on HNC cancer presenting symptoms will be presented here as this will enable assessing the efficacy and the degree of evidence-based nature of the current HNC referral guidelines presented in section 2.2.2 of this chapter. This review will also inform the data collection part of the methodology of this thesis with the aim of developing a symptom-based triage tool for HNC.

2.4.1 Head and neck cancer red flag symptoms and other risk factors

One of the first studies that quantified the nature of symptoms related to HNC was published in 1980 (Kaufman, Grabau and Loré, 1980). A total of 1026 symptoms were reported by 385 patients seen in an ENT department in New York. Sixteen symptoms accounted for over 70% of all reported symptoms (Kaufman, Grabau and Loré, 1980). The most common symptom was hoarseness (30%) for all cancers, being 95% for glottic, 71% for other laryngeal subsites and 30% for hypopharynx cancers. Dysphagia was the next most common symptom seen in 23% of patients, followed closely by neck lump (22%) and intra-oral lesion (22%) and odynophagia/persistent sore throat (21%). Looking at the HaN subsites, dysphagia was more common in hypopharynx

cancers (51%) and laryngeal cancer (except glottic) with 45%. Neck mass was most commonly seen in hypopharyngeal (38%) and oropharyngeal cancer (35%). Sudden weight loss was found in 14%, being most common in hypopharynx cancers (36%). Otalgia, dyspnoea, haemoptysis, cough, and stridor were seen in 4-9% of patients. Weakness/Fatigue was a reported symptom in 4% of patients, being most common in hypopharynx cancer with a 10% frequency. Patients with neck lumps, pharyngeal and intra-oral symptoms and signs presented earlier compared to those with voice symptoms, more general/systemic symptoms and face/jaw complaints (Kaufman, Grabau and Loré, 1980).

Since the publication of that study, many other studies have included information on symptom presentation associated with HNC. The first part of this section will cover a review of the literature on presenting symptoms relating to HNC diagnosis. As HNCs arise from several anatomical sites, they present with a variety of symptoms and signs (Mehanna *et al.*, 2010). Despite the fact that several consensus-based guidelines exist with lists of common symptoms relating to a HNC diagnosis, as will be discussed in section 2.4.3, there is no available review of the primary and secondary care literature on the HNC presenting symptoms. The objective was to perform a comprehensive collection of the terminology related to the signs and symptoms associated with HNC patient presentation to inform the design of the data collection proforma of this thesis research.

The review included observational (case-control, case-series, cohort) and interventional (statistical modelling) studies in which the focus was the presenting symptoms in patients diagnosed with HNC. The participants in the studies had to be patients with a new diagnosis of HNC or being referred to with suspected HNC. No limit was set in the time frame or study setting, hence including studies from both primary and secondary care. The studies included were published up until 01/10/2021(the date of the last update of the papers included in this chapter of the thesis). There was no language limit. An electronic database search of papers' titles and abstracts was performed on PubMed using the PubMed Advance Search Builder. Mesh terms and free text terms combinations were used associated with "head and

neck cancer", " symptoms or signs", "diagnosis".The search syntax was: (((symptoms or signs or diagnosis or referral or urgent suspected cancer or 2 weeks wait or two week wait or 2ww or USOC) and (cancer or tumour or tumor or neoplasm or malignancy 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 hypopharynx cancer or nasopharynx cancer))). The initial automated search of titles and abstracts identified 14,586 papers. After manually reading the titles and abstracts of the papers and following full-text screening for eligibility, 37 papers were identified mentioning presenting symptoms in newly diagnosed patients with HNC, and 15 studies mentioned symptoms in patients referred with suspected HNC. There were a total of 44 unique studies, as 8 studies included symptoms for both HNC patients and suspected HNC referrals (Rogers *et al.*, 2019;Talwar *et al.*, 2020;Tikka, Pracy and Paleri, 2016;Rosell Ferrer *et al.*, 2021;Allam and Nijim, 2019;Shephard, Parkinson and Hamilton, 2019;Alho, 2006;Mettias, Charlton and Ashokkumar, 2021). From each article, the following information was extracted: author, publication year, journal title, study design, study setting, type of participants, sample size, country of origin, HNC subsites, and list of symptoms and signs. The studies reported symptoms in HNC patients are listed in Table 2-3 alongside the number of HNC patients in each study and the main presenting symptoms. Less frequently reported symptoms are omitted from the table. Table 2-4 summarises the symptoms in patients presenting with suspected HNC. The studies are presented in the tables in alphabetical order.

Of the 44 included papers, 25 (56.8%) were authored by UK institutions, 4 studies were from Finland, 2 from the Netherlands, and a further 4 were from other European countries (Italy, Spain, Poland). The remainder were from institutions in the USA (n=4), Canada (n=2), Australia (n=1), Brazil (n=1) and Thailand (n=1). The majority of papers were authored in secondary care institutions (n=38, 86.4%), the rest being conducted in primary care. Only 6 (13.6%) studies had a prospective data collection methodology (Douglas *et al.*, 2018;Kassirian *et al.*, 2020;Queenan *et al.*, 2018;Amir *et al.*, 1999;Brouha *et al.*, 2005b;Rosell Ferrer *et al.*, 2021;Haikel *et al.*, 2011). All papers were written in English. Of the 37 papers focused on symptoms of patients diagnosed with HNC, 22 (59.4%) included all HNC subsites, 3 studies were focused

on laryngeal cancer, 6 studies covered oral cancer symptoms, 2 oropharyngeal cancers, with the remainder covering a combination of HNC subsites (oral cavity and oropharynx, pharynx, larynx and oropharynx, larynx, and hypopharynx).

Looking at the studies reporting symptoms in patients with HNC, one of the most commonly reported symptoms was a neck lump ranging from 1.3% - 63% (unweight mean: 29.3%). This was reported in 34 studies, including all types of HNC. It was followed in temporal order by the hoarseness symptoms, with a mean of 28.6%, reaching a high proportion above 70% in the laryngeal cancer cohorts (Brouha *et al.*, 2005b;Teppo *et al.*, 2003;Merletti *et al.*, 1990). The sore throat symptom was also reported by most studies, with an average of 24.6% of HNC patients reporting it. Dysphagia was mentioned as part of presenting symptoms in 25 studies (mean: 10.9%, range: 2% – 38% of HNC). Other common symptoms were otalgia, oral ulcer, and oral swelling, with each of these symptoms reported in 18 studies with an unweighted mean of 11%, 25% and 16.8%, respectively. A red and white patch in the mouth was mentioned in 10 papers as one of the main presenting symptoms, with a mean of 13.8%. Haemoptysis was commonly reported as a symptom in over 10 studies, but the mean percentage remained low, with 3.5% of HNC patients reporting it as one of their main presenting symptoms. Weight loss was only reported in 6 studies, but the mean percentage was high at 12.6%. Odynophagia was mentioned as presenting problem in only 6 studies (mean: 7.9%), which perhaps shows that it is usually a late presentation symptom that is not often seen at the time of initial patients' presentation (Carvalho *et al.*, 2002). Dyspnoea is also likely to be a late symptom, but despite that, it was recorded in 9 papers, with a mean proportion of 5.9%. Intermittent dysphagia was rarely mentioned as a symptom with, only present in 0.8% - 4.2% of HNC manifestation symptoms. Similarly, intermittent hoarseness was seldom present in the HNC cohort (0 -1.8%) (Singh and Warnakulasuriya, 2006;Zeitler *et al.*, 2018;Tikka, Pracy and Paleri, 2016). Other less commonly reported symptoms were HaN or intraoral pain, stridor, cough, feeling of something in the throat, unilateral nose, eye and ear symptoms, dental mobility, trismus, dyspnoea, HaN skin growths, drooling, cranial neuropathies, insomnia, high CRP, and recurrent lower respiratory tract

infections. Six studies reported symptoms based on the primary care referral, as noted in the relevant column in Table 2-3.

Fifteen studies were identified that report on percentages of symptoms in all patients referred with HaN symptoms with HNC suspicion rather than a confirmed diagnosis. Consideration and analysis of symptoms in this group of patients are as important as the symptoms profile of HNC patients. This is because any designed intervention looking to triage patients to high and low-risk categories for cancer, like the aim of this thesis, requires cancer cases as well as control patients, that is, a healthy population presenting with relevant symptoms to run probability modelling (Steyerberg *et al.*, 2004). The list of papers looking at symptom frequency in patients referred with possible HNC is seen in Table 2-4. The table includes the total number of patients in each study and the most common presenting symptoms, with the average proportion of each symptom taking into consideration the reports from all available studies. The majority of papers are again from secondary care ENT and maxillofacial units, with 2 papers reporting referral symptoms from primary care (Alho, 2006;Shephard, Parkinson and Hamilton, 2019). These two papers had the largest number of included patients, 5,867 and 4,365, respectively. The largest reported cohort from a secondary care unit was published in 2016, reporting on patients seen in 2 tertiary care UK centres with complete data for 4,715 patients (Tikka, Pracy and Paleri, 2016). The neck lump symptom was the most commonly reported, with an unweighted mean of 27%, followed by persistent hoarseness (mean: 21.8%). Oral ulcer, sore throat and weight loss were also common symptoms prompting an urgent cancer referral with mean proportions of 11.6%, 13.2% and 12.4%, respectively. Odynophagia, dysphagia and oral swelling, ulcer and red and white mucosa patches were less common presenting symptoms, with a mean proportion of less than 10%. The feeling of something in the throat (globus) symptoms was a presenting complaint in 9.8% of the total referrals from the combination of all studies. Despite that, in the HNC pooled cohort, this was the presenting complaint in only 1.4% of cases (Tikka, Pracy and Paleri, 2016;Allam and Nijim, 2019;Pitchers and Martin, 2006;Rutkowska *et al.*, 2020;Zeitler *et al.*, 2018).

Patients with more than one symptom, especially with a combination of the common red flags, had a higher probability of being diagnosed with malignancy compared to patients presenting with a single symptom (16% vs 3%), as was reported in the study by Talwar et al. (2020). But overall, the majority of HNC patients report one cardinal presenting symptom at presentation, with the percentage varying from 82% to 76% in the reporting studies (Rogers *et al.*, 2019;Queenan *et al.*, 2018;Talwar *et al.*, 2020). Common symptom combinations have been found to be local pain and neck mass for anterior oral cavity tumours, local pain and dysphagia for posterior oral cavity tumours, local pain and weight loss for retromolar trigone tumours, as well as local pain and referred otalgia. For oropharyngeal cancer, common symptom combinations are local pain and dysphagia, whereas for hypopharyngeal cancer, dysphagia and neck mass. In the supraglottic regions, hoarseness and dysphagia are seen (Dolan, Vaughan and Fuleihan, 1998).

Apart from symptoms, demographics and social history factors are also associated with HNC. The evidence available for these factors will be covered in the following paragraphs. Current or previous smoking and current or previous excess alcohol intake have been strongly linked with HNC as well as male gender (Dolan, Vaughan and Fuleihan, 1998;Rogers *et al.*, 2019). Other risk factors include a family history of HNC (Negri *et al.*, 2009), exposure to asbestos and occupational exposure to inorganic acid mists for cancer of the larynx (Straif *et al.*, 2009;Baan *et al.*, 2009)and genetic variation in the alcohol metabolism genes ADH1 and ADH7(Hashibe *et al.*, 2008). HPV infection has been linked to oropharyngeal cancer, as has passive tobacco exposure(Lee *et al.*, 2008), marijuana use(Berthiller *et al.*, 2009), a BMI less than 18.5(Gaudet *et al.*, 2010), caffeine-free diet(Galeone *et al.*, 2010), poor oral hygiene for oral cavity cancer(Guha *et al.*, 2007), less than 2 hours of physical exercise per week(Guha *et al.*, 2007) and having more than 4-lifetime oral sex partners(Guha *et al.*, 2007). All the above-mentioned risk factors have been confirmed in large-scale studies based on over 7000 HNC cases and 10.000 controls from pooled analysis within the international HNC epidemiology (INHANCE) consortium(Lee *et al.*, 2020).

A systematic review of the literature based on 41 studies found a strong association between deprivation and oral cancer with more in-depth analysis based on meta-analysis, showing higher cancer rates for low occupational social class, low education attainment and within-population of low income(Conway *et al.*, 2008). This difference remains evident in low- and high-income countries. This was a robust meta-analysis of case-control studies with a total of 15,344 cases and 33,852 controls having employed robust risk of bias analysis in the selection of included studies (Conway *et al.*, 2008).

Table 2-3. Studies reporting presenting symptoms in patients with head and neck cancer

Study	HNC	Care	Neck lump	Dysphagia	Sore throat	Hoarseness	Otalgia	Oral ulcer	Oral swelling	Patches
Alho (2006)	221	GP	6.30%	2.70%	45.10%	27.60%				
Allam and Nijim (2019)	57	2ndary	56.10%	3.50%	15.80%	17.50%	7.00%			
Amir *et al.* (1999)	186	2ndary	11.80%	8.10%	26.90%			5.40%	3.20%	
Brouha *et al.* (2005a)	189	2ndary	5.80%	4.20%	11.10%				25.90%	
Brouha *et al.* (2005b)	117	2ndary	1.70%	5.90%	10.30%	83.80%	3.40%			
Dolan, Vaughan and Fuleihan (1998)	492	2ndary	46.00%	38.00%	53.00%	44.00%	26.00%			
Douglas *et al.* (2018)	1584	2ndary	11.60%	17.90%	33.80%	33.90%		20.80%		
Flukes *et al.* (2019)	294	2ndary	36.70%	9.50%	9.90%	12.20%				0.00%
Gao *et al.* (2019)	37	2ndary	45.90%	5.00%	14.00%	16.00%	3.00%	11.00%	21.60%	11.00%
Haikel *et al.* (2011)	53	2ndary	61.00%	2.00%	15.00%	15.00%		15.00%	9.00%	4.00%
Ho, Zahurak and Koch (2004)	87	2ndary	57.00%		38.00%		6.00%			
Kassirian *et al.* (2020)	102	2ndary	38.20%		24.50%	16.70%		20.60%		
Kaufman, Grabau and Loré (1980)	385	2ndary	22.00%	23.00%		30.00%	9.00%		22.00%	
Kerdpon (2001)	155	2ndary	4.50%	5.80%	27.70%			29.60%	28.40%	
Koivunen *et al.* (2001)	84	GP	14.30%	4.80%	47.60%	2.40%	26.20%			

Kowalski *et al.* (1994)	336	2ndary	22.00%					63.00%		
Lau, Wilkinson and Moorthy (2018)	73	2ndary	36.00%	8.00%	10.00%	17.00%	3.00%	8.00%	8.00%	
Mashberg *et al.* (1989)	94	2ndary	36.00%		64.00%		29.00%	71.00%		40.00%
Merletti *et al.* (1990)	279	GP	7.50%	24.00%		74.60%	11.10%			
Metcalfe *et al.* (2019)	61	2ndary	27.10%					29.50%	24.60%	21.30%
Mettias, Charlton and Ashokkumar (2021)	66	2ndary	54.50%	6.10%	6.10%	22.70%	4.50%		6.10%	
Nieminen *et al.* (2021)	40	GP	25.00%		10.00%	5.00%				
Pitchers and Martin (2006)	69	2ndary	49.30%	2.90%	33.30%		2.90%		5.90%	
Pracy *et al.* (2013)	35	2ndary	42.80%	5.70%	14.30%	25.70%		11.40%		0.00%
Pugliano *et al.* (1999)	1010	2ndary	16.40%	22.30%	8.80%		20.50%		0.40%	
Queenan *et al.* (2018)	113	2ndary	13.30%	4.40%	24.80%	13.30%			27.40%	
Rimmer *et al.* (2012)	36	2ndary	63.00%		13.00%	19.00%	4.00%	7.00%	19.00%	0.00%
Rogers *et al.* (2019)	28	2ndary	17.90%		35.70%	10.70%		14.30%		
Rosell Ferrer *et al.* (2021)	35	2ndary	42.10%		17.10%	31.40%			5.70%	
Rutkowska *et al.* (2020)	305	2ndary	1.30%	5.60%	3.90%	2.00%		13.30%	29.50%	
Scott, Grunfeld and McGurk (2005)	245	2ndary	19.60%	9.00%	60.40%		9.80%	47.80%	30.20%	10.20%
Shephard, Parkinson and Hamilton (2019)	813	GP		5.00%	33.30%	52.00%	4.00%			
Singh and Warnakulasuriya (2006)	6	2ndary				30.10%		66.00%	16.00%	50.00%

Talwar *et al.* (2020)	6	GP	50.00%	33.30%	50.00%	33.30%	16.60%			
Teppo *et al.* (2003)	66	2ndary	6.00%		14.00%	77.00%				
Tikka, Pracy and Paleri (2016)	397	2ndary	48.30%	7.30%	4.00%	17.40%	2.00%	12.80%	5.30%	1.50%
Zeitler *et al.* (2018) *	171	2ndary	52.00%	14.60%	16.40%	36.30%	14.50%	2.90%	20.50%	
Total	8327		29.9%	10.7%	24.7%	28.4%	10.6%	24.9%	16.3%	13.8%

*The cohort of patients used in this paper was also used in 4 later published HNC papers; therefore, these were excluded from the table and any further analysis (Fingland *et al.*, 2018;Douglas, Carswell and Montgomery, 2019;Douglas *et al.*, 2021b;Tikka, Paleri and MacKenzie, 2018)

***Note**: Average values for other and less common symptoms were as follows: weight loss (10.8%); haemoptysis (3.2%); dyspnoea (5.9%); stridor (2.8%); Head and neck pain (2.9%); cough (3.3%); feeling of something in throat (1.4%); skin growth (5.6%); cranial neuropathies (1.4%), unilateral nasal symptoms (2.8%); dental mobility (0.6%).*

Table 2-4. Studies reported presenting symptoms in patients referred with suspected head and neck cancer

Study	Patients	Neck lump	Dysphagia	Hoarseness	Oral ulcer	Sore throat	Oral swelling	Patches	Globus
Alho (2006)	5867	0.40%	0.20%	1.40%		3.80%			
Allam and Nijim (2019)	790	34.20%	4.20%	18.60%		14.10%			22.90%
Fingland *et al.* (2018)	1998	31.30%	7.90%	15.90%		4.80%		0.80%	13.20%
Hobson *et al.* (2008)	177	17%	7%	28%	2%	11%		1%	7%
Kennedy *et al.* (2012)	199	35%			35%				
Mettias, Charlton and Ashokkumar (2021)	1107	23.40%	5.60%	17.80%	1.10%	12.10%	4.40%		
Miller and Hierons (2012)	108				37%		21%	31%	
Montgomery *et al.* (2019)	250	35.60%	12%			16.40%			16.40%
Pracy *et al.* (2013)	622	25.50%	4%	25%	7%	12%		4%	4%
Rogers *et al.* (2019)	390	39%		7%	6%	45%	8.20%		6.90%
Rosell Ferrer *et al.* (2021)	134	28.40%		38.10%			2.20%		
Shephard, Parkinson and Hamilton (2019)	4365		1.40%	9.90%		8.30%			
Tikka, Pracy and Paleri (2016)	4715	22.50%	4.70%	20.20%		5.80%	3.60%	3.20%	7.10%
Williams *et al.* (2014)	462	35.30%	3.70%	36.40%	4.30%	19.30%	2.20%	1.70%	
Talwar *et al.* (2020)	113	23%	7.1%	43.4%	0.8%	6.1%	15.1%	0.8%	0.8%
Total	21297	**27%**	**5.3%**	**21.8%**	**11.6%**	**13.2%**	**8.1%**	**6.1%**	**9.8%**

Note: *Average values for other/less common symptoms were as follows: intermittent hoarseness (11.2%); intermittent sore throat (6.7%); otalgia (1.6%); cough (4.8%); haemoptysis (2.6%); dyspnoea (2.1%); unilateral ear symptoms (4.1%); unilateral nose symptoms (3.5%); skin lesion (0.8%); neuropathies (1.2%); unilateral head and neck pain (4.4%).*

Lower socioeconomic status is also linked also with a higher number of 2ww referrals from these regions compared to areas of high socioeconomic class (Rogers *et al.*, 2019;Zeitler *et al.*, 2018). Nevertheless, socioeconomic status could not be blamed solely as a cause of HNC, as there is a strong positive correlation between current smoking and excess alcohol intake for low socioeconomic status areas, with a decreasing percentage of cases for the higher socioeconomic levels related to lifestyle factors differences (Zeitler *et al.*, 2018). This argument was validated in a study that found that when adjusting for smoking and alcohol consumption, the significant association between oral cancer diagnosis and social deprivation and unemployment is not maintained. Therefore deprivation should not be used as a sole factor in triaging potential referral for HNC, but rather, the focus should be towards smoking and alcohol status, which can vary across the regions and be affected by deprivation (Conway *et al.*, 2010b). Nevertheless, looking at the male population, a study found that education attainment remains a significant factor associated with a diagnosis of HNC even after smoking and alcohol are taken into consideration, but variation was seen across Europe (Conway *et al.*, 2010a). The reported variations make educational attainment not a factor that can be easily incorporated into universal risk assessment tools for HNC detection. However, the importance of inclusion of information on smoking and alcohol history in any referral triage pathway and tool should be emphasised, especially in areas with high deprivation, as incidence and mortality are higher within these populations (Taib *et al.*, 2018).

To conclude, the evidence suggests that there are symptoms strongly related to HNC diagnosis. Aside from the symptoms list, there are also social history factors linked to HNC, which are smoking and alcohol, as well as lower socioeconomic status, even though the latter is interlinked with smoking and alcohol abuse. The information from this review of the literature will be used to inform the development of the HNC risk calculator, allowing for the design of a data collection proforma to include all the above parameters mentioned in the literature that relates to a HNC diagnosis. In the following section, factors associated with an advanced-stage HNC disease will also be covered and taken into consideration in the design of the data collection proforma.

2.4.2 Symptoms and other risk factors associated with advanced-stage disease at the time of HNC diagnosis

As mentioned in the introductory chapter, currently, two-thirds of the HNC cases are diagnosed at an advanced disease stage, which is known to be associated with worse survival outcomes (Siegel et al., 2019; Thompson-Harvey et al., 2020). One of the main focuses of the NHS long-term plan is to detect 75% of cancers at an early stage by 2028 (NHS, 2019). For this to happen, it is important to understand the characteristics of the patients that are presenting with advanced disease to target these factors for early intervention approaches. The factors associated with advanced HNC will be covered in this section.

Advanced disease is associated with the presence of more than one symptom. An increase in the advanced-stage disease odds ratio from an average of 3.5 to 4.5 for the presence of 1 symptom to an average odds ratio of 23 – 35 for the presence of 3 or more symptoms has been reported (Carvalho *et al.*, 2002). Generalised symptoms of insomnia, loss of appetite and fatigue are associated with very advanced HNC that commonly leads to palliative treatment regimes (Gandhi *et al.*, 2014). Dysphagia, hoarseness and drooling for oral cavity/oropharynx tumours and the presence of odynophagia at times of initial presentation for all types of HNC has also been linked with advanced disease stage (Carvalho *et al.*, 2002). Weight loss at the time of HNC diagnosis is associated with one of the worse survival outcomes, as documented by Douglas et al. (2018). A previous study from the 90s found that neck lump, dysphagia, otalgia, and weight loss are independent factors of HNC survival, with weight loss having the strongest effect on the duration of survival (1.59 odds ratio), a patient presenting with weight loss is 1.5 times more likely to die than individuals without this symptom. Neck lump follows with an odds ratio of 1.42, dysphagia at 1.29, and otalgia being 1.28. When a patient presented with 3 or 4 of these symptoms, the mean survival duration was 31 months, whereas if only 1 symptom was present, the mean survival was 56 months, 20 months mean reduction compared to the presence of none of the 4 highlighted symptoms. The same reduction in mean survival was seen even when looking at each cancer stage separately. There has also been a suggestion for the incorporation of these symptoms in the overall cancer staging alongside the TNM

classification. Symptoms, especially those affecting survival, manifest the biological behaviour of the cancer and how it affects its host, hence adding information to the disease severity (Pugliano *et al.*, 1999). A study looking at the nature of symptoms associated with an emergency presentation with HNC and associated with advanced disease stage found specific symptomatology being associated with this presentation and late-stage disease being: airway compromise, dysphagia and malnutrition, bleeding from mouth or neck (Wilkie *et al.*, 2021).

Aside from symptoms, other factors have also been linked to advanced HNC at the time of diagnosis. In oral cancer, a study has shown that an advanced disease stage was associated with older age (>80 years for tongue and >70 years for floor of mouth), being widowed, being socially marginalised and being a current smoker or ex-smoker and heavy drinker. Patients that visited a dentist regularly were less likely to be diagnosed with advanced disease (Groome *et al.*, 2011). The majority of patients with advanced-stage disease are from areas with low socioeconomic indexes (Zeitler *et al.*, 2018;Olsen *et al.*, 2015b). With higher alcohol and smoking consumption in the socioeconomically deprived group, it has been suggested that the combined use of alcohol and smoking in the cancer cells may be promoting aggressiveness and advanced disease stage for these patients (Kaufman, Grabau and Loré, 1980). After adjusting for age and stage of disease, social deprivation is linked with worse survival outcomes within patients treated with curative intent (Rylands, Lowe and Rogers, 2016).

Given that patient-related delays (interval from symptom development to attending their GP for the first time) have not been linked with advanced cancer stage at the time of diagnosis (Vernham and Crowther, 1994), advanced disease at the point of presentation is likely to be related to tumour aggressiveness. Duration of symptoms has not been linked to the HNC stage at the time of diagnosis, nor to any specific HNC cancer subsite, alcohol use or comorbidities (Dolan, Vaughan and Fuleihan, 1998). High-grade histology was predictive of an advanced stage of diagnosis in a multivariate logistic regression analysis study that again found no difference in patient-related delay and stage of cancer at the time of diagnosis (McGurk *et al.*, 2005). This

is an important finding as any intervention aiming to improve early detection needs not only to assess for potential changes in the stage of disease at the time of diagnosis but also the long-term outcomes of these patients, that is, their overall and disease-free survival. It could be that even if a patient is diagnosed at an earlier stage following an early detection intervention, their disease is aggressive; hence it does not respond well to the available treatment to allow for improvement in long-term outcomes (Kaufman, Grabau and Loré, 1980). If a new intervention is to be introduced for HNC triaging and compared to current pathways, its long-term outcomes should be adjusted to account for the tumour aggressiveness prior to the conclusion on survival and disease outcomes.

To conclude, the evidence suggests that there are potential tumour characteristics that can relate to a later disease stage at the time of initial presentation. The information from this review will also be considered for addition to the list of recorded symptoms and risk factors to inform the design of the updated HNC risk calculator.

2.5 Chapter Summary

The first part of the literature review chapter covered the differences in the health care models across the globe and the role of primary care in cancer detection. The first healthcare contact for patients with symptoms suggestive of cancer varies depending on which part of the world and country they reside in. It can be a review by primary care doctors or other healthcare professionals or a specialist review in primary or secondary care that can be privately or state-funded. The above gives an insight into the language that will be used in the development of the HNC risk calculator, which needs to be simple enough to be understood by primary and secondary care doctors and other healthcare professionals but also potentially directly by patients. As the main focus of the thesis is the development of a risk calculator firstly for use within the UK healthcare system, developing a tool that can be used in primary and secondary care will be the main objective.

Aside from country-related differences in the structure of healthcare facilities and referral for suspected cancer arrangements that can be a barrier to early cancer

detection, other types of barriers were explored in the second section of the literature review. Lack of awareness of the cancer symptoms by primary care doctors and variations in their training that may not sufficiently cover the background knowledge and examination skills for cancer exclusion was highlighted. Large variations were also noted in the protocols, guidelines, and referral support tools available, which are mainly focused on asymptomatic population screening for common cancer rather than early detection of already symptomatic patients. Patient-related barriers to early diagnosis also exist, such as socioeconomic inequalities, lack of awareness of cancer symptoms, gender, age and culture-related misbeliefs and negative feelings. They have, however, been found to affect, to a lesser degree, the cancer early detection outcomes. Hence, to improve the early detection of HNC, that is, a cancer without currently a universally accepted screening test, the focus of this thesis will be on the development of a tool that can be used as a decision aid for the health care professionals, by summarising the red flags symptoms for cancer, standardising, and reducing the variation in the clinical assessment of patients with suspected HNC.

Next, the literature review covered how despite the fact that guidelines are in place at a continental, country, or regional level to help referral processes for suspected cancer, the list of symptoms included in the guidelines varies. They are based primarily on a low level of evidence from panels of expert opinions at organisational and government levels. Even within the UK, variations exist in the list of symptoms across the nations. And despite a sharp increase in the number of urgent suspected cancer referrals in the UK, the conversion rate is low, and the detection rate remains modest, with most cancers being diagnosed via other routes, with the percentages varying for different cancer sites. Looking specifically at HNC, international and UK-based guidelines for a HNC referral were covered. The list of symptoms included in the guidances is based on scarce primary care data and expert panel opinions. Low conversion and detection rates were seen for the urgent suspected cancer referrals despite multiple updates to the guidance being made over the years. Hence, the rationale for the development of a HNC triage algorithm is to design an evidence-based tool that will undergo a rigorous internal and external validation process for universal implementation across the UK but also with the potential for global use.

The next section of the literature review covered information for the available cancer risk calculators. The review identified a plethora of calculators available for common and rarer cancers but only 2 calculators available for symptomatic HNC population. It was noted from the review that the majority of calculators were using pre-populated registries of symptoms that were initially collected for other reasons and subsequently used for the development of cancer risk assessment models with no a priori sample size calculation and using a case-control methodology. These decisions introduce reporting bias and fail to acquire information about the true cancer and symptom incidence in the population, reducing the strength of the results and their generalisability. Cohort studies especially using primary care data, have been deemed impossible to conduct due to the volume of the consultations and the rarity of the cancer outcome as opposed to the symptoms' frequencies. Some calculators were also based on retrospective data collection, which aggravated the bias relating to reporting and recollection, as was the case of the two available HNC calculators. External validation studies are available for a minority of calculators, which reduces their generalisability. The existing calculators include symptoms and risk factors for cancer only, or they also incorporate radiological investigations and results of blood investigations or other tests. The latter are hard to implement in primary care or at the point of first patient contact to allow a one-stop clinic triaging. The addition of more specialised tests has shown only a modest increase in the prediction power of the calculators, which was not evident at all for some cancers, with any additional benefit being at the expense of needing additional visits to the primary care or secondary care referral prior to a decision being made on the risk of malignancy. The statistical analysis used in the design of the calculators is mainly multivariate logistic regression modelling, but more recently, artificial intelligence methods have been used, with the most common method being the random forest modelling. The AUC is not reported in all risk calculator papers, but when this statistic is available is usually over 0.80 for most cancers.

To inform the design of the HNC risk triaging tool, the final part of this chapter was a review of the literature, identifying all symptoms that have been linked to a HNC referral and diagnosis with no regional or time limitation, alongside information on

HNC risk factors. This information will be used for designing the data collection proforma that will inform the HNC risk model development to ensure no significant factors will be missed.

Taking all this information into consideration, the design of the HCN calculator of this thesis research work will be based on prospective data collection from secondary care consultations hence having a volume of data that is achievable to collect and analyse. Sample size calculation will be performed to ensure that the results will be a good fit for the data and allow generalisability of the results. The database of symptoms will be built based on the review of the literature associated with HNC to ensure no important symptoms will be missed. As the scope of the calculator design is for use as a triaging aid to be used at the first clinical consultation, symptoms and demographic information only will be included in the variables of the model and assess if these alone can provide good discrimination. External validation of the model will be performed to ensure its generalisability and applicability in populations other than the one it is derived from. Logistic regression and the random forest AI approach will be used to assess which method provides optimal discrimination outputs to be used in the final selected model.

Based on the knowledge gaps and clinical needs outlined above, this thesis aimed to improve early HNC diagnosis by developing and validating an updated version of a previously designed HNC risk calculator (Tikka, Pracy and Paleri, 2016) - HaNC-RC. The research objective was to improve the HaNC-RC predictive ability to be more in line with other risk calculators for common cancers, most of which have been found to have an AUC of over 80%. The first research questions were if there are new significant symptoms or relevant social history factors that can be added to the HaNC-RC and if its current variables can be refined to increase its prediction power. The second research question was how the updated version of the HaNC-RC would perform in triaging a new cohort of patients referred to HaN clinics across the UK. The methodology leading to the development of the HNC-RC v.2 calculator will be described in detail in the next chapter.

3 Methodology

This chapter will cover the study design and setting during the development and the external validation phase of the refined HNC risk calculator - HaNC-RC v.2, the selection of participants, the data collection process, and the ethical considerations. Furthermore, it will describe the data analysis methods, as well as underpinning the rationale behind selecting the prediction methods, which were informed by the literature on the development of cancer risk calculators presented in the previous chapter. As has already been mentioned in the COVID-19 Impact Statement on the introductory pages of my thesis and also in section 1.7 which provides an overview of the thesis aim and research questions, the original aim of the thesis was to develop the HNC risk calculator (phase 1). The external validation phase (phase 2) was opportunistic, being an analysis of a database which became available during the course of my PhD studies, as the ENTUK and INTEGRATE organisations collaborated in a service evaluation audit using the updated version of the calculator (HaNC-RC v.2) at the start of the COVID-19 pandemic. The methodology of the two phases of the HaNC-RC v.2 development and validation will be described in this chapter, but it needs to be clear that any decisions taken in regard to the collection of data used in the external validation phase of the calculator were led by ENTUK and INTEGRATE.

3.1 Study Design and Setting

This study was performed in two phases. In the first phase, the risk calculator was designed based on prospectively collected data from patients seen in head and neck clinics across the Greater Glasgow and Clyde region after sample size calculation. The data were collected based on a pre-designed proforma informed by the literature on HNC symptoms and risk factors and also after consultation with experts in the field. Statistical analysis was performed using logistic regression and random forest methodology to identify the best-fitted model. The second phase was the model's external validation process. It was based on a pan-UK prospectively collected cohort of patients referred with suspected HNC symptoms during the COVID-19 pandemic.

The risk calculator was used for triaging patients with low or high risk for HNC based on an initial telephone consultation due to COVID-19 pandemic constraints on face-to-face appointments. The discriminatory ability of the risk calculator was assessed based on a logistic regression analysis that was found to be the optimum methodology during the design phase of the calculator.

3.1.1 Development phase

The development and validation for the HNC risk calculator were conducted in a two-step process. The first research questions of this study were to assess if there are any new variables (symptoms, social history factors) that can be added to the HaNC-RC and if its current variables can be refined to increase its prediction power. This question was covered in the development phase of the calculator. For the development phase, data were prospectively collected from patients seen in secondary care hospitals covering the Greater Glasgow and Clyde region. These hospitals were: the Queen Elizabeth University Hospital Glasgow, the Glasgow Royal Infirmary Hospital, the New Victoria Hospital Glasgow, the Royal Alexandra Hospital Paisley, and the Inverclyde Royal Hospital. As was discussed in the previous chapter, following the literature review of other cancer risk calculators, prospective data collection is the best approach for data collection to eliminate patient recollection bias. Prospective data collection methodology was lacking in the HNC risk calculator literature, with the two available calculators having both been based on a retrospective cohort study design. The secondary care setting was used as the information from the literature had suggested that conducting cohort studies using primary care data is almost impossible in cancer research due to the rarity of the cancer event outcome and the high frequency of the reported symptoms. A case-control setting could have been selected instead and applied in primary care as it was done in other studies (Hamilton et al., 2009), but this would have precluded the true symptom prevalence estimations and the unadjusted likelihood of cancer for each symptom calculation.

3.1.2 Validation phase

The second research question of this study was to assess how the updated version of the calculator would perform in triaging a new cohort of patients with suspected HNC

symptoms. This was addressed in the validation phase of this thesis. For the validation phase of the newly developed HNC risk calculator, a UK-wide prospective validation study was conducted. The call-out for research collaboration for this nationwide external validation process was performed via the UK ENT Trainee Research Network (INTEGRATE) website (INTEGRATE, 2020) and the British Association of Otorhinolaryngology (ENT UK) (ENTUK, 2020). All UK secondary care ENT departments were invited to participate in this prospective study by emails sent out by the INTEGRATE and ENTUK organisations to their membership lists. The list of the total of 41 participating hospitals can be found in Appendix II and includes 32 hospitals in England, 6 in Scotland, 2 in Wales and 1 in Northern Ireland. As the literature review has shown, despite the fact that a plethora of risk calculators are available for common cancer, not many are externally validated, and they are usually validated in populations arising from one institution. External validation of the previous version of the HNC calculator was performed in a retrospectively collected cohort of patients seen across a region in Scotland (Tikka et al., 2018). For the external validation of the updated version of the calculator, prospective data collection was planned to eliminate recollection bias, and the patients were recruited from different regions across the UK to ensure the generalisability of the calculator, at least within the UK.

3.2 Participants

3.2.1 Development phase

For the development phase of the HNC risk calculator, newly referred patients from primary care were included and seen in all types of HaN clinics (USOC, urgent, routine) from January 2017 until December 2018 in secondary care hospitals covering the Greater Glasgow and Clyde region as listed above. Data were collected from all types of HaN clinics as the literature review suggested that over two-thirds of HNC, diagnoses are made in clinics other than the USOC appointments (Kennedy *et al.*, 2012). An 8% prevalence of HNC is assumed amongst individuals that present to their GP with HaN symptoms (Langton *et al.*, 2019). A sample size of 3602 was required assuming an 8% cancer prevalence, to achieve an 80% power with an assumed test performance of at least 80% sensitivity and 75% specificity estimated to be within +/- 5%, being in line with the available literature on risk calculator development from

other common cancers (Steyerberg and Vergouwe, 2014;Louie *et al.*, 2015). The 80% power threshold was decided as the research objective of this study was to improve the HaNC-RC predictive performance to be more in line with other common cancer risk calculators (AUC>0.8). Lack of prior sample size calculation was noted in the majority of the reviewed cancer risk calculators, which can affect the goodness of fit of any developed statistical models. This was eliminated in this research work as the sample size calculation was performed at the study's outset.

The sample was initially collected on all referrals, but after the first 18 months of data collection, the cancer incidence was lower than the expected 8% (aiming for >300 cancer cases) whilst reaching saturation of the non-cancer referrals symptoms. Hence, data collection continued only for patients with a cancer diagnosis until the targeted number of cancer cases was reached to boost cancer numbers and enable better prediction modelling. The HNC incidence, cancer diagnosis per clinic appointment, calculation of negative predictive value (NPV) and positive predictive value (PPV), both of which are dependent on the cancer prevalence in the population, and suggested re-triaging strategy following the risk calculator development, were assessed in the un-boosted cancer cohort to ensure non-contamination of sample from boosting the cancer prevalence.

3.2.2 Validation phase

The prospective validation study was performed during the first wave of the COVID-19 pandemic, from the 24th of March 2020 to the 13th of July 2020, lasting a 16-week period. All face-to-face first HaN clinic appointments were stopped during that period and replaced with telephone triage consultations. The validation study included all patients who were referred from their GPs via the suspected HNC pathway to the secondary care hospitals participating in the study across the UK. This cohort of patients was selected for the validation phase of the study, answering the second research question of the thesis, assessing the performance of the updated version of the calculator in a new cohort of patients referred to HaN clinics across the UK. All patients had an initial telephone consultation with an ENT doctor at each of the participating sites using the HNC risk calculator tool to inform patient triaging. The

patients were triaged to an urgent face-to-face appointment for the high-risk patients or to a deferred face-to-face or remote review appointment, with or without requested investigations for the low-risk patients. No a priori sample size calculation was performed during the validation phase.

3.3 Data collection

3.3.1 Development phase

The decision on what data needed to be collected was taken based on the review of literature on symptoms and risk factors associated with HNC diagnosis as described in section 2.4.1 of this thesis to ensure all literature-reported symptoms and factors are considered for collection. The first version of the HNC risk calculator (Table 3-2) was used as the initial data collection template, with the additional variables added being informed by the literature hence being expected to increase the prediction ability of the first version of the HNC calculator. The final collection proforma is seen in Table 3-3. This methodology was used to cover the first research question of the thesis assessing if there are any new symptoms or other factors that can be included in the calculator and if its current variables can be further refined. The final proforma for anonymised data collection was developed following consultation with HaN consultants in the Greater Glasgow and Clyde region to ensure that the consultation time would remain within the limits of the current clinical setup whilst all relevant symptoms, signs, demographics, and social history factors would be adequately captured. Previous HaN clinic letters were also screened to check for the data captured in usual practice. These were letters to GPs as a result of the HaN clinic consultation. This knowledge informed the development of the proforma used in the consultation with the HaN specialists. The finalised data collection form can be seen in Table 3-3. All symptoms and signs were recorded using this pre-agreed clinic proforma by the doctor at the consultation. The data were extracted from the electronic patients' notes and the final clinic letter, which included the dictation of the patients' collected data using the proforma after the end of the consultation episode for each patient. This process ensured that all potential sources of information were checked for any additional recorded symptoms keeping the data collection standardised and optimised. Data were collected using Excel, and categorical variables were numerically coded prior to transfer to the SPPS and R software for analysis.

Table 3-1 Data collection form for the first version of the HNC risk calculator (Tikka et al., 2016)

HNC risk calculator v.1 Template	
Age:	
Biological Gender: Male/Female	
Persistent hoarseness:	Yes/No
Intermittent hoarseness:	Yes/No
Oral ulcer:	Yes/No
Oral swelling:	Yes/No
Dysphagia:	Yes/No
Odynophagia:	Yes/No
Neck mass >3 weeks:	Yes/No
Neck mass <3 weeks:	Yes/No
Unexplained Otalgia:	Yes/No
Feeling of something in Throat:	Yes/No
Presence of blood in mouth/Haemoptysis:	Yes/No

Table 3-2. Risk calculator development - Data collection proforma

Head and Neck Cancer Risk Calculator Proforma			
Date:			
Patient Label			
Age:			
Biological Gender: Male/Female			
Smoking: Current/Ex/No			
Alcohol: >14U per week/<14U per week/Ex-excess		Units/week (if known):	
Symptoms			
Hoarseness:	Yes/No	Persistent/Intermittent	Duration……
Dysphagia:	Yes/No	Persistent/Intermittent	Duration……
Odynophagia:	Yes/No	Persistent/Intermittent	Duration……
Sensation of a lump in throat:	Yes/No	Persistent/Intermittent	Duration……
Sore throat/discomfort:	Yes/No	Persistent/Intermittent	
Duration……		Unilateral/Bilateral	
Pain in head and neck	Yes/No	Persistent/Intermittent	
Duration……		Unilateral/Bilateral	
Neck lump:	Yes/No	Fluctuating/persistent/increasing	Duration……
Haemoptysis:	Yes/No		Duration……
Oral/lip mucosa ulcer:	Yes/No		Duration…….
Oral mucosa/lip swelling/growth:	Yes/No		Duration…….
Unexplained otalgia:	Yes/No		Duration…….
Unintentional Weight loss:	Yes/No		Duration……..…
Other symptoms: …….			
Provisional diagnosis: ……		Cancer: Yes/No	
For investigation:	Yes/no	What investigation: …….	
Any other Comments:			

3.3.2 Validation phase

An excel spreadsheet was designed and disseminated to the local collaborator in each of the contributing sites and was also available online to download from the INTEGRATE website (INTEGRATE, 2020). The dataset included the variables found to be significant during the risk calculator development phase and included in the final model of the refined HNC risk calculator. Table 3-3 shows a capture of the excel spreadsheet that included anonymised patients' demographics, social history of smoking and alcohol and symptoms and signs based on the HNC risk calculator. These data were obtained from ENT-UK for the external validation analysis purposes. The ENTUK managing committee collated all the returns from each of the local collaborators. Following my request, I was granted access to the finalised anonymised clean database.

Table 3-3. Risk calculator validation - Data collection Excel spreadsheet

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
fore submission	Demographics		General			Voice	Airway	Swallowing				Oral		Misc			
ID	Age	Gender	Smoking	Alcohol	Unintentional weight loss	Hoarseness	Stridor	FOSIT	Sore Throat	Odynophagia	Dysphagia	Oral swelling	Oral ulcer	Unexplained unilat otalgia	Neck lump	Skin lesion	Calc result
Patient ID	Age	Gender	Do you smoke?	Do you drink alcohol?	Have you lost any weight without trying?	Do you have a hoarse voice?	Do you have noisy breathing?	Do you have a feeling of something stuck in your throat?	Do you have a pain in your throat?	Do you have pain when you swallow?	Do you have any difficulty swallowing?	Do you have a new swelling in your mouth?	Do you have a new ulcer in your mouth?	Do you have any new ear pain?	Do you have any new lumps in your neck?	Do you have a new growth on your skin on your H&N?	Outcome
Example	*45*	*Male*	*No*	*≤14 units/week*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*Low risk*

3.3.3 Missing data

The databases were cleaned of erroneous entries, and screening for missing data was undertaken. During the development phase, where missing or ambiguous data were noted, the electronic notes of these patients were checked in an attempt to find any missing data entries and update the database. During the validation phase, any missing or erroneous data enquiries were sent to the local hospital collaborators for clarification and data resubmission. When the enquiries could not be resolved, and missing data were still present during the data clearance process, the records were removed from any further analysis.

Nevertheless, in order to ensure that no significant alteration in the final results was encountered due to the deletion of the missing data, initially, a univariate analysis was performed. This was performed for each variable in the whole dataset against the outcome of interest (Cancer diagnosis) prior to missing entries deletion, alongside calculation of the odds ratio, the 95% confidence interval, and the area under the ROC curve for each feature against the cancer diagnosis. Following the deletion of the missing entries, the univariate analysis was repeated for the clean database to assess for any significant change in the univariate analysis p-values for evaluation of the effect of missingness. The results of this analysis are presented in the results section 4.2.5. Missingness did not significantly affect the analysis results.

3.4 Ethical Considerations

Approval for using the patients' details for the development phase of the HNC risk calculator was granted by the Audit department of the Greater Glasgow and Clyde NHS hospital after an audit form had been completed and signed off by the Caldicott Guardian. The data recorded by the clinicians in the clinical notes did not deviate from standard practice. No ethics committee approval was therefore required for this study. The projects were registered with Caldicott, guardian of NHS Greater Glasgow and Clyde (GGC/07/02/17), as a quality improvement project. No identifiable patient information was included in the database. The data have been transferred from the hospitals' computer to a personal computer using an NHS (National Health Service) email account, an approved way to transmit confidential patients' information (www.connectingforhealth.nhs.uk, 2014). Subsequently, a secure password unknown to third parties was set up on a personal computer to avoid unauthorised access to the patients' information.

Due to the pan-UK nature of the validation phase, data collection for each participating hospital was performed by a lead clinician at each hospital site following Caldicott's approval. The Health Research Authority (HRA) decision tool was used prior to the beginning of the study to ensure that the correct ethics approval was obtained. The study fell under the remit of service evaluation projects; therefore, no ethical approval was required (http://hra-decisiontools.org.uk/research/). ENTUK and INTEGRATE

were responsible for the information inserted on the HRA decision tool and the final decisions taken in regard to ethics approval requirements.

3.5 Data analysis

Many different analytical approaches are available for the design and validation of cancer risk calculators. Each method has pros and cons; hence it is important to present and explain the reasoning behind the selection of any final data analysis method. The following section will cover the theoretical background of cancer risk prediction modelling and justify the chosen data analysis method of this thesis.

3.5.1 Introduction to cancer risk prediction statistical modelling

Cancer risk calculators use risk prediction modelling to estimate the risk that cancer is present based on the use of specific predictors. That risk is called absolute probability, and it is measured for an individual with a particular predictor (called covariates of the model) profile. The model is based on a mathematical function that relates the binary outcome of the presence or not of cancer to a set of predictors/covariates. These covariates vary from individual characteristics such as age and gender to history and examination findings, social history factors (such as smoking and alcohol), and imaging or other test results, including blood tests and genetic markers (Johnson and Smolenski, 2007).

During the model development and validation process, several points should be considered to ensure that the final risk assessment tool is correctly developed and can be used safely in other cohorts outside the development dataset. The study participants should reflect the outcome of interest, hence including individuals at risk for the selected cancer. The sampling design needs to be ideally prospective, to reduce missing data and recall bias. Careful covariate selection is of paramount importance in model development to ensure that all possible variables that have an association and causation interaction with the outcome of interest are captured and will be assessed for final inclusion in the model. This is achieved with good knowledge of the subject matter (Steyerberg and Vergouwe, 2014). The predictors should be accurately defined and presented in a standardised way to ensure reproducibility for future use of the

generated model. All variables should be assessed for potential inclusion in the model, linearity should not be assumed for continuous variables, and it is preferable not to be turned to dichotomous data as this loses information (Moons *et al.*, 2012b). Following the model development, the model's performance should be assessed by means of discrimination and calibration statistics and by internal and external validation of the model (Moons *et al.*, 2012a).

Due to the binary profile of the outcome of interest, that is, the presence or not of cancer and the potential association with multiple covariates, multivariate logistic regression analysis is commonly used to develop cancer risk calculators. More recently, artificial intelligence has also been used with supervised and unsupervised techniques depending on knowledge or not of the outcome of interest. Usually, supervised machine learning methods are used as the outcome of interest is known, that is, the cancer status (Freedman *et al.*, 2005;Ayer *et al.*, 2010).

3.5.2 The logistic regression prediction method

Regression models are used to examine the relationship between a response variable of interest and one or more explanatory variables, often called covariates, and they are of particular use in medicine and biology research. When the outcome of interest is continuous, a linear regression model is fitted, whereas when the outcome of interest is dichotomous (binary), a logistic regression model is commonly fitted (Hosmer *et al.*, 2013). Different distribution functions have been applied in the literature to analyse binary outcome data, but binary logistic regression is more frequently used due to its simplicity and flexibility (Cox and Snell, 1989).

When we have more than one explanatory variables, a multiple logistic regression model is fitted where we consider *n* to be a collection of all independent variables denoted by the vector $\mathbf{x}' = (x_1, x_2, \ldots, x_n)$. If we denote Y as the dichotomous variable of interest, we use the quantity $\pi(x) = E(Y \mid x)$ to represent the conditional mean of Y given x when the logistic distribution is used. Y_i denotes the value of the dichotomous outcome, and x_i is the value of one of the independent variables for the ith subject from a sample of k independent observations (x_i, Y_i), $i = 1,2,\ldots,k$.

The outcome is coded as 0 or 1, indicating the absence or presence of the outcome of interest with β_0, β_1,…, β_n the unknown coefficient estimates of interest. A transformation of $\pi(x)$ which is used in logistic regression, is the *logit transformation ln[π(x)/(1-π(x))]* which leads to the regression model being defined as:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n (Equation 1)$$

Where:
$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}} \quad (Equation 2)$$

The g(x) transformation is linear in its parameters, may be continuous and may range from

$-\infty\ to\ +\infty$ depending on the values of x. The maximum likelihood method is used to estimate the regression coefficients (Hosmer et al., 2013).

The *odds ratio,* denoted as OR, is the ratio of the *odds* of the outcome of interest being present among individuals who have one of the explanatory characteristics (x_n) deviated by the *odds* of the outcome of the interest being present among individuals without the explanatory characteristic of interest (Altman, 1991).

$$OR = \frac{\frac{\pi(x_n=1)}{1-\pi(x_n=1)}}{\frac{\pi(x_n=0)}{1-\pi(x_n=0)}} (Equation 3)$$

Which equals $OR = e^{\beta_n}$ (Hosmer et al., 2013).

In order for the analysis to be valid, the fitted model needs to meet the specification and assumptions of logistic regression; otherwise, the results will not be valid. Thus, after the model fit, the presence of influential observations will be checked, as well as that the model fits the data well, and no significant interactions exist between the explanatory variables. If any problems with the model diagnostic exist, this can lead to large standard errors of the regression coefficients and invalid inferences.

The assumptions of the logistic regression are summarised below:

1. No omission of important variables
2. No inclusion of irrelevant variables
3. Data have been checked for errors

4. Independence of observations
5. No linear relationship exists between the independent explanatory variables
6. The logistic function is used to find the association between the independent variables and the conditional probabilities (Allison, 2012; www.ats.ucla.edu, 2014).

3.5.2.1 *Development of the updated version of the HaNC-RC - The selected Logistic Regression Model build method*

The aim of this research work was to find those variables that are important in the detection of HaN malignancy and try to introduce and validate a model to calculate the probability of HaN malignancy for patients being referred through the HaN clinics. As we had a dichotomous dependent variable (HNC diagnosis) and a set of independent variables, a multivariate logistic regression model – *logit* model, as described above – was fitted in the collected dataset to deal with the complex interrelationships among many variables. Initially, descriptive statistics were performed on the dependent variable (HNC diagnosis) and all independent variables. The independent variables were screened for any levels within each of the variables that do not give any information that can help develop the logistic regression model. This will be the case if no observations are present for some of the levels of a variable, with the levels requiring merging in this case or if none of the patients with a positive outcome of interest (HNC diagnosis) having presented with a variable being positive, hence identifying a non-informative variable that should not be included in any further analysis. If these variables were left in the model, they would not give any valuable information about the probability of being diagnosed with cancer as well as it would cause quasi-separation during the model building (Allison, 2012).

This was followed by univariate logistic regression analysis of each of the variables against the dependent variable, including an assessment of missingness for each independent variable as mentioned above by assessing for changes in the p-value figures prior to and after the removal of missing data cases. Those variables that had a p-value of 0.1 or less during univariate regression analysis were assessed for potential inclusion in the multivariate regression model. All potential 2-way interactions between the independent variables were also assessed for inclusion in the multivariate

model. Each interaction was tested for inclusion in the model, along with its main effects. The Bonferroni correction (α = 0.1/number of interactions tested) was used to identify the p-value threshold below which interaction is considered for inclusion in the multivariate model.

Following these steps, the initial multivariate regression model was developed with all the above selected as significant independent variables and their interactions on univariate analysis. At this stage, a backward elimination was performed sequentially, removing all non-significant variables and interaction to identify a parsimonious model with each of the finally included variables having a *P*-value of α = .05 or less.

After the model build, the estimated risk (odds ratio) for the binary dependent variable among one level of each of my independent variables was calculated in relation to the risk of another level of the same variable, using the regression coefficient of the variable. The 95% confidence interval of the risk estimate was also calculated using again the regression coefficient of each of the independent variables assuming that they have an approximately normal sample distribution (Chan, 2004).

3.5.2.2 Evaluating the Goodness of fit

The global chi-square test was used to examine if the final selected model is acceptable for the data. The Hosmer and Lemeshow Goodness-of-fit test was used to assess if there is a possibly better model than the one that has been fitted to the dataset (Hosmer and Lemeshow, 2000). The deviance chi-square test would not be a good choice to assess the goodness of fit for these data as the fitted model had many explanatory variables, but this statistic can assess the possible presence of overdispersion (Kleinbaum and Klein, 2010).

3.5.2.3 Model Diagnostic Statistics

Diagnostic statistics were generated for the finally selected model. Residual plots will be helpful in determining which observations have a poor fit by the model. For binary logistic models, the value of residuals is somehow limited as the outcome of interest can take only 2 different values, either 0 or 1. A high residual for an observation will not mean failure of the model but just that this observation had a high probability of

having one side of the outcome but ended up having the other side of the outcome instead. Influence statistics plots were also generated to check if keeping some of the observations in the model would lead to significant unwanted changes to the overall model (Allison, 2012).

3.5.3 The Artificial intelligence (AI) prediction methods

Machine learning analysis has become very popular in recent years in predicting the likelihood of a given event occurring. This approach produces an output by learning from a given dataset. The notion is of training an algorithm that is given specific variables to work with to produce an output. There are two main types of machine learning algorithms: supervised learners and unsupervised learning. The former develops predictive models from a given list of input variables with the aim of identifying and accurately predicting via modelling the relationship between the combination of the input predictor variables (Xs) and the output dependent variable (Y). In the supervised learning approach, the response variable is included in the dataset used to construct the model. The generated output can then be used to supervise the training process to identify the best combination of the predictor variables. Broadly, they can be further grouped into regression and classification subcategories. Regression problems deal with a numeric output that is on a continuous scale. On the other hand, classification problems aim to model a categorical response variable. This can be binary or multinomial (Kuhn and Johnson, 2013). This section will focus on the classification approaches as these are the models that will be used in the thesis for analysing the outcome in concern, that is, HNC diagnosis (yes/no) outcome.

Unsupervised machine learning models are not given a dependent variable as part of the training dataset. The aim is to identify groups within the data with similar attributes to create clusters or to identify variables that have similar characteristics to allow the deletion of duplicate variables, which is called dimension reduction. This technique is not a good fit for the data in this thesis as a clearly defined dependent variable is set, being the presence or absence of a cancer diagnosis. Therefore, the unsupervised machine learning models will not be discussed further.

During the data analysis process, it is difficult to know from the beginning which machine learning approach will work best for a given dataset. Hence, many approaches are initially applied prior to identifying the one that gives the best-fitted model via a thorough learning and validation process for each machine learning approach. The main objective of machine learning modelling is to find a function - f - using the given predictor variables (X) that can give an accurate prediction of dependent variables (Y). f (x) = Y (Györfi *et al.*, 2002). The following section will explain the decision that was taken to perform random forest analysis as the AI method of choice to analyse this dataset. It will initially cover decision tree AI methods as this is the method from where the random forest analysis stems, followed by a detailed review of the random forest AI method of choice and the steps followed in the data analysis informed by the literature.

3.5.3.1 Decision Tree methods

A tree-based model is a type of non-parametric matrix resembling a flowchart that partitions a database output into small groups based on similar responses grouping of the different input variables (Chen and Ishwaran, 2012). It is a supervised machine learning method because there are defined input variables and a corresponding output variable that is known from the outset. It is called a tree, as the method used for data splitting resembles leaves, nodes, and branches. The nodes are where the decision is made to split the dataset based on the presence or absence of a feature (input variable), the leaves are the input variables or final outcomes, and the branches show the combination of features that lead to a final decision. (Breiman, 2001). This method creates multiple chance nodes resulting in the final outcome of interest for each branch. The path from the first decision node – the root - (the first input variable used for splitting) to the end node – final leaf - (final input variable split) represents a classification rule. Figure 3-1 below shows the basic flow of a decision tree.

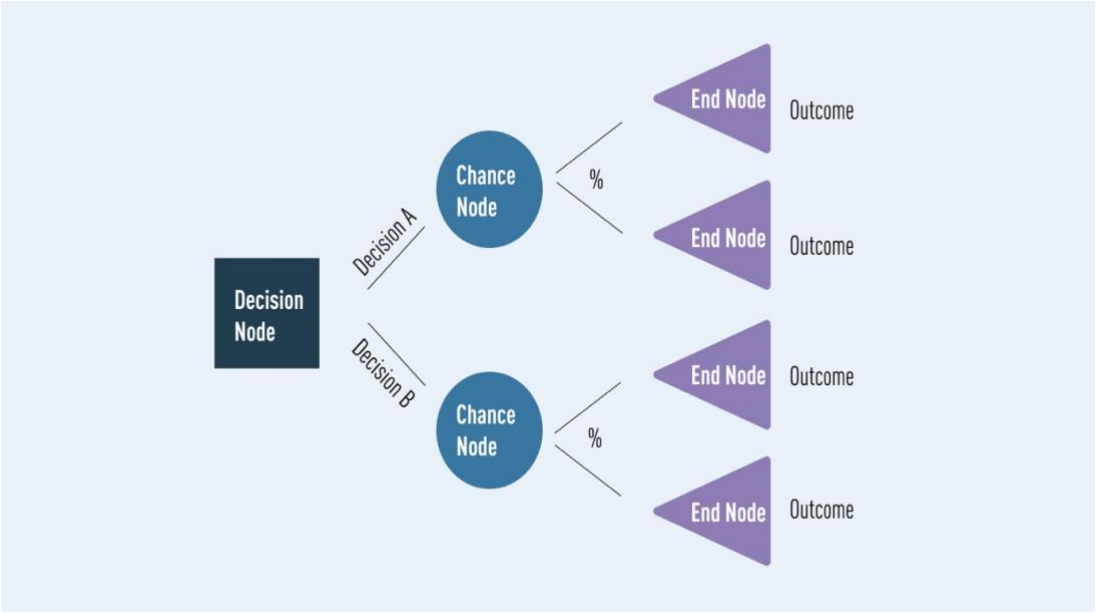


Figure 3-1. Decision tree method diagram

The most common approach used in creating a decision tree is the classification and regression tree methodology. The classification trees are used when the outcome of interest is binary. Like in this thesis, it is the presence or not of HNC diagnosis, so the decision variable is categorical. Regression trees are used when the output variable is continuous, which will not be discussed further as it is not a fit for the outcome data of this thesis.

In classification trees, this splits the training data into subgroups - nodes - based on their responses so that each group is as homogeneous as possible, hence having similar values across the predictor variables. Each subgroup is created based on a series of binary responses incorporating all the available input variables until all features are used. Upon completion of the splitting process, the final output is obtained for the given subgroup when the maximum depth of the tree is reached. The final output for the node is calculated from the average response from the lump of observations included in the subgroup and the most dominant output class within the subgroup (Breiman, 2001;Breiman, 1996).

The order of the splitting within the decision tree is important as each split is affected by the previous data division. This process is called partitioning of the training data, and the aim is each time to identify the best variable to dichotomise the data into to

minimise the possible error; that is called Gini. What is interesting in data partitioning is that the same variable may be used more than once during data splitting despite the fact that multiple variables can be available to possibly help with partitioning. This depends on how much a specific feature may dominate the decision tree algorithm. Decision tree branching can continue at great length, sometimes making the final output too difficult to comprehend. The more complex the output is, the more likely it is for data to be overfitted, precluding the generalisability of the results, hence leading to poor performance. This needs to be assessed against over-simplification of the tree that reduced its prediction power. Fine-tuning of the decision tree branches can be performed by either specifying an early stop notice in the algorithm process or allowing for the most complex possible tree to be generated, then work backwards to identify the best subtree based on the calculated error for each subtree (Strobl *et al.*, 2007).

The main issue with using simple decision trees is that the prediction power is compromised by the selection of the single best fit for the data tree. It is difficult to incorporate within one tree low bias, which can be achieved by extensive tree branching, and at the same low variance that is a feature of trees of shorter length. One way to overcome the high variance issue of extensive tree branching is to perform multiple decision trees using small samples from the same database. This is performed with a technique called bagging. Commonly, up to 500 trees need to be created, and results averaged to reduce variance, but depending on the variables included in the model, a larger number of trees may be needed. A problem with bagging is that the results from multiple decision trees can sometimes be difficult to interpret, unlike the single optimum decision tree approach (Biau, Cérou and Guyader, 2010). Therefore, this technique will not be assessed for inclusion in analysing this thesis dataset.

In logistic regression, the estimate of the variable and its p-value alongside the odds ratio calculation helps to understand how significant each variable is in the final model for the outcome prediction. With tree bagging, this is achieved by calculating the loss function reduction, which is averaged for all bootstraps (Biau, Cérou and Guyader, 2010). Variables with the largest reduction in the sum of reduction of loss of function

(SSE) are considered important. This calculation is important to be taken into consideration when interpreting the results of decision trees with bagging. This is because when a single tree is used, non-informative (significant) variables are not included in the final mode. With bagging, more variables are likely to be included in the final model, even those with small SSE. Those with high SSE are taken into consideration when the final model is interpreted; otherwise, the results will be difficult to understand and put into practice (Breiman, 1996).

Despite variance being reduced significantly with bagging, there is still room for improvement, as the main problem with this approach is that all bootstraps are being split using the same sequence of classification (Genuer, 2012). Hence trees are not completely independent of each other, which increases variance error. This is addressed with the random forest plot machine learning approach (Biau, Cérou and Guyader, 2010). Other techniques also exist, such as decision trees gradient boosting and deep learning approaches, but they extend beyond the scope of the analysis performed in this thesis.

3.5.3.2 The random forest method

Even though constructing a tree-based model can be quite simple, it can lack the accuracy of prediction, as was explained in the section above. Tree-based models tend to overfit the data they are created from because they model quite complex relationships down the branches, similar to interaction terms in logistic regression. These relationships may not be valid when the model is tested in a new dataset, so they fail to make strong predictions at external validation. (Chen and Ishwaran, 2012).

For this reason, decision tree and bagging methods will not be used for the AI part of this thesis. Instead, the random forest analysis was selected as the optimum common machine learning approach in this thesis. It uses random subsets of the data and the variables, hence creating many individual trees, which we can then average over. This overcomes the problem of overfitting, despite the fact that the relationships modelled can still be quite complex. It is a modified decision tree analysis with bagging that creates decision trees that are not related to each other hence reducing the variance error and leading to increased prediction power (Breiman, 2001). This is achieved by

a randomised splitting process each time a decision tree is built that is not correlated to previous decision tree splitting. The classification tree is performed by a random selection of variables from the total pool, followed by a selection of the best splitting point for each variable at each step of the decision tree process. This achieves high predictive power as it is benefited from both the randomness of bootstrap training set selection but also randomisation of variables used at each splitting point (Probst, Bischl and Boulesteix, 2018). The random forest method is summarised in Figure 3-2.

The random forest classification machine learning approach was used to assess if an artificial intelligence approach can be used as an alternative approach to the logistic regression analysis for improved diagnostic accuracy. The random forest package, lattice and deducer package in R was used. These packages give the "mean decrease Gini index" per variable included in the model build as well, allowing the calculation of the Youden index to assess misclassification errors using this method.

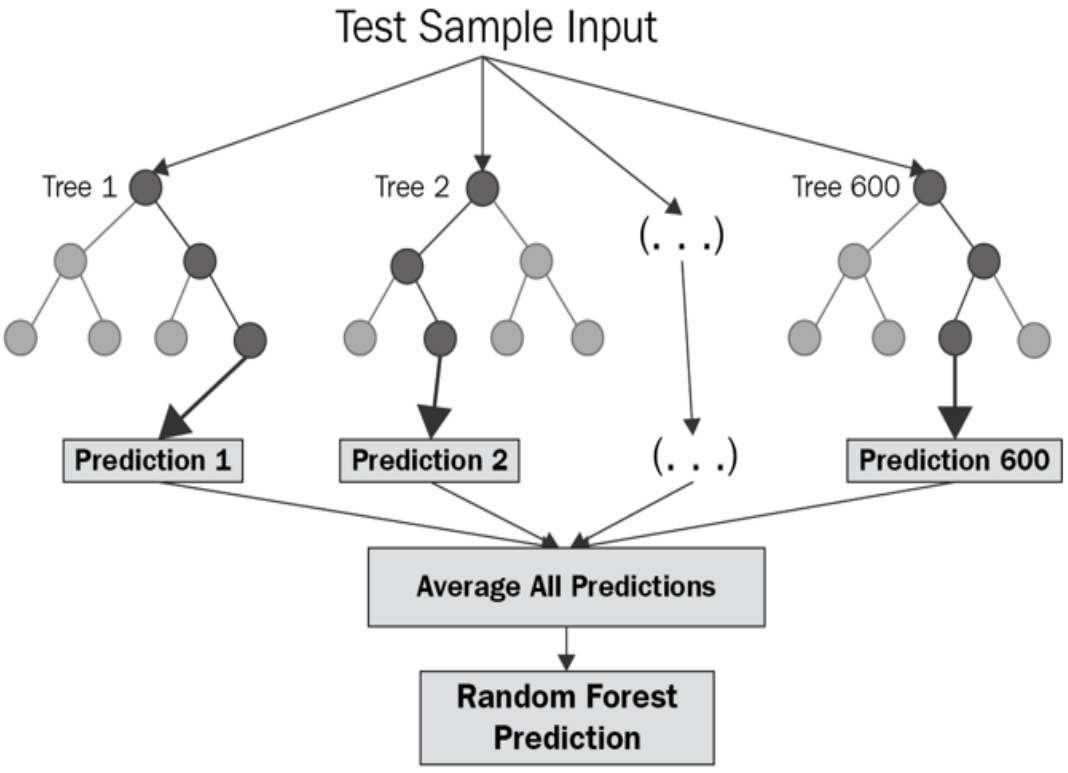


Figure 3-2. Example of the random forest prediction method

The above-mentioned machine learning packages available in R, has a pre-set combination of model parameters, that is, the total number of selected trees upon which averaging will be performed, the total number of features in each tree, the node size, tree length, the size of the node at final splitting and the total number of final nodes. A default node size value of one has been found to generate good predictions for classification (Goldstein, Polley and Briggs, 2011). The root value of the total variables number is recommended for random features selection during splitting for each tree (Kruppa *et al.*, 2014). The default random forest parameters will be used in the data analysis of this thesis using the above-mentioned pre-selected R packages. Hence, the hyperparameters used in the random forest model were 500 trees, with 5 features per node split and a node size of 1.

Post-hoc analysis of the random forest results will also be performed, assessing the difference in the important variables selection using the impurity importance variable ranking. It assesses the mean decrease in impurity (MDI), previously mentioned and also known as Gini importance. MDI counts how many times a variable is used to split a subgroup across all decision trees considering the total number of split samples and produces a final list of significant variables, with the most significant variables being illustrated at the top of the grid (Hastie, Tibshirani and Friedman, 2009).

3.5.4 Cross-validation

The development of a prediction algorithm generated by either logistic regression or a machine learning approach needs to be a good fit for the data used to generate it, and that can accurately predict future outcomes. This is ensured by model validation techniques that assess how the results of a prediction model will generalise in a new independent dataset (Hastie, Tibshirani and Friedman, 2009). For both the logistic regression and the random forest predictive models that were generated with the methodology covered in the previous sections, internal validation was performed. This was performed by splitting the data into a training and a test set. This allows the development of the model using the data in the training set, leading to the identification of the best-fit model, followed by model validation using the test set to assess the generalisability of the model. This assesses the model's performance outside the

training set, hence giving an insight into the model's performance in future datasets. This process is called generalisation error assessment (Hastie, Tibshirani and Friedman, 2009).

Data splitting is important in ensuring that enough data are available in the training set to allow adequate analysis of the model variables but also that the test set is robust enough to allow accurate assessment of model prediction power. The trade-off between model generalisability, hence reducing overfitting - and good model parameter identification is vital and affects the allocation process to training and test sets. In that regard, an 80 – 20, 70- 30 or 60 – 40 % split is often recommended in the literature. Simple random sampling can be used to split the data into training and test sets, which do not control for any particular attributes of the input data or response (Y). This approach will result in a similar training and test datasets distribution. Aside from using one training and one test set to check for model generalisability, resampling methods are often used to allow for a more robust validation assessment. This method ensures that the results of the validation did not occur just because of chance but after repetitively assessing the performance of the model on different combinations of training and set datasets. This is performed using bootstrapping or a k-fold cross-validation process (Schwarz, König and Ziegler, 2010).

The k-fold validation process divides the dataset into k-1 training sets (folds), and the validation is then applied to the one remaining fold. This procedure is repeated k times, allowing for k different validation sets to be used, each time giving a different generalisation error (ge_1, ge_2, ge_3, ….., ge_k). From this output, the mean generalisation error is computed by averaging the k – ge outputs. There is no set number given in the repetitions performed, but a k number larger than 10 has been found to minimise variability (Devroye, Györfi and Lugosi, 2013). Alternatively, bootstrapping can be performed, which creates training and test variables by allowing duplicate observations to be included in the training and validation set. This will decrease the variability of the results and can thus introduce computation bias. Evidence suggests that this can be problematic for small datasets, but it does not compromise the computation and validation results for large datasets (>1000 sample size) (Biau and

Devroye, 2010). As the dataset of this research work was over 3000 patients, the bootstrap limitations did not apply, therefore for validating the logistic regression output, internal validation of the final model was performed using 1000 bootstrap samples, each performing random splits of the data set into training and validation sets with re-sampling with a final generation of the estimated mean AUC across each of the 1000 validation sets. For the random forest validation, it has been found that when the database includes variables with many categories or when the data within the categories are not balanced, sampling without replacement is recommended, reducing selection bias by limiting tree correlations as well as the selection of split variables., (Strobl *et al.*, 2007). As the random forest in this research work includes an unbalanced outcome (cancer diagnosis), the cross-validation technique was used in the random forest data analysis splitting the dataset into training and validation sets. The R codes used for the analysis are available in Appendix I.

3.5.5 Assessing predictive ability performance

3.5.5.1 Development phase

Following the development of the models using logistic regression and random forest analysis in the training sets of the database, the assessment of the models' performance in the validation sets was assessed based on their predictive ability. This is based on the sensitivity (Se), specificity (Sp), negative predictive value (NPV), positive predictive value (PPV) and area under the receiver operating curve (AUC) of the models during the internal validation phase. The Se, Sp, NPV, and PPV were first calculated for each of the symptoms against the HNC diagnosis prior to the multivariate model development to assess the prediction that each individual symptom had for an appropriate HNC status diagnosis. This allowed for a direct comparison of the additional benefit gained when the multivariate model was used. These metrics are only possible to calculate for variables with two levels, creating a 2 x 2 table with each variable with the outcome of interest (cancer).

Their definitions based on each symptom of the dataset are presented below:

- The **sensitivity** is the proportion of true positive cancer patients (cancer=positive) that are positive for the symptom of interest (symptom=present).

- The **specificity** is the proportion of true negative cancer patients (cancer=negative) that are negative for the symptom of interest (symptom=absent).
- The **positive predictive value (PPV)** is the proportion of patients being positive for a symptom that are truly positive for cancer.
- The **negative predictive value (NPV)** is the proportion of patients that are negative for a symptom and are truly negative for cancer.
- The definition of **false positive probability** is not universal. It can be defined as the proportion of true negative for cancer patients that are positive for the symptom of interest, or it can be defined as the proportion of positive for symptom patients that are truly cancer negative. The former approach was used in calculating the false positive probability as it is used widely in logistic regression analysis.
- The definition of the **false negative probability** is, again, not universal. It is defined either as the proportion of true positives for cancer patients that are negative for the symptom of interest.

Following the calculation of the Se, Sp, PPV, and NPV per symptom and after the model development, the Receiver Operating Characteristic (ROC) curve was used to evaluate the predictive power of the selected binary outcome model. The ROC curve is a plot that shows the predictive ability of the binary classifier model (HNC calculator with yes/no cancer outcome) as its discriminatory threshold varies for the different combinations of sensitivity and specificity. On the y-axis of the ROC curve, there is the true positive rate (sensitivity) against the false positive rate (1-specificity) on the x-axis (Steyerberg, 2019). This plot is seen in Figure 3-3.

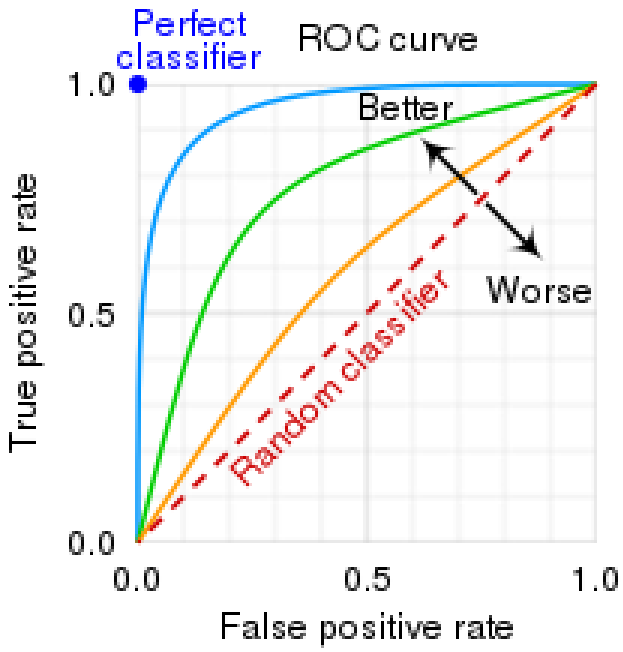


Figure 3-3. ROC curve and classification thresholds

The area under the curve calculation (AUC) provides a measure of discrimination that is the estimated probability, under the selected fitted model, that an individual with the explanatory characteristic of interests will have a higher probability of developing the outcome of interest compared to someone without these explanatory characteristics. Acceptable discrimination is considered for AUC between 0.7 and 0.8, whereas a value of more than 0.8 is considered excellent discrimination (Hosmer *et al.*, 2013).

The ROC allows the selection of optimum discrimination thresholds, which are commonly those that maximise the AUC in the development of cancer risk calculators' methodology (Steyerberg, 2019). Nevertheless, the optimum selection threshold is arbitrary depending on the scope of the triaging system that a calculator is used for and can be more in favour of sensitivity hence impacting and reducing the specificity or vice versa. A perfect classifier would have achieved fully maximising true positives and false positives. This means that a perfect classifier has the potential to identify 100% of the cancer cases whilst at the same time excluding cancer accurately in 100% of the non-cancer cases. This is considered impossible in an initial triage situation without the availability of highly specific biomarkers in the model variables (Steyerberg, 2019).

For the HNC risk model, the cut-point value of the predictive probability of the outcome of interest (y=1, i.e. the patient is cancer positive) was selected to be the predictive probability value that generates at the same time, the highest sensitivity and specificity values which are as well relatively close numerically to each other, that is located at the nearest top left corner of the ROC, as is the case in the available literature from other cancer risk calculators. This value is called the Youden index; it was calculated in the model of both the logistic regression and random forest model (Allisson, 2012).

In the logistic regression analysis, the first triaging cut-off was selected to be the probability value that generated the highest value combination of sensitivity and specificity simultaneously hence identifying the high-risk cases for potential HNC (to be used for USOC threshold). Following the exclusion of the high-risk group cases, a second cut-off threshold was also calculated using the same principle for the remaining cases, identifying another threshold that maximised the sensitivity and specificity that can hence discriminate between moderate-risk (for an urgent appointment – within 6 weeks) and low-risk cases allowing for a maximum discrimination potential (Chan, 2004). A misclassification matrix was generated based on the above classes measuring the percentage of observations that are placed on the wrong class – hence misclassified.

Chi-square analysis was also performed to compare the distribution of HNC diagnosis based on the current GP triaging and the one generated implementing the USOC and urgent thresholds on our data set. For the reclassification of referrals, the true incidence of cases was used; hence, the un-boosted cancer cases were used.

For the Random Forest model assessment, the same performance metrics were calculated for the Youden index value based on the ROC curve output alongside the misclassification matrix. The Gini index metric was also calculated for each of the variables being assessed for inclusion in the random forest model. The smaller the Gini index value, the more likely it is that the observation is coming from a given class only (Biau, Cérou and Guyader, 2010).

3.5.5.2 External validation phase

In the external validation phase, the misclassification matrix of the HNC-RC v.2 was calculated in the new cohort based on the previously defined thresholds of high-risk, moderate-risk, and low-risk groups, and it was compared to the misclassification matrix generated in the development phase. Following this, external validation was performed by applying the logistic regression algorithm in the new patient cohort and calculating the prediction power based on the negative and positive predictive power, sensitivity, specificity, and area under the receiver operating curve for the previously selected Youden index. As logistic regression was found to outperform random forest modelling during the development phase and it was the finally selected design model, random forest analysis was not re-evaluated. The output of the logistic regression at the external validation cohort was used to assess for differences in the parameter estimates in the new cohort of patients that may inform future iterations of the triage tool (Steyerberg, 2019).

3.6 Chapter Summary

In this chapter, the methodology underlying the design and validation process of the updated version of the HNC risk calculator was presented. The methodology is divided into two phases. During the development phase, data will be collected prospectively from patients seen in HaN clinics in the Greater Glasgow and Clyde region following sample size analysis. A pre-designed proforma will be utilised for the data collection. Assessment of the collected variable will be performed using univariate and multivariate regression analysis and compared with the results of the random forest AI approach. The goodness of fit of the models and the model diagnostics will be covered and compared, followed by cross-validation of the results with internal validation. The external validation is the second phase of the methodology of the HNC calculation design, collecting and analysing data from a separate prospective cohort of patients referred with suspected cancer symptoms in different hospitals across the UK. Logistic regression analysis, found to be the best-fit model technique, was used for the external validation of the calculator in this new prospective patient cohort.

4 Results

4.1 Chapter overview

This section presents the results of the development/internal validation and external validation of the refined version of the HNC risk calculator that is covered in sections 4.2 and 4.3, respectively. The presentation of the results starts with the exploration of the dependent variable (cancer outcome), followed by the exploration of the independent variables (symptoms, demographics) and then the results of the univariate and multivariate regression analysis and random forest analysis is covered. The results and predictive power of the random forest method were compared to those generated by logistic regression to decide on the finally selected model. The external validation phase results are then presented, starting with a univariate analysis of the validation cohort variables, followed by fitting the logistic regression model in the new dataset and calculating the resulting sensitivity, specificity and AUC. These metrics were then compared with the results of the development phase of the calculator.

4.2 HNC risk calculator development phase

As mentioned in the methodology chapter (section 3), the data collected during the first phase of the risk calculator development were prospectively collected from patients seen in HaN clinics in hospitals covering the Greater Glasgow and Clyde Region. A pre-designed proforma was used for the data collection process. All data were collected, transferred to an excel for analysis and were analysed by myself as part of the research work of this thesis. The sample size calculation showed that at least 3602 patients were needed, assuming a cancer prevalence of 8%, in order to achieve an 80% power and a sensitivity of 80% and specificity of 75% to within a +/- 5% range. To allow for any missing data entries, a total of 3,649 cases were finally collected.

4.2.1 Exploration of the dependent variable

The first step of the analysis involved the descriptive statistics of the dependent variable of this study. This variable is the patients' HNC status, y=cancer. This is a dichotomous categorical variable, being y=0 if a patient was found to be cancer-free following the patient's review in the clinic and completion of all the relevant assessments and investigations. Otherwise, if a patient was diagnosed with a HaN malignancy, then the value of the variable y will be 1. There were 309 (8.47%) confirmed cancer cases in the database out of a total of 3,649 cases. The HNC diagnosis included all primary cancers of the HaN regions (n = 236, 76.4%), unknown primaries to the HaN (n=10, 3.2%), metastatic cancers to the HaN from other regions, including lymphoma (n = 39, 12.6%) and cancers in neighbouring regions that manifested with HaN symptoms (n = 24, 7.8%). Table 4-1 shows the type and frequency of cancers included in the database.

The most common diagnosis was cancer of the oropharynx (n=82), followed by laryngeal malignancy (n=81), as seen in Table 4-1. At the time of the first clinical consultation, the majority of cancers were diagnosed at stage 3 or 4 (58.5%, n=128). Looking at HNCs (including the unknown primaries to the HaN), most of the cancers were also at an advanced stage at the time of diagnosis (55.7%), as seen in Table 4-2. Of the total of 3571 patients referred to the HaN clinics during the 18 months study period in the cancer un-boosted cohort, 1044 (29.2%) were referred to the USOC clinics. Only 61.2% (n=142) of the cancer cases were diagnosed through the USOC route. The cancer incidence during the data collection period, based on the un-boosted cancer cohort of 3571 patients, was 6.5% (n = 232). The distribution of cancer diagnosis per clinic type is seen in Figure 4-1.

Table 4-1. Cancer types and frequency that presented with head and neck signs and symptoms

Cancer Type	**Cancer Frequency (%)**
Oral cavity	13 (4.21%)
Oropharynx	82 (26.54%)
Hypopharynx	23 (7.44%)
Nasopharynx	1 (0.32%)
Nasal	3 (0.97%)
Larynx	81 (26.21%)
Parotid	5 (1.62%)
Other salivary glands	1 (0.32%)
Lung	10 (3.24%)
Bronchial	2 (0.65%)
Skin SCC/BCC	8 (2.59%)
Thyroid	18 (5.83%)
Unknown primary	11 (3.56%)
Lymphoma	34 (11%)
Oesophagus	11 (3.56%)
Gallbladder	1 (11%)
Metastatic breast	2 (0.65%)
Metastatic endometrial	1 (0.32%)
Metastatic melanoma	1 (0.32%)
Metastatic ovarian	1 (0.32%)
Total	307.8933

Table 4-2. Cancer stage at the time of diagnosis

Cancer Stage (ALL): n=309	HNC Staging: n=246
Stage 1: n=66 (21.4%)	Stage 1: n=60 (24.4%)
Stage 2: n=62 (20.1%)	Stage 2: n=49 (19.9%)
Stage 3: n=64 (20.7%)	Stage 3: n=53 (21.6%)
Stage 4: n=117 (37.8%)	Stage 4: n=84 (34.1%)

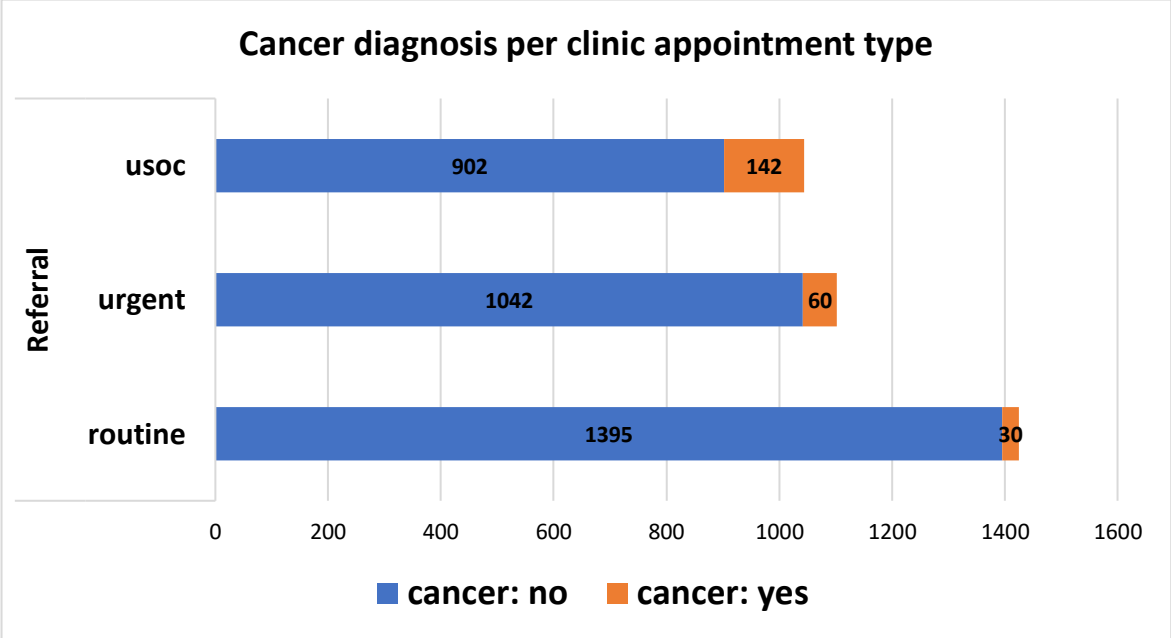


Figure 4-1. Cancer diagnosis per clinic appointment type across the cancer un-boosted database

4.2.2 Selection of independent variables

The independent variables include patients' demographics, social history factors and presenting signs and symptoms. All variables were categorical with between 2 and 5 levels, apart from age which is a continuous variable with integer numbers.

The demographic and social history (5 features) were:

- Biological Sex
- Age
- Smoking status (current smoker, ex-smoker, never smoker)
- Alcohol status (consumption of > 14 units of alcohol/week, consumes 14 or less units of alcohol per week, previous alcohol excess of > 14 units/week)
- Socio-economic status (in SIMD quintiles)

A total of 21 signs and symptoms were recorded in the database. All symptoms were present for 3 weeks or more at the point of the clinical consultation.

These were:

- Unintentional reported weight loss
- Hoarseness (persistent, intermittent, persistent following an explained cause: i.e., upper respiratory tract infection, stroke, previous surgery, intubation, voice use at work/hobby etc.).
- Sore throat (bilateral/central persistent, unilateral persistent, bilateral/central intermittent, unilateral intermittent)
- Throat discomfort or irritation
- Feeling of something (lump) in throat (FOSIT)
- Dysphagia (difficulty swallowing) - persistent, intermittent
- Regurgitation
- Odynophagia – that means pain on swallowing
- Neck pain (bilateral/central persistent, unilateral persistent, bilateral/central intermittent, unilateral intermittent)
- Neck lump (persistent/increasing, intermittent/fluctuating)
- Choking episodes/feeling
- Catarrh/mucus in throat
- Unilateral Blocked nose

- Oral swelling
- Oral ulcer
- Haemoptysis
- Unexplained unilateral otalgia with normal otoscopy
- facial pain or numbness
- Noisy breathing/stridor
- Red or white patches in the mouth
- Persistent HaN skin lesion

4.2.3 Exploration of the continuous independent variable – age

Patients' age was the single continuous independent variable in the dataset. The mean age was 57.2, with a standard deviation of 16.8. Data checking for the age variable revealed no wrong entries, with a range of values from 16 to 96 years of age. Outliers were checked via a boxplot graph of age distribution per cancer diagnosis (yes or no) that can be seen in Figure 4-2. No outliers were seen for the no cancer group, a few outliers were noted in the cancer groups, being outside the lower whisker with ages between 20 – 35 years of age (cases: 885; 2,593; 115; 1,930; 1,1015; 429).

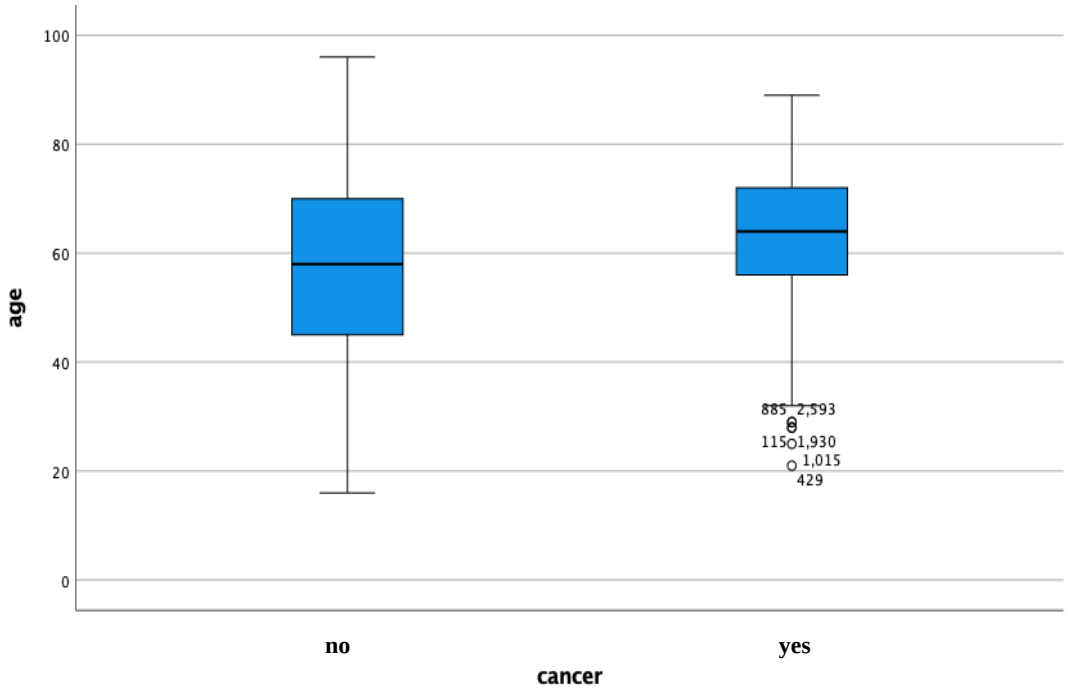


Figure 4-2. Boxplot of age versus cancer status

The normality plot for age showed a minor departure from the normality line across the whole cohort, as seen in Figure 4-3, with slight skewness to the right. The assumption of normality is not needed to proceed with the logistic regression and machine learning approaches for data analysis.

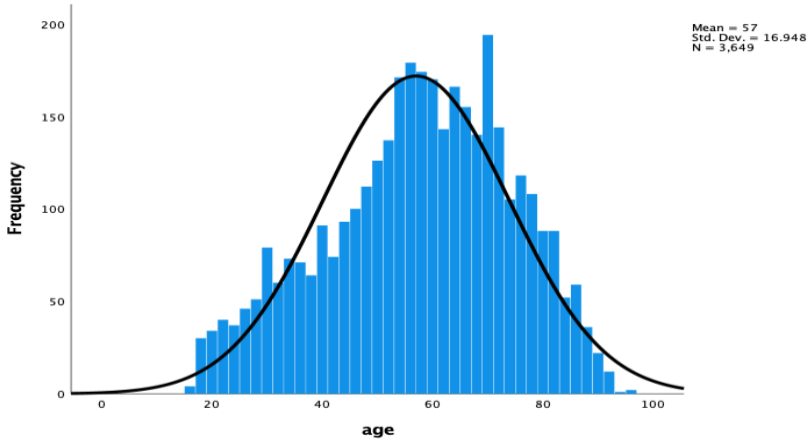


Figure 4-3. The normality plot for the age variable

When the normality plot was done against the cancer diagnosis, no distinct pattern was seen for the cancer cases which followed the normal distribution line (Figure 4-4). This pattern of distribution suggests that an attempt to convert the age variable to categorical may result in reducing the prediction ability of the multivariate modelling as no age cut-off was observed to differentiate cancer cases from the rest of the cohort. Moreover, it has been shown that converting a continuous variable to categorical can result in loss of information (Moons *et al.*, 2012b).

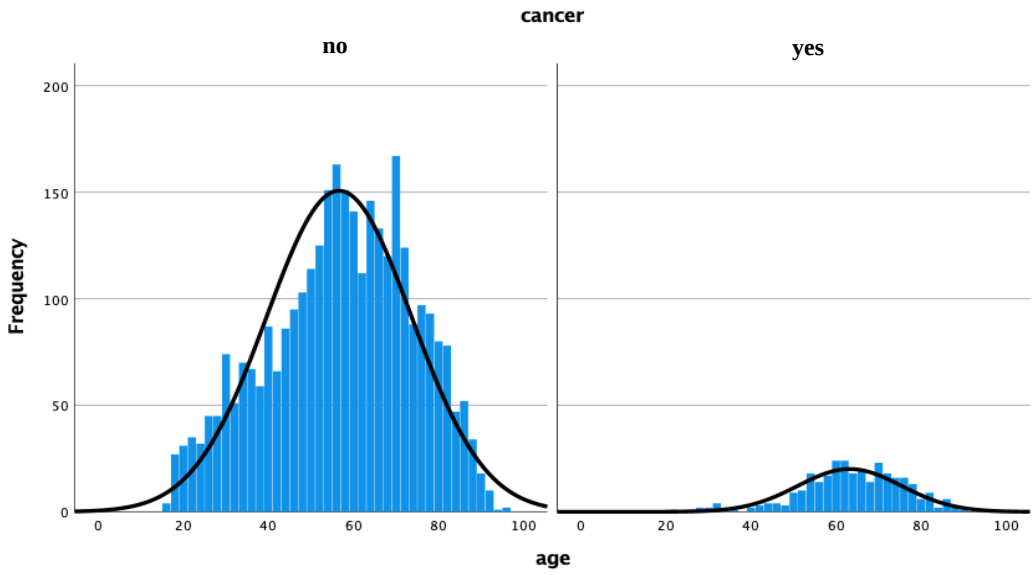


Figure 4-4. Normality plots for the age variable against a cancer diagnosis

4.2.4 Exploration of the categorical dependent variables

The independent variables list includes the patients' presenting symptoms and signs as well as biological sex, smoking and alcohol status. Examination of the categorical data will check for any unusual data values, any rare cases that may need to be excluded for computational reasons during model building, determine the frequency of the symptoms and check for any association between the variables. Table 4-3 summarises the frequency and percentages for the demographics and social history variables, and Table 4-4 includes all levels of the symptoms and signs categorical variables.

Table 4-3. Descriptive statistics for the demographics and social history factors in the total cohort of 3,649 patients and 309 cancer cases

Variable	Value	Frequency	Percentage
Age	Mean: 57 SD: 16.9 Range: 16 - 96	3649	100%
Biological Sex	Female Male	2143 1506	58.7% 41.3 %
Smoking Missing data: 10	No Yes Ex-smoker	1651 903 1085	45.4% 24.8% 29.8%
Alcohol Missing data:62	≤14 units/w >14 units/w Ex excess	3057 414 116	85.2% 11.5% 3.2%
SIMD Missing data: 75	1 (most deprived) 2 3 4 5 (least deprived)	1319 628 561 525 541	36.9% 17.6% 15.7% 14.7% 15.1%

The majority of patients were females (58.7%), and the mean age was 57 years, with a range from 16 to 96 years old. A quarter of the patients were current smokers (24.8%), and just over a quarter were ex-smokers (29.8%). Most of the cohort reported drinking alcohol in line with the UK government recommendations of equal to, or less

than, 14 units per week (85.2%). Of the rest, 11.5% were currently drinking in excess of 14 units per week, and 3.2% had a history of previous alcohol abuse. Looking at the socioeconomic deprivation of the referred cohort using SIMD quintiles, over a third (36.9%) was classed in SIMD 1, which is the most socioeconomically deprived region, with the rest of the quintiles being equally represented.

Table 4-4. Descriptive statistics for the categorical independent variables in the total cohort of 3,649 patients and 309 cancer cases

Variable	Value	Frequency	Percentage
Catarrh/mucous	No	3506	96.1%
	Yes	143	3.9%
Choking	No	3468	95%
	Yes	181	5%
Cough	No	3337	91.4%
	Yes	312	8.6%
Dysphagia	No	3140	86.1%
	Persistent	247	6.8%
	Intermittent	262	7.2%
Face pain/numbness	No	3621	99.2%
	Yes	28	0.8%
Feeling of something in throat	No	3048	83.5%
	Yes	601	16.5%
Haemoptysis	No	3571	97.9%
	Yes	78	2.1%
Hoarseness	No	2492	68.3%
	Persistent	388	10.6%
	Intermittent	681	18.7%
	Explained	88	2.4%
Neck Lump	No	2510	68.8%
	Persistent/ Increasing	851	23.3%
	Fluctuating/ Regressing	288	7.9%
Neck Pain	No	3465	95%
	Persistent bilateral	49	1.3%

Variable	Value	Frequency	Percentage
	Persistent unilateral	58	1.6%
	Intermittent bilateral	39	1.1%
	Intermittent unilateral	38	1%
Odynophagia	No	3568	97.8%
	Persistent	57	1.6%
	Intermittent	24	0.7%
Oral swelling	No	3472	95.1%
	Yes	177	4.9%
Oral ulcer	No	3623	99.3%
	Yes	26	0.7%
Otalgia	No	3550	97.3%
	Yes	99	2.7%
Persistent head and neck skin lesion	No	3610	98.9%
	Yes	39	1.1%
Red/White patches in mouth	No	3605	98.8%
	Yes	44	1.2%
Reflux	No	3457	94.7%
	Yes	192	5.3%
Regurgitation	No	3539	97%
	Yes	110	3%
Shortness of breath	No	3605	98.8%
	Yes	44	1.2%
Sore Throat	No	3171	86.9%
	Persistent bilateral	160	4.4%
	Persistent unilateral	48	1.3%
	Intermittent bilateral	247	6.8%
	Intermittent unilateral	23	0.6%
Stridor	No	3637	99.7%
	Yes	12	0.3%
Throat Clearing	No	3530	96.7%
	Yes	119	3.3%

Variable	Value	Frequency	Percentage
Throat discomfort/irritation	No	3380	92.6%
	Yes	269	7.4%
Unilateral Blocked Nose	No	3639	99.7%
	Yes	10	0.3%
Unintentional weight loss	No	3306	92.3%
Missing data: 66	Yes	277	7.7 %

The patients presented with a wide variety of symptoms which are summarised in alphabetical order in Table 4-4. The most common presenting symptom was hoarseness (31.7%), followed by neck lump (31.2%). Of the patients with hoarseness (n=1,157), the majority complained of intermittent hoarseness, with fluctuation during the day and periods of return to normal quality voice (n=681). A small number of patients had hoarseness following a recent causative event or an associated factor linked to hoarseness (n=88). These factors included previous neck surgery, stroke, diagnosis of other neurological conditions, voice overuse at work or related to hobbies or recent upper/lower respiratory tract infection. Of the patients with a neck lump (n=1,139), most had a persistent lump with reports of an increase in size or being stable since first noticed (n=851), whereas, for the remaining n=288, the lump was regressing or fluctuating in size.

4.2.5 Univariate analysis

For 5 cases, the cancer status remained unknown at the end of the data collection period, and these cases were removed from any subsequent classification analysis relating to the independent variables in the dataset, which were based on a total of 3,644 cases. Each of the variables in the database was screened to assess for any variable or level within a variable that was redundant. This process identified a non-informative level within the neck pain variable, where there was no patient with cancer when the variable level of bilateral intermittent neck pain was checked. Therefore, this subgroup was merged with the no pain level for the model build process. Similarly, within the odynophagia variable, none of the cancer patients presented with intermittent odynophagia. Thus, this group was merged with the no odynophagia group, therefore having two final groups for odynophagia: Persistent vs

No/Intermittent. The facial pain and facial numbness symptoms, as well as the blocked nose symptom, were non-informative symptoms, as no patient in the cancer group presented with these symptoms. Hence they will be excluded from the univariate and subsequent multivariate regression analysis.

Initially, a univariate analysis was performed for each variable in the whole dataset of 3,644 patients, prior to missing entries deletion, alongside calculation of the odds ratio, the 95% confidence interval, and the area under the ROC for each feature against the cancer diagnosis. This is seen in Table 4-5. Following the deletion of the missing entries, the univariate analysis was repeated for the clean database of 3,531 patients to ensure that no significant alteration in the final results would be encountered due to the deletion of the missing data. The analysis based on the clean database (n=3531) is available in Table 4-6 and Table 4-7 for the demographics/social history factors and the symptoms/signs features, respectively. No change in the p-value significance was seen following the deletion of missing data cases; hence the results were not significantly affected by missingness. In total, 118 entries were removed from the database due to missing data. These cases were excluded from the multivariate analysis. Figure 4-5 shows a flowchart of the cases excluded as each step leading to the final cohort that was used in the multivariate model.

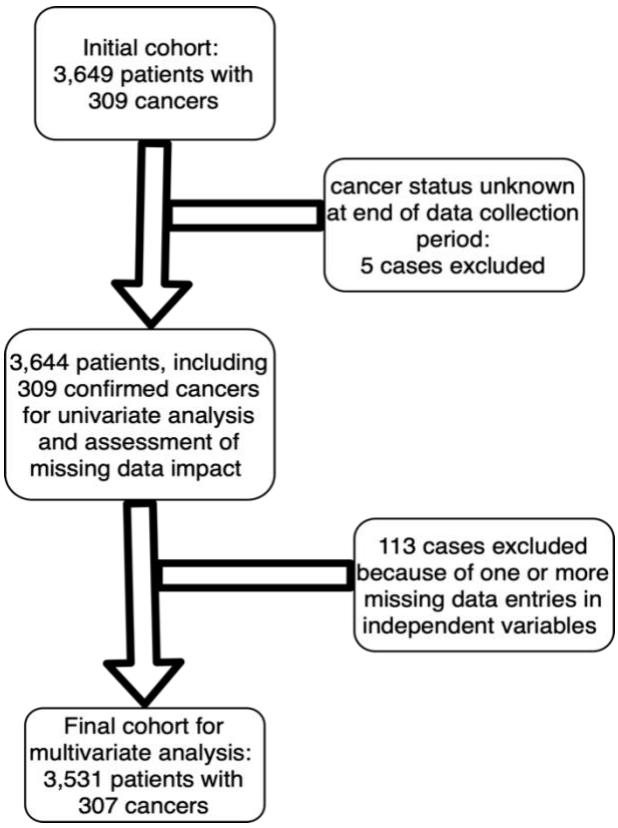


Figure 4-5. Flowchart of cases excluded during univariate analysis for multivariate cohort preparation

On univariate analysis, the demographic and social history features that were significantly associated with a cancer diagnosis at the 0.05 level of significance were age, biological sex, SIMD, smoking and alcohol. Even though over half of the cohort were females, a cancer diagnosis was 3.3 times more commonly diagnosed in males. Current smokers were 3.8 times more likely to have HNC compared to non-smokers. The odds ratio was lower for ex-smokers, having a 1.7 times higher risk of cancer than non-smokers. The difference in cancer diagnosis likelihood was more pronounced for those drinking in excess, with a 3.8 odds ratio for current and 5.4 odds ratio for previous alcohol abuse compared to those drinking less than the recommended weekly limit.

Table 4-5. Univariate analysis for all variables in the cohort of 3,644 patients and 309 cancers, prior to deletion of missing data entries from the independent variables

Variable		Total	Benign N=3335	Cancer N=309	P value
Biological Sex	Female Male	2140 (58.7%) 1504 (41.3%)	2039 (61.1%) 1296 (38.9%)	101 (32.7%) 208 (67.3%)	<0.0001
OR Male vs Female: 3.335 (2.599 – 4.281), AUC 0.6423					
Age	Median IQR	58 24	58 25	64 16	<0.001
OR 1.026 (1.018 – 1.033), AUC=0.6176					
SIMD **Missing data: 75**	1 (most deprived) 2 3 4 5 (least deprived)	1318 (36.9%) 626 (17.5%) 561 (15.7%) 524 (14.7%) 540 (15.1%)	1180 (36.1%) 571 (17.5%) 528 (16.2%) 482 (14.7%) 507 (15.5%)	138 (45.8%) 55(18.3%) 33 (11%) 42 (14%) 33 (11%)	0.003 [a]
Referral type	Routine Urgent USOC	1439 (39.5%) 1118 (30.7%) 1087 (29.8%)	1394 (41.8%) 1040 (31.2%) 901 (27%)	45 (14.6%) 78 (25.2%) 186 (60.2%)	<0.001 [a]
Smoking Missing data: 10	No Yes Ex	1649 (45.4%) 903 (24.8%) 1082 (29.8%)	1572 (47.3%) 758 (22.8%) 995 (29.9%)	77 (24.9%) 145 (46.9%) 87 (28.2%)	<0.001
OR Ex-smoker vs No: 1.739 (1.264 – 2.392) **Yes vs No: 3.840 (2.873 – 5.132)** **AUC: 0.6499**					
Alcohol Missing data: 62	≤14 units >14 units Ex excess	3052 (85.2%) 414 (11.6%) 116 (3.2%)	2859 (87.4%) 329 (10.1%) 85 (2.6%)	193 (62.5%) 85 (27.5%) 31 (10%)	<0.001
OR >14 units vs ≤14 units: 3.811 (2.881 – 5.042) **Ex excess vs ≤14 units: 5.364 (3.444 – 8.354)** **AUC 0.6259**					

Variable		Total	Benign N=3335	Cancer N=309	P value
Unintentional Weight loss Missing data: 66	No Yes	3301 (92.3%) 277 (7.7%)	3065 (93.7%) 206 (6.3%)	236 (76.9%) 71 (23.1%)	<0.001
OR Yes vs No: 4.431 (3.280 – 5.985), AUC 0.5841					
Hoarseness	Persistent Intermittent Explained No	387 (10.6%) 681 (18.7%) 88 (2.4%) 2488 (68.3%)	312 (9.4) 668 (20%) 85 (2.5%) 2270 (68.1%)	75 (24.3%) 13 (4.2%) 3(1%) 218 (70.6%)	<0.0001
OR Explained vs No: 0.385 (0.121 – 1.23) **Intermittent vs No: 0.201 (0.114 – 0.355)** **Persistent vs No: 2.544 (1.905 – 3.397)** **AUC 0.637**					
Regurgitation	No Yes	3534 (97%) 110 (3%)	3233 (96.9%) 102 (3.1%)	301 (97.4%) 8 (2.6%)	0.6503
OR Yes vs No: 0.845 (0.407 – 1.753)					
Dry/Tickly Cough	No Yes	3332 (91.4%) 312 (8.6%)	3035 (91%) 300 (9%)	297 (96.1%) 12 (3.9%)	0.0032
OR Yes vs No: 0.412 (0.228 – 0.742)					
Sore throat	Persistent bilateral Persistent unilateral Intermittent bilateral Intermittent unilateral No	160(4.4%) 48 (1.3%) 247 (6.8%) 23 (0.6%) 3166 (86.9%)	128 (3.8%) 24 (0.7%) 244 (7.3%) 22 (0.7%) 2917 (87.5%)	32 (10.4%) 24 (7.8%) 3 (1%) 1 (0.3%) 249 (80.6%)	<0.0001
OR Persistent bilateral vs No: 3.025 (2.006 – 4.561 **Persistent unilateral vs No: 13.067 (7.172 – 23.807)** **Intermittent bilateral vs No: 0.146 (0.046 – 0.459)** **Intermittent unilateral vs No: 0.520 (0.07 – 3.873)** **AUC: 0.5955**					
Neck pain					

Variable		Total	Benign N=3335	Cancer N=309	P value
Persistent bilateral		49 (1.3%)	41 (1.2%)	8 (2.6%)	
Persistent unilateral		58 (1.6%)	52 (1.6%)	6 (1.9%)	0.3642
Intermittent bilateral		39 (1.1%)	39 (1.2%)	0	
Intermittent unilateral		38 (1%)	36 (1.1%)	2 (0.6%)	
No		3460 (95%)	3167 (95%)	293 (94.8%)	
***Intermittent bilateral - non-informative category. Merged with no symptoms category ***					
Persistent bilateral		49 (1.3%)	41 (1.2%)	8 (2.6%)	
Persistent unilateral		58 (1.6%)	52 (1.6%)	6 (1.9%)	0.217
Intermittent unilateral		38 (1%)	36 (1.1%)	2 (0.6%)	
No /int bilat		3499 (96%)	3206 (96.1%)	293 (94.8%)	
Neck pain 4 categories:					
OR Persistent bilat vs No: 2.129 (0.987 – 4.592)					
Persistent unilateral vs No: 1.252 (0.533 0 2.943)					
Intermittent unilateral vs No: 0.608 (0.146 – 2.542)					
AUC: 0.5108					
Throat discomfort/ Irritation	No	3375 (92.6%)	3082 (92.4%)	293 (94.8%)	0.1423
	Yes	269 (7.4%)	253 (7.6%)	16 (5.2%)	
Yes vs No: 0.878 (0.403 – 1.140)					
Feeling of something/ Lump in throat	No	3044 (83.5%)	2744 (82.3%)	300 (97.1%)	<0.0001
	Yes	600 (16.5%)	591 (17.7%)	9 (2.9%)	
Yes vs No: 0.138 (0.071 – 0.270), 0.574					
Dysphagia	**Persistent**	247 (6.8%)	177 (5.3%)	70 (22.7%)	<0.0001
	Intermittent	262 (7.2%)	258 (7.7%)	4 (1.3%)	
	No	3135 (86%)	2900 (87%)	235 (76.1%)	
OR Intermittent vs No: 0.190 (0.07 – 0.514)					
Persistent vs No: 4.771 (3.508 – 6.489)					
AUC: 0.6105					

Variable		Total	Benign N=3335	Cancer N=309	P value
Odynophagia					<0.0001
	Persistent	57 (1.6%)	19 (0.6%)	38 (12.3%)	
	Intermittent	24 (0.7%)	24 (0.7%)	0	
	No	3563 (97.8%)	3292 (98.7%)	271 (87.7%)	
***Intermittent odynophagia - non-informative, merged with no symptoms category ***					
	Yes	57(1.6%)	19 (0.6 %)	38 (12.3%)	<0.0001
	No	3587 (98.4%)	3316 (99.4%)	271 (87.7%)	
Odynophagia 2 categories					
Persistent vs No/Intermittent: 10.967 (6.933 – 17.347), AUC: 0.556					
Choking episodes	No	3463 (95%)	3162 (94.8%)	301 (97.4%)	0.0497
	Yes	181 (5%)	173 (5.2%)	8 (2.6%)	
OR Yes vs No: 0.487 (0.237 – 0.999)					
Catarrh, mucous excess	No	3501 (96.1%)	3193 (95.7%)	308 (99.7%)	0.0099
	Yes	143 (3.9%)	142 (4.3%)	1 (0.3%)	
OR Yes vs No: 0.075 (0.010 – 0.537)					
Throat Clearing	**No**	3525 (96.7%)	3218 (96.5%)	307 (99.4%)	0.0156
	Yes	119 (3.3%)	117 (3.5%)	2 (0.6%)	
OR Yes vs No: 0.177 (0.044 – 0.721)					
Reflux	**No**	3452 (94.7%)	3147 (94.4%)	305 (98.7%)	0.0030
	Yes	192 (5.3%)	188 (5.6%)	4 (1.3%)	
OR Yes vs No: 0.221 (0.081 – 0.598)					
Neck lump	**Fluctuating/ Reducing**	287 (7.9%)	279 (8.4%)	8 (2.6%)	<0.0001
	Persistent	849 (23.3)	703 (21.1%)	146 (47.2%)	
	No	2508 (68.8%)	2353 (70.6%)	155 (50.2%)	
OR Fluctuating vs No: 0.435 (0.211 – 0.894)					
Persistent vs No: 3.101 (2.433 – 3.952)					

Variable		Total	Benign N=3335	Cancer N=309	P value
AUC: 0.6427					
Persistent Oral swelling/ growth	No Yes	3467 (95.1%) 177 (4.9%)	3214 (96.4%) 121 (3.6%)	253 (81.9%) 56 (18.1%)	<0.0001
OR Yes vs No: 6.220 (4.414 – 8.766), AUC 0.5725					
Persistent Oral ulcer	No Yes	3618 (99.3%) 26 (0.7%)	3321 (99.6%) 14 (0.4%)	297 (96.1%) 12 (3.9%)	<0.0001
OR Yes vs No: 10.040 (4.540 – 22.203), AUC: 0.5173					
Red/ White oral patch	No Yes	3600 (98.8%) 44 (1.2%)	3293 (98.7%) 42 (1.3%)	307 (99.4%) 2 (0.6%)	0.3528
OR Yes vs No: 0.509 (0.123 – 2.115), AUC: 0.5031					
Haemoptysis	No Yes	3566 (97.9%) 78 (2.1%)	3268 (98%) 67 (2%)	298 (96.4%) 11 (3.6%)	0.0747
OR Yes vs No: 1.806 (0.943 – 3.460)					
Shortness of Breath	No Yes	3600 (98.8%) 44 (1.2%)	3297 (98.9%) 38 (1.1%)	303 (98.1%) 6 (1.9%)	0.222
OR Yes vs No: 0.582 (0.244 – 1.388)					
Unilateral unexplained otalgia	No Yes	3545 (97.3%) 99 (2.7%)	3264 (97.9%) 71 (2.1%)	281 (90.9%) 28 (9.1%)	<0.0001
OR Yes vs No: 4.659 (2.951 – 7.357), AUC: 0.5347					
Stridor	No Yes	3632 (99.7%) 12 (0.3%)	3329 (99.8%) 6 (0.2%)	303 (98.1%) 6 (1.9%)	<0.0001
OR Yes vs No: 10.691 (3.427 – 33.353), AUC: 0.5088					
Persistent Head and Neck skin lesion	No Yes	3605 (98.9%) 39 (1.1%)	3304 (99.1%) 31 (0.9%)	301 (97.4%) 8 (2.6%)	0.0092
OR Yes vs No: 2.851 (1.296 – 6.274), AUC: 0.5083					

Note: *AUC: area under the curve, OR: odds ratio, [a]: Fisher's exact test.*

The persistent unilateral sore throat symptom had the highest odds ratio for HNC diagnosis, with 13 times more chance to have cancer with this symptom compared to an individual without a sore throat. The second-highest odds ratio was when persistent odynophagia was present, with an OR of 10.9, followed by an OR of 10 for oral ulcer and 6.2 for the presence of oral swelling. Having stridor was highly associated with HNC, with an odds ratio of 10.7, but it was a rare symptom manifesting in 6 patients with cancer. Persistent dysphagia also had a strong association with a cancer diagnosis, having an odds ratio of 4.7, as well as the report of unexplained otalgia and unintentional weight loss with a 4.7 and 4.4 odds ratio, respectively. A strong but negative association with cancer was found for the feeling of something in throat symptom with an OR of 0.13 ($p<0.0001$). A significant p-value with a negative HNC association was also found for the following symptoms: dry tickly cough complaints, choking episodes, catarrh and mucous secretions, reflux and throat clearing. Other symptoms did not show any significant positive or negative association with HNC diagnosis; these were regurgitation ($p=0.65$), neck pain ($p=0.36$) and throat discomfort or irritation ($p=0.14$).

Following univariate analysis, clearance of missing data entries and assessment of the impact of missingness, the cohort that will be used for the multivariate model analysis was finalised, comprised of a total of 3,531 patients, which includes 307 cancers. The p-values of each of the variables against the cancer diagnosis are seen in Table 4-6 and Table 4-7. These tables include the results of the repeat univariate analysis after the data exclusion process was complete. A cancer diagnosis was linked to the male biological sex, with 14.3% of males having cancer compared to 4.8% of females ($p=0.0001$). Older age was also a contributing factor, with a mean age of 63.7 years in the cancer groups versus 57 years in the cancer-free cohort ($p=0.0001$). Current and previous smoking status was also associated with higher cancer percentages (16.3% and 8.1%, respectively) which were statistically significant ($p=0.0001$). Similar findings were seen for those that were consuming alcohol in excess currently (20.6% cancer diagnosis) or previously (26.8% cancer incidence), $p=0.001$. The most deprived population (SIMD=1) was at the highest risk of HNC diagnosis, with an incidence of 10.7% ($p=0.005$). In this repeat univariate analysis of the finally selected cohort, the

same variables remained significant and will be included for consideration in the multivariate model.

Table 4-6. Univariate analysis of patients' demographics, smoking and alcohol as risks factors for head and neck cancer in the clean database of 307 cancers in a total cohort of 3531 patients

			Head and Neck Cancer		P-value
			Yes (N=307)	No (N=3224)	
Biological Sex	Males Females		208 (14.3%) 99 (4.8%)	1246 (85.7%) 1978 (95.2%)	0.0001
Age	Mean (SD)		63.7 (9.1)	57 (16.9)	0.0001
Smoking	Current Ex Never		145 (16.3%) 85 (8.1%) 77 (4.8%)	744 (83.7%) 963 (91.9%) 1517 (95.2%)	0.0001
Alcohol	>14 units/week Previous excess <=14 units/week		85 (20.6%) 30 (26.8%) 192 (6.4%)	327 (79.4%) 82 (73.2%) 2815 (93.6%)	0.001
Socio-economic status (SIMD quintile)		1 2 3 4 5	137 (10.7%) 55 (9.1%) 33 (6%) 41 (8.1%) 33 (6.4%)	1144 (89.3%) 550 (90.9%) 517 (94%) 468 (91.9%) 479 (93.6%)	0.005

Table 4-7. Univariate analysis of patients' presenting signs and symptoms for cancer diagnosis in the clean database

Variable		**Head and Neck Cancer**		**P-value**
		Yes	**No**	
Unintentional weight loss	Yes No	71 (25.7%) 236 (7.3%)	205 (74.3%) 3019 (92.7%)	0.0001
Hoarseness	Persistent Intermittent	75 (20.1%) 13 (1.9%)	299 (79.9%) 655 (98.1%)	0.0001

Variable		Head and Neck Cancer		P-value
		Yes	No	
	Persistent after URTI/stroke/surgery No	3 (3.7%) 216 (9%)	79 (96.3%) 2191 (91%)	
Sore Throat	Persistent Bilateral Persistent Unilateral Intermittent Bilateral Intermittent Unilateral No	32 (20.9%) 24 (53.3%) 3 (1.3%) 1 (4.3%) 247 (8%)	121 (79.1%) 21 (46.7%) 98.7%) 22 (95.7%) 2825 (92%)	0.0001
Throat Discomfort/ Irritation	Yes No	16 (6.2%) 291 (8.9%)	242 (93.8%) 2982 (91.1%)	0.142
Feeling of something in throat (FOSIT)	Yes No	9 (1.5%) 298 (10.1%)	578 (98.5%) 2646 (89.9%)	0.0001
Dysphagia	Persistent Intermittent No	70 (28.5%) 4 (1.6%) 233 (7.7%)	176 (71.5%) 253 (98.4%) 2795 (92.3%)	0.0001
Regurgitation	Yes No	8 (7.5%) 299 (8.7%)	99 (92.5%) 3125 (91.3%)	0.650
Odynophagia	Yes No	38 (30.9%) 269 (7.7%)	17 (69.1%) 3207 (92.3%)	0.001
Neck pain	Persistent bilateral Persistent unilateral Intermittent bilateral Intermittent unilateral No	8 (16.7%) 6 (10.5%) 0 2 (5.4%) 291 (8.7%)	40 (83.3%) 51 (89.5%) 39 (100%) 35 (94.6%)	0.364

Variable		Head and Neck Cancer		P-value
		Yes	No	
			3059 (91.3%)	
Neck lump	Persistent Intermittent/Regressing No	144 (17.5%) 8 (2.9%) 155 (6.4%)	681 (82.5%) 270 (97.1%) 2273 (93.6%)	0.0001
Choking episodes/ Feeling	Yes No	8 (4.5%) 299 (8.9%)	168 (95.5%) 3056 (91.1%)	0.05
Shortness of breath	Yes No	6 (14%) 301 (8.6%)	37 (86%) 3187(91.4%)	0.218
Catarrh/ mucus	Yes No	1 (0.7%) 306 (9%)	135 (99.3%) 3089 (91%)	0.01
Blocked nose	Unilateral Bilateral No	0 0 307 (8.8%)	8 (100%) 22 (100%) 3194 (91.3%)	0.999
Oral swelling	Yes No	55 (32.5%) 252 (7.5%)	114 (67.5%) 3110 (92.5%)	0.0001
Oral ulcer	Yes No	12 (48%) 295 (8.4%)	13 (52%) 2311 (91.6%)	0.0001
Haemoptysis	Yes No	11 (14.5%) 296 (8.6%)	65 (85.5%) 3159 (91.4%)	0.075
Unexplained unilateral otalgia	Yes No	28 (29.2%) 279 (8.1%)	68 (70.8%)	0.0001

Variable		Head and Neck Cancer		P-value
		Yes	No	
(normal otoscopy)			3156 (91.9%)	
Face pain/ Numbness	Yes No	0 307 (8.7%)	27 (100%) 3197 (91.2%)	0.998
Noisy breathing/ Stridor	Yes No	6 (50%) 301 (8.6%)	6 (50%) 3218 (91.4%)	0.0001
Red/ White patch in mouth	Yes No	2 (4.7%) 305 (8.7%)	41 (95.3%) 3183 (91.3%)	0.353
Persistent Head and Neck Skin lesion	Yes No	8 (21.1%) 299 (8.6%)	30 (78.9%) 3194 (91.4%)	0.009

The symptoms and signs that were significantly associated with a HNC diagnosis were unintentional weight loss, hoarseness, sore throat, feeling of something in the throat (negative association), dysphagia, odynophagia, neck lump, oral swelling, oral ulcer, unexplained otalgia, catarrh/mucous (negative association), stridor and a persistent skin lesion. These variables were significant at the 0.05 level. Additionally, choking episodes (negative association) and haemoptysis were significant at the 0.1 level. PPV, NPV, sensitivity and specificity were also calculated for the symptoms with two levels (two categories). The odynophagia symptom had the highest PPV for HNC diagnosis, being 48.1%, as well as oral ulcer, having the same value. Stridor also had a high PPV of 50%, followed by oral swelling (PPV: 32.5%), otalgia (PPV 29.2%), unintentional weight loss (PPV: 25.7%) and skin lesion (PPV: 21.1%). Even symptoms with no association or negative association with HNC had PPV over 3%. The feeling of something in the throat and the throat-clearing symptom had the smallest PPVs being 1.5% and 1.7%, respectively.

Table 4-8. Sensitivity, Specificity, and other statistics for all 2-level symptoms variables in the clean database of 307 cancer and a total of 3.531 patients

Variable	Sensitivity	Specificity	False Positive Probability	False Negative Probability	Positive Predictive Value	Negative Predictive Value
Catarrh/ mucous	0.3%	95.8%	4.2%	99.7%	0.7%	91%
Choking	2.6%	94.8%	5.2%	97.4%	4.5%	91.1%
Cough	3.9%	91%	9%	96.1%	4%	90.9%
Face pain/ numbness	0%	99.2%	0.8%	100%	0	91.2%
Feeling of something in throat	2.9%	82.1%	17.9%	97.1%	1.5%	89.9%
Haemoptysis	3.6%	98%	2%	96.4%	14.5%	91.4%
Odynophagia	12.4%	98.7%	1.3%	87.6%	48.1%	92.2%
Oral swelling	17.9%	96.5%	3.5%	82.1%	32.5%	92.5%
Oral ulcer	3.9%	99.6%	0.4%	96.1%	48%	91.6%
Otalgia	9.1%	97.9%	2.1%	90.9%	29.2%	91.9%
Persistent head and neck skin lesion	2.6%	99.1%	0.9%	97.4%	21.1%	91.4%
Red/ White patches in mouth	0.7%	98.7%	1.3%	99.3%	4.7%	91.3%
Reflux	1.3%	94.4%	5.6%	98.7%	2.2%	90.9%
Regurgitation	2.6%	96.9%	3.1%	97.4%	7.5%	91.3%
Shortness of breath	2%	98.9%	1.1%	98%	14%	91.4%
Stridor	2%	99.8%	0.2%	98%	50%	91.4%
Throat Clearing	0.7%	96.4%	3.6%	99.3%	1.7%	91.1%
Throat discomfort/ irritation	5.2%	92.5%	7.5%	94.8%	6.2%	91.1%

Variable	Sensitivity	Specificity	False Positive Probability	False Negative Probability	Positive Predictive Value	Negative Predictive Value
Unilateral Blocked Nose	0%	99.1%	0.9%	100%	0	91.2%
Unintentional weight loss	23.1%	93.6%	6.4%	76.9%	25.7%	92.7%

The cancer patients presented with various combinations of the presenting symptoms, with no single symptom being found to have a very high sensitivity. The odynophagia, oral swelling and weight loss symptoms were the only ones with double-digit sensitivities, being 12.4%, 17.9% and 23.1%, respectively. The PPV is expected to be low for the majority of symptoms due to the small number of events compared to the size of the cohort. Similarly, the NPV will be high as most of the cohort was cancer-free.

4.2.6 Multivariate Logistic Regression Analysis

4.2.6.1 The full 2-way interactions model

The multivariate analysis started by initially including in the regression all informative variables identified from univariate analysis (n=27), as seen in Table 4-9, as well as any significant interactions identified using the Bonferroni method. The SMID variable was not included in the multivariate analysis to allow its generalised application outside the Scottish population. A total of 136 interactions were tested for inclusion. The Bonferroni method threshold was set at a=0.1/136 = 0.000735. Only one interaction met the threshold for potential inclusion in the final model. This was the smoking with sore throat interaction, p=0.0002225. Hence, the logistic regression initially included all main effects and the one significant interaction of smoking with a sore throat to allow assessment of the p-values and odds ratios of all variables as in Table 4-9 and Table 4-10. A backward elimination process of all non-significant variables at the 0.05 level will follow.

Table 4-9. Logistic regression analysis including all potential main effects and the one significant interaction term

Parameter	Comparison value vs Reference value	Estimate	Standard Error	P-value
Intercept		-4.8580	0.6059	<.0001
Age		0.0278	0.00529	<.0001
Biological Sex	**Male vs female**	1.0787	0.1658	<.0001
Unintentional weight loss	**Yes vs no**	0.7852	0.2309	0.0007
Smoking	**Ex-smoker vs no**	0.3124	0.1978	0.1143
Smoking	**Yes vs no**	0.4261	0.1991	0.0323
Alcohol	**>14 units vs ≤14 units**	0.6899	0.1998	0.0006
Alcohol	**ex excess vs ≤14 units**	0.5592	0.3183	0.0789
Hoarseness	**Explained vs no**	0.4435	0.6760	0.5118
Hoarseness	**Intermittent vs no**	-0.1919	0.3434	0.5762
Hoarseness	**Persistent vs no**	1.7801	0.2327	<.0001
Regurgitation	**Yes vs no**	-0.2200	0.4817	0.6479
Cough	**Yes vs no**	-0.1699	0.3811	0.6557
Sore throat	**Persistent Bilateral vs no**	-13.1972	527.5	0.9800
Sore throat	**Persistent Unilateral vs no**	0.4963	1.1023	0.6525
Sore throat	**Intermittent bilateral vs no**	-1.1216	1.0457	0.2835
Sore throat	**Intermittent unilateral vs no**	1.3455	1.3983	0.3359
Neck pain	**intermittent unilateral vs persistent unilateral**	-0.1733	0.6388	0.7862
Neck pain	**no / intermittent bilateral vs persistent unilateral**	0.1251	0.2896	0.6657
Neck pain	**Persistent bilateral vs persistent unilateral**	0.4815	0.4581	0.2932
Throat discomfort	**Yes vs no**	0.1645	0.3379	0.6265
Feeling of something in throat	**Yes vs no**	-1.1385	0.4096	0.0054
Dysphagia	**Intermittent vs no**	-1.3540	0.6162	0.0280
Dysphagia	**Persistent vs no**	1.2548	0.2621	<.0001

odynophagia	no/ intermittent vs persistent	-1.3977	0.2167	<.0001
Choking	Yes vs no	-0.3378	0.4788	0.4806
Catarrh/mucus	Yes vs no	-1.1191	1.0593	0.2908
Neck lump	Persistent vs no	2.3663	0.2245	<.0001
Neck lump	Intermittent/fluctuating vs no	0.4960	0.4266	0.2450
Oral swelling	Yes vs no	2.2146	0.2753	<.0001
Oral ulcer	Yes vs no	1.6995	0.6359	0.0075
haemoptysis	Yes vs no	0.3905	0.4948	0.4299
Otalgia	Yes vs no	1.3343	0.3792	0.0004
Reflux	Yes vs no	-0.6915	0.5674	0.2229
Throat Clearing	Yes vs no	-0.5930	0.8020	0.4597
Stridor	Yes vs no	2.1835	0.8947	0.0147
Shortness of breath	No vs yes	-0.5683	0.2783	0.0411
Red/white patch In mouth	Yes vs no	-1.7051	0.9895	0.0849
Head&neck skin lesion	Yes vs no	2.1062	0.4854	<.0001
smoking*sore throat	Ex-smoker and persistent bilateral sore throat vs no	13.9708	527.5	0.9789
smoking*sore throat	Ex-smoker and persistent unilateral sore throat vs no	0.2893	1.6739	0.8628
smoking*sore throat	Ex-smoker and intermittent bilateral sore throat vs no	-12.4707	659.8	0.9849
smoking*sore throat	Ex-smoker and intermittent unilateral sore throat vs no	-15.6984	1773.3	0.9929
smoking*sore throat	smoker and persistent bilateral sore throat vs no	14.4488	527.5	0.9781
smoking*sore throat	smoker and persistent unilateral sore throat vs no	2.8573	1.2755	0.0251
smoking*sore throat	smoker and intermittent bilateral sore throat vs no	0.3195	1.3195	0.8087

smoking*sore throat	**smoker and intermittent unilateral sore throat vs no**	-14.8143	2232.1	0.9947

Prior to this elimination process, it can be seen that the age and biological sex variables and unintentional weight loss variables remained highly significant in the multivariate regression analysis (p<0.0001). The smoking variable is significant in the current vs no smoker level (p=0.0323) but did not reach a significance in the ex-smoker vs never smoker level (p=0.1143). Similarly, alcohol consumption over 14 units vs less than 14 units was a significant factor (p=0.0006), but the previous excess vs <=14 units per week reached significance at the 0.1 level but not the 0.05 level. The hoarseness symptom was significant when persistent hoarseness was compared against no hoarseness (p=<0.0001) with an odds ratio of 5.9. Both levels of the dysphagia symptoms and the odynophagia symptom also remained significant on multivariate analysis with high odds ratio estimates. The other variables that were significant included: oral swelling, oral ulcer, otalgia, feeling of something in the throat, neck lump at the persistent vs no lump level, stridor, shortness of breath and HaN skin lesion. Only one level of the interaction term sore throat with smoking was significant in the multivariate model, and that was the smoker with persistent unilateral sore throat vs non-smoker with no sore throat effect (p=0.0251).

Table 4-10. Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Age	1.028	1.018	1.039
Biological Sex: male vs female	2.941	2.125	4.070
Unintentional weight loss: yes vs no	2.193	1.395	3.448
Alcohol: >rec vs <=rec	1.994	1.347	2.949
Alcohol: ex excess vs <=rec	1.749	0.937	3.264
Hoarseness: explained vs no	1.558	0.414	5.861
Hoarseness: intermittent vs no	0.825	0.421	1.618
Hoarseness: persistent vs no	5.931	3.759	9.357
Regurgitation: yes vs no	0.803	0.312	2.063

Effect	Point Estimate	95% Wald Confidence Limits	
Cough: yes vs no	0.844	0.400	1.781
Neck pain: intermittent and unilateral vs persistent and unilateral	1.297	0.184	9.147
Neck pain: no / intermittent vs persistent and unilateral	1.748	0.563	5.423
Neck pain: persistent vs persistent and unilateral	2.496	0.551	11.307
Throat discomfort: yes vs no	1.179	0.608	2.286
Feeling of something in throat: yes vs no	0.320	0.144	0.715
Dysphagia: intermittent vs no	0.258	0.077	0.864
Dysphagia: persistent vs no	3.507	2.098	5.862
Odynophagia: no/intermittent vs persistent	0.061	0.026	0.143
Choking: yes vs no	0.713	0.279	1.823
Catarrh/mucus: yes vs no	0.327	0.041	2.604
Neck lump: persistent vs no	10.658	6.864	16.550
Neck lump: intermittent vs no	1.642	0.712	3.789
Oral swelling: yes vs no	9.158	5.339	15.707
Oral ulcer: yes vs no	5.471	1.573	19.027
Haemoptysis: yes vs no	1.478	0.560	3.897
Otalgia: yes vs no	3.797	1.806	7.986
Reflux: yes vs no	0.501	0.165	1.523
Throat Clearing: yes vs no	0.553	0.115	2.662
Stridor: yes vs no	8.877	1.537	51.271
Shortness of breath: no vs yes	0.321	0.108	0.955
Red/white patch throat: yes vs no	0.182	0.026	1.264
Head&neck lesion: yes vs no	8.217	3.174	21.276

4.2.6.2 The Backwards elimination process

At this stage, a backward elimination was performed sequentially, removing all non-significant variables and interaction to identify a parsimonious model with each of the finally included variables having a *P*-value of α = .05 or less.

The following table shows the summary of the backward elimination process, showing which variable was removed at every step of the backward elimination process and its corresponding p-value at each elimination step.

Table 4-11. Summary of Backward Elimination

Step	Effect Removed	Degrees of Freedom	P-value
1	Neck pain	3	0.6651
2	cough	1	0.7069
3	Throat discomfort	1	0.6315
4	regurgitation	1	0.6175
5	smoking*sore throat	8	0.5361
6	choking	1	0.4564
7	Throat Clearing	1	0.4586
8	haemoptysis	1	0.3828
9	Catarrh/mucus	1	0.2674
10	reflux	1	0.1676
11	Red/ white patch in mouth	1	0.1089
12	Shortness of breath	1	0.0618

4.2.6.3 The final selected model

The table below shows all the variables included in the final selected model, along with variables' estimates, p values and odds ratio with 95% confidence intervals. The final model includes 16 variables, compared to the 28 variables that were included in the initial regression. No interactions made it to the final model as the initially identified sore throat with smoking variable did not remain significant at the 0.05 level during the backward elimination process (p=0.5361 at the fifth round of elimination, as seen in Table 4-11). All the variables that had a p-value less than 0.05 at least one

level remained significant during the elimination process, and they were kept in the final model, apart from the shortness of breath symptoms which was removed at the last round of elimination with a p-value of 0.0618. The included variable estimates and odds ratios changed slightly compared to the initial regression, and these changes are expected as 12 variables were sequentially excluded from the model allowing for adjustments of the remaining variable estimates. The 95% confidence intervals of all the variables in the model are small that ensuring good prediction estimates with small standard errors. The variables included in the final model were: age, biological sex, alcohol and smoking, weight loss, hoarseness, dysphagia, odynophagia, sore throat, skin lesion, oral swelling, oral ulcer, neck lump, stridor, otalgia and feeling of something in the throat. Of these, the presence of a feeling of something in throat symptom was a negative indicator of HNC, as well as the explained hoarseness, intermittent dysphagia, and intermittent bilateral sore throat, having all negative variable estimates for a HNC diagnosis (Table 4-12).

Table 4-12. Final Selected Logistic Regression Model

Variable		**Estimate**	**S.E.**	**P-value**	**Odds Ratio (95% CI)**
Intercept		-6.890	0.433	<0.0001	
Age		0.028	0.005	<0.0001	1.029 (1.018 – 1.398)
Biological Sex	Male vs Female	1.031	0.163	<0.0001	2.805 (2.043 – 3.872)
Smoking	Yes vs No	0.602	0.188	0.0001	1.827 (1.265 – 2.645)
	Ex-smoker vs No	0.360	0.191	0.0588	1.434 (0.986 – 2.085)
Alcohol	>rec vs <=rec	0.753	0.194	0.0001	2.123 (1.446 – 3.098)
	Ex excess vs <=rec	0.545	0.313	0.0814	1.725 (0.919 – 3.145)
Unintentional weight loss	Yes vs No	0.778	0.228	0.0006	2.178 (1.384 – 3.383)
Hoarseness	Persistent vs No	1.813	0.227	<0.0001	6.129 (3.942 – 9.593)
	Explained vs No	-0.188	0.338	0.5791	0.829 (0.408 – 1.556)
	Intermittent vs No	0.384	0.668	0.5651	1.469 (0.315 – 4.682)

Sore Throat	Persistent Bilateral vs No Persistent Unilateral vs No Intermittent Bilateral vs No Intermittent Unilateral vs No	0.767 2.269 -1.124 0.1501	0.311 0.489 0.614 1.114	0.0136 <0.0001 0.0670 0.8929	2.154 (1.152 – 3.907) 9.678 (3.671 – 25.069) 0.325 (0.077 – 0.924) 1.162 (0.058 – 7.029)
FOSIT	Yes vs No	-1.209	0.399	0.0025	0.298 (0.127 – 0.615)
Dysphagia	Persistent vs no Intermittent vs no	1.266 -1.206	0.245 0.574	<0.0001 0.0357	3.547 (2.182 – 5.719) 0.299 (0.082 – 0.813)
Odynophagia	Yes vs No	2.604	0.216	<0.0001	13.522 (6.033-30.536)
Neck lump	Persistent vs no Intermittent/regressing vs No	2.424 0.541	0.216 0.429	<0.0001 0.2071	11.288 (7.447 – 17.395) 1.718 (0.691 – 3.785)
Oral swelling	Yes vs No	2.251	0.267	<0.0001	9.502 (5.631–16.071)
Oral ulcer	Yes vs No	1.903	0.585	0.0001	6.707 (2.107 – 20.995)
Unilateral Otalgia	Yes vs no	1.169	0.355	0.0009	3.220 (1.588 – 6.401)
Stridor	Yes vs No	2.307	0.914	0.0116	10.049 (1.414 – 57.132)
Persistent head and neck skin lesion	Yes vs No	2.193	0.475	<0.0001	8.963 (3.358 – 22.0677)

4.2.6.4 Logistic Regression Diagnostics

A series of checks will be performed in this section to ensure that the final model meets the assumptions of logistic regression and that it is a good fit for the dataset and future applications. Firstly, the linearity assumptions were assessed to check the linear relationship between the continuous variables in our model – that is, age – with the logit outcome-cancer. Figure 4-6 below shows the scatter plot between age and the logit values. The smoothed scatter line shows a relatively linear association with the cancer outcome.

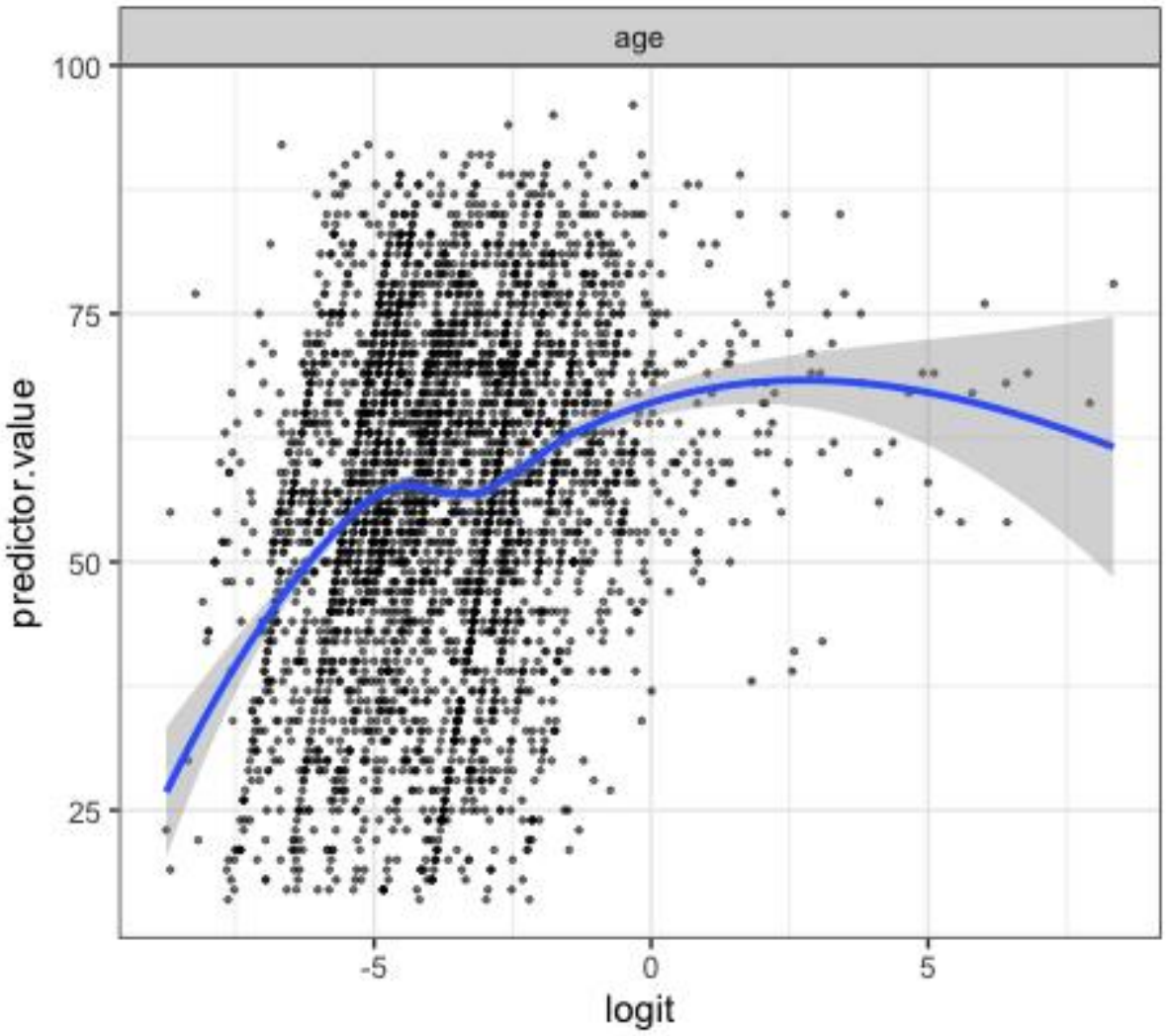


Figure 4-6. Scatter plot of age vs logit values

Influential values were checked to identify any extreme cases that can potentially significantly affect the model. The most extreme values can be examined by calculating and visualising Cook's distance values. The graph below shows the 10 most extreme values and the corresponding cases (Figure 4-7).

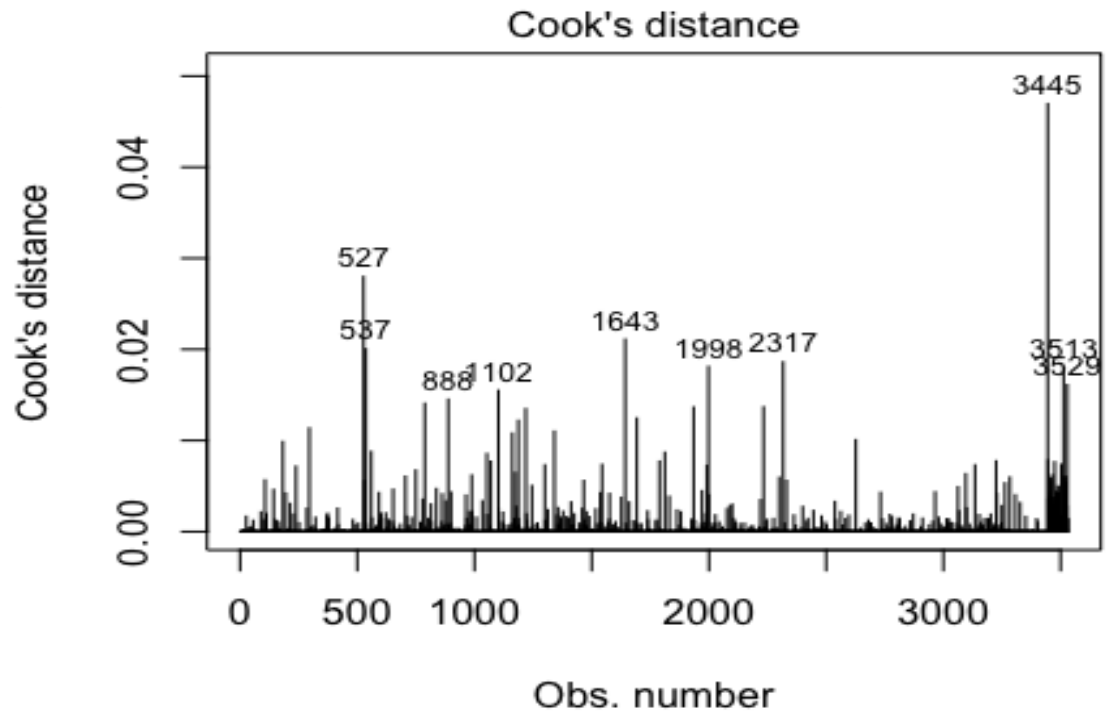


Figure 4-7. Cook's distance

Nevertheless, not all outliers are necessarily influential observations. Calculation of the standardised residual error helps in that matter. Following the calculation of the standardised residual errors and inspection, the cases with an absolute standardised residual above 3 represent possible outliers and will be checked in more detail. This process identified 4 observations that will require a detailed inspection to ensure they are not potential influential cases.

The first case was a 59-year-old male who presented with a feeling of something in the throat, was a non-smoker with no alcohol excess and was found to have cancer. The second case was a 52-year-old male smoker with again a feeling of something in his throat, and again a cancer diagnosis was made. The last two cases also had a cancer diagnosis despite not having any worrying symptoms: 62-year-old male, ex-smoker with intermittent dysphagia; 53-year-old male, non-smoker with intermittent hoarseness. These cases are correct entries, and they will be kept in the dataset. They present as outliers due to the fact that these patients were diagnosed with malignancy despite the lack of significant symptoms at the time of presentation. This can occur in very early cancer cases or as an atypical presentation. This was indeed a rare occasion in the dataset, representing 1.3% (n=4) of the cancer cases. The standardised residual

plot is seen in the following graph (Figure 4-8). It depicts the strength of the difference between the expected and observed values. No influential data points are seen, as all observations for both cancer and no cancer cases have standardised residual values within or close to three standardised residuals. The cancer cases have more widespread standardised residuals compared to the cancer-free cases, which are concentrated mainly within 1 standardised residual.

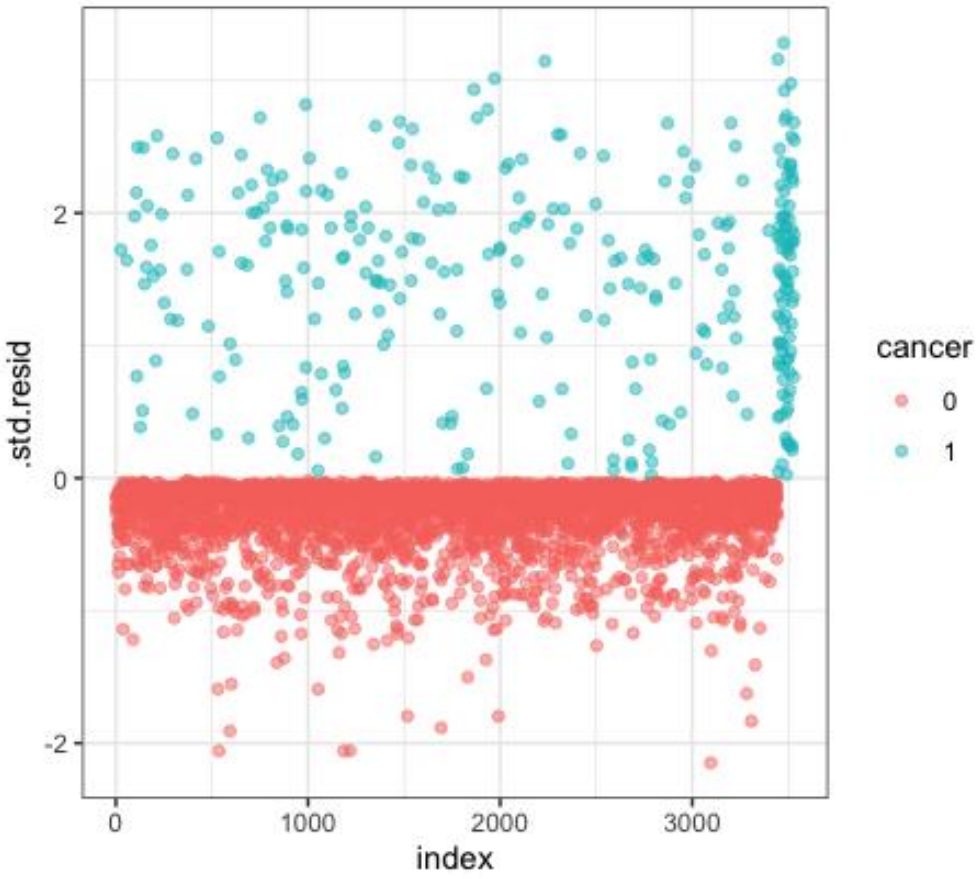


Figure 4-8. Standardised Residual Plot

Multicollinearity between the independent variables was assessed next. This is to identify any potential predictors with highly correlated values. As a rule of thumb, if the variance inflation (VIF) exceeds 5, it indicates high collinearity, which is problematic for the model build and validity. VIF was checked for all the predictor variables before any feature elimination process. No multicollinearity issues were identified, with all values well below 5. Values were also adjusted to take into consideration the weight matrix, and again low VIF values were found (Table 4-13).

Table 4-13. Multicollinearity matrix of the independent variables. VIF: variance inflation. Df: degrees of freedom. GVIF: weighted matrix VIF

Factor	VIF	Degrees of Freedom	GVIF^(1/(2*Df))
Age	1.15214	1	1.073379
Biological Sex	1.1715	1	1.082378
Unintentional weight loss	1.1902	1	1.090981
Smoking	1.2233	2	1.051698
Alcohol	1.2694	2	1.061464
hoarseness	1.7957	3	1.10249
regurgitation	1.1396	1	1.067546
Cough	1.0880	1	1.04308
Sore throat	1.3638	4	1.039548
Neck pain	1.1174	4	1.013975
Throat discomfort/irritation	1.0749	1	1.036786
foist	1.0412	1	1.020432
dysphagia	1.5465	2	1.11517
odynophagia	1.1347	1	1.065264
Choking	1.0657	1	1.032362
Catarrh/mucus	1.0275	1	1.013681
Neck lump	2.2560	2	1.225566
Oral swelling	1.1892	1	1.090512
Oral ulcer	1.0898	1	1.043973
haemoptysis	1.0557	1	1.027474
Otalgia	1.1561	1	1.075253
Reflux	1.0371	1	1.018427
Stridor	1.0487	1	1.024092
Sob	1.0704	1	1.034638
Throat Clearing	1.0246	1	1.012226
Red/white patches in mouth	1.0302	1	1.015026
Head&neck skin lesion	1.1302	1	1.063113

The goodness of fit statistics was also calculated. The chi-square goodness of fit test p-value was 0.1437, and the Hosmer and Lemeshow test p-value was 0.685. The null

hypothesis for both tests is that the model is a good fit for the data. The p-values are both over the 0.05 level of significance; hence they show that the model is a good fit for the data used to develop the model.

Finally, the calibration plot assessing the agreement between observed and actual probability values allowing estimation of the model performance at a population level showed a good alignment with the 45-degree line meaning an adequate calibration. There were two areas of deviation of the 45-degree line at the region of 45% and 85% predicted probability (Figure 4-9). In these areas, the calibration slope was below the perfect prediction line (45-degree line) where the model overpredicts in that range of predicted probabilities hence potentially falsely assigning cases to the cancer category more often.

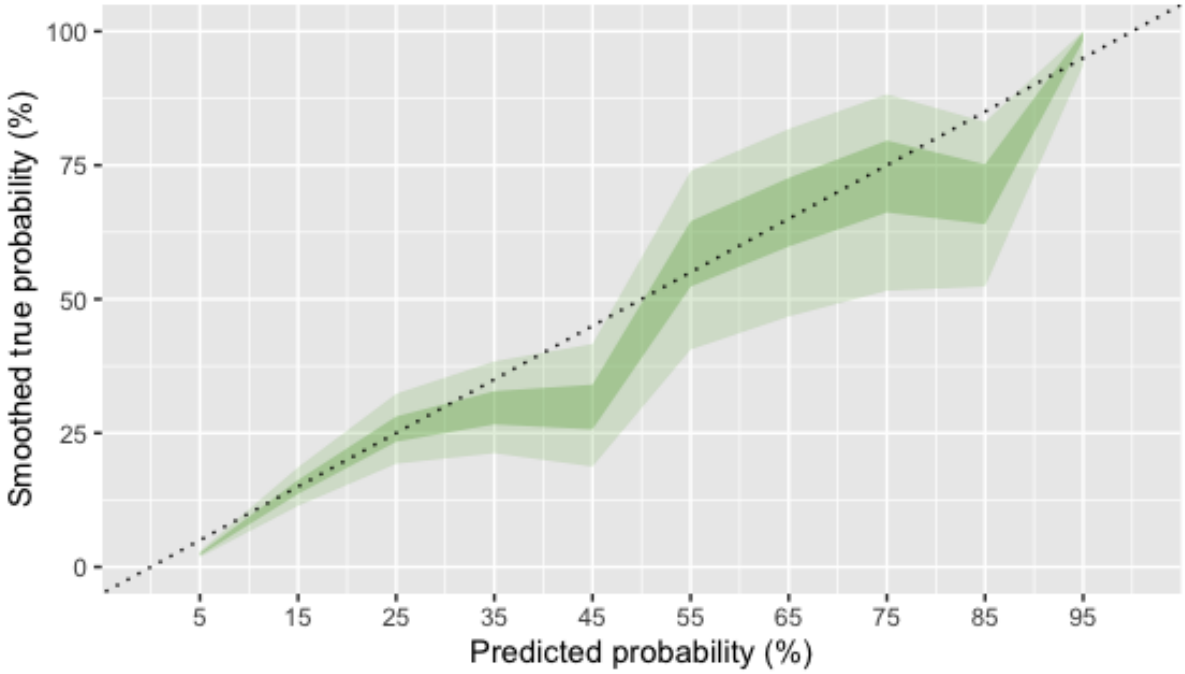


Figure 4-9. Calibration slope of the observed against the estimated cancer probability of the logistic regression model

4.2.6.5 Internal validation

Internal validation of the final model was performed using 1000 bootstrap samples, each performing random splits of the data set into training and validation sets with a final generation of the estimated mean AUC across each of the 1000 validation sets.

Following bootstrapping, the mean AUC was high at 0.8856, also corresponding to the c-statistic value (Figure 4-10) with a 0.8818 – 0.8879, 95% confidence interval. This shows very high discrimination power for the assessment of individual cases' risk. The overall sensitivity was 77.52%, with 95% CI (74.59, 82.74), and the specificity was 83.64%, with 95% CI (78.44, 86.51).

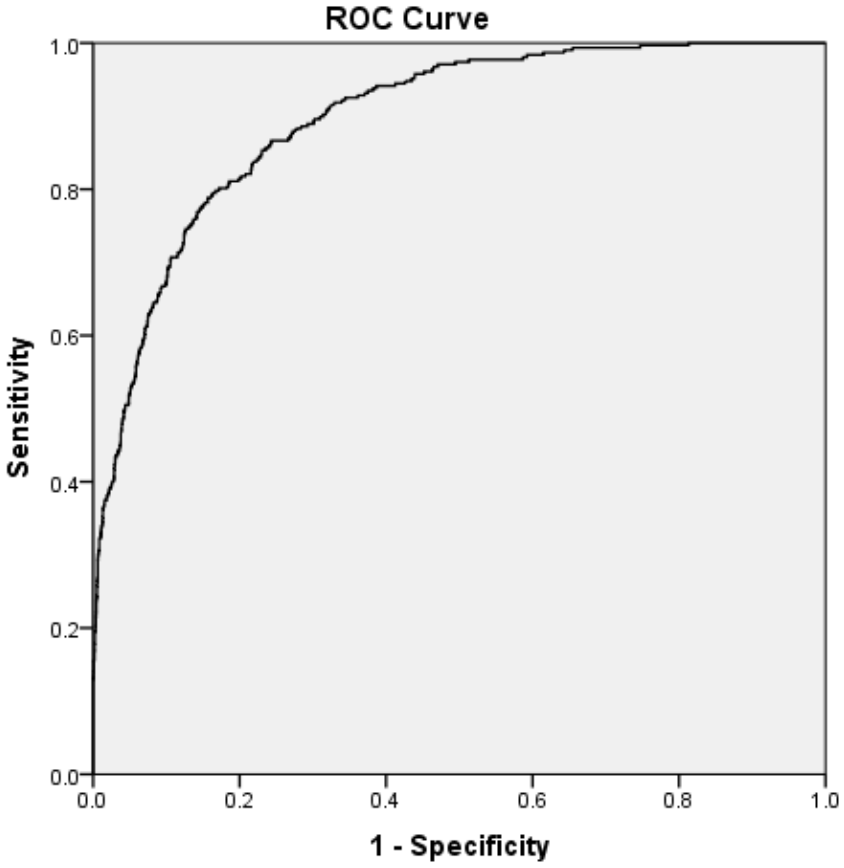


Figure 4-10. ROC curve for the generated risk calculator at internal validation. AUC: 0.897 (95% CI: 0.88 – 0.914)

4.2.6.6 Triaging probability cut-offs

The suggested probability cut-off for a USOC referral was selected to be the probability value that generated the highest value combination of sensitivity and specificity simultaneously. The 0.071 probability cut-off, as seen in Figure 4-11, maximised sensitivity and specificity at the same time, and it is recommended as a cut-off point for referral of patients in the USOC (sensitivity: 85%, specificity: 78.3% with an accuracy of 78.9%, PPV of 27.2% and NPV of 98.2%).

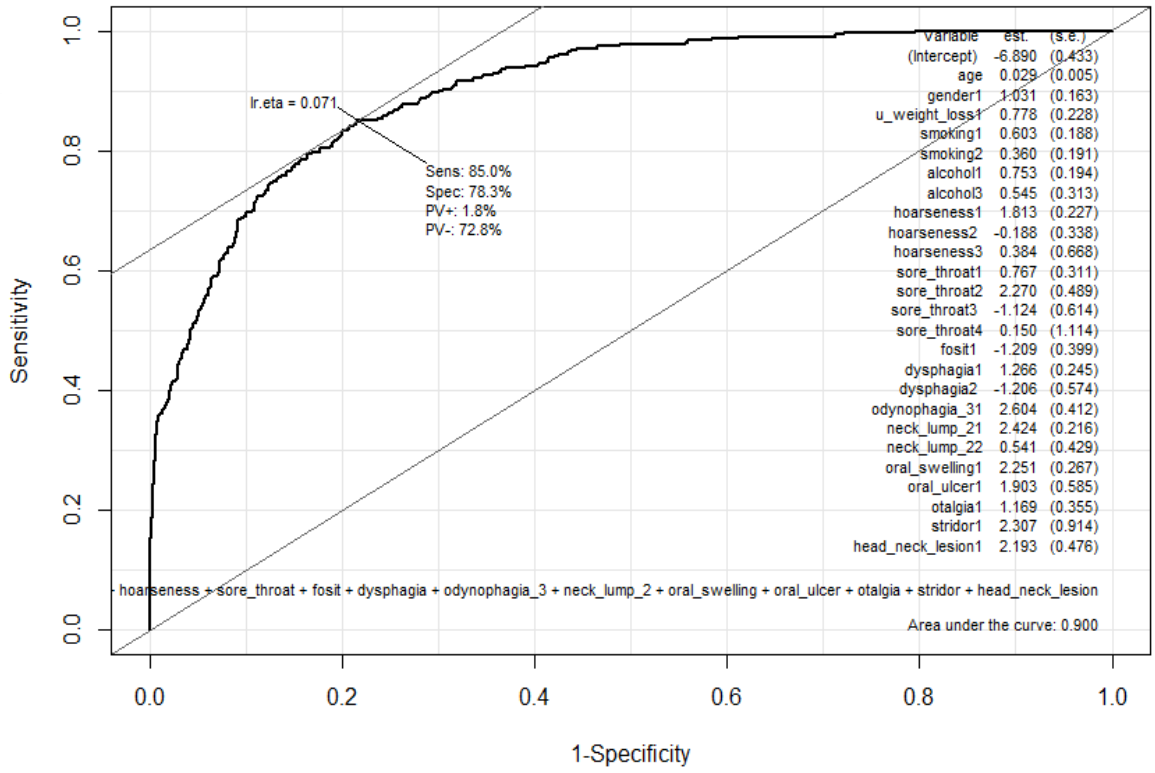


Figure 4-11. ROC Curve with suggested probability cut-off point using the whole dataset

Following the exclusion of the USOC cases, a second threshold was calculated using the same principle for the rest of the referrals, with those above the recommended cut-off being considered for an urgent (6 weeks) appointment. This second threshold was generated at 0.022 (Figure 4-12), giving a sensitivity of 97.1% and a specificity of 52.9% for the whole dataset, which maximised sensitivity and specificity again after cases with a probability of more than .071 were excluded. The PPV was 27.5%, and the NPV was 99.6% at the second cut-off point.

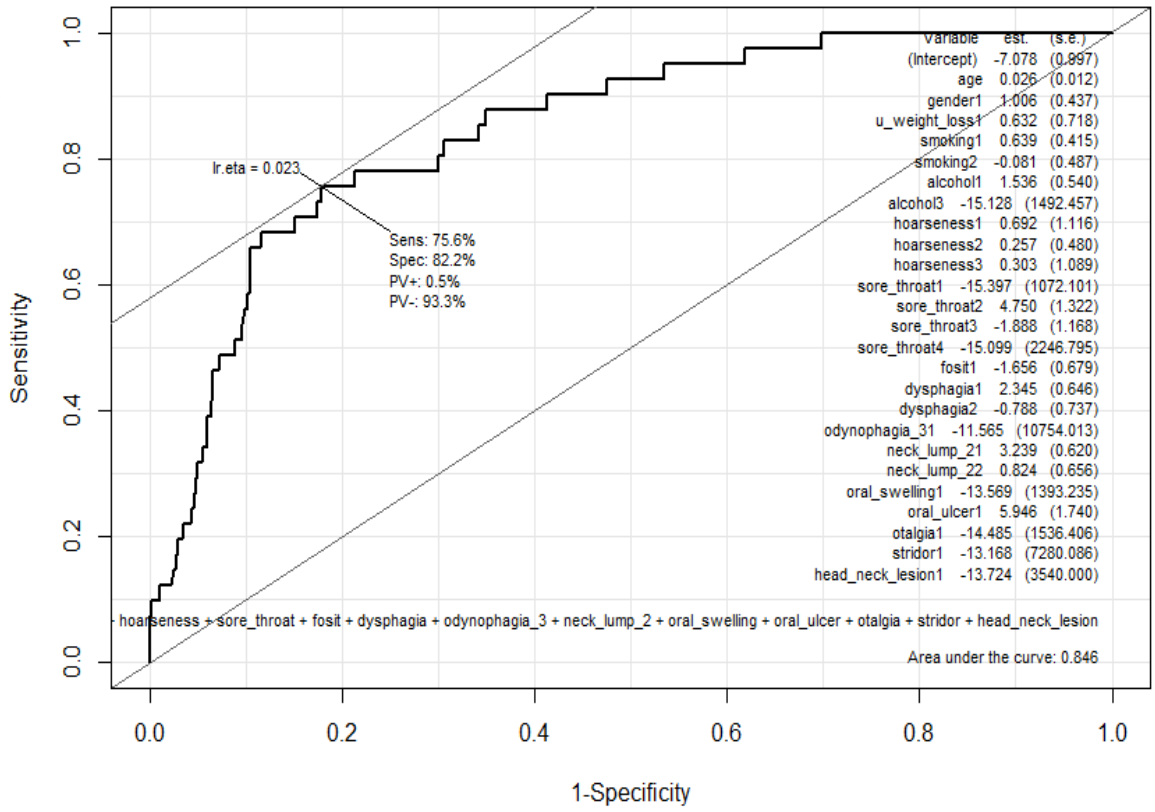


Figure 4-12. ROC curve with second suggested probability cut-off point on the whole dataset following removal of cases with a calculated HNC probability of over 7.1%.

The potential impact of the calculator on patient referrals using the two thresholds to divide patients into three groups was considered. Table 4-14 shows how the calculator would have redistributed patients to clinics, including the resulting impact on cancer detection per clinic type. The calculations were based on the un-boosted cancer population. The data suggest that the number of patients diagnosed in non-USOC clinics is significantly reduced from 39.1% (26.1% in urgent clinics; 13% in routine) to 14.8% (12.2% in urgent, 2.6% in routine) whilst the cancer detection from the USOC clinics would be significantly increased from 60.9% to 85.2%. The change in the cancer diagnosis using the HaNC-RC v.2 re-triaging was statistically significant ($P < .0001$), and this occurred whilst seeing fewer patients through the USOC and urgent route. Of the nine cancers misclassified to the routine group (low-risk category based on the HaNC-RV v.2), four (44.4%) were at cancer stage 1 (n=1 laryngeal cancer, n=2 thyroid cancers, n=1 oropharyngeal cancer), two (22.2%) cases were at stage 2 (n=2 oropharyngeal cancers) and 3 (33.3%) patients had stage 3 cancer (n=2 supraglottic cancer, n=1 oesophageal cancer).

Table 4-14. Cancer detection in head and neck clinics with current triage system compared to suggested HaNC-RC v.2. triaging

	Current triaging pathway (GP referrals)		Total	Suggested triaging (HaNC-RC v.2)		Total
	Cancer			Cancer		
	Yes	No		Yes	No	
Routine	45 (14.7%)	1339 (41.5%)	1384 (39.2%)	9 (2.9%)	1677 (52%)	1686 (47.7%)
Urgent	78 (25.4%)	1002 (31.1%)	1080 (30.6%)	32 (10.4%)	760 (23.6%)	792 (22.4%)
USOC	184 (59.9%)	883 (27.4%)	1067 (30.2%)	266 (86.6%)	787 (24.4%)	1053 (29.8%)

Finally, the sensitivity and specificity results of the revised calculator were compared with the output of the previous version of the calculator using the current cohort. Applying the older version of the HaNC-RC, the sensitivity dropped by 5% to 80.78% and the specificity by 10% to 68.08%, with an AUC of 0.801. Hence, a head-to-head comparison of the diagnostic power for the first and second versions of the calculator showed improved sensitivity and specificity values.

4.2.7 Random Forest Analysis Results

The random forest was performed in the clean dataset of 3,531 cases with 307 cancers. All possible variables (n=27) were included in the development of the random forest model, similar to the logistic regression model development, to allow for a direct comparison of the results. Interaction terms were not assessed for inclusion, as in the random forest, any interactions that are useful for prediction are part of the developed forest, so there is no need to be added explicitly as an interaction term. The table below shows the importance of the variables included in the random forest assessment in correctly identifying the cancer diagnosis based on the Mean Decrease Gini Index (Figure 4-13).

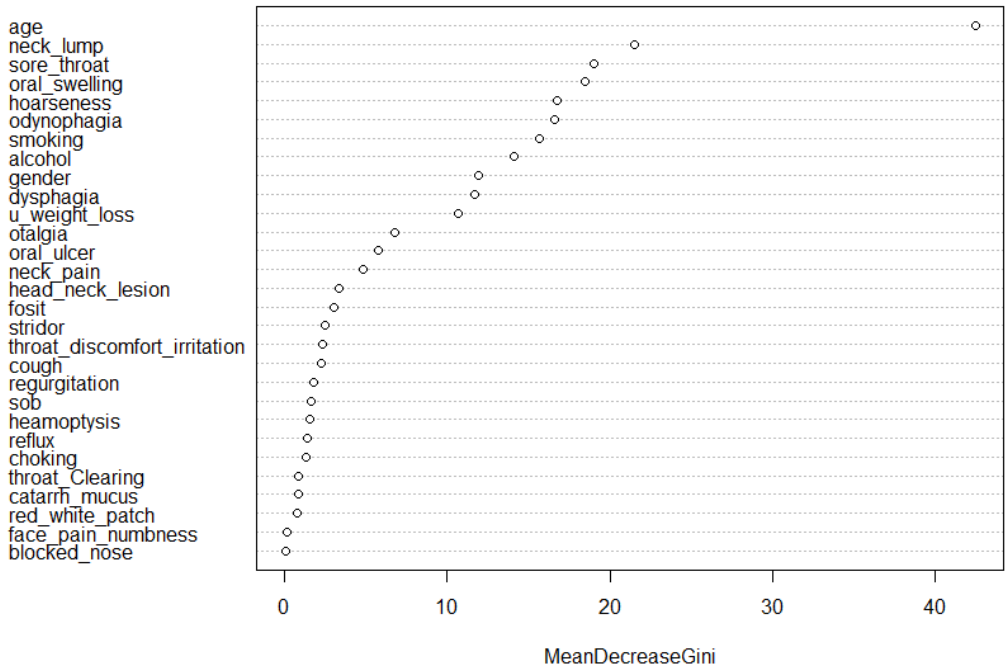


Figure 4-13. Random forest output of Mean Decrease Gini Index

The age variable was the most influential, with the largest Mean Decrease Gini index (impurity importance), followed by the neck lump, sore throat, and oral swelling variable. The first 17 variables in order of Mean Decrease Gini value were those also included in the final logistic regression model. Hence, the output of the random forest analysis agrees with the logistic regression findings.

The Youden index value was found to be 5.4% for the random forest model for triaging patients to the low- and high-risk groups for HNC diagnosis, giving the best sensitivity and specificity combination. The sensitivity at this cut-off was 87.9%, and the specificity was 89.6% (Figure 4-14).

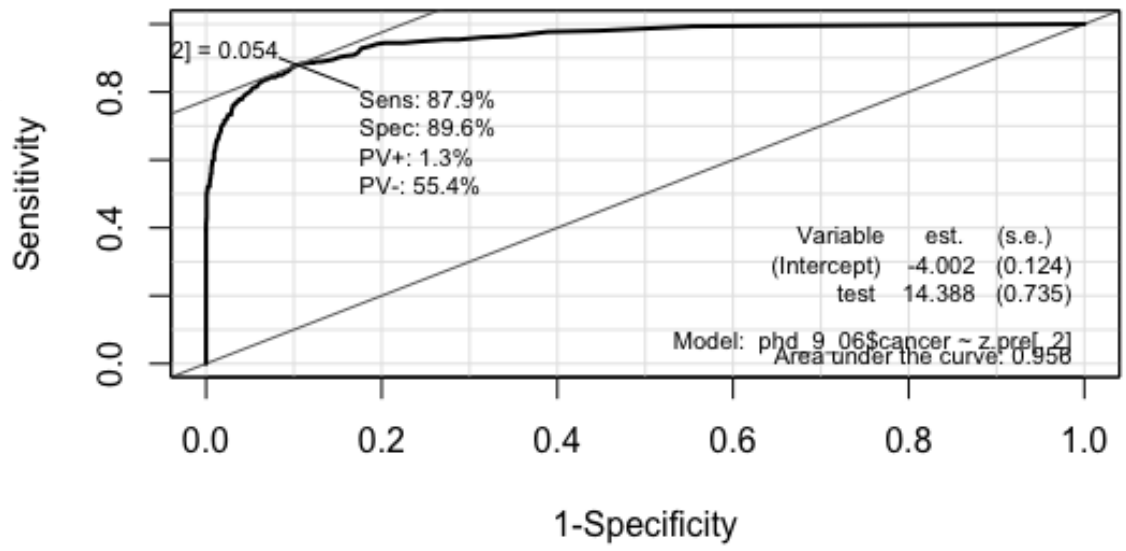


Figure 4-14. ROC Curve with Youden index cut-off point using the whole dataset

At internal validation, the model out-of-bag misclassification error was 7.4% (in the validation set), with an accuracy of 93.6%. Despite a high specificity of 99.7%, the sensitivity is low at 23.5%. The PPV was 88.8%, and the NPV was 93.6%. The AUC at internal validation was 0.829.

4.2.8 Summary of the logistic regression and random forest models results and final model selection

The AUC of the logistic regression model at internal validation was 88.5%, being higher than the AUC of the random forest model for the validation set (AUC= 82.9%). Despite the fact that the accuracy of the random forest model was very high at 93.6% at internal validation, most of the cancer cases were misclassified at internal validation with an overall sensitivity of 23.5%. On the contrary, the accuracy of the logistic regression model was lower at 76.6%, but most of the cancers were correctly classified with a sensitivity of 77.5% at internal validation. The best cut-off value assigning the maximum sensitivity and specificity combination in the dataset was a probability of 7.1% in the logistic regression model achieving a sensitivity of 77.5% and a specificity of 83.64% at internal validation. The same threshold was lower at random forest analysis, being set at 5.4%, reaching a high specificity of 99.7% but at the expense of a very low sensitivity of 23.5% at internal validation.

The selected logistic regression model includes a total of 17 significant variables at the 0.05 level of significance. All included variables have narrow odds ratios and small standard error values for the variable estimates. The model performance statistics showed a good fit for the database. The same 17 variables that were significant in the logistic regression analysis were also found to be the most influential variables in the random forest analysis. The age variable was the most influential variable, followed by the neck lump symptom, sore throat, and oral swelling. The impurity importance value was over 10, suggesting a strong influence in the random forest classification for the following variables: age, neck lump, sore throat, oral swelling, hoarseness, odynophagia, smoking, alcohol, biological sex, dysphagia, unintentional weight loss. These features were also highly significant in the logistic regression analysis.

In summary, it seems that both logistic regression and random forest are in agreement with the variables that are significant/influential in the HNC diagnosis. Random forest had very good accuracy, but it failed to identify the majority of cancer cases correctly during internal validation, achieving its high accuracy from a very high specificity value. On the other hand, the logistic regression achieved a better combination of sensitivity and specificity at internal validation. The AUC for both logistic regression and the random forest was over 80% at validation, with the logistic regression achieving a marginally higher AUC at 88.5% (vs 82.9% for the random forest).

For the above reasons, the logistic regression model was selected as the final best-fit model, as it performed better at internal validation in identifying cancer cases and had an overall higher AUC. The updated version of the calculator was called HNC risk calculator version 2 (HaNC-RC v.2) and was added to the orlhealth.com website, where the first version of the calculator was already available to clinicians. The updated version can be accessed via the following link: http://orlhealth.com/risk-calculator-2.html (www.orlhealth.com, 2019). In addition, the results of the chosen logistic regression model were published open access in the Journal of Clinical Otolaryngology in January 2020 (Tikka *et al.*, 2020).

4.3 External validation phase

4.3.1 Introduction

The COVID-19 pandemic started in the UK in February 2020, 2 months after the publication of the updated version of the HaNC-RC v.2. The ENTUK and BAHNO organisations expressed interest in the calculator in March 2020 to support the development of a remote telephone triage pathway for patients referred via the urgent HNC pathway for suspected cancer. At that point, the UK was in full lockdown, and all outpatient hospital appointments were suspended apart from emergency care clinics and necessary appointments needed for the workup of cancer cases. In collaboration with ENTUK and BAHNO, a UK-wide prospective study was initiated to evaluate the use of the HaNC-RC v.2 in triaging 2ww referrals via telephone consultations over a 16-week period. The recommendation for an urgent face-to-face clinic review was based on the individual probabilities generated by the calculator and using the recommended cut-off for a 2ww referral. Patients scoring less than 7.1% were offered a deferred outpatient appointment, or investigations were arranged based on their symptoms prior to a face-to-face clinic review, whereas those with a score of 7.1% or more were called for urgent clinic review (within 2 weeks). The second cut-off of urgent (within 4 weeks) review was not used during the service evaluation as, at that point, there were no facilities for such outpatient clinic appointments due to the first wave of the COVID-19 pandemic. Only the very high-risk patients were prioritised for face-to-face reviews. This prospective UK-wide cohort of patients was used for external validation of the HaNC-RC v.2. The predictive power of the HaNC-RC v.2 in this new cohort was assessed by calculation of negative and positive predictive power, sensitivity, specificity, and area under the receiver operating curve. The logistic regression algorithm was rerun in the new cohort to assess for differences in the parameter estimates in the new cohort of patients that may inform future iterations of the triage tool.

4.3.2 Univariate analysis

A total of 47 UK hospitals registered for participation in this nationwide study; at the end of the follow-up period, 41 hospitals submitted complete results and were included

in the data analysis. This included 32 hospitals from England, 6 Scottish centres, 2 hospitals in Wales and one from Northern Ireland. A total of 4,569 cases were registered during the study period. At the end of the data collection window, the final cancer outcome was missing for 12 cases. These cases were excluded from analysis, resulting in 4,557 cases eligible for univariate and multivariate analysis. The median number of patients included from each participating site was 99 (range: 10-337).

The majority of included patients were females (n=2,604; 57.1%), and the mean age was 56.9 years of age (1 – 98). Patient demographics are seen in Table 4-15. In the validation data, there were 3 cases with an age of less than 16 years old. None of these patients was diagnosed with cancer, and all three scored low on the HaNC-RC v.2. The demographics of the validation cohort were similar to those of the cohort used to develop the calculator with no statistically significant difference for age (p=0.379), biological sex (p=0.836), smoking (p=0.203) and alcohol (0.953) variables in the two cohorts.

Table 4-15. Patient demographics in the external validation cohort of 4557 cases

Variable	Value	Frequency	Percentage
Age	Mean: 56.9 SD: 16.4 Range: 1 - 98	4557	100%
Biological Sex	Female Male	2604 1953	57.1% 42.9 %
Smoking	No Yes Ex-smoker	2504 756 1297	54.9% 16.6% 28.5%
Alcohol	≤14 units/w >14 units/w Ex excess	4053 450 54	88.9% 9.9% 1.2%

Patient presenting symptoms are summarised in Table 4-16 below. Only the symptoms included in the calculator were recorded in the database. The dysphagia, neck lump, oral swelling, oral ulcer, skin lesion and unintentional weight loss symptoms had

similar distributions to the development cohort. The feeling of something in the throat was recorded significantly more in the validation cohort, 33.5% (=1525), compared to only 16.5% (n=601) in the development cohort. Within the hoarseness symptom, 25% complained of intermittent hoarseness, which was 18.7% in the development cohort. Pain on swallowing was recorded 10 times more in the validation cohort (10.8%) compared to the development study (1.6%). Otalgia was also more commonly reported in this cohort, being recorded in 11.2% of the patients compared to only 2.7% in the development study. The otalgia symptom in the development phase was based on a normal otoscopic examination, but this was not possible in the telephone consultation evaluation, which may have increased the number of times this symptom was reported. The sore throat symptoms also had different distributions in the symptom laterality, with 10% of patients complaining of persistent bilateral or central pain (vs 4.4% in the previous cohort), persistent unilateral in 6.4% (vs 1.3% before) with similar increases in the numbers for the intermittent unilateral and bilateral sore throat symptoms. The stridor symptom was also reported considerably more, being present in 102 patients (2.2%) versus only 12 patients (0.3%) in the earlier cohort. Potential reasons for these differences will be explored in the discussion chapter of this thesis in section 5.2.3.

Table 4-16. Patients' presenting symptoms in the external validation cohort

Variable	Value	Frequency	Percentage
Dysphagia	No	3748	82.2%
	Persistent	296	6.5%
	Intermittent	513	11.3%
Feeling of something in throat	No	3032	66.5%
	Yes	1525	33.5%
Hoarseness	No	2776	60.9%
	Persistent	541	11.9%
	Intermittent	1140	25%
	Explained	100	2.2%
Neck Lump	No	3338	73.2%
	Persistent/ Increasing	985	21.6%
	Fluctuating/ Regressing	234	5.1%

Variable	Value	Frequency	Percentage
Odynophagia	No	4067	89.2%
	Persistent	490	10.8%
Oral swelling	No	4280	93.9%
	Yes	277	6.1%
Oral ulcer	No	4406	96.7%
	Yes	151	3.3%
Otalgia	No	4047	88.8%
	Yes	510	11.2%
Persistent head and neck skin lesion	No	4511	99%
	Yes	46	1.%
Sore Throat	No	3004	65.4%
	Persistent bilateral	447	9.8%
	Persistent unilateral	292	6.4%
	Intermittent bilateral	566	12.4%
	Intermittent unilateral	248	5.6%
Stridor	No	4455	97.8%
	Yes	102	2.2%
Unintentional weight loss	No	4091	89.8%
	Yes	466	10.2 %

Cancer at the end of the 6-month follow-up period was diagnosed in 5.6% of patients (n=254). The table below (Table 4-17) shows the distribution of cancers per HaN subsite. Of these,151 cancers were of HaN origin (59.5%), 9 unknown primaries (3.5%), with the rest being in adjacent sites to HaN (n=29, 13%) or presenting with HaN metastasis from distant sites (n=10, 4%) and lymphoma/leukaemia manifesting in the HaN (n=47, 18.5%). Compared to the development cohort, there were fewer HaN origin cancers in the validation study (n=236, 76.4% in the development study), with proportionally more cancer arising from sites adjacent to the HaN (n=24, 7.8% in the development study). No results were available for the stage of cancer at the time of diagnosis.

Table 4-17. Cancer site distribution in the external validation cohort

Cancer site	Frequency (%)
Hypopharyngeal cancer	8 (3.15%)
Laryngeal cancer	31 (12.2%)
Nasal cavity	4 (1.57%)
Nasopharynx	4 (1.57%)
Oral cavity	8 (3.15%)
Oropharynx	65 (25.59%)
Salivary gland	8 (3.15%)
Skin	7 (2.76%)
Thyroid	24 (9.45%)
Unknown primary	9 (3.54%)
Metastatic breast cancer	4 (1.57%)
Metastatic colorectal cancer	1 (0.39%)
Metastatic ovarian	2 (0.79%)
Metastatic prostate	1 (0.39%)
Metastatic renal	1 (0.39%)
Metastatic liver cancer	1 (0.39%)
Oesophageal cancer	13 (5.12%)
Lung cancer	16 (6.3%)
Leukaemia	2 (0.79%)
Lymphoma	45 (17.72%)
Total	253.0002

The demographics and presenting complaints against the cancer diagnosis are summarised in the table below (Table 4-18) alongside the p-value and individual variable AUC following univariate regression analysis. The influence of the different ways of data collection became evident in the p values and odds ratio for some of the variables. The oral ulcer and oral swelling symptoms that were highly significant in univariate analysis in the development cohort failed to reach statistical significance in the current cohort with much lower odds ratio estimates. The stridor symptom was over-reported in the validation cohort. Although it was a highly significant indicator

for HNC in the previous study, it did not reach significance in the external validation cohort. The rest of the variables remained significant in this univariate analysis, but the odds ratios for many of the variables were much lower compared to those found during univariate analysis in the development dataset. For example, persistent dysphagia had an odds ratio of 3.3 and odynophagia 1.87, being 4.7 and 10.9, respectively, in the earlier cohort.

Table 4-18. Univariate analysis of the external validation cohort

Variable		Total	Benign N=4303	Cancer N=254	P value
Biological Sex	Female Male	2604 (58.7%) 1953 (41.3%)	2516 (58.5%) 1787 (41.5%)	88 (34.6%%) 166 (65.4%)	<0.001
OR Male vs Female: 2.66 (2.04 – 3.46) **AUC 0.6191**					
Age	Mean SD	56.9 16.4	56.5 16.4	64.2 13.8	<0.001
OR 1.03 (1.02 – 1.04) **AUC=0.6405**					
Smoking	No Yes Ex	2504 (54.9%) 756 (16.6%) 1297 (28.5%)	2387 (55.5%) 696 (16.2%) 1220 (28.4%)	117 (46.1%) 60 (23.6%) 77 (30.3%)	0.003
OR Ex-smoker vs No: 1.29 (0.96 – 1.73), p=0.094 **Yes vs No: 1.76 (1.27 – 2.43), p<0.001** **AUC: 0.556**					
Alcohol	≤14 units >14 units Ex excess	4053 (88.9%) 450 (9.9%) 54 (1.2%)	3849 (89.4%) 405 (9.4%) 49 (1.1%)	204 (80.3%) 45 (17.7%) 5 (2%)	<0.001
OR >14 units vs ≤14 units: 2.1 (1.49 – 2.94) – p<0.001 **Ex Excess vs ≤14 units: 1.93 (0.76 – 4.88) – p =0.168** **AUC 0.5458**					

Variable		Total	Benign N=4303	Cancer N=254	P value
Unintentional Weight loss	No Yes	4091 (89.8%) 466 (10.2%)	3884 (90.3%) 419 (9.7%)	207 (81.5%) 47 (18.5%)	<0.001
OR Yes vs No: 2.1 (1.51 – 2.93) **AUC 0.5438**					
Hoarseness	Persistent Intermittent Explained No	541 (11.9%) 1140 (25%) 100 (2.2%) 2488 (68.3%)	502 (11.7%) 1108 (25.7%) 100 (2.3%) 2270 (68.1%)	39 (15.4%) 32 (12.6%) 0 (-) 183 (72%)	<0.001
OR Explained vs No: 0 **Intermittent vs No: 0.41 (0.28 – 0.6) – p<0.001** **Persistent vs No: 1.1 (0.77 – 1.58) – p=0.97** **AUC 0.5831**					
Sore throat	Persistent bilateral Persistent unilateral Intermittent bilateral Intermittent unilateral No	447 (9.8%) 292 (6.4%) 566 (12.4%) 248 (5.4%) 3004 (65.9%)	425 (9.9%) 263 (6.1%) 558 (13%) 239 (5.6%) 2818 (65.5%)	22 (8.7%) 29 (11.4%) 8 (3.1%) 9 (3.5%) 186 (73.2%)	<0.001
OR Persistent bilateral vs No: 0.78 (0.5 – 1.23) – p=0.294 **Persistent unilateral vs No: 1.67 (1.11 – 2.52) – p=0.014** **Intermittent bilateral vs No: 0.22 (0.11 – 0.44) – p<0.001** **Intermittent unilateral vs No: 0.57 (0.29 – 1.13) – p=0.107** **AUC: 0.5864**					
Feeling of something/ Lump in throat	No Yes	3032 (66.5%) 1525 (33.5%)	2835 (65.9%) 1468 (34.1%)	197 (77.6%) 57 (22.4%)	<0.001
Yes vs No: 0.56 (0.41 – 0.76) **AUC: 0.5584**					

Variable		Total	Benign N=4303	Cancer N=254	P value
Dysphagia	Persistent Intermittent No	296 (6.5%) 513 (11.3%) 3748 (82.2%)	254 (5.9%) 495 (11.5%) 3554 (82.6%)	42 (16.5%) 18 (7.1%) 194 (76.4%)	<0.001
OR Intermittent vs No: 0.67 (0.41 – 1.09) – p=0.106 **Persistent vs No: 3.03 (2.12 – 4.33) – p <0.001** **AUC: 0.5678**					
Odynophagia	Yes No	490 (10.8%) 4067 (89.2%)	445 (10.3 %) 3858 (89.7%)	45 (17.7%) 209 (82.3%)	<0.001
OR Persistent vs No/Intermittent: 1.87 (1.33 – 2.61) **AUC: 0.5369**					
Neck lump	Fluctuating/ Reducing Persistent No	234 (5.1%) 985 (21.6%) 3338 (73.2%)	224 (5.2%) 841 (19.4%) 3238 (75.2%)	10 (3.9%) 144 (56.7%) 100 (39.4%)	<0.001
OR Fluctuating vs No: 1.45 (0.74 – 2.81) - p=0.277 **Persistent vs No: 5.54 (4.25 – 7.24) – p<0.001** **AUC: 0.6903**					
Persistent Oral swelling/ growth	No Yes	4280 (93.3%) 277 (6.1%)	4048 (94.1%) 255 (5.9%)	232 (91.3%) 22 (8.7%)	0.078
OR Yes vs No: 1.51 (0.95 – 2.37) **AUC 0.5137**					
Persistent Oral ulcer	No Yes	4406 (96.7%) 151 (3.3%)	4157 (96.6%) 146 (3.4%)	249 (98%) 5 (2%)	0.223
OR Yes vs No: 0.57 (0.23 – 1.41) **AUC: 0.5071**					
Unilateral unexplained otalgia	No Yes	4047 (88.8%) 510 (11.2%)	3832 (89.1%) 471 (10.9%)	215 (84.6%) 39 (15.4%)	0.031
OR Yes vs No: 1.48 (1.04 – 2.1) **AUC: 0.522**					

Variable		Total	Benign N=4303	Cancer N=254	P value
Stridor	No Yes	4455 (97.8%) 102 (2.2%)	4207 (97.8%) 96 (2.2%)	348 (97.6%) 6 (2.4%)	0.891
OR Yes vs No: 1.06 (0.46 – 2.44) **AUC: 0.5007**					
Persistent Head and Neck skin lesion	No Yes	4511 (99%) 46 (1%)	4264 (99.1%) 39 (0.9%)	247 (97.2%) 7 (2.8%)	0.007
OR Yes vs No: 3.1 (1.37 – 7) **AUC: 0.5092**					

Of the total cohort, 2,145 patients (47.1%) were triaged as low risk from the HNC-RC v.2 (probability <0.022). Moderate risk for HNC risk was noted for 983 patients (21.6%), and 1429 patients (31.3%) were flagged as high risk (probability >=0.071). The cancer detection rates based on the HNC-RC v.2. risk categories are seen in the table below (Table 4-19).

Table 4-19. Triaging of patients based on the HaNC-RC thresholds against the actual cancer diagnosis

	Cancer Diagnosis		Total
HNC-RC v.2. triaging	Yes	No	
Low Risk	26 (10.2%)	2,119 (49.2%)	2145 (47.1%)
Moderate Risk	42 (16.5%)	941 (21.9%)	983 (21.6%)
High Risk	186 (73.2%)	1,243 (28.9%)	1429 (31.3%)
Total	254	4303	4557

The distribution of patients to low, moderate, and high risk was similar to the cohort used to generate the HNC-RC v.2 being 47.2%, 25.4% and 27.4%, respectively, in the previous cohort.

Cancer was diagnosed in 73.2% of the high-risk group, which dropped from 85.2% in the cohort used to produce the calculator. A higher number of cancers were seen in the moderate risk (16.5%) and low risk (10.2%) groups compared to the previous cohort (12.2% and 2.6%, respectively). This difference may reflect the telephone clinic

consultation effect of the triage, as the patients' symptoms used to create the HNC-RC v.2. were all assessed in face-to-face clinics. Of the 151 HaN primary cancers, 16 (10.6%) were misclassified as low risk using the risk calculator, the rest of the missed cancers being metastatic or arising from sites adjacent to HaN (n=10), as is seen in Table 4-20.

Table 4-20. Cancers misclassified as low risk based on the HaNC-RC v.2 triaging

Cancers misclassified as low risk	Frequency
Laryngeal	4
Nasal cavity	1
Oral cavity	1
Oropharynx	5
Thyroid	4
Skin	1
Lung	4
Lymphoma	2
Oesophageal	3
Metastatic ovarian	1
Total	26

4.3.3 Multivariate analysis

The logistic regression model was run in this cohort of patients using the whole cohort as external validation of the previously generated model. The AUC, showing the predictive power of the model, remained high at 0.8396. The AUC of the risk calculator in the development phase was 0.8856, hence not differing much from the external validation set, but a 0.046 drop was noted. The sensitivity was 70.08 % (95% CI: 66.54,73.23), and the specificity was 81.09% (95% CI: 77.99,83.96).

The variables that were significant when the logistic regression model was fitted in this dataset were: age, biological sex, alcohol, hoarseness, feeling of something in the throat, dysphagia, odynophagia, neck lump and neck swelling. The rest did not reach statistical significance but are included in the model to allow direct comparison with

the HNC-RC v.2 output, so no backward elimination process was undertaken. Table 4-21 shows the p values of the logistic regression model fitted in the validation database and the odds ratios of each of the included variables.

Table 4-21. Logistic regression model output for the external validation cohort

Variable		Estimate	S.E.	P value	Odds Ratio (95% CI)
Intercept		-7.0784	0.3924	<.0001	
Age		0.0454	0.00502	<.0001	1.046 (1.036 1-.057)
Biological Sex	Male vs Female	0.9331	0.1510	<.0001	2.542 (1.891-3.418)
Unintentional weight loss		0.3605	0.2019	0.0741	1.434 (0.966 -2.130)
Smoking	Yes vs No	0.4555	0.1933	0.0184	1.577 (1.080 -2.303)
	Ex vs No	0.2234	0.1672	0.1815	1.250 (0.901 -1.735)
Alcohol	>Rec vs No	0.5048	0.2003	0.0117	1.657 (1.119 -2.453)
	Ex excess vs No	0.3091	0.5312	0.5606	1.362 (0.481 -3.859)
Hoarseness	Persistent vs No	0.00406	0.2183	0.9852	1.004 (0.655 -1.540)
	Explained persistent vs No	-14.4959	567.7	0.9796	<0.001 (<0.001->999.999)
	Intermittent vs No	-0.6059	0.2158	0.0050	0.546 (0.357 -0.833)
Sore Throat	Persistent Bilateral vs No	-0.2538	0.2751	0.3561	0.776 (0.452 -1.330)
	Persistent Unilateral vs No	0.1700	0.2727	0.5329	1.185 (0.695 -2.023)
	Intermittent Bilateral vs No	-0.8957	0.3815	0.0189	0.408 (0.193 -0.862)
	Intermittent Unilateral vs No	0.1526	0.3695	0.3695	0.858 (0.416 -1.771)
FOSIT	Yes vs No	-0.4425	0.1834	0.0158	0.642 (0.448 -0.920)
Dysphagia	Persistent vs no	0.8364	0.2395	0.0005	2.308 (1.444 -3.691)
	Intermittent vs no	-0.0366	0.2757	0.8944	0.964 (0.562 -1.655)
Odynophagia	Yes vs No	0.5681	0.2373	0.0167	1.765 (1.108 -2.810)
Neck lump	Persistent vs no	1.9980	0.1587	<0.0001	7.375 (5.403 - 10.066)
	Intermittent/ regressing vs No	0.7535	0.3545	0.0335	2.125 (1.061 -4.256)
Oral swelling	Yes vs No	1.1648	0.2690	<.0001	3.205 (1.892 -5.431)
Oral ulcer	Yes vs No	-0.7754	0.4987	0.1200	0.461 (0.173 -1.224)
Unilateral Otalgia	Yes vs No	0.4236	0.2178	0.0517	1.527 (0.997 -2.341)
Stridor	Yes vs No	0.0223	0.4667	0.9619	1.023 (0.410 -2.552)

Persistent head and neck skin lesion	Yes vs No	0.7399	0.4867	0.1285	2.096 (0.807 -5.440)

The odds ratio of the demographics and social history factors were similar to the HaNC-RC v.2 logistic regression model. The biggest difference was for alcohol consumption, where in the validation cohort, the odds for drinking in excess vs no alcohol consumption was 1.6 times, being 2.1 times in the development phase of the calculator.

Looking at the significant symptoms included in the external validation phase, persistent hoarseness was not a discriminatory level within the hoarseness variable, with the odds ratio including the value of 1. In contrast, it was highly significant during the development phase, with an odds ratio of 6.1 (95% CI – 3.9 – 9.6). This time the explained hoarseness level was significant, with no patients that reported this symptom subsequently having HNC in the development phase. This level of the hoarseness variable was not a significant discriminator. For the sore throat variable, the persistent unilateral symptom that was before highly significant (OR= 9.7 (95% CI – 3.7 – 25.1)) was now a non-significant feature with an OR of 1.18 (95% CI 0.7 – 2). The dysphagia and odynophagia symptoms remained significant, but the odds ratio of the odynophagia symptoms dropped considerably from a strong positive association of 13.5 odds ratio to a very modest odds ratio of 1.8. The oral swelling also dropped its odds ratio value to a third (from 9.5 to 3.2), and the oral ulcer symptom was not statistically significant anymore. This reflects the fact that the reporting of symptoms was done over the phone, so there was no element of examination to allow for assessment of the oral swelling and oral ulcer signs hence relying solely on patients' self-reporting. The otalgia, stridor and skin lesion symptoms were not significant in the validation cohort ($p>0.05$).

The optimal cut-off point for the classification of cases to the high-risk category for a USOC referral was 0.79 (Figure 4-15), which is very close to the 0.071 cut-off identified during the development phase of the calculator. At that cut-off point, the accuracy was 83.6%, the sensitivity 69.7% and the specificity 84.4%.

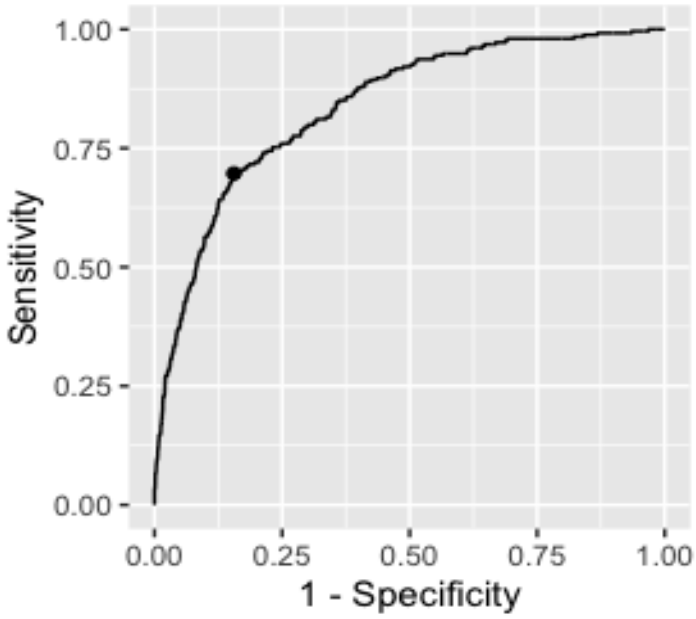


Figure 4-15. Optimal cut-off point (Youden index) in the external validation cohort

4.3.4 Summary of the external validation results

The HaNC-RC v.2 performed well in the external validation, maintaining a high AUC of 83.96% and a good combination of sensitivity and specificity of 70% and 81%, respectively. The recording of symptoms followed a different process compared to the development phase of the calculator, and this change was made to accommodate the adjustments that had to be put in place in the hospital during the lockdown resulting from the first wave of the COVID-19 pandemic in the UK. The symptoms and signs recorded during telephone triaging showed some variations compared to the development cohort. This likely reflects the way the questions were asked by the clinicians participating in the study that did not receive any training prior to running the telephone clinics. They were doctors of various grades, not necessarily with an ENT sub-specialisation, with some not having done HaN clinics previously. This was likely reflected in the way the symptoms were asked, with many patients that were subsequently found to be cancer-free reporting having symptoms previously known to be strongly associated with a cancer diagnosis, such as stridor, odynophagia, and dysphagia. The patients' demographics were directly comparable in the development and validation phase of the calculator. Symptoms that remained significant in the validation phase were unintentional weight loss, hoarseness, sore throat, neck lump,

feeling of something in the throat, dysphagia, odynophagia, neck lump and oral swelling but the level of statistical significance and their odds ratio at the different levels of each of the variable dropped showing a less strong association with a HNC diagnosis. Despite that, the combination of the symptoms in the multivariate regression model allowed very good discrimination of the malignant versus benign cases. The patients included in the validation study had already undergone triaging at the GP level based on the NICE guidelines, and they were all initially on the USOC pathway. Of the total of 4,557 patients, only 31.4% were triaged as high risk using the HaNC-RC v.2 and 21.6% as moderate risk for HNC. This resulted in a reduction of the USOC appointments by 70% during the first wave of the pandemic. Of the total of 256 cancers, 73.2% were seen in the high-risk group and 16.5% in the moderate-risk group. These figures show a much improved HNC diagnosis for the HaNC-RC v.2 triaged high and moderate-risk groups, as currently only 40% of HNC are diagnosed via the 2ww route using the NICE guidelines referral system (Langton, Siau and Bankhead, 2016). Using the HaNC-RC v.2, only twenty-six cancers (10.2%) were misclassified to the low-risk group at the end of the 6-month follow-up period. This is a 4% reduction of HNC identified via the routine route when compared to the Glasgow data used to develop the HaNC-RC v.2, as seen in Table 4-14. No information was collected on the cancer stage at diagnosis to allow for assessment of the cancer stage in the misclassified group of cancer cases. However, looking at the cancer sites, there were more cancers arising from areas adjacent to the HaN manifesting with HaN symptomatology compared to the development cohort, which also reflected in the number of primary HNCs in the cohort (76.4% in the development phase vs 59.5% in the validation phase). This difference may have also affected the symptoms presentation and the differences noted in the odds ratios.

5 Discussion

This chapter will first cover a summary of the thesis results relating to the initial aims and objectives of the study. The summary will be followed by a critical discussion of the output of this research compared to the relevant available literature. The HaNC-RC v.2 will be compared with other cancer calculators and the current guidelines for HNC referrals. The later sections will address how the results can be used in clinical practice so that the HaNC-RC v.2 can be incorporated as a triaging tool for HNC in the different clinical systems in the UK and worldwide. Since the publication of the thesis results (Tikka *et al.*, 2020), the calculator has gained popularity in the UK, and papers have been published assessing its use in clinical practice and its potential incorporation in future referral guidelines. These recent publications will be critically discussed, and the generated results and populations used for triaging will be compared to the HaNC-RC v.2 development and validation cohorts. The current limitations of the HaNC-RC v.2 and future directions will be covered in the last section of the discussion.

5.1 Summary of the study design and results

This research aimed to refine a previously developed HNC risk calculator (Tikka, Pracy and Paleri, 2016) to assess potential factors that can be added or adjusted to increase the risk calculator's predictive power. The study's design was prospective and was performed in two phases. The model development was performed in a prospectively collected dataset of 3,649 patients seen in hospitals across the Greater Glasgow and Clyde regions. The HaNC-RC v.2 included smoking and alcohol history, data that were not available in the previous iteration of the calculator. New symptoms were also included in the updated version: sore throat (with information on persistency and laterality), unintentional weight loss, stridor and HaN skin lesion. Symptoms previously presented in the HaNC-RC (v.1) were refined. The hoarseness symptom included subgroups based on symptom persistence and previous medical history that can affect hoarseness presentation. The dysphagia and neck lump symptoms were also updated to include information on persistence (intermittent or persistent). The previous version of the calculator included significant interactions between negative and

positive symptoms of cancer (feeling of something in the throat with otalgia and feeling of something in the throat with haemoptysis) (Tikka et al., 2016) which did not significantly alter the prediction power of the calculator this time; hence, they were dropped from the revised model. The addition of new symptoms and refinement of the existing variables addressed the first research question of this thesis, which was asking if such changes are possible in the updated version of the calculator. The refined HNC risk calculator (HaNC-RC v.2) had an increased predictive power compared to the first iteration of the HaNC-RC (v.1). The AUC increased from 77% to 88.6%, which was the research objective of this thesis, designing a HNC calculator in line with the predictive power of risk calculator of other cancer sites (AUC > 80%). Optimal probability thresholds were also identified for triaging patients into three risk groups (high, moderate, and low risk) based on the HaNC-RC v.2 generated HNC probability. This new triage classification aligns with the current allocation of clinic appointments across the UK to USOC, urgent and routine (NICE, 2015).

External validation was performed in a pan-UK prospective cohort of patients triaged with the HaNC-RC v.2 during the first wave of the COVID-19 pandemic via telephone hospital consultation. The triaging was performed via telephone consultations due to the restrictions on face-to-face hospital appointments during the COVID-19 pandemic. Forty-seven participating hospital sites were included in the validation database, totalling 4,557 cases. The predictive power of the HaNC-RC v.2 remained high at 83.96%, with a good sensitivity and specificity combination of 70% and 81%, respectively. The probability-based patient triaging identified 73.2% of cancers as USOC whilst reducing the total number of USOC appointments by 70%. Twenty-six cancers (10.2%) were misclassified as low risk, being much lower than the current literature average of 40% (Langton et al., 2016).

Therefore, the results of the external validation based solely on the HaNC-RC v.2 have shown a much-improved cancer detection rate compared to current standard practice whilst at the same time reducing the total volume of patients needed to be seen urgently. This answered the second research question of this thesis, which was how the HaNC-RC v.2 would perform in a new cohort of patients different to the one used

to develop the tool. Potential reasons for the improved cancer detection rate using the HaNC-RC v.2 will be covered in the following section, with an initial focus on the selection of the parameters included in the model compared to the literature of other HNC calculators but also the national and international referral guidelines as well as the reasoning behind the selected triaging thresholds.

5.2 The HaNC-RC v.2, in comparison with other cancer risk calculators

In this section, the HaNC-RC v.2 tool will be compared with other cancer risk calculators. These calculators have already been presented and covered in detail in the literature review chapter of the thesis. Here, the different aspects of a risk calculator cycle will be presented and compared for the HaNC-RC v.2 against each of the other risk tools, starting from the development and validation phase methodological aspects, followed by an exploration of the variables included in the calculators, their characteristics and interlink with model performance (AUC). The final focus will be on their applicability for use at the point of initial clinical consultation and the decision on thresholds for differentiating high-risk patients.

5.2.1 Similarities and Differences in the development phase of other cancer risk calculators compared to the HaNC-RC v.2

The data collection process and the statistical techniques used during the risk calculator's development phase will be discussed here and compared to other available cancer risk calculators. Detailed information on the other cancer risk calculators that are mentioned in this chapter is available in the literature review part (section 2.3) of this thesis.

Study design

The HaNC-RC v.2 model was designed based on a cohort study. This type of study ensured that the incidence of HNC was assessed against all potential causative factors. Similar study types were used for other HNC symptom-based calculators (Lau, Wilkinson and Moorthy, 2018;Moor, Paleri and Edwards, 2019;Tikka, Pracy and Paleri, 2016). A case-control study design has also previously been employed, such as for the RAT calculators (Hamilton, 2010). However, this design does not allow

information to be derived about cancer incidence against each causative factor. It can also limit the identification of rare causes of the outcome of interests as this would require acquiring a substantial volume of data even though this bias was minimised by using matching case-control methodology (Cole *et al.*, 2011).

Sample size evaluation

Prior to data collection, sample size calculation was performed as part of the HaNC-RC v.2 design, ensuring that the appropriate sample size was collected to produce accurate results. This step minimised standard errors. On the other hand, no a priori sample size calculation was done for any of the other HNC symptom-based calculators (Lau, Wilkinson and Moorthy, 2018;Moor, Paleri and Edwards, 2019;Tikka, Pracy and Paleri, 2016). The sample size calculation step was also omitted in other commonly used symptom-based cancer risk calculators, for example, the Qcancer risk calculator (Hippisley-Cox and Coupland, 2013a). In some cases, sample size calculation was not necessary due to the high number of cases included. For example, the Qcancer was based on a pan-UK GP registry of over 2 million patients in the derivation cohort (Hippisley-Cox and Coupland, 2013a) and therefore is a good representation of the general population. Large databases of over 20,000 patients were also used in developing each RAT cancer risk calculator (Hamilton, 2010). However, other cancer risk calculators that used similar cohort sizes to the HaNC-RC v.2 did not perform sample size calculation (Thompson, 2016; Steyerberg, 2007) (Gail *et al.*, 1989;Hoggart *et al.*, 2012;Wells *et al.*, 2014;Steffen *et al.*, 2014). For example, the ERSPC prostate cancer model is based on only 247 patients (Steyerberg *et al.*, 2007), with no sample size analysis available to confirm the suitability of the cohort for risk modelling purposes.

Data collection process

One of the main strengths of the HaNC-RC v.2 was the prospective cohort study design process, which minimises missing data and collection bias. It also reduced potential errors in data entry and interpretation. All previous HNC symptom-based calculators were based on retrospectively collected data; hence their results suffer from recall bias (Lau, Wilkinson and Moorthy, 2018;Moor, Paleri and Edwards, 2019;Tikka, Pracy and

Paleri, 2016). Well-known common cancer risk calculators also had a retrospective design process which questions the generalisability of their results such as the RAT models (Hamilton, 2010). Other cancer risk calculators have used a prospective design process, such as the Qcancer risk calculator reporting cancer probabilities for eight common cancers (Hippisley-Cox and Coupland, 2013a) as well as other single cancer-focused calculators (Thompson, 2016; Steyerberg, 2007) (Gail *et al.*, 1989;Hoggart *et al.*, 2012;Wells *et al.*, 2014;Steffen *et al.*, 2014).

Data collection proforma

The variables collected to assess for inclusion in the HaNC-RC v.2 model were included in a pre-design proforma. This was informed by a review of the literature on the common HNC symptoms and risk factors and also underwent assessment by clinicians experts to ensure that the collected data are clinically applicable, peer-reviewed and evidence-based. This robust methodology ensured that the first research question of this thesis would be covered, identifying any new potential factors for inclusion in the calculator and refining pre-existing variables. Many of the other available cancer risk calculators were developed without a predefined list of variables to be collected to inform their design (Thompson, 2016; Steyerberg, 2007) (Gail *et al.*, 1989;Hoggart *et al.*, 2012;Wells *et al.*, 2014;Steffen *et al.*, 2014). This was, for example, the case for the PCPT prostate cancer risk calculator (Thompson et al., 2006) and the subsequently introduced iteration to it (Ankerst et al., 2014) as well as the ERSPC prostate model (Steyerberg *et al.*, 2007) and the Gail and BCSC breast cancer models (Tice *et al.*, 2015) (Gail *et al.*, 1989). The use of already collected and populated databases could potentially introduce reporting bias due to the nature of data collection and interpretation. Significant features associated with a cancer diagnosis might not be reported as that was outside the scope of the initial database collection process. The available data may be non-informative, introducing inaccuracies in the model predictions. Later versions of the PCPT prostate cancer calculator addressed this issue, using large databases of many thousands of patients to update the PCPT to the currently used PCBG model (Ankerst *et al.*, 2018), and the BCSC model is based on millions of patients (Tice *et al.*, 2015).

Data analysis

Moving to the statistical analysis of the data, binary logistic regression methodology was employed to develop the HaNC-RC v.2. Logistic regression is the mainstay analysis tool for class outcome data. It provides good prediction metrics, and the results can be easily interpreted and used in clinical practice (Shmueli, 2010). This was also the statistical methodology for the previous version of the calculator and also for the second available symptom-based HNC calculator (Tikka, Pracy and Paleri, 2016;Lau, Wilkinson and Moorthy, 2018). The majority of the other risk calculators had also used binary regression analysis (Gail *et al.*, 1989;Hoggart *et al.*, 2012;Wells *et al.*, 2014;Steffen *et al.*, 2014;Steyerberg *et al.*, 2007;Thompson *et al.*, 2006) whereas nominal regression was also used when the cancer outcome had more than one categories (Ankerst et al., 2014). In logistic regression variable selection, the 0.05 level of significance has been established in medical research as the threshold for the inclusion of significant variables in the model (Zhu, 2016) and for this reason, it was also used in the HaNC-RC v.2. Interestingly, a p-value of 0.01 was used as a threshold to include variables in the most commonly used cancer risk calculator, the Qcancer model (Hippisley-Cox and Coupland, 2013a). This is not common practice as it can result in the exclusion of important parameters from the model that could have increased its prediction power (Zhu, 2016).

Aside from logistic regression, machine learning (ML) approaches have also been used in recent years (Hart *et al.*, 2020). Random forest analysis appears to have the most accurate results when the prediction power of alternative ML approaches has been compared for the development of a colorectal cancer risk calculator (Kop *et al.*, 2015). Hence it was also selected as the preferred ML method in this thesis. Nevertheless, the random forest has been criticised for focusing on improved computational abilities at the expense of the explanation of the results (Shmueli, 2010). This was also noted in the thesis when random forest analysis was attempted for the analysis of the data. Despite the fact that it achieved a high AUC of 82.9%, the variables considered most influential for the prediction outcomes were difficult to quantify as no odds ratio values were generated with this type of analysis. This difficulty in the presentation of the significant variables using ML approaches was also noted when this type of analysis

was used to develop a HNC risk calculator by Moor et al. (2019) that did not present which variables were significant in developing the HNC prediction model. The missing information makes impossible the external validation of the model (Moor, Paleri and Edwards, 2019). This is a known shortfall of machine learning, as its results are difficult to validate using a new database unless the same software is used and the coding functions are available. These are usually not well documented in the published papers as they would require a lot of text space – even as an appendix. Nevertheless, if such an effort is made by the authors of a paper, then the results can be reproduced and validated by others (Boulesteix and Strobl, 2009). The software, codes, and statistical libraries being used for the random forest analysis of this study are available in Appendix II, which enables the reproducibility of the results.

Aside from the applicability and ease of interpretation of the results issues discussed above, during HaNC-RC v.2 development, logistic regression outperformed random forest in prediction power. The AUC of the logistic regression model during the development phase was 88.5%, which was higher than the 82.9% achieved by the random forest analysis. These values may still look close enough, but the good performance of the AI approach was primarily due to a high specificity of 99.7%, but the sensitivity was very low at 23.5%. On the contrary, the high AUC of the logistic regression approach was achieved whilst maintaining a good balance between sensitivity and specificity (77.5% and 83.64%, respectively). This was also noted in the study by Moor et al. (2019), where the ML method had a very high specificity of 99.3%, but it came with a low sensitivity of 33.6% (Moor, Paleri and Edwards, 2019).

5.2.2 Similarities and differences in the validation phase of other cancer risk calculators compared to the HaNC-RC v.2.

The evidence of validation data in the available cancer risk calculators will be covered in this section and compared to the HaNC-RC v.2 validation phase. Despite numerous cancer risk calculators being available for potential use in triaging cancer referrals, a minority of them are validated in cohorts other than the development one, thus limiting their generalisability. This is one of the main strengths of the HaNC-RC v.2, as it has been externally validated by assessing how it performs in a separate cohort of patients,

which was one of the research questions of this thesis. Additionally, as discussed in section 5.5, the HaNC-RC v.2 has already been validated in another five published studies, with the calculator maintaining a high prediction power when used appropriately. The commonly used primary care triaging tools - RAT and Qcancer - as discussed in the literature review chapter, are validated in other primary care cohorts maintaining good discrimination ability (Hamilton *et al.*, 2013;Elias *et al.*, 2017;Hamilton *et al.*, 2005). Six colorectal cancer calculators (Williams *et al.*, 2016) and 2 oesophageal cancer calculators have also been externally validated (Li *et al.*, 2021). They have an excellent prediction power with AUCs above 0.89 - 0.91, even though the AUC dropped during external validations down to 0.76 for some models (Hodder *et al.*, 2005). A small drop was also seen for the HaNC-RC v.2 at external validation (88.5% to 83.9%). Unlike the HaNC-RC v.2, designed and validated in secondary care cohorts, the B-B equation model, the Fijten calculator and CAPER and Q-cancer for colon cancer models are developed and validated in primary care. All but one colorectal cancer models have been validated in primary care; hence the evidence for their use in triaging at GP level is strong (Williams *et al.*, 2016).

The small AUC drop that was seen in the validation phase of the HaNC-RC v.2 is also noted in other cancer risk calculators during the validation process, and it is to an extent expected as a different dataset is used that is not an absolute fit for the model (Hodder *et al.*, 2005;McCarthy *et al.*, 2020). The observed AUC drop could also be an after effect of the validation being done via telephone consultations, with data being recorded from doctors of variable grades in secondary care consultations. This can affect how patients' symptoms were interpreted, recorded, and subsequently used in the HaNC-RC v.2 algorithm compared to the interpretation of data by experienced HaN consultants who did the data recording during the design phase. The methodological differences between the development and validation phase of the HaNC-RC v.2 will be presented and critically discussed in the following section. It will provide insight into possible similar issues encountered when primary care validation is done and how they can be mitigated. It may be that a similar drop in AUC would arise if GPs collect data. This thesis was focusing on validation at triage in

secondary care, studies exploring validation at triage in primary care is a necessary next step for the HaNC-RC v.2 prior to widespread use.

5.2.3 The performance differences between the development and validation phase of the HaNC-RC v.2

The methodological differences between the two design phases of the HaNC-RC v.2 will be covered in this section. The validation phase of the HaNC-RC v.2 had different methods compared to the development phase due to COVID-19 pandemic-related constraints. These differences may explain the slightly worse predictive power of the HaNC-RC v.2 in the validation phase.

Firstly, there were no face-to-face triaging consultations during the calculator's validation. In comparison, ENT specialists saw all patients in HaN clinics face-to-face during the development phase. During the HaNC-RC v.2 external validation, telephone consultations precluded examination of the patient's neck, ears and oral cavity needed to identify some of the signs included in the calculator, hence relying upon patients' description of the signs if present. For example, oral examination identified an oral swelling in 5% of patients and an oral ulcer in 0.7% of cases during the development phase of the calculator. It was noted that in the validation cohort, the symptoms were reported more compared to the development cohort. For example, the oral swelling was mentioned in 6.1% of the consultations and an oral ulcer in 3.3%. Such examinations are routinely performed in primary and secondary care consultations as part of patients' initial assessments. The lack of face-to-face assessment during triaging may have contributed to missing, under or over-reporting some red flag signs that could upgrade the referral to 2ww and vice versa (i.e., oral swelling, oral ulcer, neck lump). Telephone consultation is a new method of initial patient assessment that emerged during the COVID-19 pandemic. Nevertheless, doctors did not receive prior training in this mode of patient assessment. A study has shown that only 1.6% of initial HNC consultations were done remotely before the pandemic, compared to 14% during the COVID-19 era (Tan *et al.*, 2021). Nevertheless, Tan et al. (2021) have shown that despite the increased use of telemedicine during the COVID-19 pandemic, no difference was noted in the total time

elapsed from the time of hospital referral to the time of the start of definitive treatment (Tan *et al.*, 2021).

Secondly, the telephone consultations were carried out by ENT doctors of all grades, including very junior staff, without receiving any prior training in performing the triaging. The lack of formal training and the clinical inexperience of some of the doctors could have also played a role in the lower detection rate of the calculator during the external validation process compared to the development phase when experienced ENT consultants performed the consultations. In order to mitigate this concern and ensure standardisation of the questions being asked, a guidance document was published. It was sent to all the participating sites to help during triaging, with a consultation script available for use alongside the HaNC-RC v.2 symptoms list (INTEGRATE, 2020). It was not possible to fully evaluate the impact of the published guidance on the correct use of the HaNC-RC v.2 by clinicians. But, for example, it was noted that the stridor symptom was reported considerably more in the validation cohort, being present in 2.2% of consultations versus only 0.3% in the development study. This is surprising as stridor is a rare symptom that, if truly present, needs an immediate emergency department review. The junior doctors doing the telephone consultations might not have asked or interpreted the symptoms in a similar way to that carried out by experienced clinicians. Despite the fact that the guidance was published in advance of the validation study, it may be that it was not read by all clinicians or not fully taken into consideration during the study.

Thirdly, the validation phase of the HaNC-RC v.2 was performed during the first wave of the COVID-19 pandemic, which could have affected patients' behaviours in describing their symptoms and timing their presentation to primary care during the "abnormal" pandemic health environment. Moreover, COVID-19-related symptoms could have been mistaken as HNC symptoms (i.e., sore throat). Having said that, a recent study from the USA has found no difference in the frequency and nature of the HNC presenting symptoms since the beginning of the pandemic, apart from a decline in the otalgia complaint (from 21% down to 11%) (Stevens *et al.*, 2021) so it appears that the pandemic was not a factor affecting how patients describe their symptoms.

The total number of symptoms and their duration also remained comparable with pre-COVID levels (Stevens *et al.*, 2021). So, it is anticipated that the symptoms listed in the HaNC-RC v.2 remained relevant during the pandemic waves. What did change, though, during the pandemic was the proportion of HNC diagnosed via an emergency route (Wilkie *et al.*, 2021). The number of HNC patients diagnosed following an emergency presentation has tripled since the beginning of the pandemic, with a 12.3% proportion of new HNC diagnosed via an emergency route compared to a range of 3.2 – 4.3% pre-COVID (Wilkie *et al.*, 2021). HNC cases also present in a more advanced disease stage with a larger median tumour size than pre-pandemic records (Kiong *et al.*, 2021) and an increased nodal stage (Stevens *et al.*, 2021). Similar issues with more advanced cancer at the time of initial presentation were seen in the past following natural disasters, such as after Hurricane Katrina (Carter *et al.*, 2013). The number of HaN referrals dropped during the first waves of the pandemic (Taylor *et al.*, 2020) but, over the last 12 months, has reached its highest-ever number recorded per single month(Metcalfe *et al.*, 2021) but with a larger proportion of patients diagnosed via other -emergency - routes (Wilkie *et al.*, 2021). Hence, the population of patients seen in HaN clinics is bound to be different to the pre-COVID era. This could have potentially affected the results during the validation phase of the HaNC-RC v.2. Indeed, the cancer incidence was 5.6% in the validation phase being much lower than the expected 8% (Langton et al., 2016) and the 8.4% recorded during the development phase of the calculator. The lower percentage of events (cancer) can introduce small-sample bias affecting the estimated error in the variables' estimates of the logistic regression model, reducing the generalisability of the results (King and Zeng, 2001). Therefore, the logistic regression model and generated AUC in the validation phase of the HaNC-RC v.2 should be interpreted by taking the above point into consideration. Future studies, repeating the validation process in the post-COVID era, ideally not during pandemic waves should be performed to allow for further estimations of the power of the calculator in a more stable healthcare environment.

Finally, an important difference between the two phases was that the final cancer diagnosis was reported at 6 months after the initial 2ww referrals during the validation study, whereas in the development cohort, the endpoint was at the time of initial review

and subsequent discharge from the clinic. The 6 months interval allows for further cancers to develop even if not present during the initial referral, as patients referred in the 2ww HNC clinics are at higher risk for cancer at any time (Scott *et al.*, 2020). It has been found that 4% of patients referred via the 2ww and subsequently discharged are diagnosed with cancer of any type within 5 years from that appointment. Of these, 17.5% will be in the first 6 months after their appointment. This is double the figures compared to patients seen in routine ENT clinics (Scott *et al.*, 2020). Taking these figures into consideration may explain some of the differences in missed cancers rates seen during the development and validation phase of the calculator, with 7 cancers (2.8% of the total) being identified late in patients discharged after the first telephone consultation following the combined use of the HaNC-RC v.2 and clinical judgement (Hardman et al., 2020), which was the same cohort used for the validation phase of the calculator. As one of the main concerns of any new intervention is the potential of a missed cancer diagnosis using telemedicine, online triage tools or other forms of triaging, it needs to be taken into consideration that there is currently a known percentage of late cancers emerging within 6 months (up to 5 years) following a standard 2ww HNC clinic review. The 4% 5-year cancer potential, 17.5% of this within 6 months (Scott *et al.*, 2020), should be included in the expected future cancer detection and not be pointed out as a failure of future triaging interventions such as the HaNC-RC v.2. Perhaps safety netting of these high-risk patients, who were found to be cancer free at the time of their 2ww appointment and they were subsequently discharged, could be an option. This could be a telephone or face-to-face followed by either their primary care doctor or secondary care after such an agreement on the preferred pathway is reached. Patients should be made aware of being at higher risk than the general population of a future HNC diagnosis and be alerted on the red flag symptoms (potentially given an advice sheet) that would require a further referral for specialist assessment.

To summarise, methodological differences between the development and validation of the HaNC-RC v.2 existed mainly due to the COVID-19 pandemic during the validation phase. In contrast to the face-to-face consultant-led consultations performed during the development phase of the risk calculator, the HaNC-RC v.2 was validated in remote

consultations done by ENT doctors of various grades without the availability of physical examination and no prior training was given in the use of the tool. A literature review did not identify any other risk calculator going through validation during similar unforeseen pandemic-like events. Further studies are needed to validate the tool in the post-pandemic era to allow a more representative comparison of its prediction power at external validation.

5.2.4 Types of data variables included in the cancer risk calculators

Aside from the methodological similarities and differences of the various cancer risk calculators, it is also important to focus on the variables included in the various cancer risk assessment models and how these compare to the HaNC-RC v.2. As these calculators are used for identifying cancer in patients based on various characteristics, the type of factors included in the model will be compared based on their attributes. Calculators that include only symptoms and demographic information will be covered first, followed by those that include results of investigations. This categorisation is important to appreciate how easily the calculators can be used in various clinical settings and the time scale needed for the collection of the information needed to use the calculators. The trade-off between prediction power and types of variables included in the HaNC-RC v.2 model will then be critically discussed based on the above comparisons to understand if the potential addition of more time and cost-demanding data could translate into higher prediction outcomes for the HaNC-RC.

5.2.4.1 The Symptoms and Demographics-only calculators (age, gender, social history factors)

The HaNC-RC v.2 calculator will be compared here against other cancer risk calculators that include symptoms and demographics variables only. The most important feature that will be used to measure the HaNC-RC v.2 performance against the other cancer symptom-based calculators will be the AUC value and the results of any external validation. When the comparison is made against the other available HNC risk calculators, the symptoms that are included in the calculators and the process resulting in their selection will also be critically discussed to assess if these decisions have affected the models' performance. Sensitivity and specificity combinations, statistical analysis methods and the presence of a validation process when available

for comparison will also be mentioned for the HNC calculators as a head-to-head comparison is more relevant.

The HaNC-RC v.2 is one of the few cancer risk calculators based only on symptoms and patients' demographics (age, gender, smoking and alcohol history). The HaNC-RC v.2 achieved a high AUC of 88.6 with a strong sensitivity and specificity combination of 85% and 78.3%, respectively, which remained high at external validation (AUC: 83.9%, sensitivity: 69.7%, specificity 84.4%) at the selected cut-off of 7.1% for USOC referrals. Its AUC and sensitivity and specificity combinations performed favourably compared to the Lau et al. (2018) HNC symptom-based calculator, which achieved a lower AUC of 0.79 (Lau, Wilkinson and Moorthy, 2018) based solely on patients' symptoms and demographics. It also had a very low sensitivity of only 31% despite a high specificity of 92%. This limits its use as a triaging tool as it cannot identify cancer cases adequately. Its selected probability cut-off was not mentioned, and no external validation has been performed (Lau, Wilkinson and Moorthy, 2018). One of the reasons for the low sensitivity figure could be that this calculator does not include important symptoms associated with a HNC diagnosis. As discussed in the literature review, dysphagia, odynophagia, and oral swelling symptoms are noted as red flags in many of the studies reporting symptoms in patients with HNC (Talwar et al., 2020; Queenan et al., 2018; Kaufman, Grabau and Lore, 1980; Dolam, Vaughan and Fuleihan, 1998; Tikka et al., 2016; Tikka et al., 2018). These symptoms were not part of the Lau et al. (2018) calculator. The retrospective data collection method in the Lau et al. (2018) study may have introduced reporting bias, resulting in missing information on the incidence of symptoms. Decisions taken during symptom grouping and coding could have also affected the diagnostic power of the tool. It is noted that some of the symptoms being coded as separate variables in the Lau et al. (2018) calculator could have potentially been grouped together, allowing more robust predictions. Such adjustments could have reduced the model's complexity and errors related to high correlation within the variables included in the model. To that end, collinearity assessment was not mentioned being performed in the methods section of the paper (Lau et al., 2018).

The methodology for developing the HaNC-RC v.2 was carefully designed to ensure that all possible symptoms related to HNC would be available for analysis. This was ensured by the design of a proforma for prospective data collection during the development phase. In the proforma, all symptoms that were identified as significant following the literature review of symptoms related to HNC were included. The addition of other symptoms, as free text, was also included in the proforma to ensure no other potential symptoms were missed. This approach resulted in common symptoms being included in the analysis, such as neck lump, hoarseness, sore throat, dysphagia, otalgia, oral ulcer and oral ulcer as mentioned in the literature (Brouha *et al.*, 2005b;Teppo *et al.*, 2003;Merletti *et al.*, 1990) but also less commonly reported symptoms that have a strong link with advanced disease and worse survival outcomes such as odynophagia and weight loss (Douglas *et al.*, 2018). This process answered the first research question of the thesis and resulted in an improved AUC and sensitivity/specificity combination of the HaNC-RC v.2 compared to its earlier version that had achieved an AUC of 0.77 (Sensitivity:74.8%; specificity:65.9%) at internal validation and an AUC of 0.81 (sensitivity: 79.3%; specificity:68.6%) at external validation (Tikka, Pracy and Paleri, 2016) (Tikka, Paleri and MacKenzie, 2018). Due to the retrospective nature of the previous version of the HaNC-RC, the weight loss symptom, HaN skin lesion, and stridor symptoms were not previously part of the model. One of the new additions to the HaNC-RC v.2 was also the sore throat symptom. As was mentioned in the literature review, this addition is supported by recent studies showing unilateral sore throat to have a 9.5% PPV in identifying HNC (Allam and Nijim, 2019). Some cohort studies reported sore throat percentages as high as 50% (Talwar *et al.*, 2020), with an average percentage of 24.7% (Table 2-5), based on the cumulative results following the literature review covered in the second chapter of this thesis. The multivariate model of this thesis found that patients with unilateral persistent sore throat were 9.7 times more likely to have cancer than individuals without a sore throat. Finally, previously identified important interactions between symptoms (otalgia and feeling of something in throat; haemoptysis and feeling of something in throat) that were part of the HaNC-RC v.1 were removed from the updated model, as now more significant indicators were identified, making the above interaction non statistically significant. The hoarseness symptom included subgroups

based on symptom persistency and previous medical history that can affect hoarseness presentation. The dysphagia and neck lump symptoms were also updated to include information on frequency. All these changes resulted in an increase in the AUC from 0.77 to 0.88 (Tikka *et al.*, 2020).

Aside from the HNC symptom-based risk calculators that were covered above, which are also presented in detail in the literature review chapter (section 2.3.3.12), the breadth of available symptom-based risk calculators for other cancers have been previously presented in section 2.3 of the literature review. Such calculators are available for oesophageal, bladder, colorectal cancer and prostate cancer (Williams *et al.*, 2016). The HaNC-RC v.2 compares favourably to these calculators as it has resulted in one of the highest AUCs for a symptom-based cancer risk calculator of 0.886 at internal and 0.839 at external validation, as was mentioned earlier. Only the colorectal calculators have achieved a slightly higher AUC of over 0.89 at external validation (Williams et al., 2016), whereas the bladder cancer models have achieved slightly lower AUCs of 0.79- 0.83 at external validation (Loo *et al.*, 2013; Tan *et al.*, 2019; Matulewicz, Rademaker and Meeks, 2020). The oesophageal cancer models lack external validation, but their AUC ranged from 0.68 and 0.894 during the design phase, being, on average lower than the HaNC-RC v.2 performance during the design phase (Xie and Lagergren, 2016). The first version of the ERSCP prostate cancer risk calculator is also based only on symptoms and demographics, but the AUC of this model has not been recorded (Steyerberg et al., 2007); (SWOP, 2022), hence direct comparison with the HaNC-RC v.2 is not possible. Further information on these calculators has already been covered in section 2.3.

In addition to symptoms and patients' age and gender, some calculators incorporate additional demographic factors: social deprivation status and family history information. For example, the commonly used primary care Q cancer tool includes information on deprivation status (Hippisley-Cox and Coupland, 2013a). Nevertheless, using a locally adopted deprivation score - such as the Townsend deprivation index used in Qcancer, which is mainly employed in the UK - could limit the widespread use and adoption of the tool (Adams, Ryan and White, 2005). Social

deprivation status is known to be linked with the presence of HNC (Conway et al., 2008) and also the advanced cancer stage at diagnosis (Olsen et al., 2015b). A decision was made not to include the Scottish index of multiple deprivation (SIMD) variable in the HaNC-RC v.2 models. This was based on the knowledge that a strong correlation exists between SIMD and smoking and alcohol (Zeitler et al., 2018). Conway et al. (2010b) noted that the effect of social deprivation in HNC diagnosis is not maintained after controlling for the smoking and alcohol status. Keeping SIMD in the model would have introduced multicollinearity bias resulting in a high variance of the variables' coefficients in the model with potentially large standard errors. This can result in unpredictable changes in the parameters' estimates, even after minor changes in the data (Greene, 2012). Moreover, the SIMD was the only index available for recording social deprivation during data collection. Including the SIMD in the calculator would have limited the use of the HaNC-RC v.2 to within Scotland only, whereas the study aimed to develop a triage tool that can be implemented internationally.

Further demographic additions to the HaNC-RC v.2 included smoking and alcohol status, which were also found to be significant predictors of HNC in the Lau et al. (2018) model, and have been found to have a strong link to HNC in large-scale studies (Lee *et al.*, 2020;Rogers *et al.*, 2019). Many other social history factors are linked with HNC presentation, with drug abuse, family history of HNC, and exposure to chemicals and asbestos being among the most common factors (Baan et al., 2009; Straif et al., 2009; Berthiller et al., 2009). Nevertheless, these factors are not commonly reported in consultation with patients with suspected HNC clinics (Douglas *et al.*, 2021a). Other factors considered but not included in the HaNC-RC v.2 were family history of cancer or past medical history. Such data are included in the Qcancer risk calculator (Hippisley-Cox and Coupland, 2013a) as well as some other cancer tools, such as the ERSCP prostate cancer risk calculator (Steyerberg *et al.*, 2007). This information was not collected during the development of the HaNC-RC v.2. This decision was taken after consultation with the participating consultants in the study. It was felt that this information is not commonly collected during HaN clinics, and collecting this additional data could increase the time per consultation that could not be

accommodated in the current clinical practice. It could also be a commonly forgotten point during data collection, resulting in a large volume of missing data. Nevertheless, the collection of such data may be useful during future updates of the HaNC-RC tool if the clinical setting allows it. It has been found that odds ratios of 10 or more are usually needed for the AUC of a model to change significantly when new parameters are added (Pepe *et al.*, 2004). Such univariate analysis should be performed first in future studies before attempts are made to adjust the HaNC-RC v.2 further.

In summary, the HaNC-RC v.2 AUC is directly comparable or better to other cancer risk calculators based primarily on patients' symptoms which are available for prostate, colorectal, oesophageal, ovarian and bladder cancer. The HaNC-RC v.2 also achieved a much higher AUC and sensitivity/specificity combination when compared to its previous version and also to the only other available HNC symptom-based calculator, answering the second research question of the thesis. The variables included in the HaNC-RC v.2 were carefully selected to ensure that they are clinically applicable and informed by the literature, steps that were lacking in the previously published HNC models.

5.2.4.2 Rick calculators incorporating radiological investigations and blood and other test results

Apart from the limited number of symptom-based only risk calculators described above, the others incorporate data from investigations in their algorithms to achieve high prediction power. The HaNC-RC v.2 achieved an AUC of 0.886 at internal validation and 0.839 at external validation based solely on symptoms and demographic variables. Other HNC calculators exist that incorporate variables other than symptoms and demographics, with their AUC being comparable to the HaNC-RC v.2. In particular, results of ultrasound findings are included in risk triaging for thyroid cancer (Choi et al., 2015; Kwak et al., 2013; Shen et al., 2019). These models have achieved a slightly higher AUC compared to the HaNC-RC v.2, ranging from 0.869 – to 0.903 but require secondary care referral for specialist assessment and neck ultrasound. Similarly, histopathology results are needed to use developed oral cancer prediction tools for assessing leukoplakia lesions that, despite that have achieved an AUC of 0.763 - 0.88, being lower or equal to the HaNC-RC v.2 (Zhang et al., 2017; Pereira et

al., 2016). Only the model by Liu et al. (2017) achieved a clearly better performance than the HaNC-RC v.2, having perfect discrimination with an AUC of 1, but excision of the suspected leukoplakia lesion and assessment by histopathologists is required, which precludes its use in primary care (Liu *et al.*, 2017b). Other tools require results of circulating tumour DNA (Arantes et al., 2018) or breath analysis of organic compounds (Dharmawarana et al., 2020), but their AUCs were 0.8 - 0.821 being lower than the HaNC-RC v.2 performance. Although the above-mentioned tests are only currently available in a research setting, they may become available in the years to come in primary or secondary care. However, it is unknown when such tests will be available on a sufficiently large scale to allow prospective assessment of their use in triaging facilities. A decision was taken not to include in the HaNC-RC v.2 results of investigations, such as blood tests or radiological examinations. This was to allow for the use of the tool at the time of the first clinical appointment without waiting for the results of any investigations. The HaNC-RC v.2 reached a similarly high AUC of 88.6% compared to the above-mentioned tools at internal validation, without the addition of investigations, hence reducing time delays for onward referral for specialist assessment and enabling its use in all clinical settings without significantly compromising prediction power.

The majority of risk calculators for other cancer sites are also based on a combination of symptoms and other factors. For example, the RATs are based on symptoms and haematological investigation results that can be collected in primary care, depending on the specific cancer being evaluated. No AUC values are available for direct comparison with the HaNC-RC v.2. Apart from the addition of investigations, another difference when compared to the HaNC-RC v.2 is that the RATs' predictions are based on the PPV combination of only 2 symptoms/results at a time (Hamilton, 2010). Nevertheless, it is known that only about half of the cancers present with obvious red-flag symptoms and that a single or combination of two symptoms has a modest PPV for cancer diagnosis (Ingebrigtsen *et al.*, 2013;Jones *et al.*, 2007). Therefore it has been suggested that the decision on referral should be based on combinations of more than two symptoms if present, and patients' social history and demographics (Ingebrigtsen *et al.*, 2013;Jones *et al.*, 2007). This is addressed in the revised HaNC-RC v.2 with

information on age, gender, smoking and alcohol, as well as a more detailed list of symptoms. Therefore, the RAT diagnostic capacities are limited due to the designed methodology, allowing predictions based on combinations of only two symptoms/investigation results at a time and no AUCs are reported.

The variables included in cancer site-specific risk calculators have been presented in detail in chapter 2 (section 2.3.3). Here, the calculators that include investigations will be briefly compared against the HaNC-RC v.2 based on their prediction performance and ease of use in clinical practice. The 6 most validated prostate cancer calculators require haematological and biopsy investigations (Louie *et al.*, 2015), making these tools difficult to use as a triage tool, but the ERSPC and PCPT model have simplified versions requiring only the PSA blood test, with results available within a day. Hence it could be used in primary or secondary care triage clinics. (Steyerberg *et al.*, 2007;Louie *et al.*, 2015).The AUCs of these models range from 0.66 to 0.79 in the validation cohorts (Louie *et al.*, 2015). Hence they perform worse than the HaNC-RC v.2 with an AUC of 0.839 AUC at validation, based only on symptoms and demographics. AUCs of 0.8-0.9, being similar to the HaNC-RC v.2 performance, have been seen for colorectal cancer calculators that include results of blood tests that can be done easily in primary care or at the point of the first review in secondary care (MCV and haemoglobin count) (Williams *et al.*, 2016). The HaNC-RC v.2 has directly comparable results to the best performing lung cancer risk calculators, being the Hoggart (AUC: 0.843) (Hoggart *et al.*, 2012) and the PLCO model (AUC:0.859) (Tammemägi *et al.*, 2013) but without requiring prior patient selection based on imaging. The lung cancer calculator could be used for initial patient triaging as it is based only on epidemiological factors but after a CXR has shown a nodule (Gray *et al.*, 2016). For other cancer calculators, such as for endometrial cancer (Tingulstad *et al.*, 1996) and kidney cancer (Harrison *et al.*, 2021), the variables included in the models are too complex and require scans and blood tests that require input from secondary care with a review of results by experienced clinicians. Some of these calculators have achieved a very high AUC of over 0.9 but at the expense of needing targeted specialised investigations (Morrissey *et al.*, 2015).

So, in summary, a good number of cancer risk calculators exist that have good prediction power, but this comes at the expense of the requirement of clinical investigation, which can limit their use in some clinical settings. The HaNC-RC v.2 can be used in all settings as it is based only on symptoms and demographics, having achieved comparable prediction power with other more complex cancer risk models. Apart from the AUC outcomes, the triaging ability of the cancer risk calculators is important in determining optimal thresholds for the allocation of patients in different risk groups. The availability of this last feature will be discussed next for the HaNC-RC v.2 and compared with other risk calculators.

5.2.5 Clinical applicability of the cancer risk calculators in patients' triaging

A common problem seen in the cancer risk calculator literature is that limited calculators exist that have made any strong recommendations for the optimum probability threshold to be used to aid decisions for an urgent cancer referral. In prostate cancer, the PCPT tool, which is one of the most well-researched calculators, has suggested the clinicians decide the threshold for referral, with a note in the discussion to consider a referral for a cancer probability of over 25%, without mentioning the sensitivity and specificity combinations in this arbitrarily selected threshold (Thompson *et al.*, 2006). Like the prostate cancer risk calculators, lung cancer prediction tools do not have information on optimal thresholds for discrimination of cancer cases (Gray *et al.*, 2016). In colorectal cancer risk calculators, cut-off thresholds are not mentioned at all or variably selected at different points per study (i.e. 0.5%, 16% or 40% of high-risk patients) with no firm recommendations of the optimum cut-off for urgent triaging (Adelstein *et al.*, 2010). The 1% threshold was recommended by an externally validated bladder cancer risk calculator for optimum discrimination and an AUC of 0.79 but without mention of the reason behind this selection (Matulewicz, Rademaker and Meeks, 2020).

A slightly different approach was used in Qcancer, providing model performance statistics for three risk thresholds based on the 90th, 95th and 99th centile, giving the threshold for the highest risk of cancer groups. The centiles corresponded to 4%, 7% and 19% cancer probabilities across all cancers, with different thresholds for individual

cancers. The first two thresholds are similar to those generated by the HaNC-RC v.2 methodology, but on all occasions, the combination of sensitivity and specificity favouring specificity, with the highest value for sensitivity being 59% in Qcancer (Hippisley-Cox and Coupland, 2013a). This is much lower than the 77.5% sensitivity achieved by the HaNC-RC v.2 whilst maintaining a high specificity value. Unlike the HaNC-RC v.2 user platform, the Qcancer tool does not provide any recommendations for the type of referral based on the probabilities generated (Hippisley-Cox and Coupland, 2013a). On the other hand, the HaNC-RC v.2 has clearly defined recommendations for urgent cancer referral thresholds based on the best possible combinations of sensitivity and specificity, with the optimum thresholds at the 2.2% and 7.1% cut-offs.

To summarise, HaNC-RC v.2 provides concise recommendations about triaging patients to low, moderate and high risk for HNC. It is based on statistical methodology optimising sensitivity and specificity concurrently based on logistic regression analysis. Some of the other risk calculators give less clear triaging instructions, making it difficult for clinicians to decide on the best probability threshold for a referral without compromising diagnosis or causing over-referrals of healthy individuals.

5.3 HNC referral guidelines compared to the HaNC-RC v.2

This section will look at how the HaNC-RC v.2 compares to commonly used HNC referral guidelines, with a particular focus on the NICE and SIGN guidelines. The prediction power of the HaNC-RC v.2 was significantly better than the NICE guidelines detection rates used to compare the calculator's results in the UK-wide external validation study. The HaNC-RC v.2 also compared favourably to the SIGN guidelines used as the comparative arm at the internal validation phase. Potential reasons for these differences will be discussed in the following paragraphs.

The UK USOC guidelines designed by NICE were one of the first guidelines for streamlining urgent referral to secondary care for suspected cancer cases in the worldwide literature, and many other later generated guidelines from other countries have been based on the NICE guidelines outline, namely Australia and New Zealand

(Cancer Council Victoria and Department of Health Victoria, 2021). A summary of the symptoms included in the NICE guidelines and the other worldwide referral guidelines is available in Table 2-1. The NICE HNC guidelines are the most audited, but many concerns have been raised over the years about their effectiveness in HNC diagnosis and that they are consensus-based (NICE, 2015). Audits have shown that the red and white patches in oral cavity symptom is not a good indicator of HNC (Shah, Williams and Irvine, 2006) despite that it is still included in the guidelines. The results of the HaNC-RC v.2 also showed that it is not a significant predictor of HNC. The presence of red/white oral cavity patches continues to be part of many other international guidelines (Cancer Council Victoria and Department of Health Victoria, 2021, European Head & Neck Society, 2020). The New Zealand referral committee has made an effort to make this symptom more specific to allow for better triaging of referrals. The guidance suggests an urgent referral with a red/white oral cavity patch only when associated with localised pain, swelling or bleeding (NZGG, 2009). Apart from the inclusion of non-significant or broadly described symptoms in the NICE guidelines, it has also been noted that other symptoms with well-reported links with HNC are not included in the guidelines, such as the dysphagia symptom (Talwar et al., 2020, Dolan et al., 1998, Pugliano et al., 1999) odynophagia (Mody *et al.*, 2021), unilateral otalgia (Merletti *et al.*, 1990), (Koivunen *et al.*, 2001) and weight loss (Douglas *et al.*, 2018). All these symptoms have been found to be significant for HNC diagnosis in this thesis and are included in the HaNC-RC v.2. They are also included in other international and national guidelines (Cancer Council Victoria and Department of Health Victoria, 2021, NZGG, 2009, European Head & Neck Society, 2020).

HaNC-RC v.2 also has favourable outcomes when compared to the Scottish referral guidelines. Even though the Scottish guidelines include the odynophagia symptom, other significant symptoms such as dysphagia and weight loss are not included in the guidance, whereas the red/white patch in mouth symptom is still in the guidance (cancerreferral.scot.nhs.uk, 2019). Audits have shown that the HNC detection rate is better in the Scottish cohorts, with a rate of 12% (Douglas, Carswell and Montgomery, 2019) compared to a modest 8% reported in the English cohorts, based on the NICE

guidelines (Langton, Siau and Bankhead, 2016). It is possible that the inclusion of the odynophagia symptom in the Scottish referrals has played a role in the improved detection, as well as the fact that up until the most recent iteration of the Scottish guidelines, dysphagia was also part of the referral guidelines but was removed in the 2019 review with no reason given for this (www.cancerreferral.scot.nhs.uk, 2019). The dysphagia symptom is present in many international and national guidelines, such as the Australian and European guidelines (European Head & Neck Society, 2020). Nevertheless, there is much debate about the pathway of referral for patients presenting with dysphagia; as such, patients are usually referred initially to an urgent gastroenterology clinic to first rule out oesophageal malignancy. Nevertheless, dysphagia was the presenting symptom in many large-volume studies of thousands of patients looking at presenting symptoms in HNC cohorts (Tikka, Pracy and Paleri, 2016;Zeitler *et al.*, 2018;Douglas *et al.*, 2018). These results agree with the HaNC-RC v.2 results with dysphagia having a high PPV for HNC diagnosis and being a significant parameter in the HNC calculator.

Other highlighted issues are that the NICE and Scottish HNC guidelines are too broad, with no details about the symptoms' frequency or laterality. More details for each symptom could help perform the triaging more effectively, resulting in fewer patients being referred urgently (Shah, Williams and Irvine, 2006). Similar problems with a very broad description of symptoms, easily prompting an urgent referral, have been found for international and other national guidelines across the globe. For example, the American Cancer Society mentions “eating difficulties” as one of the red flag symptoms with no mention of the chronicity and frequency of this symptom, which can have a wide range of benign aetiologies (cancer.org, 2021). Another example is the recommendations of the society of Surgical oncology that include “persistent throat irritation” as a symptom to prompt referral for exclusion of laryngeal cancer (Shaha, Byers and Terz, 1997a). This is a very broad term that can be linked with common benign diagnoses such as globus resulting in a considerable increase in the number of patients referred urgently without this translating into a higher HNC detection. This was taken into consideration when the HaNC-RC v.2 was developed. The very commonly reported hoarseness symptom was subgrouped to account for its frequency

(persistent/intermittent) and any preceding benign event linked with hoarseness. The dysphagia symptom was also subgrouped to persistent or intermittent and the neck lump to fluctuating/regressing or persistent. Attempts to make the referral symptoms more specific have been made by the New Zealand and Australian referral guidelines committees (Cancer Council Victoria and Department of Health Victoria, 2021). For example, the Australian guidelines specify that a referral for persistent sore throat should be expedited when there is also the presence of otalgia and nasal blockage associated with double vision or eye swelling (Cancer Council Victoria and Department of Health Victoria, 2021).

The reason for the discrepancies in the included symptoms in the various national and international guidelines is that they are based on expert panel consensus. The reasoning for that is that there are limited studies from the primary care setting for patients presented in the GPs with HaN symptoms suspicious of malignancy to inform the guidelines (NICE, 2015, (Alho, 2006). It is mentioned in the guidelines that any available study from secondary care on the HNC presenting symptoms was not considered as part of the evidence-based pool. Nevertheless, there is no evidence in the literature that the way the patients mention the characteristics and frequencies of their symptoms in primary care differs from a secondary care review. On the contrary, when standardised proformas are used for the collection of patients' symptoms, signs and social history factors, there is a good proportion of concordance between the GP-filled forms and the subsequent secondary care-filled forms, with a small proportion of inappropriate referrals of less than 20% (Rosell Ferrer *et al.*, 2021). A slight tendency of some symptoms being over-ranked by GPs is noticed compared to the ranking at the specialist review. (Rosell Ferrer *et al.*, 2021). The HaNC-RC v.2 triaging tool, with its available online platform, can help towards this as it has a clearly described list of symptoms to be asked by the triaging health care professional.

Another argument for not including secondary care studies to inform the HNC referral guidelines is that patients' symptoms at the time of initial GP presentation may be different from those reported at the hospital at the time of diagnosis (NICE, 2015). Nevertheless, an older study from the 90s has looked into this particular matter and

found no significant difference in the distribution of symptoms between the first report time and at the time of diagnosis (Merletti *et al.*, 1990). The general patterns were similar, implying a limited effect of recall bias. For laryngeal cancers, the most common symptom was hoarseness, whereas supraglottic and hypopharyngeal cancers commonly presented with dysphagia, odynophagia, adenopathy, and otalgia. For each symptom, the ratio of the number of times it was reported at first presentation against the number reported at the time of diagnosis was 70-90% for hoarseness and dysphagia, 50% for odynophagia, otalgia, neck lump and cough, being low only for the dyspnoea symptom at 36% (Merletti *et al.*, 1990).

Since the latest update of the NICE referral guidelines, only one primary care-based study has been published, looking at the symptoms relating to laryngeal cancer diagnosis. This study is likely to be included in future iterations of the NICE guidelines, but it has been noticed that common red flag symptoms for laryngeal cancer are not part of the study results (Shephard, Parkinson and Hamilton, 2019). More general symptoms are included instead, which are commonly reported in primary care but have not been found in other studies to correlate with a HNC diagnosis. Such symptoms and signs are insomnia, second presentation with a chest infection and raised CRP (Shephard, Parkinson and Hamilton, 2019). Some of these symptoms are non-specific and quite general, and they can also overlap, acting as confounding factors that, if together added to a statistical model, may cause erroneous results. Symptoms such as neck lump, stridor and weight loss were not considered using this primary care dataset, which is interesting given the fact that these symptoms are commonly reported in secondary care studies (Zeitler et al., 2018) and also significantly linked with worse morbidity, especially for the weight loss symptom (Douglas *et al.*, 2018). The methods section of this paper did not mention the assessment of the multicollinearity of the included variables in the model, which questions the validity of their results (Shephard, Parkinson and Hamilton, 2019). Insomnia and chest infections were not significant indicators of HNC in the HaNC-RC v.2 analysis nor in other large-scale studies (Tikka, Pracy and Paleri, 2016;Zeitler *et al.*, 2018;Douglas *et al.*, 2018). High CRP was not part of the data collection for the

development of the HaNC-RC v.2 as its intended use is for triaging at the point of first clinical consultation without the need for results of investigations.

Therefore, future iterations of the HNC referral guidelines should consider taking into account studies from secondary care to better align the symptoms list to the significant findings of primary and secondary care. This is important as the list of recommended symptoms that are included in the guidelines can affect the strength of cancer detection. The results of this thesis based on a large cohort of patients referred to secondary care with the identification of statistically significant symptoms following univariate and multivariate analysis can add useful information for future versions of the HNC referral guidelines. Another suggestion is to abolish the concept of referral guidelines completely and instead develop a referral protocol based on the clinician's assessment and the HaNC-RC v.2 value. This protocol can then guide referral streams to routine, urgent and USOC clinics. Further work will need to be undertaken to ensure that this concept is clinically viable, but it has the potential to reduce the confusion related to numerous versions of referral guidelines that are also outdated on many occasions.

Apart from this significant step forward in improving the list of symptoms included in the HNC referral guidelines informed by the literature, it is also important to appreciate how the guidelines are used in primary care for identifying cancer cases. The selection of the HNC guidelines referral thresholds will be discussed next, how they are currently implemented in primary care practice and how they compare with the HaNC-RC v.2 referral thresholds.

5.4 HaNC-RC v.2 triaging thresholds and comparison with the current referral guidelines

This section will critically compare the thresholds for a referral based on the current HNC referral guidelines against the HaNC-RC v.2 triage thresholds. It has been argued that the NICE cancer guideline network has adopted an angle of encouraging GPs to refer as many patients as possible to secondary care, using an imprecise and relatively broad and non-specific list of symptoms and signs in an attempt to capture malignant

cases, accepting a very low specificity level (Rogers, Dailey and Langton, 2021). Nevertheless, as a counter-argument, NICE's aim when developing the current guidelines was to ensure no cancer is missed without assessing the potential effect of the low referral threshold on the overall number of referrals (NICE, 2015).

The HaNC-RC v.2 referral threshold for a USOC clinic was 7.1%, resulting in a sensitivity of 85% and a specificity of 78.3% at internal validation at that cut-off. This is much higher than the outcomes from a recent pan-UK study for all cancer types using the USOC NICE referral guidelines (Burton et al., 2017). The study looked at the sensitivity and specificity outcomes of all USOC referrals across all English GPs. They found a mean sensitivity of 47.4% (95% CI 47.2 %– 47.5%) with a specificity of 87.8% (95% CI 87.7% - 88%) in the total of 5,479 practices. One-fifth of the GPs (n=1091) were outliers by having a much lower referral threshold (high referral rate), and 16.9% (n=928) had a higher than the mean referral threshold, hence a lower referral rate interpreted by high specificity. Over half of the GPs (n=63.2%) were within the mean values for both sensitivity and specificity. The referral threshold for a USOC referral currently lies at a PPV value of 3% based on individual symptoms as per the NICE referral guidelines (NICE, 2015). Nevertheless, as seen by the Burton et al. (2017) study across all English GPs, the guidance thresholds have failed to achieve high sensitivity, reaching a modest value of 47%, showing that most cancers are diagnosed via routes other than the USOC clinics. This agrees with Langton et al.'s (2016) findings that only 40% of HNC are diagnosed via 2ww clinics.

The NICE cancer guidance group arbitrarily decided on the 3% threshold in an attempt to accommodate the population demand for investigating any individual with a 1% or more potential of malignancy (based on (Banks *et al.*, 2014), balancing it against the economic costs of referral and investigating almost all patients referred with symptoms even remotely linked with cancer (NICE, 2015). The 3%-4% threshold has been found optimal in other diseases for initiation of treatment (Patel *et al.*, 2021), but no such research exists specifically for cancer diagnosis. However, based on optimal learning curves analysis of decision thresholds, it has been suggested that these thresholds likely have transferable results for other conditions, including cancer (Patel *et al.*,

2021). These studies formed the basis of the 3% PPV cut-off for the NICE referral guidelines across all cancers. A recent study looking at the impact of lowering further cancer referral thresholds from 3% to 2% has shown a calculated 8% increase in the patients requiring referral which was deemed manageable but without a noticeable improvement in cancer detection. Reducing the threshold further to 1% will lead to a 136% increase in referral rates (Moore *et al.*, 2021). Further research has shown that lowering the referral threshold further can result in an increased burden on the secondary care capacity of urgent cancer clinics without significantly improving the cancer yield (Kostopoulou *et al.*, 2019). This is in agreement with the results of this thesis, as the PPV threshold was much higher - at levels over 20% for each of the included symptoms in the HaNC-RC v.2.

A HNC NICE referral guidelines audit by McKie et al. (2008) has highlighted an increase in waiting times for 2ww appointments from 7 to 9.5 days and an increase in the number of GP referrals not meeting the NICE criteria. Despite the high sensitivity of 83.9%, the specificity of the NICE guidelines was found to be low, making it an inappropriate tool for effective screening. A proposal for further changes is highlighted with the possibility of a 10% PPV cut-off for a symptom to be included in any new update of the guidance(McKie *et al.*, 2008). This was found to slightly reduce sensitivity (from 83.9% to 79.1%) but increase specificity to 48.2% (from 30%), with other suggested adjustments including taking into consideration smoking status (current or previous smokers) for patients referred with hoarseness (McKie *et al.*, 2008). The development of the HaNC-RC v.2 has addressed some of these points. Smoking status is included in the model, and the PPV of included symptoms is much higher than the recommended 3%. Symptoms inclusion in the final version of the model was not based on their PPV but on the significance threshold (p-value) at multivariate analysis. Nevertheless, it was identified that even the symptoms not statistically associated with a cancer diagnosis had a PPV over 3%, as can be seen in Table 4-8. The PPV for the non-significant symptoms ranged between 0 and 14.5%, whereas the statistically significant symptoms had PPV between 21.1% and 50%. Hence, the above evidence suggests that there is a need for a shift away from arbitrarily decided low PPV values per individual symptom and a possible move towards

algorithms that incorporate multiple statistically significant factors enabling cancer diagnosis or at least a higher PPV value threshold for each individual symptom that is informed by the literature.

PPV values above the 3% threshold have been found in other USOC HNC studies (Allam and Nijim, 2019; Tikka et al., 2016) hence again questioning if the PPV threshold of 3% is appropriate for discrimination of HNC referrals. The PPV of the symptoms currently in the guidelines may be inflated, as some symptoms may be recorded only in consultation with patients that GPs suspect from the outset of a cancer diagnosis (Emery and Vedsted, 2015). This bias was eliminated in this thesis by using a set proforma of symptoms to be considered in all consultations. Introducing proformas as part of the primary care referral process has been found to increase the accuracy of symptom recording and reduce the proportion of missing data, especially concerning social history factors for HNC (Rosell Ferrer *et al.*, 2021).

The above findings show that the PPV symptoms thresholds to inform HNC USOC referral have not led to the expected outcomes for high cancer detection; hence it should not be used as a marker to assess the adoption of other triaging methods. The HaNC-RC v.2 has excluded symptoms that had PPV over 3% but were not found to be significant for HNC diagnosis following univariate and multivariate analysis. Doing that has achieved a higher sensitivity and specificity combination at the 7.1% probability cut-off compared to the NICE guidelines triaging (Burton et al., 2017). The overall PPV of the HaNC-RC v.2 at the USOC threshold was 27.2%, much higher than the modest NICE symptoms-based PPV of 3%, without affecting the sensitivity and specificity power of the tool. The two PPV values cannot be directly compared as the former is a combination of multiple variables resulting from the logistic regression analysis, whereas the latter corresponds to individual symptoms' PPV thresholds. It can give a scale of the HaNC-RC v.2 tool's positive prediction power, which has a much higher positive cancer yield, with over a quarter of the patients having a probability threshold of over 7.1% being diagnosed with cancer. On the contrary, the NICE guidelines require 100 patients to be seen to identify 3 cancer cases per reported symptom. Balancing the use of secondary care resources and early cancer detection is

the key to using the guidance in each full potential and any other similar triage resources. The above shortcomings of the design of the referral guidelines based on PPVs support the need for a change of the referral pathways to a more standardised referral protocol, guided by the clinician's evaluation of the patient and the cancer probability value generated by the HaNC-RC v.2.

There is a pressing need for timely cancer detection, especially after the introduction of the 28-day rule by the NHS Long-term plan, indicating that patients with cancer should get their diagnosis within 28 days from referral to the hospital (NHS England, 2019). With an ever-increasing number of inappropriate 2ww referrals, this is difficult to be achieved, causing a significant burden to the secondary care facilities (Araghi, Harris and Kyzas, 2020;NHS England, 2019). The rise of the referrals to secondary care is evident from both GPs and dental practices, and it is expected to rise further due to fears over litigation and an increase in fitness to practice cases, leading to defensive referring of an ever-growing number of patients without red-flag symptoms and signs (Tajmehr, 2019;Roy and Anjum, 2018;Grey and Walsh, 2019). Reports have also started linking the higher volume of referrals to increasing patient-related anxiety for a potential cancer diagnosis as more patients are now eligible for referral via the 2ww pathway, as well as a high burden caused to the hospital resources from a large number of referrals (Rogers, Dailey and Langton, 2021). The current results of the HaNC-RC v.2 have shown that its use as an aid for HNC triaging can reduce to a third the number of 2ww referrals, rationalising referral thresholds, and using an evidence-based approach without compromising the cancer detection rates. This can help in reducing the waiting times for 2ww appointments.

As mentioned in earlier sections, secondary care triaging by ENT doctors was introduced across the UK during the first wave of the pandemic, resulting in the database used for the external validation of the HaNC-RC v.2. Since then, other authors have also published their experience of using the tool in clinical practice. The next session will review these studies to gain further insight into the real-life clinical applicability of the calculator.

5.5 Direct uses of the HaNC-RC v.2: How the HaNC-RC v.2 is currently used as a triaging aid in the literature

This section will cover the available evidence of the applicability and effectiveness of the HaNC-RC v.2 following the publication of the first phase of its development in January 2020 (Tikka *et al.*, 2020). A few researchers attempted this triaging model during the COVID-19 pandemic, and the results of their studies will be discussed below.

The results of the development phase of the HaNC-RC v.2 were published in January 2020, just a few months prior to the first wave of COVID-19 in the UK (Tikka *et al.*, 2020). The multivariate analysis of the validation phase of the HaNC-RC v.2 has not yet been published. However, the same cohort of 4,557 patients that was used for the validation phase of this thesis was also the basis of a published pan-UK audit led by ENTUK and BAHNO assessing the use of the HaNC-RC v.2 in conjunction with clinical judgement in remote triaging of patients during the first wave of the COVID-19 pandemic (Hardman *et al.*, 2021). The results of this study have only briefly been mentioned in 5.2.3, where the impact of the clinician's overruling of the HaNC-RC v.2 tool was discussed. Here, the performance of the HaNC-RC v.2 triaging with and without clinicians' overruling decisions will be presented in comparison to the external validation results of this thesis, followed by how this is translating into clinical outcomes of efficacy. In the pan-UK validation study, only two of the HaNC-RC v.2 triage outcomes were used for patients' classification: either high risk or low risk for HNC, where the low-risk group incorporated the moderate and low risk of HNC groups as per the initial HaNC-RC v.2 publication (Tikka *et al.*, 2020) due to the COVID-19 constraints. Because of this difference, the performance of the tool was different to the external validation results of this thesis. The HNC outcomes were reported 6 months from the beginning of the study. The HaNC-RC v.2 achieved a sensitivity of 73.2%, specificity of 71.1% and NPV of 97.8%. In comparison, the external validation multivariate regression analysis of this thesis has shown a sensitivity of 70.08 % (95% CI: 66.54 - 73.23) and a specificity of 81.09% (95% CI: 77.99 - 83.96). The sensitivity of the tool in the Hardman et al. (2020) study was inside the 95% confidence interval

found during the multivariate regression analysis of the same cohort, but the specificity fell outside this interval by 6.8%. With the addition of the clinicians' decision for an urgent patient's review based on the HaNC-RC v.2 outcome and clinical judgement, the sensitivity increased to 92.5%, with a lower specificity at 49.1%, with only 19 cancers being missed (7.5%). Of these, seven cancers (2.8%) were diagnosed in patients that were discharged with no planned follow-up after the initial telephone consultation, and 12 cancers (4.7%) were seen in patients that had a deferred appointment being initially triaged as low risk for cancer with no face-to-face appointment (Hardman *et al.*, 2021). This audit has shown that combined use of the HaNC-RC v.2 as a triaging aid alongside clinical judgment has reduced the 2ww wait face-to-face reviews by almost 50% during the first COVID-19 wave, whilst achieving a high sensitivity of over 90% in HNC diagnosis. A harm analysis has shown that potential harm from a non-urgent review was caused in 11 cases (0.24% of the total patients triaged) (Hardman *et al.*, 2021). Having not been in a pandemic environment, the moderate risk group based on the HaNC-RC v.2 triage would have been seen within 4 weeks, which would have reduced the number of cancers seen late from 68 based on the calculator alone down to 26, as seen in the validation results of this thesis (see Table 4-19). The use of HaNC-RC v.2 as a triaging aid with clinicians' final decision based on the calculator result and their clinical experience has shown a missed cancer percentage of 7.5%, being much lower than the 40.1% found during the data collection stage of this thesis that also agrees with the results of the systematic review by Langton et al. (2016) showing that 40% of HNC are diagnosed by non-USOC routes.

Since that initial pan-UK validation study, a few other papers have been published looking at the prediction power of the HaNC-RC v.2 during the COVID-19 first and subsequent waves (Banerjee *et al.*, 2021),(Kaddour *et al.*, 2021) (Li *et al.*, 2022). Kaddour et al. (2021) reported their experience using the HaNC-RC v.2 for remote triaging in a relatively large cohort of 412 patients. Only 20% of patients were in the high-risk groups, with the majority being classed in the low-risk group (60%). No cancer was found in the low-risk group following a 12-month review. Of the total of 28 cancers, 20 were in the high-risk group and the remainder in the intermediate-risk

group. This gave a conversion rate of almost 1 in 4 for cancer in the high-risk group (20/86 patients - 23.3%). (Kaddour *et al.*, 2021). It should be noted here that in this study, the remote triaging was performed only by consultants, hence ensuring that the HaNC-RC v.2 questions were asked and interpreted by doctors experienced in assessing patients with HaN symptoms. This may be one of the reasons that no cancers were found in the low-risk group in this study. An interesting cost analysis done as part of the Kaddour et al. (2021) study showed that using the HaNC-RC v.2 for remote triaging and patient prioritisation for 2ww clinics has resulted in £200 savings per clinic, extrapolated to £10,300 for the 412 patients assessed in their study and to around £50,000 per year assuming an average of 2000 referrals(Kaddour *et al.*, 2021). These resources could be potentially directed to increase slots for radiological or other investigations and to set up one-stop HNC clinics. Banerjee et al. (2021) reported the HaNC-RC v.2 telephone triage outcome in a small cohort of 64 patients. Over half of the cohort was stratified to the low-risk group (51.6%), avoiding an urgent face-to-face appointment during the COVID-19 first-wave restrictions. Unfortunately, there are a few methodological issues in this study which limit the generalisability of their results. The number of cases is too small, with only 3 cancers diagnosed in their cohort. (Banerjee *et al.*, 2021). Moreover, the remote triaging was performed by doctors of any grade, in contrast to the study by Kaddour et al. (2021). This difference might have contributed to one out of three cancers being triaged into the low-risk group, as the remote triaging was done by relatively inexperienced doctors in assessing HaN referrals. No data were available about the subsite of HNC in the one patient triaged in the low-risk group, his symptoms, social history factors and stage at diagnosis. These are important facts that would be useful to be reported in future studies looking at the performance of the HaNC-RC v.2, as these will form the basis of further iterations of the calculator to attempt to increase its accuracy further. Further larger-scale prospective studies are required prior to recommendations for a UK-wide use of the HaNC-RC v.2 tool for secondary care-based triaging but the results of the above-mentioned initial studies show a very good triaging performance and a positive impact on resource allocation.

Aside from the above-mentioned studies that all performed HaNC-RC v.2 aided triaging in secondary care, the first study of primary care use of the HaNC-RC v.2 tool has been recently published (Li *et al.*, 2022). In this study, cancer outcomes of 1,110 referrals prior to the introduction of the HaNC-RC v. 2-based triaging were compared to 913 prospectively collected referrals after its addition to the referral pathway (Li *et al.*, 2022). No cancers were missed (100% sensitivity) compared to 7 missed cancers in the comparison cohort prior to the use of the tool (sensitivity: 76.6%). The results also revealed that the GPs used the tool similar to the NICE guidelines recommendation, with the tool having a PPV of 4.3%, similar to the NICE recommended threshold of 3%. This means that the GPs, despite using the tool, identified most patients as being high-risk and referred them to the 2ww pathway. This reduced the specificity of the tool to 13.7% compared to the high number achieved at internal and external validation based on secondary care triaging (specificity: 81 - 83.6%) (Li *et al.*, 2022) and compared to the above-mentioned secondary care-based studies. These findings highlight again the need for training and engagement of the primary healthcare professionals in the use of the HaNC-RC v.2, followed by further larger-scale prospective studies prior to its potential rollout in primary care. Table 5-1 summarises the studies reporting triaging outcomes using the HaNC-RC v.2.

Table 5-1. Studies reporting HaNC-RC v.2 triage outcomes

Studies	Total patients	High-risk %	Low-risk %	Cancers	Low-risk group cancers	High-risk group cancers	Triaging by
Warner et al. (2020)	48	33% 16	31% 15				ENT Consultants (Secondary care)
Kaddour et al, (2021)	412	20.9% 86	60.2% 248	24	1	20	ENT Consultants (Secondary care)
Banerjee et al. (2021)	64	34.4% 22	51.6% 33	3	0	1	ENT Doctors, any grade (Secondary care)
Hardman et al. 2021	4557	31.3% 1429	47.1% 2145	254	26*	186	ENT Doctors, any grade (Secondary care)
Metcalfe et al. (2021)	340	54.4% 185	45.6% 155	32	1	31	ENT Consultants (Secondary care)
Li et al., (2022)	913	83.8% 765	11.7% 148	37	0	37	GPs (Primary care)

* The study reported low and intermediate-risk group cancers together (n=68). The study data were obtained for further analysis as part of this thesis. Twenty-six cancer patients were found in the low-risk group. The dataset from this study was used for the validation phase of the HaNC-RC v.2.

The HaNC-RC v.2 tool was also part of a recently published review of the literature on HNC risk calculators (Smith *et al.*, 2022). It was graded as having a low risk of bias across all assessment domains (participants, predictors, outcome, analysis, overall), being also the only well-performing model to include symptoms and signs as predictive factors. Only one more model, by Koyanagi et al. (2017), was given a low risk of bias across all domains. This model and all others included in that review of the literature have also been presented in detail in section 2.3.3.12 of this thesis. The HaNC-RC v.2 was also within the best three overall performing models when AUC measurement, validation and overall quality assessment were considered, having marginally higher AUC compared to the other two models (AUC:0.8856 vs 0.87 for the other two models, by Amarasinghe et al., 2010 and Total et al., 2019). The only

criticism of the HaNC-RC v.2 was that the calibration plot of the model was not mentioned in the publication. This was indeed the case due to the limited availability of space and graphs that could be included in the manuscript. The HaNC-RC v.2 calibration plot showing good discrimination is available in section 4.2.6.4 (Figure 4-9).

The HaNC-RC v.2 publication has also helped form protocols and recommendation papers for the management of HNC referrals during and after the pandemic (Warner *et al.*, 2020;Metcalfe *et al.*, 2021;Doll, Braden and Thibeault, 2021) and was included in the Canadian Society of Otolaryngology position paper for management of HaN referral during the pandemic(Chan *et al.*, 2020) as well as the ENTUK and BAHNO associations (ENTUK, 2020). The HaNC-RC v.2 tools were also included in a recent systematic review of the literature on HNC risk calculators being rated amongst the highest tools in performance and quality of methodology, achieving low-risk bias (Smith *et al.*, 2022). The other tools were not based on symptoms and were assessing the future risk of HNC development in an asymptomatic population. These tools have also been covered in section 2.3.3.12 of this thesis. The use of the HaNC-RC v.2 has also been added to a protocol for managing HNC referrals during the pandemic in a large tertiary care centre in London (Warner *et al.*, 2020). In this protocol, low-risk patients can be discharged to GP with safety-net to seek further referral if symptoms worsen. For the intermediate-risk group, clinical judgement should decide if a patient can be discharged or if imaging is required before or after a face-to-face review (Figure 5-1). High-risk patients will have appropriate imaging as per their symptoms following clinical review. Preliminary results have shown that the protocol was feasible, it did not cause an increase in the radiology department workload for urgent scans, no cancers were missed, and a third of patients were safely discharged without imaging (Warner *et al.*, 2020). As with the previous study by Banerjee et al. (2021), the sample size in the Warner et al. study (2020) was very small, with only 48 patients and no cancers being diagnosed during their audit period. Therefore, no strong conclusions about the applicability of the HaNC-RC v.2 can be made from these studies.

A larger sample size study from the West Midlands, including 340 patients, has introduced a new pathway for managing HaN referrals, with hoarseness being the primary complaint, based on the HaNC-RC v.2 results (Metcalfe *et al.*, 2021). All low-risk patients were allocated to a nurse-led clinic, where clinical assessment, including flexible endoscopy, is performed by a trained advanced nurse practitioner (ANP). The authors defined this pathway as the "telescopic referral pathway" for low-risk patients, as seen in Figure 5-2 (Metcalfe *et al.*, 2021). High and intermediate-risk patients were assessed face-to-face by ENT consultants. Patients with neck lumps as part of their symptoms were assessed with USS prior to any further decision plan (Metcalfe *et al.*, 2021). The new pathway allowed high-risk patients to be seen within 5 days of referral, while the low-risk group was also seen quickly, with a mean waiting time of 15 days. All but 1 of the cancers were diagnosed in the high-risk group, with a conversion rate of 17% for the HaNC-RC v.2 high-risk group (Metcalfe *et al.*, 2021). The "telescoping pathway" images were stored by the ANPs and reviewed by consultants within 48 hours as a safety net to ensure no abnormal findings would be missed. This study has provided the first real-life scenario of using the calculator in triaging patients in secondary care with the integration of allied health care professionals for a review of the low-risk group of patients. Their structured pathway is seen in Figure 5-2 and can be used by other units to reform the HaN clinics in secondary care, optimising each unit's available resources.

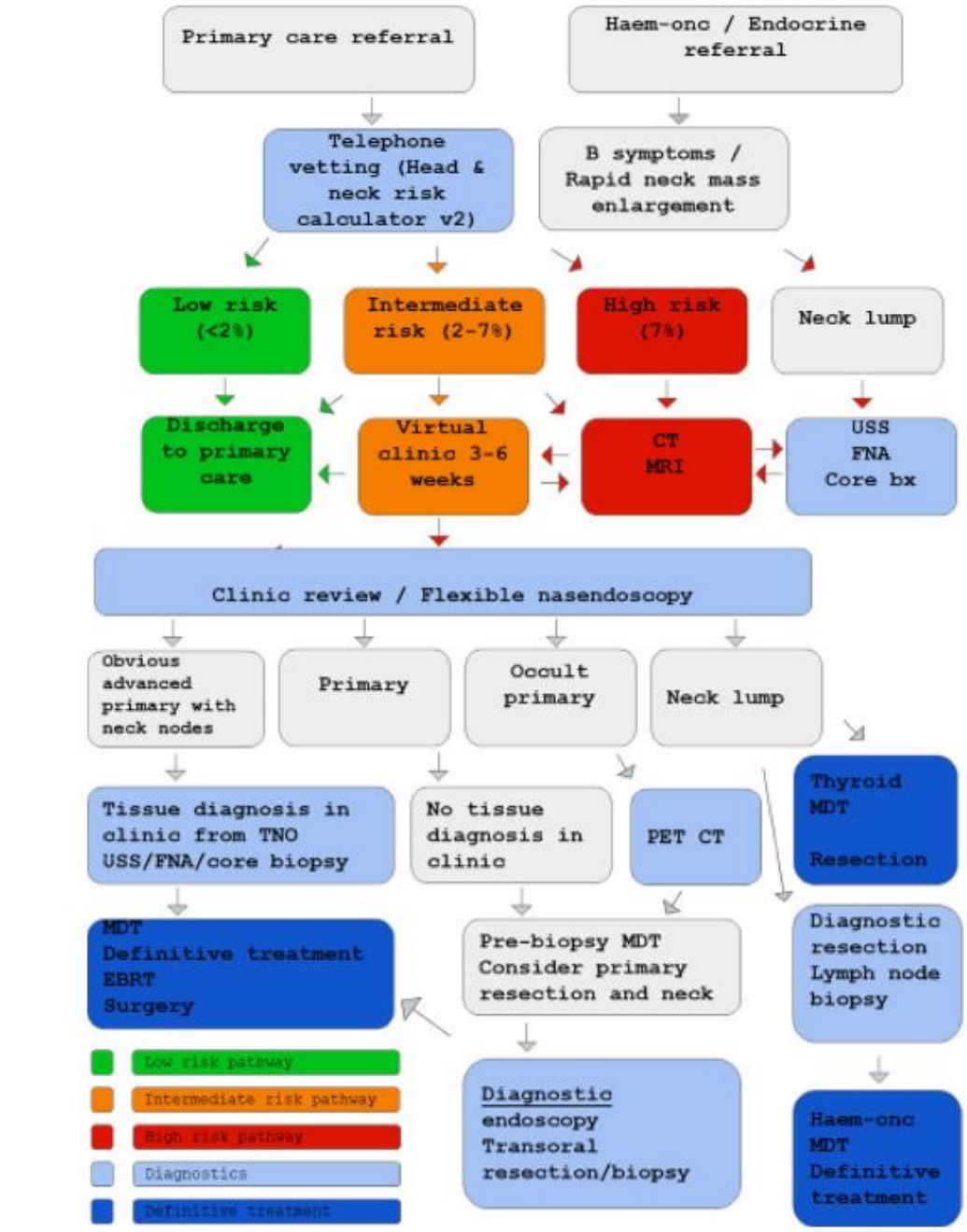


Fig. 1. Flow chart of the North East London Covid-19 protocol for diagnostics in two-week wait head and neck cancer patients. Haem-onc = haematology-oncology; v2 = version 2; CT = computed tomography; MRI = magnetic resonance imaging; USS = ultrasound scan; FNA = fine needle aspiration; TNO = transnasal oesophagoscopy; PET-CT = positron emission tomography computed tomography; MDT = multidisciplinary team; EBRT = external beam radiotherapy

Figure 5-1. The North East London COVID-19 protocol for diagnostics in 2ww pathway patients (Warner et al., 2021)

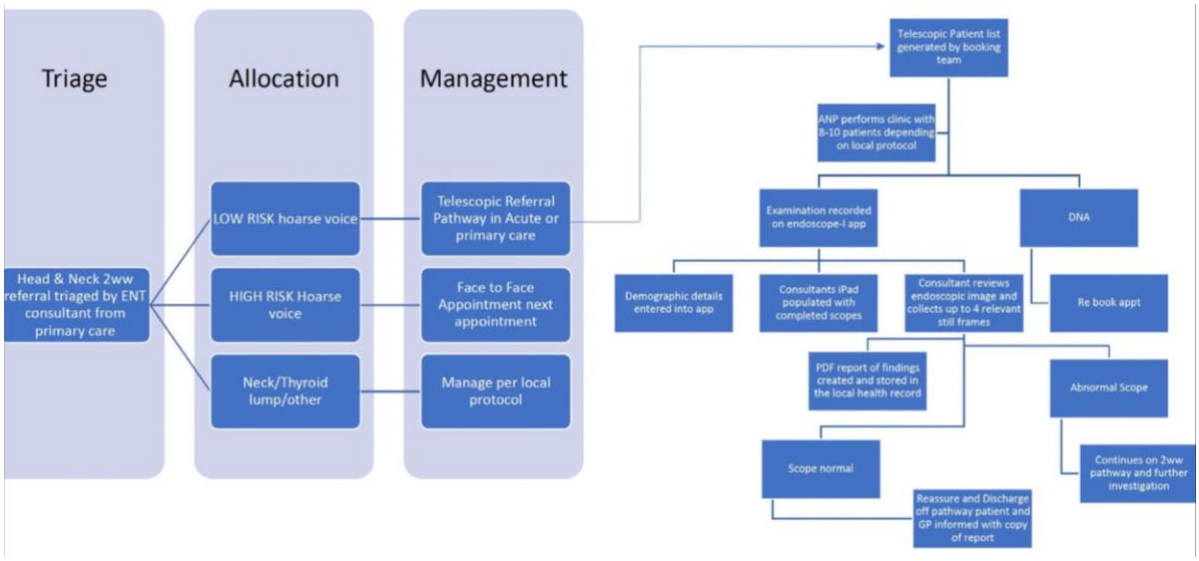


Figure 5-2. HNC referral pathway for head and neck referrals (Metcalfe et al., 2021)

Apart from the performance of the HaNC-RC v.2 in identifying cancer cases and reducing the burden for 2ww clinics, an important aspect that has not been studied so far is how patients perceive remote triaging as opposed to an initial face-to-face consultation. One study was identified looking at patients' satisfaction with remote triaging using the HaNC-RC v.2 with outcomes compared to a group of patients going through the standard face-to-face pathway (Zhu *et al.*, 2021). The study has shown that patients' satisfaction was between satisfied and very satisfied for all domains for both the remote triage and face-to-face groups, apart from the accessibility to healthcare services and effectiveness of consultation, which was between neutral and satisfied for the remote triage group. Over half of the patients were happy to receive phone consultations beyond the pandemic. The overall satisfaction received a score of 4.29/5 for the remote triage clinics compared to 4.54/5 for face-to-face consultation, being a non-statistically significant difference (p=0.24) (Zhu *et al.*, 2021). Other studies looking at patients' satisfaction with remote triage ENT clinics, not explicitly using the HaNC-RC v.2, also reported high overall patient satisfaction (87% - 98%) (Fieux *et al.*, 2020;Watters *et al.*, 2021) with 83% of patients open to the option of virtual clinics beyond the pandemic (Watters *et al.*, 2021). These are promising findings for implementing such clinics beyond the pandemic, but further studies will be required to look at any differences across different patient age groups and final diagnosis. Clinicians' satisfaction with remote triaging also needs to be explored, and the

amalgamation of these findings to be used for reforming HaN consultations in the years to come.

A multitude of studies, thus, demonstrate that the HaNC-RC v.2 can be an effective tool to support HNC referrals. The evidence so far supports the use of the tool in triaging referrals in secondary care, with the tool being used by ENT doctors, preferably experienced consultant clinicians. The use of the tool by GPs has also been audited in one study with no adverse outcomes. Further prospective cohort studies and non-inferiority randomised control trials will be needed prior to recommending any national-wide change with the incorporation of the HaNC-RC v.2 in primary or secondary care level triaging. The following section will discuss the possible scenario and rationale of incorporating the HaNC-RC v.2 into primary care triaging and the potential benefits and difficulties that may be encountered. Later sections will discuss potential scenarios for the implementation of the tool in alternative triaging settings. It should be made clear here that sections 5.6 - 5.7 cover hypothetical scenarios and aspirations of how the HaNC-RC v.2 could be potentially used in the future. Robust evidence from randomised controlled trials is needed prior to decisions for incorporation of the triage tool in any of the below-discussed clinical settings.

5.6 Primary care triaging of HNC referrals and the potential future use of HaNC-RC v.2

In previous sections, the HaNC-RC v.2 performance was compared against other cancer risk calculators and also against the current referral criteria for HNC. This section will focus on how the HaNC-RC v.2 could be implemented in clinical practice to improve outcomes. The HNC calculator could help remind primary care doctors of the ‘HNC red flag symptoms’ and ensure all questions relating to HNC are adequately covered during consultations. The reasons that an aid memoir is needed for HNC referrals and how the HaNC-RC v.2 could be used in helping GPs triage patients in primary care will be discussed. This section will also cover a discussion on possible difficulties in implementing the HaNC-RC v.2 in primary care consultations and potential solutions.

In the UK, Australia, New Zealand, and most European countries, a referral for suspected HNC is carried out in primary care as GPs are the first point of contact for any health-related problem (Boerma, 2003;Groenewegen, Schellevis and Boerma, 2016). It is envisaged that the HaNC-RC v.2 can be incorporated into the GPs' online software to help triage suspected HNC consultations to improve the referral rates and the HNC detection rate. Many studies have noted that the undergraduate medical school curriculum and the postgraduate GP training are limited in HNC teaching (Shah, Williams and Irvine, 2006). There is a lack of in-depth education for GPs on the signs and symptoms of HNC, resulting in a failure to understand and appropriately follow the HNC referral guidelines (Shah, Williams and Irvine, 2006). The undergraduate ENT/maxillofacial medical school curriculum is known to be very limited, followed at best by a short four to six-month hospital rotation or, in some instances, there is no formal post-graduate ENT rotation training, apart from scheduled didactic lectures (Langton *et al.*, 2019). This increases GPs' anxiety about missing a HNC diagnosis and highlights their lack of ENT knowledge and clinical experience (Langton *et al.*, 2019). Concerns are raised from primary care studies about the inexperience of GPs in appropriately identifying high-risk patients for HNC, mentioning that on average, a GP will see 3.2 HNC during a 30-year career, with 1 new cancer diagnosis being made for every 6000 patients or 12,500 GP visits (Nieminen *et al.*, 2021). Hence, using the HaNC-RC v.2 triage tool as a guidance of the red flag symptoms during the GP consultation would help ensure the right questions are asked to allow safe triaging of referrals. This could enable the development of a new referral protocol process based on GPs assessment and the probability value of the HaNC-RC v.2. The online version of the HaNC-RC v.2 includes a list of all significant HNC symptoms allowing the clinicians to ask them in a checklist format, reducing the chances of important questions being missed. Using this evidence-based pre-set list of symptoms and social history factors during the consultation should also help to overcome the problem of lack of knowledge of the HNC red flags.

Misinterpreting the frequency or quality of a symptom is another major problem in the primary care HNC referral pathway (Mettias, Charlton and Ashokkumar, 2021). A

prime example is the symptom of globus that many GPs interpret as dysphagia, and patients are accordingly referred via the 2ww pathway for suspected malignancy (Montgomery *et al.*, 2019). The HaNC-RC v.2 differentiates these two symptoms ensuring that the globus symptom will be asked separately from the dysphagia symptom to reduce confusion. The differentiation between these two symptoms is very important, as globus is negatively associated with HNC, whereas dysphagia is strongly positively associated with HNC diagnosis. Hence, asking these questions appropriately has an important effect on the triaging outcome. It has been shown that when patients are asked in an open-question format survey about their symptoms, they can describe the "feeling of something/lump in the throat", which is distinct from the GPs referral letter mentioning dysphagia (Montgomery *et al.*, 2019). This highlights the need for GPs' education on more appropriate interpretation and documentation of patients' symptoms, and the HaNC-RC v.2 can help to do this by including these symptoms as separate entities. To ensure that this will not lead to confusion, GPs will likely require further education on how to differentiate them, listening carefully to how the patients describe the symptoms using their own words. Further research will need to focus on extracting the symptoms information directly from the source -the patients- using this information in future versions of the HaNC-RC. This is currently investigated by the Evolution of a patiEnt-REported symptom-based risk stratification sySTe to redesign the suspected Head and Neck cancer referral pathway (EVEREST-HN) trial. This study aims to develop a patient-reported symptom-based inventory enabling the design of a primary care focused version of the head and neck cancer risk calculator, that will be compared in a non-inferiority trial setting to current standard practice. The HaNC-RC v.2 algorithm methodology will be used as the starting point to assess for addition and refinement of the variables to be included in this new patient-focused head and neck cancer triage tool (https://fundingawards.nihr.ac.uk, 2022).

It has been noted by several audits that there is a large number of referrals that are not appropriately completed, with information on symptoms or social history factors not included in the 2ww referral form (Miller and Hierons, 2012;Rimmer *et al.*, 2012;Ea, Harding and Courtney, 2008). If the correct symptoms are not asked, or the examination does not include an assessment of the throat or the neck, important

symptoms and signs may be missed and not included in the referral letters. The HaNC-RC v.2. standardises the referral process, introducing a mandatory tick-box approach to include all referral symptoms and important history factors; otherwise, no individual probabilities can be generated. This can reduce the frequency of missing data and ensure that all the important questions are being asked during the consultation. Such a solution was proposed back in 2006 by Duvvi et al., who also proposed that the referrals checklist should have a direct electronic link with secondary care. Telemedicine and IT resources as adjuncts to the 2WW referral pathway were also suggested by Singh and Warnakulasuriya (2006). The HaNC-RC v.2 has covered these suggestions. Its online version is freely available (www. Orlhealth.com), alerting the user of individual cancer probability prompting USOC or urgent referrals for the high-risk groups. The results can also be printed and attached to the GP referral letter.

Increased likelihood of non-compliance with the NICE HNC guidelines has been noticed in more recent years, with 5% vs 20% non-compliance in a 2-year cycle audit for earlier vs more recent years)(Haikel *et al.*, 2011). An audit of the quality of referrals has shown that symptoms were reported in 56% of referrals, and 55% noted symptom duration. Social history of alcohol and smoking was present in under half of the referrals (Hong *et al.*, 2016). Following a review of the referral and patient by a HaN specialist, it was found that only about 50% could be justified as 2ww referrals (Hong *et al.*, 2016). Hence, using a nation-wide proforma of referral symptoms is becoming very relevant, being designed to include only those symptoms and other referral factors with proven correlation with a new diagnosis of a HNC (Haikel *et al.*, 2011). The HaNC-RC v.2 list of symptoms, with its robust methodology, addresses this point, as all included parameters in the model are significantly associated with HNC diagnosis based on the multivariate regression analysis and have undergone internal and pan-UK external validation. The results of this thesis have shown that the use of the HaNC-RC v.2 standardised proforma and probability value for triaging has the potential to reduce by 70% the referrals that need to be seen within 2 weeks, as can be seen in Table 4-14 of the results section. Using the HaNC-RC v.2 to identify patients with low risk for cancer can also help alleviate patients' anxiety concerning referral for possible HNC. A study has shown that 77% of patients referred via the 2ww referral pathway felt

anxious about being referred to a HNC specialist. With a NICE HNC 2ww referral conversion rate of 8% for HNC (ranging from 5.6% to 12.4%), the large proportion of patients with an avoidably significant emotional burden can be reduced by a refined triaging of the referrals (Fingland *et al.*, 2018).

The above issues have the potential to be resolved using the HaNC-RC v.2 as a triaging aid. Nevertheless, studies exploring the use of other cancer risk calculators in primary care have shown a low pick-up from the GPs (Price *et al.*, 2019). Even though GPs are aware of the cancer triage tools, concerns have been raised about the lack of education in using the triage tools and IT problems making integration with the current GP software difficult, as well as concerns about GPs' time commitment due to tight time slots per GP consultation to allow for the use of such triaging tools (Bradley *et al.*, 2021). If the HaNC-RC v.2 is to be added to the current GP software, training of primary care professionals should be provided to ensure that the list of symptoms and social history factors is adequately covered, and the questions are being asked appropriately but also the necessary extra time is allocated to the consultations. HaN specialists should likely provide such training using a lecture-based teaching format or interactive group sessions, or the teaching material can be available in a standardised simple download package provided to GPs as part of their continuous professional development activities that can be accessed at any time.

Other reasons that are currently limiting the use of cancer risk calculators are concerns that the triage aids will not be endorsed by the secondary care resulting in the dismissal of the referrals that have used these tools as triage aids (Bradley *et al.*, 2021). However, recent evidence during the pandemic has shown that secondary care clinics were, in fact, prepared to use risk calculators themselves, with the HaNC-RC v.2 being endorsed by ENTUK and BAHNO for triaging HNC referrals(BAHNO, 2020). Forty-one secondary care centres across the UK took part in the validation process of the HaNC-RC v.2, and doctors of all grades in secondary care became familiar with the triaging tool (Hardman et al., 2021). Informal feedback during the audit period was positive from the doctors and other healthcare professionals using the calculator for triaging, as is evident by the wide adoption of the HaNC-RC v.2 during the pandemic

(Hardman *et al.*, 2021). However, no publications are available to date that formally report on healthcare professionals' satisfaction with the tool. An audit of primary care use of the HaNC-RC v.2 in Scotland has been recently published and was discussed in 5.5 of this thesis. No information was available on users' satisfaction or other qualitative feedback from its use, but the results suggest that the tool was used by GPs similarly to the national guidelines, as no reduction but rather an increase was noted in the total number of urgent cancer referrals. Nevertheless, no cancers were missed. (Li *et al.*, 2022). Agreement, training and active dialogue between primary and secondary care are of paramount importance in discussing referral rates and achieving a consensus on how to optimally triage the referrals using developed risk calculators. Otherwise, a non-constructive triaging process can only lead to over-referral to secondary care by GPs focusing only on releasing pressure from the primary care sector without using the developed triaging aids at their full potential (O'Donnell, 2000). Over-referrals can also result from pressure from the patients to be referred via the urgent cancer pathway and concerns that non-urgent referrals will take a considerable amount of time before secondary care review (Dodds *et al.*, 2004). The individually generated probability of HNC can be discussed with the patients, and the reason for an urgent or routine referral can be explored with them and explained. A low chance of cancer can help alleviate concerns and fears of a cancer diagnosis. This approach has already been used by some GPs when using cancer risk calculators (Bradley et al., 2020), and it can work well when GPs are using these tools or other guidelines as decision-aid tools rather than an obligatory activity prior to the referral process (Price, Abel and Hamilton, 2021).

Other issues being mentioned are that the currently available cancer risk calculators have lists of symptoms that are not in sync with the NICE guidelines. As was discussed in detail in the previous section of the discussion, the NICE guidelines are consensus-based due to a lack of primary care studies to inform the guidelines (NICE, 2015, (Alho, 2006). However, as was mentioned in the discussion above, it has been shown that secondary care results should be taken into consideration in future iterations of the NICE HNC guidance (Rossell-Ferrer et al., 2021, Merletti et al., 1990). It is hoped that the results of the HaNC-RC v.2 study will inform future updates in the HNC NICE

guidance, reducing GPs' concerns about the use of the tool. Finally, a great concern from the GP community is the fear that using a cancer risk calculator will replace their clinical judgement and make their role redundant. The design of the HaNC-RC v.2 has no intention to replace the GPs' role in referring patients with suspected HNC to secondary care but rather to work as an aid memoir of the HNC symptoms helping the clinicians in their referral decisions. Its result can always be overruled by the clinicians who will make the final call for a referral. This was noted on many occasions during the HaNC-RC v.2 real-life pan-UK audit during the first wave of the COVID-19 pandemic, where the clinicians overruled the HaNC-RC v.2 alerts in 21.9% of cases, requesting an urgent assessment as part of the triaging. In the audit, no moderate risk group was available due to the constraints in face-to-face appointments during the pandemic (Hardman et al., 2020). Of the total of 4,557 patients in the audit, 1,429 were triaged as high risk as per the HaNC-RC v.2. The cancer yield in this group was 13% (n=186, being 73.2% of total cancer diagnosed). Clinician's overruling resulted in additional 946 patients being triaged as urgent, with the cancer yield in this group being 5% (n=49) (Hardman *et al.*, 2021). The majority of these patients would have fallen in the moderate risk group (n=42 cancers, n=983 moderate risk patients) if that option had been available during the audit, as can be seen in Table 4-19. These figures can likely reflect the additional effect of GPs' assessment and intuition in cancer diagnosis using the HaNC-RC v.2 as a clinical aid.

To summarise, HaNC-RC v.2 can help overcome some of the current problems of the primary care triaging of suspected HNC cases stemming from GPs' inexperience in HNC consultations. Nevertheless, some concerns relating to the widespread use of such tools in primary care may take a long time to address. Prospective cohort studies and randomised control trials will be required to assess the use of the tool prior to any attempts for widespread implementation in primary care referral pathways. The results of the EVEREST-HN, as mentioned earlier, will provide evidence from a randomised controlled trial setting in the use of a refined version of the tool for patient-level use in primary care. Qualitative studies will also be needed to assess the views of patients, clinicians and other stakeholders on the use of the tools for primary care triaging. The following section will discuss alternative clinical settings for using the HaNC-RC v.2.

5.7 Alternative clinical settings for patient triaging and how the HaNC-RC v.2 can support them

This section will cover potential uses of the HaNC-RC v.2 in triaging settings other than GP-led, which is discussed above. I would like to make clear here that these are hypothetical scenarios discussing aspirations of the use of the tool in various clinical settings. Implementation of the tool in any of the following scenarios will first require carefully designed research to allow proof of concept with pilot studies, followed by robust methodology evidence from prospective cohort studies and randomised control trials, as well as qualitative studies.

The increased number of HNC referrals due to GPs' inexperience and fear of litigation has resulted in a communication gap between primary and secondary care in the UK (Langton *et al.*, 2019). Other European countries have noted better communication avenues, with direct discussions of possible urgent cancer cases over the phone with the referring hospital (Langton *et al.*, 2019). The HaNC-RC v.2 can be used in primary care, but such an endeavour may take considerable time to be implemented due to administrative and practical issues. Other avenues for using the HaNC-RC v.2 will be discussed to explore different options in HNC triaging. These alternatives could also be well suited for countries where GPs do not act as gatekeepers to secondary care. This is the case in the USA, Canada, and some European countries, as was discussed extensively in the literature review chapter of this thesis.

Secondary care triaging services can be established, taking as a paradigm the HaNC-RC v.2 implementation during the pan-UK service evaluation audit (Hardman *et al.*, 2021). Theoretical proposals for such a concept are not new but have not been clinically adopted before the pandemic. Talwar et al. (2020) have previously suggested implementing secondary care triaging that can be run by appropriately trained health care practitioners, allowing re-triaging of the GP referrals to 2ww or other clinics using the referral letter for triaging rather than telephone consultations. It was argued that the volume of potential HNC cancers seen in a GP practice is too small to allow for expertise to develop in triaging referrals in the community. As a result, GPs become

overcautious in an attempt to reduce the chances of missing cancer to the minimum, causing a significant burden to the secondary care urgent cancer services (Talwar *et al.*, 2020). The HaNC-RC v.2 can be used to re-evaluate the referral priority based on the GP referral letter, this time being done by a secondary care service. This service can be run by ENT doctors or HaN specialists as well as trained ENT nurses and other health care professionals such as speech and language therapists. To allow this to be achieved, the information in the referral letter needs to have all the necessary information on patients' symptoms and social history for the HaNC-RC v.2 to generate cancer probability allowing for appropriate prioritisation of HaN referrals. Alternatively, virtual or face-to-face clinics can run, with the secondary care professionals going through the questions in the HaNC-RC v.2 with the patient, then allocating them to 2ww, urgent or routine clinics. The latter was employed in the external validation of the HaNC-RC v.2, with virtual clinics run by ENT doctors, due to the COVID-19 constraints. High-risk patients based on the virtual consultation using the HaNC-RC v.2 were re-prioritised for a 2ww face-to-face appointment (Hardman et al., 2020).

5.7.1 Speech and Language therapy led clinics for triaging HNC referrals

This section will cover the evidence for SLT-led patient triaging and how the HaNC-RC v.2 could contribute to such services. Attempts to establish HNC triage clinics run by trained SLT professionals have been made in recent years outside the UK (Payten *et al.*, 2020). An Australian study has established 2 separate pathway clinics run by SLT for dysphagia and dysphonia symptoms(Payten *et al.*, 2020). Patients were allocated to these clinics following an initial clinical screening of the referral letters by the ENT team, and no triaging aid was used, allocating to the healthcare professional clinics patients deemed to be at low risk of malignancy or other complex pathology (Payten *et al.*, 2020). This new service was found to have significantly reduced the clinic waiting times by an average of 277 days to a significantly lower average of 68 days. In particular, 72% of patients seen in the dysphagia clinic and 81% of dysphonia patients were discharged after an average of 2.2 clinical consultations with no recorded adverse effects, with fewer than 10% of patients having an organic pathology. 42% of patients needed input by ENT at some during their clinic pathway, and only 4% were

re-categorised as a high priority (Payten *et al.*, 2020). Similar results have been reported by another Australian unit study, with two-thirds of patients being managed and discharged after SLT-led clinic review without the need for ENT clinic consultation (Seabrook *et al.*, 2019). The HaNC-RC v.2 may be a useful addition to this new clinical service, allowing for a more informed triaging process and perhaps allowing the triaging to be done solely by the health care professionals rather than requiring the initial input of ENT doctors. The lack of enough information in the referral letters for risk stratification of patients from the outset, without ENT doctors' input, was highlighted as one of the difficulties encountered in the study. A checklist proforma was suggested to be included in the future (Payten *et al.*, 2020). The HaNC-RC v.2 questions list can help towards this, either being incorporated in the GP referral processor via an initial telephone or face-to-face consultation with an allied health care professional. Using the calculator may help reduce the number of patients needing ENT input during their pathway with better allocation of resources. The allied professionals' clinics were running parallel to ENT consultant clinics that were available for advice or second opinion (Payten *et al.*, 2020). Input by ENT was needed on 30-42% of occasions (Seabrook *et al.*, 2019;Payten *et al.*, 2020). Less disruption to the ENT clinics could have been possibly achieved if the HaNC-RC v.2 was used during the initial triaging.

Similar attempts have only recently started in the UK, even though it has been known for many years that only a small number of malignancies are diagnosed via hoarseness clinics in the UK (Moore *et al.*, 2004). A throat clinic runs in Aberdeen led by SLT, which sees patients with globus symptoms (feeling of something in the throat), with an ENT consultant working closely with them to review flexible endoscopy images and be available to discuss patients' management if there are any uncertainties. Just over 5% of patients seen in this clinic required further investigations to establish a diagnosis and initiate treatment. No cancer incidence report was available in the published abstract (Asimakopoulos *et al.*, 2014). SLT-led pilot hoarseness clinics have also reported encouraging results for patients referred with low-risk factors of HNC malignancy (Occomore-Kent and Slade, 2021). Associated pain, dysphagia, neck lump, smokers over 55 years of age and family history of HNC were deemed high-risk

features and excluded patients from the clinics. Similar to the Australian studies, the SLT-led clinic runs parallel to an ENT consultant clinic, highlighting the importance of an MDT approach in managing these referrals. Only a minority of patients (20%) had structural abnormalities and were referred to a HaN consultant (Occomore-Kent and Slade, 2021). The authors have also welcomed the possibility of using HNC risk stratification tools in the future to aid the allocation of low-risk patients to SLT-led clinics following initial GP referral (Occomore-Kent and Slade, 2021). The HaNC-RC v.2 will be well placed to help with this, and it would be interesting to assess the calculator's contribution to this new pathway if future service audits are planned.

SLTs' reviews in establishing such clinics have been cautiously welcomed (Occomore-Kent, Hatch and Cruice, 2021). Although they realise that most patients complaining of hoarseness require SLT input rather than an ENT specialist, and being seen first by SLT speeds up patients' management pathway, there is an ongoing concern of a HNC missed diagnosis (Occomore-Kent, Hatch and Cruice, 2021). The authors mentioned that the risks could be mitigated by adequate training, supervision, and close contact with ENT specialists, as well as the use of technological advancements to risk-stratify patients to low and high risk of HNC, with allocation to appropriate clinics(Occomore-Kent, Hatch and Cruice, 2021). This is clearly where the HaNC-RC v.2 could have an important role. Other issues that need to be addressed are professional indemnity for the extended role of SLTs running such clinics, additional credentials and a formulation of a standardised pathway of training for running SLT-led 2ww clinics (Occomore-Kent, Hatch and Cruice, 2021). When similar clinics were run in the Australian studies, SLTs underwent further training in performing flexible laryngoscopy and received additional training and professional credentials for their extended service{Payten, 2020 #1493. A similar path is likely to be needed in the UK if SLTs perform flexible laryngoscopy as part of an extended role for 2ww SLT-led triaging. A second paper that included interviews with UK-based SLTs identified similar patterns of responses (Bradley and Patterson, 2021). It also highlighted issues of poor communication between ENT and SLT departments in some hospitals that are barriers to the implementation of SLT-led clinics (Bradley and Patterson, 2021). There were issues pointed out regarding the funding source for such clinics and the

availability of interested ENT consultants with voice experience available for the SLTs to consult (Bradley and Patterson, 2021).

5.7.2 Nurse-led clinics for triaging of HNC referrals

Apart from the evidence available for SLT-led clinics for dysphonia, globus and dysphagia referrals, the literature covering nurse-led triaging of HNC referrals will be covered and critically discussed here and how the HaNC-RC v.2 could support such service. Recent studies report nurse-led clinics in oral medicine (Spellman *et al.*, 2020;Spellman, Kanatas and Ong, 2018). Referrals were triaged to low and high risk by the maxillofacial doctors, with low-risk patients being allocated to the nurse-led clinics run by band 6 nurses receiving 1-2 years of training in oral pathology by maxillofacial consultants, then start seeing patients independently with consultants available for case discussion running parallel clinics (Spellman *et al.*, 2020;Spellman, Kanatas and Ong, 2018).

Aside from issues relating to professional indemnity, additional credentials, funding resources and establishing new referral pathways if nurses are running triage clinics, as mentioned earlier for SLTs(Occomore-Kent, Hatch and Cruice, 2021), there is also a significant burden from the time required for training (Ong, Spellman and Kanatas, 2020). It has been noted that healthcare professionals require a lengthy training program prior to independently running clinics that require consultant time allocation to the training program and clearly set objectives and portfolio of training(Ong, Spellman and Kanatas, 2020). The impact on the training of registrars has also been pointed out by Kyzas et al. (2021) review of the literature on the role of nurse-led oral and maxillofacial oncology clinics. Concerns were raised about the quality of care delivered by nurses only trained over a short period to perform history, examination, and diagnosis of conditions that, up until now, require a medical and dental degree and extensive further surgical and medical training. The costs associated with such training are challenging to establish and the potential of missed diagnosis in cases not discussed with a consultant (Kyzas, 2021). Another highlighted concern is that it is difficult to draw a line when a patient does not require a review by a specialist and can be safely managed by allied healthcare professionals (Kyzas, 2021). A good solution would

perhaps be to focus the initial attempts to train nurses to use the HaNC-RC v.2, being the only validated HNC referral tool, for allocation to low-risk clinics run by general ENT doctors and high-risk clinics by specialist HNC consultants. Training in the use of the tool is unlikely to take as long as training nurses to run HNC clinics independently. This should be researched in future studies and assess if it will impact waiting times for 2ww clinics, the total cost for running nurse-led triage clinics, the time needed for training and the impact of nurse-led re-triaging in cancer detection rates. Re-triaging of referrals by ENT specialists, following the initial 2ww GP referral, has been found to increase the cancer conversion rate as high as 24% but it comes associated with high costs of using expensive consultant time for the triaging (Breeze *et al.*, 2009). These costs could be potentially reduced significantly by the tasks being performed by trained nurses using the HaNC-RC v.2 as a triaging aid.

5.7.3 ENT Doctors - led clinics for triaging of HNC referrals

Triaging of referrals performed by ENT doctors will be discussed in this section. This is a suggestion that resembles the methodology of the validation phase of HaNC-RC v.2. The triaging can be done by ENT Registrars and consultants working in the hospital or in private ENT clinics for the countries where such a setting is in place. HaNC-RC v.2 triaging can be established, requiring little additional resources or training. This is because the ENT doctors are already familiar with the HNC red flags and can ask the questions included in the calculator without issues related to misunderstanding the reported symptoms. The calculator is available freely online, so no integration with other software is required. Onwards referral to the HaN specialist clinics can then be made based on the clinical suspicion and the HaNC-RC v.2 triage aid information. The clinics can be either virtual via telephone consultations, face-to-face or a desktop review of GP referral letters with the extraction of the necessary information.

In the public healthcare sector, if doctor-led triaging is performed, employing more ENT consultants and specialist doctors in each department will be required, as the current number of the ENT medical workforce is unable to meet such demand (Brocklehurst *et al.*, 2012). So, this will remain an issue if the triaging is done using

ENT doctors' time. Downgrading referral to low risk for possible HNC was performed by HaN doctors in the study by Payten et al. (2020) based on the primary care referral letters. However, issues were noted due to important information missing from the letters to triaging of referrals. ENT specialists had to manually read the letters before downgrading the referrals, which was time-consuming (Payten *et al.*, 2020). Implementing the HaNC-RC v.2 as part of the process as a checklist approach with all significant symptoms, signs, and social history of patients available to generate HNC probabilities could aid in this approach and can help in risk stratification of patients to low and high-risk for HNC. The use of such tools has been welcomed as long as their sensitivity and specificity are higher than the current predictive power of the NICE guidelines, which is known to be low (Araghi, Harris and Kyzas, 2020). The results of this thesis have shown the superiority of the HaNC-RC v.2 predictive power compared to currently established referral guidelines.

Employing any of the above-mentioned triage options also using the HaNC-RC v.2 can potentially make cost-effective the nationwide implementation of one-stop consultant/specialist-led HNC clinics, which are currently considered unjustified due to the associated costs, the low number of patients requiring investigations (only 15% needing an FNAc and 12.6% an ultrasound scan) and low pick up rates via the 2ww pathway (Pracy *et al.*, 2013). One-stop clinics with the availability of radiologists for USS scans and urgent CT scan slots, and pathologists for same-day reporting of FNA/core biopsy samples are currently considered feasible only in neck lump clinics (Sood *et al.*, 2021). Auditing has shown that GPs can successfully identify a neck lump in over 80% of 2ww neck lump referrals, and of these patients, about three quarters needed an ultrasound scan hence justifying the presence of a radiologist in the clinic, with FNA required on average in 40% of cases (Sood *et al.*, 2021). Increasing the cancer detection rates across the high-risk HNC clinics using the HaNC-RC v.2 could allow for one-stop clinics to be implemented not only for patients that have a neck lump but also for those with other red flag symptoms. High-risk patients with persistent hoarseness, dysphagia, and odynophagia would benefit from one-stop clinics with CT scan slots available on the same day and awake fibreoptic channelled biopsies to be performed and reported during the initial consultation.

To summarise, there are many potential uses of the HaNC-RC v.2 for patients' triaging in the secondary care setting. It can be done by ENT doctors or trained health care professionals such as SLTs and nurses, followed by allocating low-risk patients to either generalist ENT clinics or allied health care professionals-led clinics. However, issues are raised regarding the time needed to train non-doctor healthcare professionals to run such clinics independently. Availability of parallel-run ENT consultant clinics as a safety net during SLT- led and nurse-led 2ww clinics is currently a pre-requisite in the new suggested pathways. Using the HaNC-RC v.2 for triaging cases to low and high risk for cancer could free up more time for the consultants to spend training the health care professionals or performing other clinical tasks, especially in times of significant staff shortages. Finally, the use of HaNC-RC v.2 for triaging to high-risk group clinics could potentially allow for the development of cost-effective one-stop clinics with a justifiable cost for the availability of radiologists, pathologists, and allocated time slots for USS and cross-sectional imaging. All the above-discussed scenarios are currently hypothetical ways for the use of the tool for patients' triaging. Prior to any attempt for implementation into direct patient-care pathways, studies looking at comparing standard practice with any proposed alternative triaging pathway are required in a randomised control trial setting. Qualitative studies are also currently lacking and are needed to look into the views of involved parties in the introduction of such triage pathways.

The following section will cover the limitations noted during the development and validation of the HaNC-RC v.2. It will help set the directions for future studies to improve its performance alongside any information collected for published studies using the HaNC-RC v.2.

5.8 Limitations of the thesis and future directions

This section will cover the limitations of the current work and future research directions for the HaNC-RC v.2. Acknowledging and discussing the study limitations is important in planning future research.

To begin with, the development phase of the calculator was based on data collected from a single region in Scotland, which could limit the generalisability of the model. However, previous research has shown directly comparable results of Scottish and English cohorts in the presentation of HNC symptoms, demographics and cancer incidence, making the calculator relevant for use across the UK (Tikka, Paleri and MacKenzie, 2018). Data recording was performed by different HaN consultants, which could introduce reporting bias, but it was assumed that the consultants were clinicians with similar background knowledge and ran their clinics in a comparable manner. A data recording proforma was also used to ensure uniformity of data capture. Additionally, during the development phase, the data collection was performed by a single researcher, which precluded assessment by a second reviewer for any errors or inconsistencies in data interpretation and entry on the database. A second assessor would have been beneficial but was not feasible as this study was performed as part of unfunded PhD research work, and a large amount of data had to be evaluated.

As discussed in an earlier section, the validation phase of the HaNC-RC v.2 had different methods compared to the development phase due to the constraints of the COVID-19 pandemic. Even though attempts were made to try to mitigate this as much as possible with the publication of detailed guidance on the telephone triage use of the HaNC-RC v.2 (INTEGRATE, 2020), differences were unavoidable due to the pandemic.

Taking the above into consideration, further validation studies are required in a non-pandemic environment that will allow for a more representative cohort of HaN referrals. Data recording and collection should be standardised, ensuring prior training in the use of the calculator is given to all participating doctors, and independent assessors should overlook data recording and analysis. The development and validation phase of the HaNC-RC v.2 has been performed on secondary care cohorts with the triaging performed by ENT doctors. Triaging in primary care is also needed prior to recommendations for incorporating the HaNC-RC v.2 in primary care online systems. Depending on the healthcare system that the HaNC-RC v.2 might be applied to, future studies could also focus on triaging being performed in secondary care by

allied healthcare professionals, such as ANPs or SLTs, to assess the feasibility of such an approach.

In hospitals where the HaNC-RC v.2 has been incorporated into the referral system since the COVID-19 pandemic, it is important that prospective data collection continues to include all patients being triaged using the calculator. Additional significant symptoms, demographics and other medical history factors may arise following ongoing clinical use of the calculator. The collection of such information will also evaluate the inclusion of other potentially significant factors in future iterations of the calculator.

Future work should also focus on using the HaNC-RC v.2 in a more patient-friendly format for consideration of patient-led triaging. The language of the calculator could be adjusted to allow patients to use it without input being required from healthcare professionals, and future research is needed to ensure all relevant symptoms are included in a language that patients are using and understand when describing their symptoms. These alterations could allow the development of an online system used directly by patients advising them what to do if they have concerns about their symptoms, being a similar concept to one of the online prostate cancer calculators (SWOP, 2022). The HaNC-RC v.2 could also be developed further based on patients' and clinicians' focused groups and thematic analysis, enabling triaging of patients on the basis of self-completed questionnaires following the initial GP referral. As has already been mentioned earlier, some of these suggestions will be addressed by the EVEREST-HN trial, which stemmed from the HaNC-RC v.2 telephone triage audit {Hardman, 2021 #1814}. The EVEREST-HN trial aims to develop a patient-focused triage tool for head and neck cancer, developing a primary care-based version of the risk calculator in a non-inferiority randomised control trial setting (https://fundingawards.nihr.ac.uk, 2022).

Such studies are currently lacking and are urgently needed in order to robustly compare the current standard pathway of referrals with a new pathway incorporating triaging using the HaNC-RC v.2 or any future refined version of it. Prospective comparison of

outcomes both on cancer detection rates and long-term outcomes, including disease-free survival and treatment outcomes, should be the focus of future research. Ethical considerations relating to the acceptable level of a missed cancer diagnosis should also be addressed during the evaluation of the current versus any suggested new pathway, taking into consideration the currently reported 40% HNC diagnosis via USOC routes, with the rest being diagnosed by other routes (Langton, Siau and Bankhead, 2016), the public demand of investigating any patient with a 1% cancer probability (Banks *et al.*, 2014) and the 3% PPV symptom threshold suggested by experts (NICE, 2021).

6 Conclusions

In conclusion, this study succeeded in achieving its aims of developing and validating an updated version of a previously designed HNC risk calculator. The HaNC-RC v.2 had an improved prediction power compared to its earlier version, with an AUC of 88.5% at internal validation and 83.96% at external validation. It was developed taking into consideration the available evidence base on HNC and risk calculators from other cancer sites, ensuring that the shortfalls in other tools were considered and addressed during the design of the HaNC-RC v.2.

The tool was clinically validated during the first wave of the COVID-19 pandemic making a significant contribution in triaging patients with suspected HNC symptoms to the limited available face-to-face clinic services that were impacted by the pandemic constraints. The HaNC-RC v.2 is the first structured assessment tool that has been robustly generated, validated and rapidly implemented for use in triaging HNC referrals. This can be the groundwork for other similar cancer tools to be developed using a similarly robust design process.

The HaNC-RC v.2 is still used in many units across the UK, and its effectiveness has also been shown in studies published recently. The implementation of the HNC-RC V.2 as an aid for triaging HNC referrals has shown that such an approach can be potentially used nationwide to identify high-risk patients for HNC. These patients can be targeted for expedited face-to-face specialist review as well as urgent arrangements of investigations even prior to the initial consultation expediting the cancer diagnostic pathway. Future studies could assess the incorporation of the HaNC-RC v.2 in the primary care referral software or other secondary care clinical service models potentially used as an alternative to the current referral guidelines but also in an online questionnaire format that can be adjusted for patient self-completion. Further iterations of the HaNC-RC are also possible and can be based on data collected directly from patients and also explore the addition of biomarkers as model variables to boost further the prediction power of the tool that can have different versions depending on the clinical setting being used and the availability and cost of more specialised investigations.

7 References

Abel, G. *et al.* (2018) 'Variation and statistical reliability of publicly reported primary care diagnostic activity indicators for cancer: a cross-sectional ecological study of routine data'. *BMJ Quality & Safety,* 27 (1), pp. 21-30.
Abrahão, R. *et al.* (2020) 'Predictors of Survival After Head and Neck Squamous Cell Carcinoma in South America: The InterCHANGE Study'. *JCO Global Oncology,* (6), pp. 486-499.
Adams, J., Ryan, V. and White, M. (2005) 'How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data'. *Journal of public health (Oxford, England),* 27 101-106.
Adelstein, B.-A. *et al.* (2010) 'Who needs colonoscopy to identify colorectal cancer? Bowel symptoms do not add substantially to age and other medical history'. *Alimentary Pharmacology & Therapeutics,* 32 (2), pp. 270-281.
Ahmad, A.S., Ormiston-Smith, N. and Sasieni, P.D. (2015) 'Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960'. *British Journal of Cancer,* 112 (5), pp. 943-947.
Alblas, M. *et al.* (2018) 'Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review'. *Crit Rev Oncol Hematol,* 126 92-99.
Alcázar, J.L. and Jurado, M. (1998) 'Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings'. *Gynecol Oncol,* 69 (2), pp. 146-150.
Alho, O.P. (2006) 'Head and neck cancer in primary care: presenting symptoms and the effect of delayed diagnosis of cancer cases'. *Canadian Medical Association Journal,* 174 (6), pp. 779-784.
Allam, A. and Nijim, H. (2019) 'Persistent Unilateral Sore Throat: Should It Be Included in the 2-Week Wait Referral Criteria by NICE'. *International Journal of Otolaryngology,* 2019 1-6.
Allen, J. *et al.* (2002) 'The European definitions of the key features of the discipline of general practice: the role of the GP and core competencies'. *Br J Gen Pract,* 52 (479), pp. 526-527.
Amarasinghe, H.K. *et al.* (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), pp. 303-309.
Amir *et al.* (1999) 'Diagnostic delays in head and neck cancers'. *European Journal of Cancer Care,* 8 (4), pp. 198-203.
Amir, E. *et al.* (2003) 'Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme'. *J Med Genet,* 40 (11), pp. 807-814.
Andersen, M.R. *et al.* (2008) 'Combining a symptoms index with CA 125 to improve detection of ovarian cancer'. *Cancer,* 113 (3), pp. 484-489.
Ankerst, D.P. *et al.* (2019) 'Incorporation of Urinary Prostate Cancer Antigen 3 and TMPRSS2:ERG into Prostate Cancer Prevention Trial Risk Calculator'. *Eur Urol Focus,* 5 (1), pp. 54-61.

Ankerst, D.P. *et al.* (2008) 'Predicting prostate cancer risk through incorporation of prostate cancer gene 3'. *J Urol*, 180 (4), pp. 1303-1308; discussion 1308.
Ankerst, D.P. *et al.* (2013) 'The impact of prostate volume, number of biopsy cores and American Urological Association symptom score on the sensitivity of cancer detection using the Prostate Cancer Prevention Trial risk calculator'. *The Journal of urology*, 190 70-76.
Ankerst, D.P. *et al.* (2014) 'Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer'. *Urology*, 83 (6), pp. 1362-1367.
Ankerst, D.P. *et al.* (2018) 'A Contemporary Prostate Biopsy Risk Calculator Based on Multiple Heterogeneous Cohorts'. *Eur Urol*, 74 (2), pp. 197-203.
Antoniou, A. *et al.* (2003) 'Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies'. *Am J Hum Genet*, 72 (5), pp. 1117-1130.
Antoniou, A.C. *et al.* (2008) 'The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions'. *Br J Cancer*, 98 (8), pp. 1457-1466.
Antoniou, A.C. *et al.* (2002) 'A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes'. *Br J Cancer*, 86 (1), pp. 76-83.
Araghi, A.S., Harris, Y. and Kyzas, P. (2020) 'An audit to analyse the two-week wait pathway at an oral cancer specialist district general hospital'. *British Dental Journal*, 228 (9), pp. 693-699.
Arantes, L. *et al.* (2018) 'Serum, plasma and saliva biomarkers for head and neck cancer'. *Expert Rev Mol Diagn*, 18 (1), pp. 85-112.
Asimakopoulos, P. *et al.* (2014) *Management of globus pharyngeus in the "throat therapy" clinic: A two cycle audit.*
Auffenberg, G.B. *et al.* (2017) 'Evaluation of Prostate Cancer Risk Calculators for Shared Decision Making Across Diverse Urology Practices in Michigan'. *Urology*, 104 137-142.
Ayer, T. *et al.* (2010) 'Breast cancer risk estimation with artificial neural networks revisited'. *Cancer*, 116 (14), pp. 3310-3321.
Azzani, M., Roslani, A.C. and Su, T.T. (2015) 'The perceived cancer-related financial hardship among patients and their families: a systematic review'. *Support Care Cancer*, 23 (3), pp. 889-898.
Baan, R. *et al.* (2009) 'A review of human carcinogens--Part F: chemical agents and related occupations'. *Lancet Oncol*, 10 (12), pp. 1143-1144.
Bagepalli Srinivas, S. *et al.* (2020) 'A Novel Risk-Scoring Model for Prediction of Premalignant and Malignant Lesions of Uterine Endometrium Among Symptomatic Premenopausal Women'. *Int J Womens Health*, 12 883-891.
BAHNO (2020) *ENTUK2WW triage.* Available at: https://bahno.org.uk/entuk2ww_triage.aspx.
Balagurunathan, Y. *et al.* (2019) 'Quantitative Imaging features Improve Discrimination of Malignancy in Pulmonary nodules'. *Sci Rep*, 9 (1), pp. 8528.
Bandala-Jacques, A. *et al.* (2021) 'Prostate Cancer Risk Calculators for Healthy Populations: Systematic Review'. *JMIR Cancer*, 7 (3), pp. e30430.
Banerjee, S. *et al.* (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 (3), pp. 241-245.
Banks, J. *et al.* (2014) 'Preferences for cancer investigation: a vignette-based study of primary-care attendees'. *The Lancet Oncology*, 15 (2), pp. 232-240.

Barbieri, C.E. *et al.* (2012) 'Decision curve analysis assessing the clinical benefit of NMP22 in the detection of bladder cancer: secondary analysis of a prospective trial'. *BJU Int,* 109 (5), pp. 685-690.
Baughan, P., Keatings, J. and O'Neill, B. (2011) 'Urgent suspected cancer referrals from general practice: audit of compliance with guidelines and referral outcomes'. *Br J Gen Pract,* 61 (592), pp. e700-706.
Beral, V. and Peto, R. (2010) 'UK cancer survival statistics'. *Bmj,* 341 c4112.
Bernier, J. *et al.* (2004) 'Postoperative Irradiation with or without Concomitant Chemotherapy for Locally Advanced Head and Neck Cancer'. *New England Journal of Medicine,* 350 (19), pp. 1945-1952.
Berrino, F. *et al.* (2007) 'Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EUROCARE-4 study'. *Lancet Oncol,* 8 (9), pp. 773-783.
Berrino, F. *et al.* (1998) 'The EUROCARE II study'. *European Journal of Cancer,* 34 (14), pp. 2139-2153.
Berrino, F. *et al.* (1995) 'Survival of cancer patients in europe?The EUROCARE study'. *Cancer Causes and Control,* 6 (5), pp. 469-470.
Berry, D.A. *et al.* (2002) 'BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes'. *J Clin Oncol,* 20 (11), pp. 2701-2712.
Berry, D.A. *et al.* (1997) 'Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history'. *J Natl Cancer Inst,* 89 (3), pp. 227-238.
Berthiller, J. *et al.* (2009) 'Marijuana smoking and the risk of head and neck cancer: pooled analysis in the INHANCE consortium'. *Cancer Epidemiol Biomarkers Prev,* 18 (5), pp. 1544-1551.
Beukers, W. *et al.* (2013) 'The Use of Molecular Analyses in Voided Urine for the Assessment of Patients with Hematuria'. *PLoS ONE,* 8 (10), pp. e77657.
Biau, G., Cérou, F. and Guyader, A. (2010) 'On the Rate of Convergence of the Bagged Nearest Neighbor Estimate'. *Journal of Machine Learning Research,* 11 (2), pp.
Biau, G. and Devroye, L. (2010) 'On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification'. *Journal of Multivariate Analysis,* 101 (10), pp. 2499-2518.
Biswas, S. *et al.* (2012) 'Assessing the added value of breast tumor markers in genetic risk prediction model BRCAPRO'. *Breast Cancer Res Treat,* 133 (1), pp. 347-355.
Blank, L. *et al.* (2014) 'Referral interventions from primary to specialist care: a systematic review of international evidence'. *Br J Gen Pract,* 64 (629), pp. e765-774.
Boerma, W.G.W. (2003) Published. 'Profiles of general practice in Europe : an international study of variation in the tasks of general practitioners'. 2003.
Boulesteix, A.-L. and Strobl, C. (2009) 'Optimal classifier selection and negative bias in error rate estimation: an empirical study on high-dimensional prediction'. *BMC medical research methodology,* 9 (1), pp. 1-14.
Bradley, P.T. *et al.* (2021) 'Factors shaping the implementation and use of Clinical Cancer Decision Tools by GPs in primary care: a qualitative framework synthesis'. *BMJ Open,* 11 (2), pp. e043338.
Bradley, P.T. and Patterson, J. (2021) 'Attitudes to the Implementation of Speech and Language Therapist Led Low Risk Two Week Wait Clinic in the UK: A Survey Exploration Using Normalization Process Theory'. *Journal of Voice*.

Bray, F. *et al.* (2012) 'Global cancer transitions according to the Human Development Index (2008-2030): a population-based study'. *Lancet Oncol*, 13 (8), pp. 790-801.
Breast Cancer Surveillance Consortium Risk Calculator (2017) *Risk Calculator V2*. Available at: https://tools.bcsc-scc.org/BC5yearRisk/calculator.htm.
Breeze, J. *et al.* (2009) 'Can early detection rates for cancer referred to an oral and maxillofacial surgery department be improved by consultant triage of referral letters'. *Oral Surgery*, 2 77-79.
Breiman, L. (1996) 'Bagging predictors'. *Machine learning*, 24 (2), pp. 123-140.
Breiman, L. (2001) 'Random forests'. *Machine learning*, 45 (1), pp. 5-32.
Brentnall, A.R. *et al.* (2019) 'A case-control study to add volumetric or clinical mammographic density into the Tyrer-Cuzick breast cancer risk model'. *Journal of breast imaging*, 1 (2), pp. 99-106.
Brentnall, A.R. and Cuzick, J. (2020) 'Risk models for breast cancer and their validation'. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 35 (1), pp. 14.
Brentnall, A.R. *et al.* (2018) 'Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density'. *JAMA Oncol*, 4 (9), pp. e180174.
Brentnall, A.R. *et al.* (2015) 'Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort'. *Breast Cancer Research*, 17 (1), pp. 1-10.
Brentnall, A.R. *et al.* (2020) 'A case-control evaluation of 143 single nucleotide polymorphisms for breast cancer risk stratification with classical factors and mammographic density'. *Int J Cancer*, 146 (8), pp. 2122-2129.
Brocklehurst, P. *et al.* (2012) 'Analysis of the impact of deprivation on urgent suspected head and neck cancer referrals in the Mersey region between January 2004 to December 2006'. *Br J Oral Maxillofac Surg*, 50 (3), pp. 215-220.
Brouha, X.D. *et al.* (2005a) 'Oral and pharyngeal cancer: analysis of patient delay at different tumor stages'. *Head Neck*, 27 (11), pp. 939-945.
Brouha, X.D.R. *et al.* (2005b) 'Laryngeal cancer patients: Analysis of patient delay at different tumor stages'. *Head & Neck*, 27 (4), pp. 289-295.
Brown, S. *et al.* (2014) 'How might healthcare systems influence speed of cancer diagnosis: A narrative review'. *Social Science & Medicine*, 116 56-63.
Burton, C. *et al.* (2020) 'Contribution of primary care organisation and specialist care provider to variation in GP referrals for suspected cancer: ecological analysis of national data'. *BMJ Quality & Safety*, 29 (4), pp. 296-303.
Butler, J. *et al.* (2013) 'The International Cancer Benchmarking Partnership: An international collaboration to inform cancer policy in Australia, Canada, Denmark, Norway, Sweden and the United Kingdom'. *Health Policy*, 112 (1-2), pp. 148-155.
Byrd, J.K. and Ferris, R.L. (2016) 'Is There a Role for Robotic Surgery in the Treatment of Head and Neck Cancer?'. *Current Treatment Options in Oncology*, 17 (6), pp.
Cancer Council Victoria and Department of Health Victoria (2021) *Optimal care pathway for people with head and neck cancer*. 2nd edition edn. Melbourne: Cancer Council Victoria.
Cancer Patient Experience survey, England. (2016) Available at: http://www.ncpes.co.uk/index.php/reports/2016-reports (Accessed: 09/05/2019).
Cancer patient experience survey, England. 2016. (2016) Available at: www.ncpes.co.uk/index.php/reports/2016-reports. (Accessed: 09/05/21).

cancer.org (2020) *Signs and Symptoms of Cancer.* Available at: https://www.cancer.org/treatment/understanding-your-diagnosis/signs-and-symptoms-of-cancer.html (Accessed: 05/03).
cancer.org (2021) *American Cancer Society Guidelines for the Early Detection of Cancer.* Available at: https://www.cancer.org/cancer/laryngeal-and-hypopharyngeal-cancer/detection-diagnosis-staging/signs-symptoms.html (Accessed: 02/09).
cancerreferral.scot.nhs.uk (2019) *Scottish Referral Guidelines for Suspected cancer: Head and neck cancers.* Available at: http://www.cancerreferral.scot.nhs.uk/head-and-neck-cancers/ (Accessed: 03/05).
CancerResearchUK (2020) *Skin Cancer Toolkit.* Available at: https://www.cancerresearchuk.org/health-professional/learning-and-support/online-learning/toolkits#toolkits1.
Cao, B. *et al.* (2017) 'Benchmarking life expectancy and cancer mortality: global comparison with cardiovascular disease 1981-2010'. *BMJ,* j2765.
Caras, R.J. and Sterbis, J.R. (2014) 'Prostate cancer nomograms: a review of their use in cancer detection and treatment'. *Curr Urol Rep,* 15 (3), pp. 391.
Carter, J.M. *et al.* (2013) 'A faith- and community-based approach to identifying the individual at risk for head and neck cancer in an inner city'. *Laryngoscope,* 123 (6), pp. 1439-1443.
Carter, N. (2021) 'Mouth cancer: the challenges ahead'. *BDJ In Practice,* 34 (11), pp. 20-21.
Carvalho, A.L. *et al.* (2002) 'Predictive Factors for Diagnosis of Advanced-Stage Squamous Cell Carcinoma of the Head and Neck'. *Archives of Otolaryngology–Head & Neck Surgery,* 128 (3), pp. 313.
Cassidy, A. *et al.* (2008) 'The LLP risk model: an individual risk prediction model for lung cancer'. *Br J Cancer,* 98 (2), pp. 270-276.
Cavadas, V. *et al.* (2010) 'Prostate cancer prevention trial and European randomized study of screening for prostate cancer risk calculators: a performance comparison in a contemporary screened cohort'. *Eur Urol,* 58 (4), pp. 551-558.
CCA (2021) *Guidelines: Clinical practice guidelines developed by cancer council Australia.* Available at: https://wiki.cancer.org.au/australia/Guidelines.
Cha, E.K. *et al.* (2012a) 'Immunocytology is a strong predictor of bladder cancer presence in patients with painless hematuria: a multicentre study'. *Eur Urol,* 61 (1), pp. 185-192.
Cha, E.K. *et al.* (2012b) 'Accurate risk assessment of patients with asymptomatic hematuria for the presence of bladder cancer'. *World J Urol,* 30 (6), pp. 847-852.
Chan, Y. *et al.* (2020) 'CSO (Canadian Society of Otolaryngology - Head & Neck Surgery) position paper on rhinologic and skull base surgery during the COVID-19 pandemic'. *Journal of Otolaryngology - Head & Neck Surgery,* 49 (1), pp.
Chang, J. *et al.* (2013) 'Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population'. *Carcinogenesis,* 34 (8), pp. 1782-1786.
Chen, F. *et al.* (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 (14), pp. 2543-2548.
Chen, X. and Ishwaran, H. (2012) 'Random forests for genomic data analysis'. *Genomics,* 99 (6), pp. 323-329.

Cheng, E.S. *et al.* (2018) 'Cancer burden and control in Australia: lessons learnt and challenges remaining'. *Annals of Cancer Epidemiology*, 2.
Cheung, L.C. *et al.* (2021) 'Risk-Based Selection of Individuals for Oral Cancer Screening'. *Journal of Clinical Oncology*, 39 (6), pp. 663-674.
Chima, S. *et al.* (2019) 'Decision support tools to improve cancer diagnostic decision making in primary care: a systematic review'. *Br J Gen Pract*, 69 (689), pp. e809-e818.
Choi, Y.J. *et al.* (2015) 'Web-Based Malignancy Risk Estimation for Thyroid Nodules Using Ultrasonography Characteristics: Development and Validation of a Predictive Model'. *Thyroid*, 25 (12), pp. 1306-1312.
Chojnacka-Szawłowska, G. *et al.* (2017) 'Knowledge of cancer symptoms and anxiety affect patient delay in seeking diagnosis in patients with heterogeneous cancer locations'. *Curr Probl Cancer*, 41 (1), pp. 64-70.
Cintolo-Gonzalez, J.A. *et al.* (2017) 'Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications'. *Breast cancer research and treatment*, 164 (2), pp. 263-284.
Claus, E.B., Risch, N. and Thompson, W.D. (1991) 'Genetic analysis of breast cancer in the cancer and steroid hormone study'. *Am J Hum Genet*, 48 (2), pp. 232-242.
Claus, E.B., Risch, N. and Thompson, W.D. (1993) 'The calculation of breast cancer risk for women with a first degree family history of ovarian cancer'. *Breast Cancer Res Treat*, 28 (2), pp. 115-120.
Claus, E.B., Risch, N. and Thompson, W.D. (1994) 'Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction'. *Cancer*, 73 (3), pp. 643-651.
Colditz, G.A. and Rosner, B. (2000) 'Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study'. *Am J Epidemiol*, 152 (10), pp. 950-964.
Cole, J.A. *et al.* (2011) 'Reducing selection bias in case-control studies from rare disease registries'. *Orphanet Journal of Rare Diseases*, 6 (1), pp. 61.
Coleman, M. *et al.* (2011) 'Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data'. *The Lancet*, 377 (9760), pp. 127-138.
Coleman, M.P. *et al.* (2003) 'EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century'. *Ann Oncol*, 14 Suppl 5 v128-149.
Collins, G.S. and Altman, D.G. (2012) 'Identifying patients with undetected colorectal cancer: an independent validation of QCancer (Colorectal)'. *British Journal of Cancer*, 107 (2), pp. 260-265.
Collins, G.S. and Altman, D.G. (2013) 'Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer® (Pancreas)'. *British Journal of General Practice*, 63 (614), pp. e636-e642.
Cone, E.B. *et al.* (2020) 'Assessment of Time-to-Treatment Initiation and Survival in a Cohort of Patients With Common Cancers'. *JAMA Network Open*, 3 (12), pp. e2030072-e2030072.
Conway, D.I. *et al.* (2010a) 'Socioeconomic factors associated with risk of upper aerodigestive tract cancer in Europe'. *Eur J Cancer*, 46 (3), pp. 588-598.

Conway, D.I. *et al.* (2010b) 'Components of socioeconomic risk associated with head and neck cancer: a population-based case-control study in Scotland'. *Br J Oral Maxillofac Surg,* 48 (1), pp. 11-17.
Conway, D.I. *et al.* (2008) 'Socioeconomic inequalities and oral cancer risk: a systematic review and meta-analysis of case-control studies'. *Int J Cancer,* 122 (12), pp. 2811-2819.
Costantino, J.P. *et al.* (1999) 'Validation studies for models projecting the risk of invasive and total breast cancer incidence'. *J Natl Cancer Inst,* 91 (18), pp. 1541-1548.
Curado, M. and Bezerra de Souza, D.L. (2014) 'Cancer Burden in Latin America and the Caribbean'. *Annals of Global Health,* 80 (5), pp. 370.
Cust, A.E. *et al.* (2018) 'Assessing the incremental contribution of common genomic variants to melanoma risk prediction in two population-based studies'. *Journal of Investigative Dermatology,* 138 (12), pp. 2617-2624.
de Almeida, J.R. *et al.* (2014) 'A systematic review of transoral robotic surgery and radiotherapy for early oropharynx cancer: a systematic review'. *Laryngoscope,* 124 (9), pp. 2096-2102.
De Angelis, R. *et al.* (2014) 'Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5—a population-based study'. *The Lancet Oncology,* 15 (1), pp. 23-34.
Del Mar, C.B. and Green, A.C. (1995) 'Aid to diagnosis of melanoma in primary medical care'. *BMJ,* 310 (6978), pp. 492-495.
Delivering cancer waiting times: a good practical guide. (2015) Available at: https://www.england.nhs.uk/wp-content/uploads/2015/03/delivering-cancer-wait-times.pdf (Accessed: 10/09/2019).
Dembrower, K. *et al.* (2020) 'Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction'. *Radiology,* 294 (2), pp. 265-272.
DePriest, P.D. *et al.* (1993) 'A morphology index based on sonographic findings in ovarian cancer'. *Gynecol Oncol,* 51 (1), pp. 7-11.
Deschler, D.G., Moore, M.G. and Smith, R.V. (2014) 'Quick reference guide to TNM staging of head and neck cancer and neck dissection classification'. *American Academy of Otolaryngology–Head and Neck Surgery Foundation.[Internet],* 1-34.
Devroye, L., Györfi, L. and Lugosi, G. (2013) *A probabilistic theory of pattern recognition.* Springer Science & Business Media.
DH (2007) 'Cancer Reform Strategy'.
DH (2011) 'Improving Outcomes: A strategy for Cancer'. *Department of Health.*
Dharmawardana, N. *et al.* (2020) 'Development of a non-invasive exhaled breath test for the diagnosis of head and neck cancer'. *British Journal of Cancer,* 123 (12), pp. 1775-1781.
Dikomitis, L., Green, T. and Macleod, U. (2015) 'Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences'. *Primary Health Care Research & Development,* 16 (6), pp. 548-555.
Djulbegovic, B. *et al.* (2015) 'Rationality, practice variation and person-centred health policy: a threshold hypothesis'. *J Eval Clin Pract,* 21 (6), pp. 1121-1124.
Dodds, W. *et al.* (2004) 'Implementing the 2-week wait rule for cancer referral in the UK: general practitioners' views and practices'. *Eur J Cancer Care (Engl),* 13 (1), pp. 82-87.

DoH (2000) *Referral Guidelines for Suspected Cancer.* London: Department of Health.
Dolan, R.W., Vaughan, C.W. and Fuleihan, N. (1998) 'Symptoms in early head and neck cancer: an inadequate indicator'. *Otolaryngol Head Neck Surg,* 119 (5), pp. 463-467.
Doll, E.J., Braden, M.N. and Thibeault, S.L. (2021) 'COVID-19 and Speech-Language Pathology Clinical Practice of Voice and Upper Airway Disorders'. *Am J Speech Lang Pathol,* 30 (1), pp. 63-74.
Dommett, R.M. *et al.* (2013) 'Risk of childhood cancer with symptoms in primary care: a population-based case-control study'. *British Journal of General Practice,* 63 (606), pp. e22-e29.
Douglas, C., Carswell, V. and Montgomery, J. (2019) 'Outcomes of urgent suspicion of head and neck cancer referrals in Glasgow'. *The Annals of The Royal College of Surgeons of England,* 101 (2), pp. 103-106.
Douglas, C.M. *et al.* (2018) 'Presenting symptoms and long-term survival in head and neck cancer'. *Clin Otolaryngol,* 43 (3), pp. 795-804.
Douglas, C.M. *et al.* (2021a) 'Patterns of urgent hoarseness referrals to ENT-When should we be suspicious of cancer?'. *Clin Otolaryngol,* 46 (3), pp. 562-569.
Douglas, C.M. *et al.* (2021b) 'Patterns of urgent hoarseness referrals to ENT—When should we be suspicious of cancer?'. *Clinical Otolaryngology,* 46 (3), pp. 562-569.
Duvvi, S.K. *et al.* (2006) 'Two-week rule for suspected Head and neck cancer. A study of compliance and effectiveness'. *Journal of Evaluation in Clinical Practice,* 12 (6), pp. 591-594.
Ea, G., Harding, S. and Courtney, D. (2008) *Study of head and neck cancer two-week wait referral pathway.*
East, S.C., Stocker, J.L. and Avery, B.S. (2005) 'Is the two week rule of any benefit to patients with oral cancer?'. *British Journal of Oral and Maxillofacial Surgery,* 43 (6), pp. 511-512.
Elias, S.G. *et al.* (2017) 'Published diagnostic models safely excluded colorectal cancer in an independent primary care validation study'. *Journal of Clinical Epidemiology,* 82 149-157.e148.
Elliss-Brookes, L. *et al.* (2012) 'Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets'. *British Journal of Cancer,* 107 (8), pp. 1220-1226.
Emery, J. and Vedsted, P. (2015) 'New NICE guidance on diagnosing cancer in general practice'. *British Journal of General Practice,* 65 (638), pp. 446-447.
Emery, J.D. *et al.* (2017) 'The Improving Rural Cancer Outcomes Trial: a cluster-randomised controlled trial of a complex intervention to reduce time to diagnosis in rural cancer patients in Western Australia'. *British Journal of Cancer,* 117 (10), pp. 1459-1469.
Emery, J.D. *et al.* (2014) 'The role of primary care in early detection and follow-up of cancer'. *Nature Reviews Clinical Oncology,* 11 (1), pp. 38-48.
English, D.R. *et al.* (2003) 'Evaluation of aid to diagnosis of pigmented skin lesions in general practice: controlled trial randomised by practice'. *BMJ,* 327 (7411), pp. 375.
ENTUK (2020) *ENT UK Cancer Telephone Triage: Service Evaluation.* Available at: https://www.entuk.org/ent-uk-cancer-telephone-triage-service-evaluation (Accessed: 20/4/21).

European Head & Neck Society (2020) *Make sense campaign–head & neck cance.* Available at: makesensecampaign.eu/en/cancer-information/head-neck-cancer (Accessed: 03/09).
Eyre, S.J. *et al.* (2009) 'Validation in a multiple urology practice cohort of the Prostate Cancer Prevention Trial calculator for predicting prostate cancer detection'. *J Urol,* 182 (6), pp. 2653-2658.
Federatie Medisch Specialisten (2016) *Prostate Carcinoma.* Available at: https://richtlijnendatabase.nl/richtlijn/prostaatcarcinoom/algemeen.html (Accessed: 16/10/20).
Ferris, R.L. *et al.* (2020) 'Transoral robotic surgical resection followed by randomization to low- or standard-dose IMRT in resectable p16+ locally advanced oropharynx cancer: A trial of the ECOG-ACRIN Cancer Research Group (E3311)'. *Journal of Clinical Oncology,* 38 (15_suppl), pp. 6500-6500.
Fieux, M. *et al.* (2020) 'Telemedicine for ENT: Effect on quality of care during Covid-19 pandemic'. *European annals of otorhinolaryngology, head and neck diseases,* 137 (4), pp. 257-261.
Fijten, G.H. *et al.* (1995) 'Predictive value of signs and symptoms for colorectal cancer in patients with rectal bleeding in general practice'. *Family Practice,* 12 (3), pp. 279-286.
Fingland, P. *et al.* (2018) 'The value of chest X-ray in the Scottish Referral Guidelines for suspected head and neck cancer in 2144 patients'. *The Journal of Laryngology & Otology,* 132 (5), pp. 434-438.
Fischer, C. *et al.* (2013) 'Evaluating the performance of the breast cancer genetic risk models BOADICEA, IBIS, BRCAPRO and Claus for predicting BRCA1/2 mutation carrier probabilities: a study based on 7352 families from the German Hereditary Breast and Ovarian Cancer Consortium'. *J Med Genet,* 50 (6), pp. 360-367.
Flukes, S. *et al.* (2019) 'Pre-treatment wait time for head and neck cancer patients in Western Australia: description of a new metric and examination of predictive factors'. *ANZ Journal of Surgery,* 89 (7-8), pp. 858-862.
Foley, R.W. *et al.* (2015) 'Predicting prostate cancer: analysing the clinical efficacy of prostate cancer risk calculators in a referral population'. *Ir J Med Sci,* 184 (3), pp. 701-706.
Foley, R.W. *et al.* (2016) 'European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators significantly outperform the Prostate Cancer Prevention Trial (PCPT) 2.0 in the prediction of prostate cancer: a multi-institutional study'. *BJU Int,* 118 (5), pp. 706-713.
Fontanillas, P. *et al.* (2021) 'Disease risk scores for skin cancers'. *Nat Commun,* 12 (1), pp. 160.
Fortner, R.T. *et al.* (2017) 'Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort'. *International Journal of Cancer,* 140 (6), pp. 1317-1323.
Frank, T.S. *et al.* (2002) 'Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals'. *J Clin Oncol,* 20 (6), pp. 1480-1490.
Frank, T.S. *et al.* (1998) 'Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk'. *J Clin Oncol,* 16 (7), pp. 2417-2425.

Frantzi, M. *et al.* (2014) 'Discovery and validation of urinary biomarkers for detection of renal cell carcinoma'. *J Proteomics,* 98 44-58.
Freedman, A.N. *et al.* (2005) 'Cancer Risk Prediction Models: A Workshop on Development, Evaluation, and Application'. *JNCI: Journal of the National Cancer Institute,* 97 (10), pp. 715-723.
Gail, M.H. *et al.* (1989) 'Projecting individualized probabilities of developing breast cancer for white females who are being examined annually'. *J Natl Cancer Inst,* 81 (24), pp. 1879-1886.
Galeone, C. *et al.* (2010) 'Coffee and tea intake and risk of head and neck cancer: pooled analysis in the international head and neck cancer epidemiology consortium'. *Cancer Epidemiol Biomarkers Prev,* 19 (7), pp. 1723-1736.
Gandhi, A.K. *et al.* (2014) 'Symptom Burden and Quality of Life in Advanced Head and Neck Cancer Patients: AIIMS Study of 100 Patients'. *Indian J Palliat Care,* 20 (3), pp. 189-193.
Gao, C. *et al.* (2019) 'Two week wait referral criteria – heading in the right direction?'. *The Journal of Laryngology & Otology,* 133 (8), pp. 704-712.
Gatta, G. *et al.* (2015) 'Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study'. *European Journal of Cancer,* 51 (15), pp. 2130-2143.
Gatta, G. *et al.* (2017) 'Burden and centralised treatment in Europe of rare tumours: results of RARECAREnet—a population-based study'. *The Lancet Oncology,* 18 (8), pp. 1022-1039.
Gaudet, M.M. *et al.* (2010) '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 (4), pp. 1091-1102.
Genuer, R. (2012) 'Variance reduction in purely random forests'. *Journal of Nonparametric Statistics,* 24 (3), pp. 543-562.
Geomini, P. *et al.* (2009) 'The accuracy of risk scores in predicting ovarian malignancy: a systematic review'. *Obstet Gynecol,* 113 (2 Pt 1), pp. 384-394.
Giannella, L. *et al.* (2014) 'A risk-scoring model for the prediction of endometrial cancer among symptomatic postmenopausal women with endometrial thickness > 4 mm'. *Biomed Res Int,* 2014 130569.
Girardi, F.M., Silva, L.M.D. and Flores, C.D. (2019) 'A predictive model to distinguish malignant and benign thyroid nodules based on age, gender and ultrasonographic features'. *Braz J Otorhinolaryngol,* 85 (1), pp. 24-31.
Goff, B.A. *et al.* (2007) 'Development of an ovarian cancer symptom index'. *Cancer,* 109 (2), pp. 221-227.
Goldstein, B.A., Polley, E.C. and Briggs, F.B. (2011) 'Random forests for genetic association studies'. *Statistical applications in genetics and molecular biology,* 10 (1), pp.
Goy, J. *et al.* (2009) 'Diagnostic delay and disease stage in head and neck cancer: a systematic review'. *Laryngoscope,* 119 (5), pp. 889-898.
Gray, E.P. *et al.* (2016) 'Risk Prediction Models for Lung Cancer: A Systematic Review'. *Clin Lung Cancer,* 17 (2), pp. 95-106.
Green, T. *et al.* (2014) 'Exploring GPs' experiences of using diagnostic tools for cancer: a qualitative study in primary care'. *Family Practice,* 32 (1), pp. 101-105.
Greene, H.W. (2012) *Econometric Anlysis. Chapter 4: The Least Squares Estimators.* 7th edition. edn. Pearson.

Grey, R. and Walsh, S. (2019) 'A review of the electronic two-week rule referrals for head and neck cancer to Western Sussex Hospitals NHS Foundation Trust'. *Br Dent J*, 227 (5), pp. 403-409.
Grimes, D., Patel, J. and Avery, C. (2017) 'New NICE referral guidance for oral cancer: does it risk delay in diagnosis?'. *British Journal of Oral and Maxillofacial Surgery*, 55 (4), pp. 404-406.
Groenewegen, P., Schellevis, F. and Boerma, W. (2016) *Primary care in 34 countries:perspectives of general practitioners and their patients.* Utrecht, The Nevetherlands: NIVEL.
Groome, P.A. *et al.* (2011) 'A population-based study of factors associated with early versus late stage oral cavity cancer diagnoses'. *Oral Oncol*, 47 (7), pp. 642-647.
Gu, F. *et al.* (2018) 'Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma'. *Hum Mol Genet*, 27 (23), pp. 4145-4156.
Guha, N. *et al.* (2007) 'Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies'. *Am J Epidemiol*, 166 (10), pp. 1159-1173.
Guizard, A.N. *et al.* (2017) 'Diagnosis and management of head and neck cancers in a high-incidence area in France: A population-based study'. *Medicine (Baltimore)*, 96 (26), pp. e7285.
Gulati, A. *et al.* (2015) 'Is an online skin cancer toolkit an effective way to educate primary care physicians about skin cancer diagnosis and referral?'. *Journal of the European Academy of Dermatology and Venereology*, 29 (11), pp. 2152-2159.
Györfi, L. *et al.* (2002) *A distribution-free theory of nonparametric regression.* Springer.
Ha, S.M. *et al.* (2017) 'Validation of Three Scoring Risk-Stratification Models for Thyroid Nodules'. *Thyroid*, 27 (12), pp. 1550-1557.
Haikel, S. *et al.* (2011) 'The Effect of Increasing Two-Week Wait Referrals for Head and Neck Cancer in East Kent'. *The Bulletin of the Royal College of Surgeons of England*, 93 (6), pp. 217-220.
Hamilton, W. (2009) 'The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients'. *Br J Cancer*, 101 Suppl 2 (Suppl 2), pp. S80-86.
Hamilton, W. (2010) 'Cancer diagnosis in primary care'. *British Journal of General Practice*, 60 (571), pp. 121-128.
Hamilton, W. *et al.* (2013) 'Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study'. *British Journal of General Practice*, 63 (606), pp. e30-e36.
Hamilton, W. *et al.* (2009) 'The risk of colorectal cancer with symptoms at different ages and between the sexes: a case-control study'. *BMC Medicine*, 7 (1), pp. 17.
Hamilton, W. *et al.* (2005) 'Clinical features of colorectal cancer before diagnosis: a population-based case–control study'. *British Journal of Cancer*, 93 (4), pp. 399-405.
Hansen, R.P. *et al.* (2011) 'General practitioner characteristics and delay in cancer diagnosis. a population-based cohort study'. *BMC Family Practice*, 12 (1), pp. 100.
Hardman, J.C. *et al.* (2021) 'Remote triage incorporating symptom-based risk stratification for suspected head and neck cancer referrals: A prospective population-based study'. *Cancer*.

Harris, M. *et al.* (2018) 'Identifying important health system factors that influence primary care practitioners' referrals for cancer suspicion: a European cross-sectional survey'. *BMJ Open,* 8 (9), pp. e022904.
Harrison, H. *et al.* (2021) 'Risk Prediction Models for Kidney Cancer: A Systematic Review'. *Eur Urol Focus,* 7 (6), pp. 1380-1390.
Harrison, H. *et al.* (2022) 'Risk prediction models for symptomatic patients with bladder and kidney cancer: a systematic review'. *British Journal of General Practice,* 72 (714), pp. e11-e18.
Hart, G.R. *et al.* (2020) 'Population-Based Screening for Endometrial Cancer: Human vs. Machine Intelligence'. *Front Artif Intell,* 3 539879.
Hashibe, M. *et al.* (2008) 'Multiple ADH genes are associated with upper aerodigestive cancers'. *Nat Genet,* 40 (6), pp. 707-709.
Hastie, T., Tibshirani, R. and Friedman, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.
healthwatch (2022) *Our position on NHS dentistry.* Available at: https://www.healthwatch.co.uk/news/2022-10-12/our-position-nhs-dentistry (Accessed: 01/11/2022).
Hee, T.G. *et al.* (2013) 'Stratifying patients with haematuria into high or low risk groups for bladder cancer: a novel clinical scoring system'. *Asian Pac J Cancer Prev,* 14 (11), pp. 6327-6330.
Hernandez, D.J. *et al.* (2009) 'Predicting the outcome of prostate biopsy: comparison of a novel logistic regression-based model, the prostate cancer risk calculator, and prostate-specific antigen level alone'. *BJU Int,* 103 (5), pp. 609-614.
Hippisley-Cox, J. and Coupland, C. (2011) 'Identifying patients with suspected gastro-oesophageal cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 61 (592), pp. e707-e714.
Hippisley-Cox, J. and Coupland, C. (2012a) 'Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 62 (594), pp. e29-e37.
Hippisley-Cox, J. and Coupland, C. (2012b) 'Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 62 (594), pp. e38-e45.
Hippisley-Cox, J. and Coupland, C. (2012c) 'Identifying patients with suspected renal tract cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 62 (597), pp. e251-e260.
Hippisley-Cox, J. and Coupland, C. (2012d) 'Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm'. *BMJ,* 344 d8009.
Hippisley-Cox, J. and Coupland, C. (2013a) 'Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 63 (606), pp. e1-e10.
Hippisley-Cox, J. and Coupland, C. (2013b) 'Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm'. *British Journal of General Practice,* 63 (606), pp. e11-e21.
Ho, T., Zahurak, M. and Koch, W.M. (2004) 'Prognostic Significance of Presentation-to-Diagnosis Interval in Patients With Oropharyngeal Carcinoma'. *Archives of Otolaryngology–Head & Neck Surgery,* 130 (1), pp. 45.

Hobson, J.C. *et al.* (2008) 'Outcomes for patients referred urgently with suspected head and neck cancer'. *The Journal of Laryngology & Otology*, 122 (11), pp. 1241-1244.
Hodder, R.J. *et al.* (2005) 'Pitfalls in the construction of cancer guidelines demonstrated by the analyses of colorectal referrals'. *Annals of the Royal College of Surgeons of England*, 87 6 419-426.
Hofmarcher, T. *et al.* (2019) 'Comparator Report on Cancer in Europe 2019 - Disease Burden, Costs and Access to medicines'. *IHE Report*, 2019:7. IHE:lUND, Sweden.
Hoggart, C. *et al.* (2012) 'A risk model for lung cancer incidence'. *Cancer Prev Res (Phila)*, 5 (6), pp. 834-846.
Holtedahl, K. *et al.* (2018) 'Abdominal symptoms and cancer in the abdomen: prospective cohort study in European primary care'. *British Journal of General Practice*, 68 (670), pp. e301-e310.
Hong, B. *et al.* (2016) 'Two-week wait false alarms? A prospective investigation of 2WW head and neck cancer referrals'. *British Dental Journal*, 220 (10), pp. 521-526.
https://fundingawards.nihr.ac.uk (2022) *Research award: EVolution of a patiEnt-REported symptom-based risk stratification sySTem to redesign the suspected Head and Neck cancer referral pathway (EVEREST-HN).* Available at: https://fundingawards.nihr.ac.uk/award/NIHR202862 (Accessed: 21/05/22).
Hudečková, M. *et al.* (2021) 'Gene Mutations in Circulating Tumour DNA as a Diagnostic and Prognostic Marker in Head and Neck Cancer—A Systematic Review'. *Biomedicines*, 9 (11), pp. 1548.
Humphrys, E. *et al.* (2019) 'The influence of health literacy on the timely diagnosis of symptomatic cancer: A systematic review'. *Eur J Cancer Care (Engl)*, 28 (1), pp. e12920.
Hüsing, A. *et al.* (2016) 'An epidemiological model for prediction of endometrial cancer risk in Europe'. *Eur J Epidemiol*, 31 (1), pp. 51-60.
Hutt, S. *et al.* (2021) 'Statistical Meta-Analysis of Risk Factors for Endometrial Cancer and Development of a Risk Prediction Model Using an Artificial Neural Network Algorithm'. *Cancers (Basel)*, 13 (15), pp.
IARC (2020) *Cancer Today.* Available at: https://gco.iarc.fr/today/home (Accessed: 02/03/2020).
Ingebrigtsen, S.G. *et al.* (2013) 'Frequency of 'warning signs of cancer' in Norwegian general practice, with prospective recording of subsequent cancer'. *Fam Pract*, 30 (2), pp. 153-160.
INTEGRATE (2020) *ENT UK HN CANCER TELEPHONE TRIAGE: SERVICE EVALUATION*
. Available at: https://entintegrate.co.uk/entuk2wwtt (Accessed: 20/5/21).
Internal data analysis of Cancer Incidence for Scotland in 2015. (2015) Available at: www.isdscotland.
org/Health-Topics/Cancer/Cancer-Statistics (Accessed: 03/01/2019).
Internal data analysis of cancer incidence statistics for England in 2015. (2015) Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddi
seases/bulletins/cancerregistrationstatisticsengland/previousRe
leases. (Accessed: 02/02/2019).
Internal data analysis of Cancer Incidence Statistics for Northern Ireland in 2015. (2015) Available at: www.qub.ac.uk/research-

centres/nicr/CancerInformation/officialstatistics/BySite/all-cancers (Accessed: 03/01/2019).
Internal data analysis of Cancer Incidence Statistics for Wales in 2015. (2015) Available at: www.wcisu.wales.nhs.uk/cancer-incidence-in-wales-1 ((Accessed: 03/01/2019).
International data analysis of Cancer Incidence statistics for Egland in 2015. (2015) Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/previousReleases. (Accessed: 05/09/2019).
Iskandarsyah, A. *et al.* (2014) 'Psychosocial and cultural reasons for delay in seeking help and nonadherence to treatment in Indonesian women with breast cancer: a qualitative study'. *Health Psychol,* 33 (3), pp. 214-221.
Iyen-Omofoman, B. *et al.* (2013) 'Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier'. *Thorax,* 68 (5), pp. 451-459.
Jackson, A. *et al.* (1998) 'Can primary prevention or selective screening for melanoma be more precisely targeted through general practice? A prospective study to validate a self administered risk score'. *Bmj,* 316 (7124), pp. 34-38; discussion 38-39.
Jacobs, I. *et al.* (1990) 'A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer'. *Br J Obstet Gynaecol,* 97 (10), pp. 922-929.
Janssen, K.J. *et al.* (2008) 'Updating methods improved the performance of a clinical prediction model in new patients'. *J Clin Epidemiol,* 61 (1), pp. 76-86.
Johnson, C.M. and Smolenski, D. (2007) 'Risk assessment models to estimate cancer probabilities'. *Current Oncology Reports,* 9 (6), pp. 503-508.
Jones, R. *et al.* (2007) 'Alarm symptoms in early diagnosis of cancer in primary care: cohort study using General Practice Research Database'. *Bmj,* 334 (7602), pp. 1040.
Joshi, H. and McPartlin, D. (2012) 'Aiming to reduce late detection of head and neck cancer'. *Br J Gen Pract,* 62 (597), pp. 182.
Kaddour, H. *et al.* (2021) 'Remote triaging of urgent suspected head and neck cancer referrals: our experience during the first wave of the COVID-19 pandemic'. *European Archives of Oto-Rhino-Laryngology*.
Kaijser, J. *et al.* (2014) 'Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis'. *Hum Reprod Update,* 20 (3), pp. 449-462.
Kaiser, I. *et al.* (2020) 'Risk Prediction Models for Melanoma: A Systematic Review on the Heterogeneity in Model Development and Validation'. *Int J Environ Res Public Health,* 17 (21), pp.
Kaplan, D.J. *et al.* (2010) 'Evaluation of the Prostate Cancer Prevention Trial Risk calculator in a high-risk screening population'. *BJU Int,* 105 (3), pp. 334-337.
Kassirian, S. *et al.* (2020) 'Delay in Diagnosis of Patients with Head-and-Neck Cancer in Canada: Impact of Patient and Provider Delay'. *Current Oncology,* 27 (5), pp. 467-477.
Kattan, M.W. *et al.* (2003) 'Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors'. *J Urol,* 170 (5), pp. 1792-1797.

Kaufman, S., Grabau, J.C. and Loré, J.M., Jr. (1980) 'Symptomatology in head and neck cancer: a quantitative review of 385 cases'. *Am J Public Health,* 70 (5), pp. 520-522.
Keane, M.G. *et al.* (2014) 'A case–control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer'. *BMJ Open,* 4 (11), pp. e005720.
Kennedy, A.-M. *et al.* (2012) 'Do GP referral guidelines really work? Audit of an electronic urgent referral system for suspected head and neck cancer'. *European Archives of Oto-Rhino-Laryngology,* 269 (5), pp. 1509-1512.
Kerdpon, D. (2001) 'Factors related to advanced stage oral squamous cell carcinoma in southern Thailand'. *Oral Oncology,* 37 (3), pp. 216-221.
Kim, M.-K. *et al.* (2009) 'A hospital-based case-control study of identifying ovarian cancer using symptom index'. *Journal of Gynecologic Oncology,* 20 (4), pp. 238.
Kiong, K.L. *et al.* (2021) 'The impact of COVID -19 on head and neck cancer diagnosis and disease extent'. *Head & Neck,* 43 (6), pp. 1890-1897.
Kitson, S.J., Evans, D.G. and Crosbie, E.J. (2017) 'Identifying High-Risk Women for Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model'. *Cancer Prevention Research,* 10 (1), pp. 1-13.
Koivunen, P. *et al.* (2001) 'The impact of patient and professional diagnostic delays on survival in pharyngeal cancer'. *Cancer,* 92 (11), pp. 2885-2891.
Koo, M.M. *et al.* (2018) 'Symptom Signatures and Diagnostic Timeliness in Cancer Patients: A Review of Current Evidence'. *Neoplasia,* 20 (2), pp. 165-174.
Koo, M.M. *et al.* (2021) 'Conceptual Framework to Guide Early Diagnosis Programs for Symptomatic Cancer as Part of Global Cancer Control'. *JCO Global Oncology,* (7), pp. 35-45.
Kop, R. *et al.* (2015) 'On the Advantage of Using Dedicated Data Mining Techniques to Predict Colorectal Cancer'. Springer International Publishing, pp. 133-142.
Kostopoulou, O. *et al.* (2019) 'Referral Decision Making of General Practitioners: A Signal Detection Study'. *Med Decis Making,* 39 (1), pp. 21-31.
Kowalski, L.P. *et al.* (1994) 'Lateness of diagnosis of oral and oropharyngeal carcinoma: factors related to the tumour, the patient and health professionals'. *Eur J Cancer B Oral Oncol,* 30b (3), pp. 167-173.
Koyanagi, Y.N. *et al.* (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 (1), pp. 38-47.
Krishna Rao, S. *et al.* (2016) 'A screening model for oral cancer using risk scores: development and validation'. *Community Dentistry and Oral Epidemiology,* 44 (1), pp. 76-84.
Kruppa, J. *et al.* (2014) 'Probability estimation with machine learning methods for dichotomous and multicategory outcome: theory'. *Biometrical Journal,* 56 (4), pp. 534-563.
Kuhn, M. and Johnson, K. (2013) *Applied predictive modeling.* Springer.
Kumar, R. *et al.* (2012) 'Efficacy of the Two-Week Wait Referral System for Head and Neck Cancer: A Systematic Review'. *The Bulletin of the Royal College of Surgeons of England,* 94 (3), pp. 101-105.
Kwak, J.Y. *et al.* (2013) 'Image reporting and characterization system for ultrasound features of thyroid nodules: multicentric Korean retrospective study'. *Korean J Radiol,* 14 (1), pp. 110-117.

Kyzas, P. (2021) 'Nurse-led oral and maxillofacial oncology clinics: a review'. *British Journal of Oral and Maxillofacial Surgery,* 59 (2), pp. 151-155.
Langton, S. *et al.* (2019) 'Referrals for head and neck cancer in England and The Netherlands: an international qualitative study of the views of secondary-care surgical specialists'. *British Journal of Oral and Maxillofacial Surgery,* 57 (2), pp. 116-124.
Langton, S., Siau, D. and Bankhead, C. (2016) 'Two-week rule in head and neck cancer 2000-14: a systematic review'. *Br J Oral Maxillofac Surg,* 54 (2), pp. 120-131.
Lau, K., Wilkinson, J. and Moorthy, R. (2018) 'A web-based prediction score for head and neck cancer referrals'. *Clin Otolaryngol,* 43 (4), pp. 1043-1049.
Lee, F. *et al.* (2018) 'An evaluation of in-office flexible fiber-optic biopsies for laryngopharyngeal lesions'. *J Otolaryngol Head Neck Surg,* 47 (1), pp. 31.
Lee, Y.-C.A. *et al.* (2020) 'Risk Prediction Models for Head and Neck Cancer in the US Population From the INHANCE Consortium'. *American Journal of Epidemiology,* 189 (4), pp. 330-342.
Lee, Y.C. *et al.* (2008) 'Involuntary smoking and head and neck cancer risk: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium'. *Cancer Epidemiol Biomarkers Prev,* 17 (8), pp. 1974-1981.
Leeman JE, K.N., Wong, RJ, Lee NY, Romesser PB. (2020) 'Cancer of the Head and Neck'. In*:* Niederhuber JE, A.J., Doroshow JH, Kastan MB, Tepper JE, eds. (ed.) *Abeloff's Clinical Oncology.* 6th edn. Philadelphia, Pa: Elsevier, pp.
Leeners, B. *et al.* (1996) 'Colour Doppler sonography improves the pre-operative diagnosis of ovarian tumours made using conventional transvaginal sonography'. *Eur J Obstet Gynecol Reprod Biol,* 64 (1), pp. 79-85.
Lewis, N.R., Le Jeune, I. and Baldwin, D.R. (2005) 'Under utilisation of the 2-week wait initiative for lung cancer by primary care and its effect on the urgent referral pathway'. *British Journal of Cancer,* 93 (8), pp. 905-908.
Li, H. *et al.* (2021) 'Risk prediction models for esophageal cancer: A systematic review and critical appraisal'. *Cancer Med,* 10 (20), pp. 7265-7276.
Li, L.Q. *et al.* (2022) 'Primary care triaging of head and neck cancer referrals using the HaNC-RC v.2: Impact on a Tertiary Head and Neck Service'. *J Laryngol Otol,* 1-24.
Lingwood, R.J. *et al.* (2008) 'The challenge of cancer control in Africa'. *Nat Rev Cancer,* 8 (5), pp. 398-403.
Liu, M. *et al.* (2017a) 'A Model To Identify Individuals at High Risk for Esophageal Squamous Cell Carcinoma and Precancerous Lesions in Regions of High Prevalence in China'. *Clin Gastroenterol Hepatol,* 15 (10), pp. 1538-1546 e1537.
Liu, Y. *et al.* (2017b) 'Quantitative prediction of oral cancer risk in patients with oral leukoplakia'. *Oncotarget,* 8 (28), pp. 46057-46064.
Loeb, S. and Dani, H. (2017) 'Whom to Biopsy: Prediagnostic Risk Stratification with Biomarkers, Nomograms, and Risk Calculators'. *Urol Clin North Am,* 44 (4), pp. 517-524.
Loo, R.K. *et al.* (2013) 'Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria'. *Mayo Clin Proc,* 88 (2), pp. 129-138.
Lortet-Tieulent, J. *et al.* (2020) 'Profiling global cancer incidence and mortality by socioeconomic development'. *International Journal of Cancer,* 147 (11), pp. 3029-3036.
Lotan, Y. *et al.* (2009) 'Impact of clinical factors, including a point-of-care nuclear matrix protein-22 assay and cytology, on bladder cancer detection'. *BJU Int,* 103 (10), pp. 1368-1374.

Loud, J.T. and Murphy, J. (2017) 'Cancer Screening and Early Detection in the 21(st) Century'. *Semin Oncol Nurs*, 33 (2), pp. 121-128.
Louie, K.S. *et al.* (2015) 'Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis'. *Annals of Oncology*, 26 (5), pp. 848-864.
Luryi, A.L. *et al.* (2014) 'Public Awareness of Head and Neck Cancers: A Cross-Sectional Survey'. *JAMA Otolaryngology–Head & Neck Surgery*, 140 (7), pp. 639-646.
Lyons, M. *et al.* (2004) 'Audit of referrals for head and neck cancer - the effect of the 2-week, fast track referral system'. *Clinical Otolaryngology and Allied Sciences*, 29 (2), pp. 143-145.
Lyratzopoulos, G. *et al.* (2018) 'Associations between diagnostic activity and measures of patient experience in primary care: a cross-sectional ecological study of English general practices'. *Br J Gen Pract*, 68 (666), pp. e9-e17.
Lyratzopoulos, G. *et al.* (2012) 'Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England'. *The Lancet Oncology*, 13 (4), pp. 353-365.
MacKie, R.M. (1990) 'Clinical recognition of early invasive malignant melanoma'. *Bmj*, 301 (6759), pp. 1005-1006.
MacKie, R.M., Freudenberger, T. and Aitchison, T.C. (1989) 'Personal risk-factor chart for cutaneous melanoma'. *Lancet*, 2 (8661), pp. 487-490.
Maclean, R. *et al.* (2015) 'Primary care characteristics and stage of cancer at diagnosis using data from the national cancer registration service, quality outcomes framework and general practice information'. *BMC Cancer*, 15 (1), pp. 500.
Macleod, U. *et al.* (2009) 'Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers'. *Br J Cancer*, 101 Suppl 2 (Suppl 2), pp. S92-s101.
MacMillan Cancer Support (2018) *Understanding Head and Neck Cancers.*
Maddams, J., Utley, M. and Møller, H. (2012) 'Projections of cancer prevalence in the United Kingdom, 2010–2040'. *British Journal of Cancer*, 107 (7), pp. 1195-1202.
Marshall, T. *et al.* (2011) 'The diagnostic performance of scoring systems to identify symptomatic colorectal cancer compared to current referral guidance'. *Gut*, 60 (9), pp. 1242-1248.
Martínez, M.T. *et al.* (2015) 'Implementation and assessment of a fast-track programme to improve communication between primary and specialized care in patients with suspected cancer: how to shorten time between initial symptoms of cancer, diagnosis and initiation of treatment'. *Clin Transl Oncol*, 17 (2), pp. 167-172.
Martínez, M.T. *et al.* (2021) 'Ten-year assessment of a cancer fast-track programme to connect primary care with oncology: reducing time from initial symptoms to diagnosis and treatment initiation'. *ESMO Open*, 6 (3), pp. 100148.
Mashberg, A. *et al.* (1989) 'Appearance, site of occurrence, and physical and clinical characteristics of oral carcinoma in Torino, Italy'. *Cancer*, 63 (12), pp. 2522-2527.
Mathoulin-Pélissier, S. and Pritchard-Jones, K. (2019) 'Evidence-based data and rare cancers: The need for a new methodological approach in research and investigation'. *European Journal of Surgical Oncology*, 45 (1), pp. 22-30.

Matulewicz, R.S., Rademaker, A. and Meeks, J.J. (2020) 'A simplified nomogram to assess risk of bladder cancer in patients with a new diagnosis of microscopic hematuria'. *Urol Oncol,* 38 (4), pp. 240-246.
Mazzola, E. *et al.* (2015) 'Recent Enhancements to the Genetic Risk Prediction Model BRCAPRO'. *Cancer Inform,* 14 (Suppl 2), pp. 147-157.
McCarthy, C. *et al.* (2020) 'Development and validation of a multivariable risk prediction model for head and neck cancer using the UK Biobank'. *International Journal of Oncology*.
McCowan, C. *et al.* (2011) 'Identifying suspected breast cancer: development and validation of a clinical prediction rule'. *British Journal of General Practice,* 61 (586), pp. e205-e214.
McCutchan, G. *et al.* (2016) 'Barriers to cancer symptom presentation among people from low socioeconomic groups: a qualitative study'. *BMC Public Health,* 16 (1), pp. 1052.
McCutchan, G.M. *et al.* (2015) 'Influences of cancer symptom knowledge, beliefs and barriers on cancer symptom presentation in relation to socioeconomic deprivation: a systematic review'. *BMC cancer,* 15 1000-1000.
McDonald, K. *et al.* (2013) 'Patient Safety Strategies Targeted at Diagnostic Errors'. *Annals of Internal Medicine,* 158 (5_Part_2), pp. 381-389.
McGeoch, L. *et al.* (2019) 'Risk Prediction Models for Colorectal Cancer Incorporating Common Genetic Variants: A Systematic Review'. *Cancer Epidemiol Biomarkers Prev,* 28 (10), pp. 1580-1593.
McGurk, M. *et al.* (2005) 'Delay in diagnosis and its effect on outcome in head and neck cancer'. *Br J Oral Maxillofac Surg,* 43 (4), pp. 281-284.
McKee, M., Healy, J. and Falkingham, J. (2002) *Health care in central Asia.* Philadelphia, USA: Open University Press.
McKie, C. *et al.* (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 Oncology,* 44 (9), pp. 851-856.
McRonald, F.E. *et al.* (2014) 'The UK Lung Screen (UKLS): demographic profile of first 88,897 approaches provides recommendations for population screening'. *Cancer Prev Res (Phila),* 7 (3), pp. 362-371.
McWilliams, A. *et al.* (2013) 'Probability of cancer in pulmonary nodules detected on first screening CT'. *N Engl J Med,* 369 (10), pp. 910-919.
Meads, C., Ahmed, I. and Riley, R.D. (2012) 'A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance'. *Breast Cancer Res Treat,* 132 (2), pp. 365-377.
Meechan, D. *et al.* (2012) 'Variation in use of the 2-week referral pathway for suspected cancer: A cross-sectional analysis'. *British Journal of General Practice,* 62 (602), pp. e590-e597.
Mehanna, H. *et al.* (2010) 'Head and neck cancer—Part 1: Epidemiology, presentation, and prevention'. *BMJ,* 341 c4684.
Mendonca, S.C. *et al.* (2019) 'Associations between general practice characteristics with use of urgent referrals for suspected cancer and endoscopies: a cross-sectional ecological study'. *Family Practice,* 36 (5), pp. 573-580.
Merletti, F. *et al.* (1990) 'Topographic classification, clinical characteristics, and diagnostic delay of cancer of the larynx/hypopharynx in Torino, Italy'. *Cancer,* 66 (8), pp. 1711-1716.

Metcalfe, C. *et al.* (2019) 'Introduction of a referral pathway guide for general dental practitioners in Cheshire & Merseyside: the effect on two-week suspected cancer referrals'. *British Dental Journal,* 227 (12), pp. 1058-1062.
Metcalfe, C. *et al.* (2021) 'Introduction of a novel telescopic pathway to streamline 2-week-wait suspected head and neck cancer referrals and improve efficiency: A prospective service evaluation'. *Laryngoscope Investigative Otolaryngology*.
Mettias, B., Charlton, A. and 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 (10), pp. 869-873.
Meys, E.M. *et al.* (2016) 'Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis'. *Eur J Cancer,* 58 17-29.
Miller, C.C. and Hierons, R.J. (2012) 'Two Audits of the Diagnosis of Oral Cancer and the Two-Week Rule following Referrals from Primary Care Practitioners in Newcastle'. *Primary Dental Care,* os19 (2), pp. 63-68.
Minaretzis, D. *et al.* (1994) 'Ovarian tumors: prediction of the probability of malignancy by using patient's age and tumor morphologic features with a logistic model'. *Gynecol Obstet Invest,* 38 (2), pp. 140-144.
Mody, M.D. *et al.* (2021) 'Head and neck cancer'. *The Lancet*.
Moffat, J. *et al.* (2015) 'The impact of national cancer awareness campaigns for bowel and lung cancer symptoms on sociodemographic inequalities in immediate key symptom awareness and GP attendances'. *Br J Cancer,* 112 Suppl 1 (Suppl 1), pp. S14-21.
Moffat, J., Ironmonger, L. and Green, T. (2014) *Clinical Decision Support Tool for Cancer (CDS) Project: Evaluation Report to the Department of Health.* Available at: https://www.cancerresearchuk.org/sites/default/files/cds_final_310714.pdf (Accessed: 19/12/2021).
Møller, H. *et al.* (2015) 'Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study'. *BMJ,* h5102.
Møller, H. *et al.* (2010) 'Breast cancer survival in England, Norway and Sweden: a population-based comparison'. *International Journal of Cancer,* 127 (11), pp. 2630-2638.
Moloney, J. and Stassen, L.F. (2010) 'An audit of the quality of referral letters received by the Department of Oral and Maxillofacial Surgery, Dublin Dental School and Hospital'. *J Ir Dent Assoc,* 56 (5), pp. 221-223.
Montgomery, J. *et al.* (2019) "Urgent suspicion of cancer' referrals to a head and neck clinic - what do patients expect?'. *J Laryngol Otol,* 133 (9), pp. 782-787.
Moons, K.G.M. *et al.* (2012a) 'Risk prediction models: II. External validation, model updating, and impact assessment'. *Heart,* 98 (9), pp. 691.
Moons, K.G.M. *et al.* (2012b) 'Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker'. *Heart,* 98 (9), pp. 683.
Moor, J.W., Paleri, V. and Edwards, J. (2019) 'Patient classification of two-week wait referrals for suspected head and neck cancer: a machine learning approach'. *J Laryngol Otol,* 133 (10), pp. 875-878.
Moore, P.L.A. *et al.* (2004) 'Detection of laryngeal carcinoma and epithelial hyperplastic laryngeal lesions via a rapid-access dysphonia clinic'. *The Journal of Laryngology & Otology,* 118 (8), pp. 633-636.

Moore, S.F. *et al.* (2021) 'The impact of changing risk thresholds on the number of people in England eligible for urgent investigation for possible cancer: an observational cross-sectional study'. *Br J Cancer,* 1-5.
Morhason-Bello, I.O. *et al.* (2013) 'Challenges and opportunities in cancer control in Africa: a perspective from the African Organisation for Research and Training in Cancer'. *The Lancet Oncology,* 14 (4), pp. e142-e151.
Morisod, B. and Simon, C. (2016) 'Meta-analysis on survival of patients treated with transoral surgery versus radiotherapy for early-stage squamous cell carcinoma of the oropharynx'. *Head Neck,* 38 Suppl 1 E2143-2150.
Morrissey, J.J. *et al.* (2015) 'Evaluation of Urine Aquaporin-1 and Perilipin-2 Concentrations as Biomarkers to Screen for Renal Cell Carcinoma: A Prospective Cohort Study'. *JAMA Oncol,* 1 (2), pp. 204-212.
Moyer, V.A. (2014) 'Screening for oral cancer: U.S. Preventive Services Task Force recommendation statement'. *Ann Intern Med,* 160 (1), pp. 55-60.
Murchie, P. *et al.* (2012) 'Comparing diagnostic delay in cancer: a cross-sectional study in three European countries with primary care-led health care systems'. *Family Practice,* 29 (1), pp. 69-78.
Murchie, P. *et al.* (2015) 'General practice performance in referral for suspected cancer: influence of number of cases and case-mix on publicly reported data'. *British Journal of Cancer,* 112 (11), pp. 1791-1798.
Muris, J.W. *et al.* (1995) 'Non-acute abdominal complaints in general practice: diagnostic value of signs and symptoms'. *Br J Gen Pract,* 45 (395), pp. 313-316.
Musonda, P. *et al.* (2011) 'Comparing the performance of two clinical models in estimating the risk of endometrial cancer in symptomatic postmenopausal women'. *Eur J Obstet Gynecol Reprod Biol,* 159 (2), pp. 433-438.
myhealth.london.nhs.uk (2017) *Pan-London Suspected Cancer Referral Guide - Head & Neck.* Available at: https://myhealth.london.nhs.uk/wp-content/uploads/2019/04/Pan-London-Suspected-Cancer-Referral-Guide-Head-and-Neck.pdf.
National Cancer Intitute (2017) '*The Breast Cancer Risk Assessment Tool*'.
National Comprehensive Cancer Network (2020) '*NCCN Clinical Practice Guidelines in Oncology: Breast Cancer Risk Reduction. Version 1.2020*'.
NCRAS (2016) *Routes to diagnosis;2006 - 2013.* National Cancer Registration and Analysis service.
NCRI (2010) *Cancer by deprIvation in England 1996 - 2011.* (Accessed: 07/10/2021).
Neal, R.D. (2009) 'Do diagnostic delays in cancer matter?'. *British Journal of Cancer,* 101 (S2), pp. S9-S12.
Neal, R.D. *et al.* (2014) 'Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database'. *British Journal of Cancer,* 110 (3), pp. 584-592.
Negri, E. *et al.* (2009) 'Family history of cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium'. *Int J Cancer,* 124 (2), pp. 394-401.
Nguyen, C.T. *et al.* (2010) 'Performance of prostate cancer prevention trial risk calculator in a contemporary cohort screened for prostate cancer and diagnosed by extended prostate biopsy'. *J Urol,* 183 (2), pp. 529-533.

NHS (2019) *The NHS Long Term Plan.* Available at: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ (Accessed: 20/12/21).
NHS England (2019) *The NHS Long Term Plan.* Available at: www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf (Accessed: 05/03/2020).
NHSScotland (2019) *Scottish Referral Guidelines for Suspected Cancer.* Available at: http://www.cancerreferral.scot.nhs.uk/.
NICaN (2021) *Northern Ireland Referral Guidance for Suspected Cancer – Red Flag Criteria.* Available at: https://nican.hscni.net/info-for-professionals/primary-care/primary-care-guidance/ (Accessed: 05/09).
NICE (2004) *Guidance on Cancer Services: Improving outcomes in head and neck cancers. The Manual.* Available at: www.nice.org.uk/guidance/csghn/resources/improvingoutcomes-in-head-and-neck-cancers-the-manual2
NICE (2011) *Ovarian cancer: recognition and initial management. Clinical guideline [CG122].* Available at: https://www.nice.org.uk/guidance/cg122 (Accessed: 15/11/20).
NICE (2015) *Suspected cancer:recognition and referral.*
NICE (2021) *Suspected cancer:recognition and referral.* (Accessed: 03/03/2021).
NICR (2016) *Northern Irelnd Cancer Registry, Queens University Belfast, Incidence by stage 2010-2014.*
Niederhuber, J. (2006) 'The nation's investment in cancer research: A plan and budget proposal for fiscal year 2008'. *Bethesda (MD): US Department of Health and Human Services, National Institutes of Health, National Cancer Institute*.
Nieminen, M. *et al.* (2021) 'Challenges in diagnosing head and neck cancer in primary health care'. *Annals of Medicine*, 53 (1), pp. 26-33.
Nolte, E. and McKee, C.M. (2008) 'Measuring the health of nations: updating an earlier analysis'. *Health Aff (Millwood)*, 27 (1), pp. 58-71.
Nørrelund, N. and Nørrelund, H. (1996) 'Colorectal cancer and polyps in patients aged 40 years and over who consult a GP with rectal bleeding'. *Family Practice*, 13 (2), pp. 160-165.
Norwegian Directorate of health (2015) *Entrance to the package for head and neck cancer: Suspicion.* Available at: https://www.helsedirektoratet.no/pakkeforlop/hode-halskreft/inngang-til-pakkeforlop-for-hode-halskreft#mistanke.
NZGG (2009) *Suspected cancer in primary care: Guidelines for investigation, referral and reducing ethnic disparities.* Wellington, New Zealand: Ministry of Health.
O'Donnell, C.A. (2000) 'Variation in GP referral rates: what can we learn from the literature?'. *Family Practice*, 17 (6), pp. 462-471.
Occomore-Kent, L. and Slade, S. (2021) 'Speech and language therapy assessment of patients on the 2-week wait head and neck cancer referral pathway: evidence and opinion following a pilot study'. *Curr Opin Otolaryngol Head Neck Surg*, 29 (3), pp. 179-186.
Occomore-Kent, L.C., Hatch, E. and Cruice, M. (2021) 'Scoping opinion: Speech and language therapists' views on extending their role to the urgent ear, nose and throat pathway'. *International Journal of Language & Communication Disorders*.
Ohlstein, J.F. *et al.* (2015) 'Initial Experience of a Patient Navigation Model for Head and Neck Cancer'. *JAMA Otolaryngology–Head & Neck Surgery*, 141 (9), pp. 804.

Oliveira, M. *et al.* (2011) 'Head-to-head comparison of two online nomograms for prostate biopsy outcome prediction'. *BJU Int,* 107 (11), pp. 1780-1783.
Olsen, C.M. *et al.* (2015a) 'Independent validation of six melanoma risk prediction models'. *J Invest Dermatol,* 135 (5), pp. 1377-1384.
Olsen, C.M. *et al.* (2018) 'Risk Stratification for Melanoma: Models Derived and Validated in a Purpose-Designed Prospective Cohort'. *J Natl Cancer Inst,* 110 (10), pp. 1075-1083.
Olsen, M.H. *et al.* (2015b) 'Socioeconomic position and stage at diagnosis of head and neck cancer - a nationwide study from DAHANCA'. *Acta Oncol,* 54 (5), pp. 759-766.
Ong, H., Spellman, J. and Kanatas, A. (2020) 'Skills and competences needed by nurses to allow them to deliver a safe nurse-led oral and maxillofacial oncology clinic: the Leeds experience'. *British Journal of Oral and Maxillofacial Surgery,* 58 (9), pp. e3-e5.
Osses, D.F. *et al.* (2018) 'Multivariable risk-based patient selection for prostate biopsy in a primary health care setting: referral rate and biopsy results from a urology outpatient clinic'. *Transl Androl Urol,* 7 (1), pp. 27-33.
Pan, C., Issaeva, N. and Yarbrough, W.G. (2018) 'HPV-driven oropharyngeal cancer: current knowledge of molecular biology and mechanisms of carcinogenesis'. *Cancers of the Head & Neck,* 3 (1), pp.
Parekh, D.J. *et al.* (2006) 'External validation of the Prostate Cancer Prevention Trial risk calculator in a screened population'. *Urology,* 68 (6), pp. 1152-1155.
Patel, B.S. *et al.* (2021) 'Learning decision thresholds for risk stratification models from aggregate clinician behavior'. *Journal of the American Medical Informatics Association,* 28 (10), pp. 2258-2264.
Patel, M., Khan, A.Q. and Thiruchelvam, J. (2011) 'Importance of quality in referral letters sent for potentially malignant oral, head and neck lesions'. *Dent Update,* 38 (3), pp. 192-194, 196-198.
Patel, S. *et al.* (2020) 'Patients undergoing primary surgery for oral and oropharyngeal cancer: how many are referred on the two-week wait pathway and by whom?'. *Ann R Coll Surg Engl,* 102 (7), pp. 532-535.
Payten, C.L. *et al.* (2020) 'Outcomes of a multidisciplinary Ear, Nose and Throat Allied Health Primary Contact outpatient assessment service'. *Clinical Otolaryngology,* 45 (6), pp. 904-913.
PCPT (2006-2018) *PCPT Risk Calculator.* Available at: https://riskcalc.org/PCPTRC/ (Accessed: 20/11/20).
Pepe, M.S. *et al.* (2004) 'Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker'. *Am J Epidemiol,* 159 (9), pp. 882-890.
Perandini, S. *et al.* (2017) 'Multicenter external validation of two malignancy risk prediction models in patients undergoing 18F-FDG-PET for solitary pulmonary nodule evaluation'. *Eur Radiol,* 27 (5), pp. 2042-2046.
Pereira, L.H. *et al.* (2016) 'Risk Stratification System for Oral Cancer Screening'. *Cancer Prev Res (Phila),* 9 (6), pp. 445-455.
Pereira-Azevedo, N. *et al.* (2017) 'Rotterdam Prostate Cancer Risk Calculator: Development and Usability Testing of the Mobile Phone App'. *JMIR Cancer,* 3 (1), pp. e1.
Pfeiffer, R.M. *et al.* (2013) 'Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies'. *PLoS Med,* 10 (7), pp. e1001492.

Piggott, L. (2015) 'Comparison of the diagnostic yield of medical and dental referrals an oral and maxillofacial 2 weeks rule clinic'. *Natl J Maxillofac Surg,* 6 (1), pp. 52-54.
Pignon, J.P. *et al.* (2009) 'Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients'. *Radiother Oncol,* 92 (1), pp. 4-14.
Pitchers, M. and Martin, C. (2006) 'Delay in referral of oropharyngeal squamous cell carcinoma to secondary care correlates with a more advanced stage at presentation, and is associated with poorer survival'. *British Journal of Cancer,* 94 (7), pp. 955-958.
Poyet, C. *et al.* (2016) 'Prostate cancer risk prediction using the novel versions of the European Randomised Study for Screening of Prostate Cancer (ERSPC) and Prostate Cancer Prevention Trial (PCPT) risk calculators: independent validation and comparison in a contemporary European cohort'. *BJU Int,* 117 (3), pp. 401-408.
Pracy, P. *et al.* (2013) 'The 2-Week Wait Head & Neck Cancer Referrals: Is this System Working?'. 6 184-188.
Prades, J. *et al.* (2011) 'Implementing a Cancer Fast-track Programme between primary and specialised care in Catalonia (Spain): a mixed methods study'. *Br J Cancer,* 105 (6), pp. 753-759.
Price, S., Abel, G.A. and Hamilton, W. (2021) 'Guideline interval: A new time interval in the diagnostic pathway for symptomatic cancer'. *Cancer Epidemiol,* 73 101969.
Price, S. *et al.* (2019) 'Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care'. *British Journal of General Practice,* 69 (684), pp. e437-e443.
Price, S.J. *et al.* (2014) 'Non-visible versus visible haematuria and bladder cancer risk: a study of electronic records in primary care'. *Br J Gen Pract,* 64 (626), pp. e584-589.
Probst, H.B., Hussain, Z.B. and Andersen, O. (2012) 'Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians—A national Danish project'. *Health Policy,* 105 (1), pp. 65-70.
Probst, P., Bischl, B. and Boulesteix, A.-L. (2018) 'Tunability: Importance of hyperparameters of machine learning algorithms'. *arXiv preprint arXiv:1802.09596*.
Pugliano, F.A. *et al.* (1999) 'Symptoms as an Index of Biologic Behavior in Head and Neck Cancer'. *Otolaryngology–Head and Neck Surgery,* 120 (3), pp. 380-386.
Qcancer.org (2017) *The QCancer scores.* Available at: https://www.qcancer.org/ (Accessed: 03/06).
Queenan, J.A. *et al.* (2018) 'Symptom appraisal, help seeking, and lay consultancy for symptoms of head and neck cancer'. *Psycho-Oncology,* 27 (1), pp. 286-294.
Rice, M.S. *et al.* (2017) 'Breast cancer risk prediction: an update to the Rosner-Colditz breast cancer incidence model'. *Breast Cancer Res Treat,* 166 (1), pp. 227-240.
Richards, M.A. (2009) 'The size of the prize for earlier diagnosis of cancer in England'. *British Journal of Cancer,* 101 (S2), pp. S125-S129.
Richards, M.A. *et al.* (1999) 'Influence of delay on survival in patients with breast cancer: a systematic review'. *Lancet,* 353 (9159), pp. 1119-1126.
Ridic, G., Gleason, S. and Ridic, O. (2012) 'Comparisons of health care systems in the United States, Germany and Canada'. *Materia socio-medica,* 24 (2), pp. 112-120.
Rimmer, J. *et al.* (2012) 'A Head and Neck 'Two-Week Wait' Clinic: Cancer Referrals or the Worried Well ?'. *The Bulletin of the Royal College of Surgeons of England,* 94 (6), pp. 1-4.

Rockhill, B. *et al.* (2001) 'Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention'. *J Natl Cancer Inst,* 93 (5), pp. 358-366.
Roennegaard, A.B. *et al.* (2018) 'The Danish Head and Neck Cancer fast-track program: a tertiary cancer centre experience'. *European Journal of Cancer,* 90 133-139.
Rogers, S.N., Dailey, Y. and Langton, S. (2021) 'Re: re: Two-week rule: suspected head and neck cancer referrals from a general medical practice perspective'. *British Journal of Oral and Maxillofacial Surgery,* 59 (5), pp. 612-613.
Rogers, S.N. *et al.* (2019) 'Audit of the two-week pathway for patients with suspected cancer of the head and neck and the influence of socioeconomic status'. *British Journal of Oral and Maxillofacial Surgery,* 57 (5), pp. 419-424.
Roobol, M.J. and Schröder, F.H. (2003) 'European Randomized Study of Screening for Prostate Cancer: achievements and presentation'. *BJU Int,* 92 Suppl 2 117-122.
Rosell Ferrer, R. *et al.* (2021) 'Validity of a rapid diagnostic programme for head and neck cancer'. *Acta Otorrinolaringologica (English Edition),* 72 (5), pp. 271-279.
Rosner, B. *et al.* (2013) 'Validation of Rosner–Colditz breast cancer incidence model using an independent data set, the California Teachers Study'. *Breast cancer research and treatment,* 142 (1), pp. 187-202.
Rosner, B. and Colditz, G.A. (1996) 'Nurses' health study: log-incidence mathematical model of breast cancer incidence'. *J Natl Cancer Inst,* 88 (6), pp. 359-364.
Rossing, M.A. *et al.* (2010) 'Predictive Value of Symptoms for Early Detection of Ovarian Cancer'. *JNCI: Journal of the National Cancer Institute,* 102 (4), pp. 222-229.
Round, T. *et al.* (2021) 'Cancer detection via primary care urgent referral in England 2009/10 to 2018/19 and the association with practice characteristics'. *British Journal of General Practice,* BJGP.2020.1030.
Round, T. *et al.* (2020) 'Association between use of urgent suspected cancer referral and mortality and stage at diagnosis: a 5-year national cohort study'. *British Journal of General Practice,* 70 (695), pp. e389-e398.
Roy, S. and Anjum, K. (2018) 'The two-week wait – a qualitative analysis of suspected head and neck cancer referrals'. *British Dental Journal,* 225 (2), pp. 159-163.
Rutkowska, M. *et al.* (2020) 'Oral cancer: The first symptoms and reasons for delaying correct diagnosis and appropriate treatment'. *Advances in Clinical and Experimental Medicine,* 29 (6), pp. 735-743.
Rylands, J., Lowe, D. and Rogers, S.N. (2016) 'Outcomes by area of residence deprivation in a cohort of oral cancer patients: Survival, health-related quality of life, and place of death'. *Oral Oncol,* 52 30-36.
Saltman, R., Rico, A. and Boerma, W. (2006) *Primary care in the driver's seat?* Maidenhead: Open University Press.
Sankaranarayanan, R., Ramadas, K. and Qiao, Y.-L. (2014) 'Managing the changing burden of cancer in Asia'. *BMC Medicine,* 12 (1), pp. 3.
Sankaranarayanan, R. *et al.* (2010) 'Cancer survival in Africa, Asia, and Central America: a population-based study'. *Lancet Oncol,* 11 (2), pp. 165-173.
Sarma, E.A., Kobrin, S.C. and Thompson, M.J. (2020) 'A Proposal to Improve the Early Diagnosis of Symptomatic Cancers in the United States'. *Cancer Prev Res (Phila),* 13 (9), pp. 715-720.

Sassone, A.M. *et al.* (1991) 'Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy'. *Obstet Gynecol,* 78 (1), pp. 70-76.
Scelo, G. *et al.* (2018) 'KIM-1 as a Blood-Based Marker for Early Detection of Kidney Cancer: A Prospective Nested Case-Control Study'. *Clin Cancer Res,* 24 (22), pp. 5594-5601.
Schmidt-Hansen, M. *et al.* (2017) 'Lung cancer in symptomatic patients presenting in primary care: a systematic review of risk prediction tools'. *British Journal of General Practice,* 67 (659), pp. e396-e404.
Schwarz, D.F., König, I.R. and Ziegler, A. (2010) 'On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data'. *Bioinformatics,* 26 (14), pp. 1752-1758.
Scott, S.E., Grunfeld, E.A. and McGurk, M. (2005) 'The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma'. *Oral Oncology,* 41 (4), pp. 396-403.
Scott, S.E. *et al.* (2020) 'Tracking cancer occurrence in the 5 years after referral for suspected head and neck cancer'. *Oral Oncology,* 109 104955.
Seabrook, M. *et al.* (2019) 'Implementation of an extended scope of practice speech-language pathology allied health practitioner service: An evaluation of service impacts and outcomes'. *Int J Speech Lang Pathol,* 21 (1), pp. 65-74.
Sengoku, K. *et al.* (1994) 'Evaluation of transvaginal color Doppler sonography, transvaginal sonography and CA 125 for prediction of ovarian malignancy'. *Int J Gynaecol Obstet,* 46 (1), pp. 39-43.
Shah, H.V., Williams, R.W. and Irvine, G.H. (2006) 'Fast-track referrals for oral lesions: A prospective study'. *British Journal of Oral and Maxillofacial Surgery,* 44 (3), pp. 207-208.
Shaha, A., Byers, R. and Terz, J. (1997a) 'Society of Surgical Oncology Practice Guidelines. Laryngeal cancer surgical practice guidelines'. *Oncology,* 11 (8), pp. 1215-1216.
Shaha, A., Byers, R. and Terz, J. (1997b) 'Society of Surgical Oncology Practice Guidelines. Oropharyngeal and oral cavity cancer surgical practice guidelines'. *Oncology* 11 (8), pp. 1211–1212.
Sharma, N. and Om, H. (2015) 'Usage of Probabilistic and General Regression Neural Network for Early Detection and Prevention of Oral Cancer'. *The Scientific World Journal,* 2015 1-11.
Shattuck-Eidens, D. *et al.* (1997) 'BRCA1 sequence analysis in women at high risk for susceptibility mutations: risk factor analysis and implications for genetic testing'. *Jama,* 278 (15), pp. 1242-1250.
Shen, Y. *et al.* (2019) 'Comparison of Different Risk-Stratification Systems for the Diagnosis of Benign and Malignant Thyroid Nodules'. *Front Oncol,* 9 378.
Shephard, E. *et al.* (2013) 'Clinical features of kidney cancer in primary care: a case-control study using primary care records'. *British Journal of General Practice,* 63 (609), pp. e250-e255.
Shephard, E.A., Parkinson, M.A. and Hamilton, W.T. (2019) 'Recognising laryngeal cancer in primary care: a large case–control study using electronic records'. *British Journal of General Practice,* 69 (679), pp. e127-e133.

Shetty, A. *et al.* (2021) 'Clinical utility of skin cancer and melanoma risk scores for population screening: TRoPICS study'. *Journal of the European Academy of Dermatology and Venereology,* 35 (5), pp. 1094-1098.
Shin, A. *et al.* (2014) 'Risk prediction model for colorectal cancer: National Health Insurance Corporation study, Korea'. *PLoS One,* 9 (2), pp. e88079.
Shmueli, G. (2010) 'To explain or to predict?'. *Statistical science,* 25 (3), pp. 289-310.
Siegel, R.L., Miller, K.D. and Jemal, A. (2020) 'Cancer statistics, 2020'. *CA: A Cancer Journal for Clinicians,* 70 (1), pp. 7-30.
Simon, C. *et al.* (2020) 'Best Practice in Surgical Treatment of Malignant Head and Neck Tumors'. *Frontiers in Oncology,* 10 (140), pp.
Singh, P. and Warnakulasuriya, S. (2006) 'The two-week wait cancer initiative on oral cancer; the predictive value of urgent referrals to an oral medicine unit'. *British Dental Journal,* 201 717-720.
Sloan, F. and Gelband, H. (2007) *Cancer Control Opportunities in Low- and Middle-Income Countries.* Washington DC: National Academies Press.
Smith, C.D.L. *et al.* (2022) 'Risk prediction models for head and neck cancer: A rapid review'. *Laryngoscope Investigative Otolaryngology*.
Smith, T. *et al.* (2019) 'Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: a systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies'. *Gut,* 68 (4), pp. 672-683.
Smits, S. *et al.* (2018) 'Development of a Behavior Change Intervention to Encourage Timely Cancer Symptom Presentation Among People Living in Deprived Communities Using the Behavior Change Wheel'. *Ann Behav Med,* 52 (6), pp. 474-488.
Soardi, G.A. *et al.* (2017) 'Multicentre external validation of the BIMC model for solid solitary pulmonary nodule malignancy prediction'. *Eur Radiol,* 27 (5), pp. 1929-1933.
Soardi, G.A. *et al.* (2015) 'Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features'. *Eur Radiol,* 25 (1), pp. 155-162.
Sood, T. *et al.* (2021) 'One stop neck lump clinic: a boon for quick diagnosis and early management'. *European Archives of Oto-Rhino-Laryngology,* 278 (12), pp. 4951-4954.
Sorensen, J.R. *et al.* (2014) 'A "package solution" fast track program can reduce the diagnostic waiting time in head and neck cancer'. *Eur Arch Otorhinolaryngol,* 271 (5), pp. 1163-1170.
Spellman, J., Kanatas, A. and Ong, T.K. (2018) 'Early experience of a nurse-led clinic in a tertiary centre'. *Br J Oral Maxillofac Surg,* 56 (4), pp. 338-339.
Spellman, J. *et al.* (2020) 'Our experience of a nurse-led oral and maxillofacial surgical clinic in a tertiary centre: two years on'. *Br J Oral Maxillofac Surg,* 58 (1), pp. 99-101.
Stapleton, C.P. *et al.* (2019) 'Polygenic risk score as a determinant of risk of non-melanoma skin cancer in a European-descent renal transplant cohort'. *Am J Transplant,* 19 (3), pp. 801-810.
Steffen, A. *et al.* (2014) 'Development and validation of a risk score predicting risk of colorectal cancer'. *Cancer Epidemiol Biomarkers Prev,* 23 (11), pp. 2543-2552.
Stevens, M.N. *et al.* (2021) 'Impact of COVID-19 on presentation, staging, and treatment of head and neck mucosal squamous cell carcinoma'. *Am J Otolaryngol,* 43 (1), pp. 103263.

Steyerberg, E. *et al.* (2004) 'Validation and updating of predictive logistic regression models: A study on sample size and shrinkage'. *Statistics in Medicine,* 23 (16), pp. 2567-2586.
Steyerberg, E.W. (2019) *Clinical prediction models.* Springer.
Steyerberg, E.W. *et al.* (2013) 'Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research'. *PLOS Medicine,* 10 (2), pp. e1001381.
Steyerberg, E.W. *et al.* (2007) 'Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram'. *J Urol,* 177 (1), pp. 107-112; discussion 112.
Steyerberg, E.W. and Vergouwe, Y. (2014) 'Towards better clinical prediction models: seven steps for development and an ABCD for validation'. *Eur Heart J,* 35 (29), pp. 1925-1931.
Stjernswärd, J. (1985) 'Cancer control: Strategies and priorities. World Health Forum'. *World Health Forum,* (6), pp. 160–164.
Straif, K. *et al.* (2009) 'A review of human carcinogens--Part C: metals, arsenic, dusts, and fibres'. *Lancet Oncol,* 10 (5), pp. 453-454.
Strobl, C. *et al.* (2007) 'Bias in random forest variable importance measures: Illustrations, sources and a solution'. *BMC bioinformatics,* 8 (1), pp. 1-21.
Stukan, M., Badocha, M. and Ratajczak, K. (2019) 'Development and validation of a model that includes two ultrasound parameters and the plasma D-dimer level for predicting malignancy in adnexal masses: an observational study'. *BMC Cancer,* 19 (1), pp. 564.
Su Kim, D. *et al.* (2013) 'Composite three-marker assay for early detection of kidney cancer'. *Cancer Epidemiol Biomarkers Prev,* 22 (3), pp. 390-398.
Sullivan, C.O. *et al.* (2005) 'Case-mix and variation in specialist referrals in general practice'. *British Journal of General Practice,* 55 (516), pp. 529-533.
Sun, Z. *et al.* (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), pp. 2549-2557.
Sung, H. *et al.* (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 (3), pp. 209-249.
Swisher-McClure, S. *et al.* (2020) 'A Phase 2 Trial of Alternative Volumes of Oropharyngeal Irradiation for De-intensification (AVOID): Omission of the Resected Primary Tumor Bed After Transoral Robotic Surgery for Human Papilloma Virus-Related Squamous Cell Carcinoma of the Oropharynx'. *Int J Radiat Oncol Biol Phys,* 106 (4), pp. 725-732.
SWOP (2022) *The Prostate Cancer Risk Calculators.* Available at: https://www.prostatecancer-riskcalculator.com/seven-prostate-cancer-risk-calculators (Accessed: 11/05/20).
Taib, B.G. *et al.* (2018) 'Socioeconomic deprivation and the burden of head and neck cancer-Regional variations of incidence and mortality in Merseyside and Cheshire, North West, England'. *Clin Otolaryngol,* 43 (3), pp. 846-853.
Tajmehr, N. (2019) 'Appropriateness of referrals on the suspected cancer referral pathway - a secondary care perspective'. *British Dental Journal,* 226 (8), pp. 539-540.
Talwar, C. *et al.* (2020) 'Two-week rule: suspected head and neck cancer referrals from a general medical practice perspective'. *British Journal of Oral and Maxillofacial Surgery,* 58 (8), pp. 981-985.

Tammemägi, M.C. *et al.* (2013) 'Selection criteria for lung-cancer screening'. *N Engl J Med,* 368 (8), pp. 728-736.
Tan, H. *et al.* (2021) 'COVID-19 did not delay time from referral to definitive management for head and neck cancer patients in a regional Victorian centre'. *ANZ J Surg,* 91 (7-8), pp. 1364-1368.
Tan, W.S. *et al.* (2019) 'Development and validation of a haematuria cancer risk score to identify patients at risk of harbouring cancer'. *J Intern Med,* 285 (4), pp. 436-445.
Taylor, R. *et al.* (2020) 'The impact of coronavirus disease 2019 on head and neck cancer services: a UK tertiary centre study'. *J Laryngol Otol,* 134 (8), pp. 684-687.
Teppo, H. *et al.* (2003) 'Diagnostic delays in laryngeal carcinoma: Professional diagnostic delay is a strong independent predictor of survival'. *Head & Neck,* 25 (5), pp. 389-394.
Terry, M.B. *et al.* (2019) '10-year performance of four models of breast cancer risk: a validation study'. *Lancet Oncol,* 20 (4), pp. 504-517.
The National Collaborating Centre for Primary Care (2005) *Referral guidelines for suspected cancer, CG27.* Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20060715141954/http:/nice.org.uk/nicemedia/pdf/cg027niceguideline.pdf (Accessed: 05/05).
Thompson, I.M. *et al.* (2006) 'Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial'. *J Natl Cancer Inst,* 98 (8), pp. 529-534.
Thompson-Harvey, A. *et al.* (2020) 'Rising incidence of late-stage head and neck cancer in the United States'. *Cancer,* 126 (5), pp. 1090-1101.
Thomson, C.S. and Forman, D. (2009) 'Cancer survival in England and the influence of early diagnosis: what can we learn from recent EUROCARE results?'. *British Journal of Cancer,* 101 (S2), pp. S102-S109.
Thun, M. *et al.* (2017) *Cancer Epidemiology and Prevention.* 4th edn. Oxford: Oxford University Press.
Tice, J.A. *et al.* (2019) 'Validation of the breast cancer surveillance consortium model of breast cancer risk'. *Breast Cancer Res Treat,* 175 (2), pp. 519-523.
Tice, J.A. *et al.* (2015) 'Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer'. *J Clin Oncol,* 33 (28), pp. 3137-3143.
Tikka, T. *et al.* (2020) 'Head and neck cancer risk calculator (HaNC-RC)—V.2. Adjustments and addition of symptoms and social history factors'. *Clinical Otolaryngology,* 45 (3), pp. 380-388.
Tikka, T., Paleri, V. and MacKenzie, K. (2018) 'External validation of a cancer risk prediction model for suspected head and neck cancer referrals'. *Clin Otolaryngol,* 43 (2), pp. 714-717.
Tikka, T., Pracy, P. and Paleri, V. (2016) 'Refining the head and neck cancer referral guidelines: a two-centre analysis of 4715 referrals'. *Clinical Otolaryngology,* 41 (1), pp. 66-75.
Timmerman, D. *et al.* (2008) 'Simple ultrasound-based rules for the diagnosis of ovarian cancer'. *Ultrasound Obstet Gynecol,* 31 (6), pp. 681-690.
Timmerman, D. *et al.* (1999) 'Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses'. *Ultrasound in Obstetrics and Gynecology,* 13 (1), pp. 17-25.

Tingulstad, S. *et al.* (1996) 'Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses'. *Br J Obstet Gynaecol,* 103 (8), pp. 826-831.
Tokez, S. *et al.* (2020) 'Predicting keratinocyte carcinoma in patients with actinic keratosis: development and internal validation of a multivariable risk-prediction model'. *Br J Dermatol,* 183 (3), pp. 495-502.
Toronto Central Regional Cancer Program (2016) *Head & Neck Cancer.* Available at: https://www.trcp.ca/en/diagnosis-management/Pages/head-and-neck.aspx (Accessed: 03/11/2021).
Tota, J.E. *et al.* (2019) 'Development and validation of an individualized risk prediction model for oropharynx cancer in the US population'. *Cancer,* 125 (24), pp. 4407-4416.
Toustrup, K. *et al.* (2011) 'Reduction in waiting time for diagnosis and treatment of head and neck cancer – a fast track study'. *Acta Oncologica,* 50 (5), pp. 636-641.
Trottier, G. *et al.* (2011) 'Comparison of risk calculators from the Prostate Cancer Prevention Trial and the European Randomized Study of Screening for Prostate Cancer in a contemporary Canadian cohort'. *BJU Int,* 108 (8 Pt 2), pp. E237-244.
Usher-Smith, J. *et al.* (2015) 'Risk prediction tools for cancer in primary care'. *Br J Cancer,* 113 (12), pp. 1645-1650.
Usher-Smith, J.A. *et al.* (2014) 'Risk prediction models for melanoma: a systematic review'. *Cancer Epidemiol Biomarkers Prev,* 23 (8), pp. 1450-1463.
Usher-Smith, J.A. *et al.* (2017) 'Identifying people at higher risk of melanoma across the U.K.: a primary-care-based electronic survey'. *British Journal of Dermatology,* 176 (4), pp. 939-948.
Usher-Smith, J.A. *et al.* (2016) 'Risk Prediction Models for Colorectal Cancer: A Systematic Review'. *Cancer Prevention Research,* 9 (1), pp. 13-26.
Vachon, C.M. *et al.* (2015a) 'The contributions of breast density and common genetic variation to breast cancer risk'. *J Natl Cancer Inst,* 107 (5), pp.
Vachon, C.M. *et al.* (2015b) 'The contributions of breast density and common genetic variation to breast cancer risk'. *J Natl Cancer Inst,* 107 (5), pp. dju397.
van Hanegem, N. *et al.* (2012) 'Prediction models in women with postmenopausal bleeding: a systematic review'. *Womens Health (Lond),* 8 (3), pp. 251-262.
van Vugt, H.A. *et al.* (2011) 'Prediction of prostate cancer in unscreened men: external validation of a risk calculator'. *Eur J Cancer,* 47 (6), pp. 903-909.
Vedsted, P. and Olesen, F. (2011) 'Are the serious problems in cancer survival partly rooted in gatekeeper principles? An ecologic study'. *British Journal of General Practice,* 61 (589), pp. e508-e512.
Verheij, R.A. *et al.* (2018) 'Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse'. *Journal of Medical Internet Research,* 20 (5), pp. e185.
Verkouteren, J.A.C. *et al.* (2015) 'Predicting the Risk of a Second Basal Cell Carcinoma'. *J Invest Dermatol,* 135 (11), pp. 2649-2656.
Vernham, G.A. and Crowther, J.A. (1994) 'Head and neck carcinoma--stage at presentation'. *Clin Otolaryngol Allied Sci,* 19 (2), pp. 120-124.
Vilmun, B.M. *et al.* (2020) 'Impact of adding breast density to breast cancer risk models: A systematic review'. *Eur J Radiol,* 127 109019.
Vineis, P. and Wild, C.P. (2014) 'Global cancer patterns: causes and prevention'. *The Lancet,* 383 (9916), pp. 549-557.

Vos, T. *et al.* (2017) 'Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016'. *The Lancet,* 390 (10100), pp. 1211-1259.
Vuong, K. *et al.* (2014) 'Risk prediction models for incident primary cutaneous melanoma: a systematic review'. *JAMA Dermatol,* 150 (4), pp. 434-444.
Walker, J.G. *et al.* (2015) 'Cancer Risk Assessment Tools in Primary Care: A Systematic Review of Randomized Controlled Trials'. *The Annals of Family Medicine,* 13 (5), pp. 480-489.
Walter, F. *et al.* (2012) 'The Andersen Model of Total Patient Delay: A Systematic Review of Its Application in Cancer Diagnosis'. *Journal of Health Services Research & Policy,* 17 (2), pp. 110-118.
Walter, F.M. *et al.* (2013) 'Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: a diagnostic validation study'. *Br J Gen Pract,* 63 (610), pp. e345-353.
Wang, H.H. *et al.* (2019) 'Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer'. *JAMA Dermatol,* 155 (11), pp. 1277-1283.
Wang, W. *et al.* (2018) 'A Prediction Tool to Facilitate Risk-Stratified Screening for Squamous Cell Skin Cancer'. *J Invest Dermatol,* 138 (12), pp. 2589-2594.
Warner, E. *et al.* (2020) 'North East London coronavirus disease 2019 protocol for diagnostics in two-week wait head and neck cancer patients'. *The Journal of Laryngology & Otology,* 134 (8), pp. 680-683.
Watters, C. *et al.* (2021) 'Virtual voice clinics in the COVID-19 era: have they been helpful?'. *European Archives of Oto-Rhino-Laryngology,* 278 (10), pp. 4113-4118.
Weber, R.S. *et al.* (2003) 'Outcome of salvage total laryngectomy following organ preservation therapy: the Radiation Therapy Oncology Group trial 91-11'. *Arch Otolaryngol Head Neck Surg,* 129 (1), pp. 44-49.
Weller, D. *et al.* (2012) 'The Aarhus statement: improving design and reporting of studies on early cancer diagnosis'. *Br J Cancer,* 106 (7), pp. 1262-1267.
Wells, B.J. *et al.* (2014) 'Colorectal cancer predicted risk online (CRC-PRO) calculator using data from the multi-ethnic cohort study'. *J Am Board Fam Med,* 27 (1), pp. 42-55.
Whitaker, K.L. *et al.* (2015) 'Help seeking for cancer 'alarm' symptoms: a qualitative interview study of primary care patients in the UK'. *Br J Gen Pract,* 65 (631), pp. e96-e105.
White, C.S. *et al.* (2017) 'The Vancouver Lung Cancer Risk Prediction Model: Assessment by Using a Subset of the National Lung Screening Trial Cohort'. *Radiology,* 283 (1), pp. 264-272.
White, D.A. *et al.* (2004) 'Facilitators and barriers to improving the quality of referrals for potential oral cancer'. *Br Dent J,* 197 (9), pp. 537-540.
White, H.N. (2013) 'Learning Curve for Transoral Robotic Surgery<subtitle>A 4-Year Analysis</subtitle><alt-title>Learning Curve for Transoral Robotic Surgery</alt-title>'. *JAMA Otolaryngology–Head & Neck Surgery,* 1.
Whiteman, D.C. *et al.* (2016) 'A Model to Predict the Risk of Keratinocyte Carcinomas'. *Journal of Investigative Dermatology,* 136 (6), pp. 1247-1254.
WHO (2001) *Assessment of national capacity for noncommunicable disease prevention and control. The report of a global survey.* Geneva: WHO.

WHO (2002) *National Cancer Control Programmes. Policies and Managerial Guidelines.* Geneva: Switzerland: WHO.
WHO (2007) *Cancer control: early detection. WHO Guide for effective programmes.* Geneva.
WHO (2013) *Health literacy: the solid facts. .* Copenhagen: WHO Regional Office for Europe.
WHO (2016) *Atlas of african health statistics 2016: health situation analysis of the African Region.* World Health Organization.
WHO (2017) *Guide to cancer early detection.* Geneva, Switzerland: WHO Document production Services.
WHO (2020) *WHO report on cancer: setting priorities, investing wisely and providing care for all.* Geneva: WHO.
Wilkens, J. *et al.* (2016) 'The 2015 National Cancer Program in Sweden: Introducing standardized care pathways in a decentralized system'. *Health Policy,* 120 (12), pp. 1378-1382.
Wilkie, M.D. *et al.* (2021) 'Emergency presentations of head and neck cancer: Our experience in the wake of the COVID-19 pandemic'. *Clinical Otolaryngology,* 46 (6), pp. 1237-1241.
Williams, C. *et al.* (2014) 'Two-week referrals for suspected head and neck cancer: two cycles of audit, 10 years apart, in a district general hospital'. *J Laryngol Otol,* 128 (8), pp. 720-724.
Williams, L.H. *et al.* (2011) 'Identifying Persons at Highest Risk of Melanoma Using Self-Assessed Risk Factors'. *J Clin Exp Dermatol Res,* 2 (6), pp.
Williams, T.G. *et al.* (2016) 'Risk prediction models for colorectal cancer in people with symptoms: a systematic review'. *BMC Gastroenterol,* 16 (1), pp. 63.
Win, A.K. *et al.* (2012) 'Risk prediction models for colorectal cancer: a review'. *Cancer Epidemiol Biomarkers Prev,* 21 (3), pp. 398-410.
Woods, L.M. *et al.* (2011) 'Evidence against the proposition that "UK cancer survival statistics are misleading": simulation study with National Cancer Registry data'. *BMJ,* 342 (jun09 1), pp. d3399-d3399.
Wu, Y. *et al.* (2016a) 'A serum-circulating long noncoding RNA signature can discriminate between patients with clear cell renal cell carcinoma and healthy controls'. *Oncogenesis,* 5 (2), pp. e192.
Wu, Y. *et al.* (2016b) 'Genetic scores based on risk-associated single nucleotide polymorphisms (SNPs) can reveal inherited risk of renal cell carcinoma'. *Oncotarget,* 7 (14), pp. 18631-18637.
www.hra.nhs.uk (2021) *What approvals and decisions do I need?* Available at: https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/ (Accessed: 01/01/21).
www.ons.gov.uk (2016) *Cancer survival in England. One- and five-year net survival (%), for adults (aged 15 to 99 years) diagnosed between 2011 and 2015: England, by sex, Common cancers. .* Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinengland/adultstageatdiagnosisandchildhoodpatientsfollowedupto2016 (Accessed: 05/03/2019).

www.orlhealth.com (2019) *HaNC-RC v.2 (2019): Symptom Based Risk Calculator for Head And Neck Cancer Referrals v2.* Available at: http://orlhealth.com/risk-calculator-2.html (Accessed: May 2021).
www.surgonc.org (2021) *Society of Surgical Oncology.* Available at: https://www.surgonc.org/.
Xie, S.H. and Lagergren, J. (2016) 'A model for predicting individuals' absolute risk of esophageal adenocarcinoma: Moving toward tailored screening and prevention'. *Int J Cancer,* 138 (12), pp. 2813-2819.
Xie, S.H. *et al.* (2018) 'Assessing the feasibility of targeted screening for esophageal adenocarcinoma based on individual risk assessment in a population-based cohort study in Norway (The HUNT Study)'. *Am J Gastroenterol,* 113 (6), pp. 829-835.
Yala, A. *et al.* (2019) 'A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction'. *Radiology,* 292 (1), pp. 60-66.
Yang, L. *et al.* (2021) 'Reporting and Performance of Hepatocellular Carcinoma Risk Prediction Models: Based on TRIPOD Statement and Meta-Analysis'. *Can J Gastroenterol Hepatol,* 2021 9996358.
Yu, M.-Z. *et al.* (2022) 'Risk Prediction Models for Patients with Head and Neck Cancer among the Taiwanese Population'. *Cancers,* 14 (21), pp. 5338.
Zeitler, M. *et al.* (2018) 'Deprivation in relation to urgent suspicion of head and neck cancer referrals in Glasgow'. *Clinical Otolaryngology,* 43 (3), pp. 861-867.
Zhang, X. *et al.* (2017) 'Nomogram for risk prediction of malignant transformation in oral leukoplakia patients using combined biomarkers'. *Oral Oncol,* 72 132-139.
Zhou, Y. *et al.* (2018) 'Variation in 'fast-track' referrals for suspected cancer by patient characteristic and cancer diagnosis: evidence from 670 000 patients with cancers of 35 different sites'. *British Journal of Cancer,* 118 (1), pp. 24-31.
Zhu, W. (2016) 'p < 0.05, < 0.01, < 0.001, < 0.0001, < 0.00001, < 0.000001, or < 0.0000001 …'. *J Sport Health Sci,* 5 (1), pp. 77-79.
Zhu, Y. *et al.* (2021) 'Patient Satisfaction With the Head and Neck Cancer Telephone Triage Service During the COVID-19 Pandemic'. *Cureus*.
Oshyvalova, O., Ziukov, O.L. and Gurianov, V.G. (2019) 'Prognostic model of skin cancer risk assessment'. *Wiad Lek,* 72 (5 cz 1), pp. 817-822.

8 Appendix I

Table 8-1. List of participating hospitals in the validation phase of the HaNC-RC v.2

Centre	**Trust**
Aberdeen Royal Infirmary	NHS Grampian
Aintree University Hospital	Liverpool University Hospitals NHS Foundation Trust
Antrim Area Hospital	Northern Health and Social Care Trust
Birmingham City Hospital	Sandwell and West Birmingham Hospitals NHS Trust
Blackpool Victoria Hospital	Blackpool Teaching Hospitals NHS Foundation Trust
Broomfield Hospital, Chelmsford	Mid Essex Hospital Services NHS Trust
Charing Cross Hospital, London	Imperial College Healthcare NHS Trust
Chase Farm Hospital, London	Royal Free London NHS Foundation Trust
Countess of Chester Hospital	Countess of Chester NHS Foundation Trust
Cumberland Infirmary, Carlisle	North Cumbria University Hospitals NHS Trust
East Surrey Hospital, Redhill	Surrey and Sussex Healthcare NHS Trust
Glangwili General Hospital, Carmarthen	Hywel Dda University Health Board
Glasgow Royal Infirmary	NHS Greater Glasgow and Clyde
Guy's Hospital	Guy's and St Thomas' NHS Foundation Trust
Hinchingbrooke Hospital, Huntingdon	North West Anglia NHS Foundation Trust
Kent & Canterbury Hospital	East Kent Hospitals University NHS Foundation Trust
Manchester Royal Infirmary (MRI)	Manchester University NHS Foundation Trust
Milton Keynes University Hospital	Milton Keynes University Hospital NHS Foundation Trust
Ninewells Hospital, Dundee	NHS Tayside
Northampton General Hospital	Northampton General Hospital NHS Trust
Northwick Park Hospital, London	London North West University Healthcare NHS Trust
Pinderfields Hospital, Wakefield	The Mid Yorkshire Hospitals NHS Trust
Princess Alexandra Hospital, Harlow	Princess Alexandra Hospital NHS Trust
Queen Elizabeth Hospital Birmingham	University Hospitals Birmingham NHS Foundation Trust

Royal Albert Edward Infirmary, Wigan	Wrightington, Wigan and Leigh NHS Foundation Trust
Royal Blackburn Hospital	East Lancashire Hospitals NHS Trust
Royal Preston Hospital	Lancashire Teaching Hospitals NHS Foundation Trust
St John's Hospital, Livingston	NHS Lothian
Stepping Hill Hospital, Greater Manchester	Stockport NHS Foundation Trust
Sunderland Royal Hospital	South Tyneside and Sunderland Foundation NHS Trust
The Royal Liverpool University Hospital	Liverpool University Hospitals NHS Foundation Trust
The Royal Marsden Hospital	The Royal Marsden NHS Foundation Trust
University College London Hospital	University College London Hospitals NHS Foundation Trust
University Hospital Coventry and Warwickshire (UHCW)	University Hospitals Coventry and Warwickshire NHS Trust
University Hospital Crosshouse, Kilmarnock	NHS Ayrshire & Arran
University Hospital Monklands, Airdrie	NHS Lanarkshire
University Hospital of Wales (UHW), Cardiff	Cardiff & Vale University Health Board
Walsall Manor Hospital	Walsall Healthcare NHS Trust
Warrington Hospital	Warrington and Halton Teaching Hospitals NHS Foundation Trust
West Suffolk Hospital, Bury St Edmunds	West Suffolk NHS Foundation Trust
Wythenshawe Hospital, Greater Manchester	South Manchester NHS Foundation Trust

9 Appendix II - R codes

9.1 Libraries

```
library (Deducer)
library (pROC)
library (randomForest)
library (OptimalCutpoints)
library (dplyr)
library (ggplot2)
library(rsample)
library(caret)
library(cutpointr)
library(h2o)
h2o.init()
library(randomForest)
library(Epi)
library(lattice)
```

9.2 Logistic regression

```
phd_9_06$simd16_5<-factor(phd_9_06$simd16_5)
phd_9_06$gender<-factor(phd_9_06$gender)
phd_9_06$referral<-factor(phd_9_06$referral)
phd_9_06$u_weight_loss<-factor(phd_9_06$u_weight_loss)
phd_9_06$neck_lump<-factor(phd_9_06$neck_lump)
phd_9_06$smoking<-factor(phd_9_06$smoking)
phd_9_06$alcohol<-factor(phd_9_06$alcohol)
phd_9_06$hoarseness<-factor(phd_9_06$hoarseness)
phd_9_06$regurgitation<-factor(phd_9_06$regurgitation)
phd_9_06$cough<-factor(phd_9_06$cough)
phd_9_06$sore_throat<-factor(phd_9_06$sore_throat)
```

```
phd_9_06$neck_pain<-factor(phd_9_06$neck_pain)
phd_9_06$throat_discomfort_irritation<-
factor(phd_9_06$throat_discomfort_irritation)
phd_9_06$fosit<-factor(phd_9_06$fosit)
phd_9_06$throat_Clearing<-factor(phd_9_06$throat_Clearing)
phd_9_06$dysphagia<-factor(phd_9_06$dysphagia)
phd_9_06$odynophagia_3<-factor(phd_9_06$odynophagia_3)
phd_9_06$odynophagia<-factor(phd_9_06$odynophagia)
phd_9_06$choking<-factor(phd_9_06$choking)
phd_9_06$catarrh_mucus<-factor(phd_9_06$catarrh_mucus)
phd_9_06$blocked_nose<-factor(phd_9_06$blocked_nose)
phd_9_06$neck_lump<-factor(phd_9_06$neck_lump)
phd_9_06$oral_swelling<-factor(phd_9_06$oral_swelling)
phd_9_06$oral_ulcer<-factor(phd_9_06$oral_ulcer)
phd_9_06$heamoptysis<-factor(phd_9_06$heamoptysis)
phd_9_06$otalgia<-factor(phd_9_06$otalgia)
phd_9_06$face_pain_numbness<-factor(phd_9_06$face_pain_numbness)
phd_9_06$reflux<-factor(phd_9_06$reflux)
phd_9_06$throat_Clearing<-factor(phd_9_06$throat_Clearing)
phd_9_06$stridor<-factor(phd_9_06$stridor)
phd_9_06$sob<-factor(phd_9_06$sob)
phd_9_06$red_white_patch<-factor(phd_9_06$red_white_patch)
phd_9_06$head_neck_lesion<-factor(phd_9_06$head_neck_lesion)
phd_9_06$cancer<-factor(phd_9_06$cancer)
phd_9_06$simd16_20<-factor(phd_9_06$simd16_20)
phd_9_06$simd16_10<-factor(phd_9_06$simd16_10)
phd_9_06$extra_cases<-factor(phd_9_06$extra_cases)

summary (phd_9_06)
model<-glm(cancer~ age + gender + u_weight_loss + smoking + alcohol + hoarseness
+ sore_throat
        + fosit + dysphagia + odynophagia_3 + neck_lump + oral_swelling + oral_ulcer
```

```
        + otalgia + stridor + head_neck_lesion
        , family="binomial", data=phd_9_06)
summary(model)

ROC(form=cancer~ age + gender + u_weight_loss + smoking + alcohol + hoarseness
+ sore_throat
   + fosit + dysphagia + odynophagia_3 + neck_lump + oral_swelling + oral_ulcer
   + otalgia + stridor + head_neck_lesion
   , family="binomial", data=phd_9_06, na.action=na.omit)

prob = predict (model, type=c("response"))
phd_9_06$prob=prob
library(pROC)
g <- roc(cancer~prob, data=phd_9_06)
plot(g)

rocplot(model)
```

9.3 Logistic regression bootstrapping

```
library(Epi)
library(ROCR)

z.df<-phd_9_06

z.df<-subset(z.df, !is.na(z.df$cancer))

m <- 1000
auc <- rep(NA,m)
sp <- rep(NA,m)
se <- rep(NA,m)
for(j in 1:m){
```

```
 z.k <- 10
 z.sel <- rep(1:z.k,length=nrow(z.df))
 z.sel <- sample(z.sel,length(z.sel),replace=FALSE)
 head(z.sel,10)
 if (exists("z.out"))  rm(z.out)
 for (i in 1:z.k) {
  z.df.train <- subset(z.df,z.sel !=i)
  z.df.test <- subset(z.df,z.sel == i)


  z <- glm(cancer~ age + gender + u_weight_loss + smoking + alcohol + hoarseness
+ sore_throat
        + fosit + dysphagia + odynophagia_3 + neck_lump + oral_swelling +
oral_ulcer
        + otalgia + stridor + head_neck_lesion
        ,data=z.df.train, family="binomial")

  z.pred.test <- predict(z,newdata=z.df.test)
  z.res <- data.frame(LP=z.pred.test, cancer = z.df.test$cancer)
  if (exists("z.out")) z.out <- rbind(z.out,z.res) else z.out <- z.res
 }
 z.out$p <- exp(z.out$LP)/(1+exp(z.out$LP))
 pred <- prediction(z.out$LP, z.out$cancer)
 perf <- performance (pred,"auc")
 auc[j] <- round(perf@y.values[[1]],4)
 perf <- performance(pred, "sens","spec")
 sp[j]                                                                    <-
round(perf@x.values[[1]][which.max(perf@x.values[[1]]+perf@y.values[[1]])]*100,
2)
 se[j]                                                                    <-
round(perf@y.values[[1]][which.max(perf@x.values[[1]]+perf@y.values[[1]])]*100,
2)
```

```
}
zz <- cbind.data.frame(round(quantile(auc,0.5),4)

,paste("(",round(quantile(auc,0.025),4),",",round(quantile(auc,0.975),4),")",sep="")
                       ,round(quantile(se,0.5),2)

,paste("(",round(quantile(se,0.025),2),",",round(quantile(se,0.975),2),")",sep="")
                       ,round(quantile(sp,0.5),2)

,paste("(",round(quantile(sp,0.025),2),",",round(quantile(sp,0.975),2),")",sep=""))
names(zz) <- c("AUC", "AUC 95%CI", "SE", "SE 95%CI", "SP", "SP 95%CI")
if (exists("z.r")) z.r <- rbind(z.r,zz) else z.r <- zz
```

9.4 Logistic Regression Validation

```
library (magrittr)
library (tidyverse)
library (broom)

probabilities <- predict(model, type = "response")
predicted.classes <- ifelse(probabilities > 0.5, "pos", "neg")
head(predicted.classes)



# Select only numeric predictors
mydata <- phd_9_06 %>%
  dplyr::select_if(is.numeric)
predictors <- colnames(mydata)
# Bind the logit and tidying the data for plot
mydata <- mydata %>%
  mutate(logit = log(probabilities/(1-probabilities))) %>%
  gather(key = "predictors", value = "predictor.value", -logit)
```

```
ggplot(mydata, aes(logit, predictor.value))+
  geom_point(size = 0.5, alpha = 0.5) +
  geom_smooth(method = "loess") +
  theme_bw() +
  facet_wrap(~predictors, scales = "free_y")

plot(model, which = 4, id.n = 10)



# Extract model results
model.data <- augment(model) %>%
  mutate(index = 1:n())

model.data %>% top_n(3, .cooksd)

ggplot(model.data, aes(index, .std.resid)) +
  geom_point(aes(color = cancer), alpha = .5) +
  theme_bw()

model.data %>%
  filter(abs(.std.resid) > 3)

car::vif(model)

library(InformationValue)
optCutOff <- optimalCutoff(phd_9_06$cancer, probabilities)[1]
#=> 0.071

# Optimal cut-off Information Value
misClassError(phd_9_06$cancer, probabilities, threshold = 0.47976)
```

```
# Optimal cut-off Youden's Index
misClassError(phd_9_06$cancer, probabilities, threshold = 0.071)

Concordance(phd_9_06$cancer, probabilities)


confusionMatrix(phd_9_06$cancer, probabilities, threshold = 0.47976)

library(performance)

performance_hosmer(model)
```

9.5 Random forest

```
z.df<-data.frame(phd_9_06)

set.seed(100123)
z.rows.sel <- rbinom(nrow(z.df),1,prob = 0.33)
table(z.rows.sel)

z.df.train <- subset(z.df,z.rows.sel ==0)
z.df.test <- subset(z.df,z.rows.sel ==1)

z <- randomForest(formula = factor(cancer)~ age + gender + u_weight_loss + smoking
+ alcohol +hoarseness + regurgitation + cough + sore_throat + neck_pain +
throat_discomfort_irritation + fosit + dysphagia + odynophagia_3 + choking +
catarrh_mucus + blocked_nose + oral_swelling + oral_ulcer + heamoptysis + otalgia
+face_pain_numbness + reflux + throat_Clearing + stridor + sob + red_white_patch +
head_neck_lesion + neck_lump,                    data=z.df.train, na.action = na.omit)
z
importance(z)
varImpPlot(z,main="")
```

```
z.pre <- predict(z,newdata=z.df.test,type="prob")
head(z.pre)
dim(z.pre)

ROC(test=z.pre[,2], stat=z.df.test$cancer, plot="ROC")

cutoff <- log(0.09)

z.df.test$cancer.character <- ifelse(z.df.test$cancer==1,"Cases", "Controls")
histogram(~log(z.pre[,2])|z.df.test$cancer.character,layout=c(1,2),xlab="Log
Prediction Probability",
        panel=function(x,...){
          panel.histogram(x,...)
          panel.abline(v=cutoff,lty=2,col="dark blue",lwd=2)})
```