Accuracy of Imaging Modalities for Detecting Extracapsular Spread of Cervical Lymph Node Metastases in HPV-Associated Oropharyngeal Cancer: A Systematic Review

A thesis submitted by

Tristan John Morey

JBI

Faculty of Health and Medical Sciences

University of Adelaide

Australia

Submitted to the Adelaide Graduate Centre, University of Adelaide in partial fulfillment of the requirements for the degree of Master of Clinical Science on 21 March 2022.

CONTENTS

ABBREVIATIONS

AJCC	American Joint Committee on Cancer
AUC	Area under the receiver operating characteristic (ROC) curve
CASP	Critical appraisal skills programme
CCND1	Cyclin D1 gene
CECT	Contrast-enhanced computed tomography
CIs	Confidence intervals
CT	Computed tomography
DTA	Diagnostic test accuracy
EBHC	Evidence-based healthcare
EBM	Evidence-based medicine
ECS	Extracapsular spread
EGFR	Epidermal growth factor gene
EHSN	European Head and Neck Society
ENE	Extra-nodal extension
ESMO	European Society for Medical Oncology
ESTRO	European Society for Therapeutic Radiology
FAME	Feasibility, appropriateness, meaningfulness and effectiveness
FIG	Figure
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
H&NC	Head & Neck Cancer
HPV	Human papillomavirus
INOS	Inducible nitric oxide synthase
JBI	Joanna Briggs Institute
K	Cohen's kappa coefficient
MRI	Magnetic resonance imaging
N	Number
NCNN	National Comprehensive Cancer Network
NPV	Negative predictive value
OPSCC	Oropharyngeal squamous cell carcinoma
PENE	Pathological extra-nodal extension
PET	Positron emission tomography

PIRD	Population, index test, reference standard, and diagnosis of interest
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
RCT	Randomised controlled trials
ROC	Receiver operating characteristic
RSN	Radiographically suspicious nodes
SCC	Squamous cell carcinoma
SIGN	Scottish intercollegiate guidelines network
SN	Sensitivity
SoF	Summary of Findings
SP	Specificity
SROC	Summary receiver operating characteristic curve
STARD	Standards for the reporting of diagnostic accuracy studies
SUMARI	JBI System for the Unified Management, Assessment and Review of Information
TLM	Transoral laser microsurgery
TN	True negative
TORS	Transoral robotic surgery
TP	True positive
US	Ultrasound
+	Positive
-	Negative

ABSTRACT

Background

Extracapsular spread (ECS) of lymph node metastases is associated with poor prognosis and its detection in head and neck cancer (H&NC) is crucial for treatment planning. Commonly used imaging modalities to detect ECS in H&NC include computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and ultrasonography (US). Currently there is no gold standard imaging modality to detect ECS in H&NC, leaving clinicians to rely heavily on clinical examination and surgical histopathology for ECS based treatment decisions.

The purpose of this study was to undertake a systematic review using Joanna Briggs Institute (JBI) methodology that aimed at identifying and synthesising the best available evidence regarding the accuracy of conventional imaging modalities and their abilities to detect ECS in the specific population group of patients with human papillomavirus positive (HPV+) oropharyngeal squamous cell carcinoma (OPSCC).

Inclusion criteria

Type of participants

Participants were those with a confirmed diagnosis of HPV+ OPSCC and suspected diagnosis of cervical lymph node metastases and ECS. Participants were not excluded due to age, sex, race or education status.

Type of index tests

This review examined studies that utilised a conventional imaging modality to detect radiologic ECS in HPV+ OPSCC.

Type of reference test

This review examined studies that utilised surgical histopathology as the reference standard for the diagnosis of ECS (gold standard for ECS detection).

Type of outcomes

This review examined two primary and four secondary outcomes of interest. The primary outcomes of interest included: sensitivity and specificity measures with 95% confidence intervals for each imaging modality used to detect ECS in HPV+ OPSCC. The secondary

outcomes included: positive predictive value (PPV), negative predictive value (NPV), area under the ROC curve (AUC) and interobserver agreements (K) (where applicable) for the different imaging modalities.

Diagnosis of interest

The phenomena of interest in this review was ECS of cervical lymph node metastases (also known as extra-nodal extension (ENE)).

Type of studies

This review examined published studies that examined the diagnostic accuracy (including sensitivity and specificity) of an imaging modality used to detect ECS in HPV+ OPSCC. Diagnostic cohort studies were the preferred study design for inclusion. All six of the included studies were retrospective cohort studies.

Methods

Methodological approach

The methodological approach to the review was based on JBI guidance for systematic reviews involving diagnostic test accuracy (DTA) studies.

Search strategy

A comprehensive search using a three-phased approach was conducted across four databases, one clinical trials register, as well as a manual search for primary studies (published) in the reference lists of all included studies. There was no restriction on publication date, however. only studies in English were included in the review.

Methodological quality

Two reviewers assessed the methodological quality of the included studies using the QUADAS-2 tool. The QUADAS-2 tool is structured around assessing for risk of bias in four domains; Patient Selection, Index Test, Reference Standard, and Flow and Timing.

Data extraction

Quantitative data was extracted using the JBI data extraction tool for DTA studies.

Data analysis

Meta-analysis and assessment of heterogeneity was conducted on four CT studies using a random-effects model. The remaining two studies underwent a narrative synthesis. The GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach was used to assess the certainty in the evidence.

Results

Out of 1772 hits, six retrospective cohort studies were included in the review and four underwent meta-analysis. Four investigated the diagnostic ability of CT, one investigated PET/CT, and one investigated 'CT and MRI' (with no separation of index test to outcomes). Meta-analysis of the four CT studies showed CT had an overall sensitivity of 77% (60-94%) and specificity of 60% (47-73%). PET/CT had a sensitivity of 86% (73-94%) and specificity of 76% (61-87%). 'CT and MRI' had a sensitivity of 62% (53-70%) and specificity of 78% (70-84%). No meta-analysis or comparison meta-regression could be performed on the PET/CT or 'CT and MRI' studies.

Conclusions

The findings of this review imply pooled CT specificity values (60%) are too low to suggest clinical value for CT as a diagnostic tool to detect ECS in HPV+ OPSCC. Pending further research, the use of CT and PET/CT however might have clinically acceptable sensitivity and negative predictive values to help confirm the absence of radiologic ECS in HPV+ OPSCC. No studies on MRI or US were identified for assessment of ECS in HPV+ OPSCC.

Implications for practice

There is insufficient evidence to suggest CT or PET/CT are reliable diagnostic tools for radiological detection of ECS in HPV+ OPSCC. Further studies of high-quality involving CT, PET/CT, MRI and US are required to help establish clinical guidance and a gold standard for radiologic ECS detection in HPV+ OPSCC.

Keywords

Extra-capsular spread; human papillomavirus; imaging modalities; oropharynx; squamous cell carcinoma.

DECLARATION OF ORIGINALITY OF WORK

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

_____

Tristan John Morey

ACKNOWLEDGEMENT

I would like to take this opportunity to thank my supervisors Dr. Cindy Stern, Dr. Andrew Foreman and Dr. John-Charles Hodge for their guidance and support throughout this project. Additionally, I would like to thank Vikki Langton (Scientific Librarian at The University of Adelaide) for helping with the development of the search strategy, and Suzanne Edwards (Statistician at The University of Adelaide) for helping with the data analysis.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The focus of this thesis is the presentation of a systematic review following the Joanna Briggs Institute (JBI) methodology for reviewing evidence on the diagnostic accuracy of commonly used radiological tests that are used to detect a phenomena of interest in a particular patient demographic. The review sought to identify and synthesise the best available evidence on the diagnostic accuracy of conventional imaging modalities and their abilities to detect extracapsular spread (ECS) in patients with human papillomavirus-positive oropharyngeal squamous cell carcinoma (HPV+ OPSCC) and suspected localised (cervical) lymph node metastases. The aim was to explore the accuracy of each available imaging modality using sensitivity, specificity, positive predictive, negative predictive, area under the receiver operating characteristic curves (AUC), and interobserver agreements (K)(where applicable) values.

1.1 Thesis structure

This thesis is organised into the following five chapters:

Chapter 1: Introduction:
In chapter one, an introduction to the health problem (HPV+ OPSCC) and topic of interest (accuracy of conventional imaging modalities at detecting ECS in HPV+ OPSCC) is provided. This chapter will include an overview of the current research, along with a rationale for undertaking a systematic review on the topic.

Chapter 2: Methodology:
In chapter two, the methodological principles upon which the systematic review of international literature is based are addressed. This includes a description of the development and origins of evidence-based healthcare (EBHC), evidence synthesis and the systematic review.

Chapter 3: Systematic review methods:
In chapter three, the methodological processes undertaken in the systematic review are outlined including the review question, inclusion criteria, search and selection process, appraisal process for methodological quality, process for data extraction, method of data synthesis and assessment of the certainty in the findings.

Chapter 4: Results:
In chapter four, the search results and the methodological quality and characteristics of the included studies are described. The findings of the review are also presented in this chapter.

Chapter 5: Discussion, conclusions and implications for practice and research:
In the final chapter the main findings generated from the systematic review, the limitations of the review and the implications for practice and research are discussed.

1.2 Overview of chapter 1

The remaining part of chapter one of this thesis is broken down into the following sections: a contextual overview of extracapsular spread (ECS) (section 1.3), contextual overview of human papillomavirus (HPV) and ECS in H&NC (section 1.4), recognition and screening of ECS in H&NC (section 1.5), management and treatment of ECS in HPV+ OPSCC (section 1.6), clinical implications for the diagnosis of ECS in H&NC (section 1.7), statement of the review question (section 1.8), scope and state of current literature on the topic (section 1.9) and relationship between the existing literature and the proposed systematic review (section 1.10).

1.3 Contextual overview of extracapsular spread (ECS)

Although currently there is an ongoing debate regarding a standardised definition of extracapsular spread (ECS), ECS is most commonly defined as the spread of cancer cells beyond a lymph node capsule in lymph node metastasis.[1] Extracapsular spread of lymph node metastasis is a well-known phenomenon in cancers such as head and neck, breast, and bladder cancer indicating the progression of disease. Although more widely established in breast and bladder cancer, the prognostic implications of ECS in head and neck cancer (H&NC) remains an ongoing area for research.[2]

The first description of ECS was made by Willis in 1930 from autopsies of patients with advanced stage H&NC.[3] Then subsequently in 1971, Bennett and his colleagues were the first to discover a poor prognostic link between ECS and subtypes of H&NC including cancers of the larynx and hypopharynx.[4] Clinically to date, the detection of ECS remains one of the most important adverse prognostic factors for local recurrence, distant metastasis and survival particularly in H&NC.[5,6] As such, to improve tumour classification and prognosis, the ECS

characteristic was incorporated into the 2017 H&NC Staging Manual in the American Joint Committee on Cancer (AJCC) 8th Edition.[7] The addition of ECS to the H&NC staging manual has allowed clinicians to more accurately discuss prognosis and treatment goals in the various subtypes of H&NC. To many clinicians, the recognition of ECS has been considered a gamechanger to the management of patients with H&NC.

1.4 Contextual overview of human papillomavirus (HPV) and ECS in H&NC

Worldwide over 500,000 new cases of H&NC are reported annually with around 300,000 deaths occurring each year.[8] In Australia, the most recent five-year prevalence statistics of H&NC (including lip) was 17,220 for both males and females from 2012 to 2016,[9,10] while the annual incidence and mortality for males and females was 5,104 and 1,201, respectively, in 2021.[11,12] Of the total H&NC's about 90% are a type of cancer known as squamous cell carcinoma (SCC).[13] Oropharyngeal squamous cell carcinoma (OPSCC) is the most common type of H&NC in the western world.[14] Previously it has been well reported that sustained exposure to tobacco, tobacco related products and alcohol increases the risk of developing OPSCC.[15] Interestingly however in more recent times, there has been a major shift in the type of SCC and risk factors responsible for the vast majority of current OPSCC's.[16]

P16-positive (P16+) OPSCC has now emerged as the new challenging OPSCC for clinicians since tobacco-related OPSCC has been on a decline. The primary risk factor for developing P16+ OPSCC is oral human papillomavirus (HPV) infection and the majority of oral HPV-infections are acquired by oral sex with infected oral or anogenital sites.[16] P16 is a surrogate marker for HPV infection and therefore P16+ and human papillomavirus positive (HPV+) OPSCC are often used interchangeably in clinical practice. The steep rise in the incidence of HPV+ OPSCC has occurred on a global scale. From 1984-2004 there was a 225% increase in the prevalence of HPV+ OPSCC reported in the United States.[17] Similar trends have been reported in Australia and New Zealand. In Australia, Hong et al. reported more than a three-fold increase in the percentage of HPV+ OPSSC in the last two decades from 20% to 63%.[18]

Currently, HPV+ OPSCC has a male predominance with men suffering a three to five times higher incidence than women worldwide.[19] The higher rates of HPV+ OPSCC in men has been found to be associated with men performing oral sex on women with the female genitalia carrying a higher HPV burden than male genitalia. In addition, women have been found to have a higher seroconversion rate after genital HPV exposure, leading to some

relative protection against oral HPV infection. Whilst the introduction of school-based HPV vaccine programmes initially targeted at young women, has provided women with more protection against HPV disease than men.[19]

Interestingly HPV+ OPSCC is thought to be a distinct clinical and molecular entity with unique histopathological features as compared to HPV negative (HPV-) OPSCC.[20] In general, HPV+ OPSCC has been found to have a better prognosis with improved survival and enhanced response to treatment compared to OPSCC due to other risk factors such as tobacco.[21,22] Nodal characteristics such as the presence of ECS however still reflect an aggressive type of cancer requiring multimodal treatment regardless of cancer type.[21-23] Radiographically, HPV+ lymph nodes tend to present with large cystic masses, with a clinically and radiographically smaller or even occult tumour compared to HPV- disease.[24] Also unique to the HPV+ related cancers is the propensity for peritumoral desmoplasia (fibrosis) which can mimic ECS making the radiological diagnosis of ECS and subsequent treatment planning in HPV+ OPSCC more problematic compared to other types of H&NC.[25]

1.5 Recognition and screening of ECS in H&NC

The major symptoms of oropharyngeal cancer include a sore throat, odynophagia (painful swallowing), dysphagia (difficulty swallowing), a persistent white or red patch in the oral cavity, haemoptysis (coughing up blood), lymphadenopathy (palpable lump) and unexpected weight loss.[26]

Routinely in current practice, patients with a suspected diagnosis of H&NC undergo a thorough assessment which may feature a combination or all of the following: history taking concerning a patients symptoms and medical history, clinical exam including visual inspection for signs of disease and palpation of lymph nodes, radiological tests sometimes referred to as imaging or 'staging scans' to assess tumour characteristics and evidence of metastases, and histopathology of any suspicious lesions through core or excisional biopsy. The presence of ECS in H&NC can either be diagnosed through radiological findings (radiologic ECS), or surgical histopathology (pathologic ECS, the gold standard). Frequently the findings of the assessment (including ECS status) will then be presented and discussed at a Multidisciplinary Team meeting, which involves a team consisting (but not limited to) of surgeons, medical oncologists, radiation oncologists, radiologists, pathologists, palliative care nurses and speech therapists to help tailor an individual's treatment plan.[27]

The increased global availability of radiological tests has allowed for a more accurate assessment of patients with various types of cancer, where in the past suspicious lesions may have been missed clinically or resulted in the patient needing surgery for histopathology. The use of radiological tests allows clinicians to more accurately visualise the size of a lesion, assess lesion characteristics (i.e. whether benign or malignant), identify the precise location of a lesion, and assess the spread of the disease (i.e. local, regional or distant metastasis) which will help inform tumour staging and treatment planning.

Most frequently the imaging modality of choice for staging in H&NC is contrast-enhanced computed tomography (ceCT).[28] The increasing reliability on different imaging modalities to detect ECS however is an ongoing disputed topic. Imaging techniques including computed tomography (CT) (plain or contrast-enhanced), magnetic resonance imaging (MRI), positron emission tomography (PET), combination PET/CT and ultrasound (US) can all be utilised to detect and diagnose ECS from a radiological perspective.[29] The choice of imaging technique for assessment of radiologic ECS largely comes down to clinician preference as currently there is no gold-standard imaging technique for detecting ECS in H&NC. An overview of the various imaging techniques and their use in H&NC is presented in Table 1.

Imaging	**Principle**	**Advantages**	**Disadvantages**
ceCT	Uses ionizing radiation to create cross-sectional (slices) pictures of the body from different angles.	Considered the gold standard for routine tumour-node-metastasis staging, as well as assessing bone detail and involvement. Has a rapid acquisition time.	Uses radiation, risk of contrast-induced nephropathy.
MRI	Uses powerful magnets and radio waves to create cross-sectional (slices) pictures of the body from different angles.	Superior ability to characterise localised disease extent and soft tissue involvement for most primary H&NC subtypes. Does not use radiation.	Longer acquisition time than ceCT (difficult for patients with claustrophobia). Multiple contraindications: cardiac implantable electronic devices (i.e. pacemakers, defibrillators), metallic foreign bodies, cochlear implants.
PET/CT	Uses radiolabelled 18-fluorodeoxyglucose tracer to	Useful for identifying primary or metastatic disease, differentiating cancerous tissue from	High cost and time consuming, interpretation can be non-specific (malignancy and

	demonstrate abnormal increased metabolic activity, in combination with whole body CT.	post-treatment oedema/scarring/benign lesions.	inflammatory processes demonstrate increased metabolic activity). Contraindications: pregnancy, uncontrolled diabetes.
US	Uses localised high frequency sound waves and their echoes to create pictures of an area of interest.	Provides excellent non-invasive soft tissue characterisation of superficial primary sites, lymph nodes, thyroid nodules and salivary gland lesions.	Technique is operator dependent, inability to assess deeper structures due to limited soft tissue penetration.

Table 1: Overview of the various imaging techniques and their use in H&NC.[30]

1.6 Management and treatment of ECS in HPV+ OPSCC

In terms of treatment protocols for patients with HPV+ OPSCC, a variety of techniques such as surgery, radiotherapy and chemotherapy are often utilised.

In more recent times the recognition of HPV+ OPSCC as a biologically distinct disease compared to other subtypes of H&NC resulted in a major shift of treatment from radiotherapy (with or without chemotherapy) to newer techniques including transoral laser microsurgery (TLM) and transoral robotic surgery (TORS) coupled with appropriate adjuvant therapy (radiotherapy or chemoradiotherapy) depending on factors such as age and surgical findings. Studies in the field of H&NC have reported that the addition of these minimally invasive surgical techniques (TLM and TORS) has allowed patients to benefit from superior functional outcomes compared to traditional open surgery, and reduced toxicity compared to chemoradiotherapy through treatment de-intensification despite comparable oncological outcomes.[31,32]

Based on the 2021 USA National Comprehensive Cancer Network (NCNN)[33] and the 2020 European Head and Neck Society (EHNS)-European Society for Medical Oncology (ESMO)-European Society for Therapeutic Radiology (ESTRO)[34] treatment guidelines, the current recommendations for primary and adjuvant treatment in HPV+ OPSCC utilising radiology for guidance are as follows:

- Patients with HPV+ OPSCC without signs of nodal disease are recommended single modality therapy i.e. surgery (TLM/TORS/open surgery +/- lymph node dissection) or definitive radiotherapy.

If ECS is detected in surgical histopathology, patients are then recommended adjuvant chemoradiotherapy.

- Patients with presumed nodal metastasis and ECS negative status are recommended either surgery (TLM/TORS/open surgery +/- lymph node dissection) and adjuvant radiotherapy, or primary non-surgical treatment with chemoradiotherapy (both treatment options known as 'bimodal therapy').
  If ECS is detected in surgical histopathology, patients are then recommended adjuvant chemoradiotherapy (collectively known as 'trimodal therapy').
- Patients with presumed nodal metastasis and radiologic ECS are recommended up front (primary) chemoradiotherapy.
- Patients who have been diagnosed with advanced stage cancer that is not considered curable (especially in the case of distant metastatic disease at diagnosis), are considered fully disabled, or score poorly on a calculated risk assessment are recommended supportive (palliative) care as management for their disease which may include palliative chemotherapy/radiotherapy/immunotherapy.

For each stage of disease, in addition to surgery, radiotherapy or chemotherapy, a 'clinical trial' is 'strongly supported' as an option (if available and the patient deemed suitable) in the NCNN treatment guidelines,[33] reflecting an ongoing evolving approach to the management of patients with HPV+ OPSCC. Whilst current research and trials are heavily focused on treatment de-intensification protocols to ensure oncological outcomes continue to balance the side-effects associated with surgery and chemoradiotherapy.[31,32]

Overall the findings from the 2021 NCNN and 2020 EHNS-ESMO-ESTRO treatment guidelines continue to show ECS along with positive surgical margins remain an indication for adjuvant chemoradiotherapy, emphasising the concerning presence of ECS.[33,34]

Of concern in HPV+ OPSCC treatment, is the fact many patients with radiologic ECS will undergo treatment (chemoradiotherapy) based on imaging without histopathologic confirmation (the gold standard).[33,34] The implications of this including clinical consequences and patient mismanagement will be discussed in section 1.7 and chapter five of the review.

1.7 Clinical implications for the diagnosis of ECS in H&NC

Of potential clinical concern is the value placed on radiological techniques to detect ECS in H&NC with no consistently reliable imaging modality reported in the literature. Whilst as HPV+ SCC is a complex disease with a propensity to develop peritumoral desmoplasia

(fibrosis mimicking radiologic ECS), some patients with HPV+ OPSCC are recommended primary chemoradiotherapy, when the radiological diagnosis of ECS may be a false positive and a treatment path consisting of surgery and adjuvant radiotherapy may be more appropriate.[25,31,32]

Consequences of a pathway involving chemoradiotherapy include a significantly higher incidence and severity of acute treatment related toxicities (such as nausea, vomiting, pain, mucositis, dysphagia and odynophagia) compared to those undergoing surgery and adjuvant radiotherapy.[35] Studies have found that up to 77% of patients that undergo adjuvant chemoradiotherapy for pathologic ECS experience severe toxicities compared to a pathway of surgery and adjuvant radiotherapy.[36] Whilst studies have also found that up to 25% of patients who undergo concurrent chemoradiotherapy for H&NC will require unplanned admissions to hospital to manage treatment related toxicities in comparison to those undergoing surgery and radiotherapy.[37] The use of an appropriate imaging modality to establish a pre-treatment diagnosis of ECS is therefore crucial to treatment planning in H&NC and a patient's quality of life. Particularly whether or not a patient will be recommended surgery (and adjuvant therapy) or primary chemoradiotherapy.[33,34]

Unfortunately given that current imaging modalities have limitations in their abilities to detect microscopic ECS, there is an ongoing discord between radiologic and pathologic detected ECS. Consequentially there is no gold standard imaging modality for detecting ECS in H&NC and ECS remains a surgical histopathologic diagnosis after lymph node dissection. If a consistently accurate and reliable imaging modality could be utilised to diagnose ECS without the need for surgery and histopathologic confirmation, this would be invaluable to patients and clinicians in the H&NC field. Whilst psychologically patients could be counselled more appropriately at time of their diagnosis.

1.8 Statement of the review question

To synthesize the best available evidence related to the diagnostic test accuracy of conventional imaging modalities used to detect ECS of cervical lymph node metastases in HPV+ OPSCC.

The primary outcomes included sensitivity and specificity measures (with 95% confidence intervals) for each imaging modality used to detect ECS in HPV+ OPSCC.

Secondary outcomes included PPV, NPV, area under the receiver operating characteristic (ROC) curves (AUC), and interobserver agreements (K)(where applicable) for the different imaging modalities.

1.9 Scope and state of current literature on the topic

A preliminary search of PROSPERO, MEDLINE, the Cochrane Database of Systematic Reviews and JBI Evidence Synthesis was conducted and no other systematic reviews on the topic were identified as being underway.

Ultimately there are very few studies addressing ECS in the HPV+ OPSCC population, and the following research is the first systematic review reporting on the accuracy of imaging modalities at detecting ECS exclusively in HPV+ OPSCC. The lack of studies is possibly due to the unique clinical and radiologic features of HPV+ ECS, and the result of numerous studies to date not identifying or disclosing participants P16 status in their data. Unfortunately the lack of studies makes it extremely difficult to establish any clinical guidance for optimal imaging regarding ECS detection specifically in the unique HPV+ OPSCC population. Given that the management of HPV+ OPSCCC differs from HPV- OPSCC and other forms of H&NC,[33,34] studies reporting on ECS detection in the HPV+ OPSCC population are crucial.

1.10 Relationship between the existing literature and the proposed systematic review

Several reviews have been published on the accuracy of imaging modalities to detect ECS however in the broader population of H&NC.[29,38,39] The population of H&NC includes various different cancers such as OPSCC, nasopharyngeal, laryngeal, thyroid, skin and salivary gland malignancies that behave completely differently to HPV+ OPSCC. Reviews in the more broad H&NC population should therefore be treated cautiously and separately to HPV+ OPSCC studies.

However for an understanding of the current literature on the topic, the reviews in H&NC are the closest research to the following systematic review. Amongst these reviews, it was found that either combination PET/CT or MRI has the greatest potential for clinical use in detecting ECS in H&NC. Albeit large variabilities in predictive values depending on the study, imaging criteria and radiologists involved.[29,38,39]

In 2015, a systematic review by Su et al.[29] involving 15 studies (n=1155 participants) investigated various imaging techniques used to detect ECS in all H&NC types. The findings revealed CT to be appropriately specific (85%) however poorly sensitive (77%). MRI had potential for superior sensitivity (85%) whilst similar specificity (85%) to CT. US and combination PET/CT showed no evidence for their use in detecting radiologic ECS. Su et al.[29] concluded MRI had potential to be the imaging modality of choice for detecting ECS in clinical practice however the review had several limitations such as selection bias which impacted on their conclusions.

In 2020, a systematic review was conducted by Park et al.[38] involving 22 studies (n=2478 participants) on the imaging techniques of CT and MRI and their abilities to detect ECS in SCC-related H&NC. Contrary to the Su et al.[29] review findings, CT was found to have a higher sensitivity of 73% compared to 60% for MRI. Whilst MRI was found to have a higher specificity of 96% compared to 83% for CT. The review also included a subgroup analysis involving five studies on CT accuracy in the HPV+ OPSCC population which showed an overall poorer performance compared to the SCC-related H&NC analysis (sensitivity 65%, specificity 74% P = 0.01). The subgroup analysis by Park et al.[38] is the only analysis to assess radiological accuracy and ECS detection specifically in the HPV+ OPSCC population.

Most recently in 2021, Abdel-Halim et al.[39] conducted a systematic review involving 25 studies (n=1995 participants) on ECS detection in patients with any form of SCC-related H&NC. Contrary to the Su et al.[29] review findings, Abdel-Halim et al.[39] reported that PET/CT had the highest diagnostic values with sensitivity and specificity values of 80% and 83% respectively, compared with CT (76%, 77%) and MRI (72%, 78%) for detecting ECS. Despite these findings, Abdel-Halim et al.[39] concluded that PET/CT, CT and MRI showed no significant differences in their diagnostic ability at detecting ECS however PET/CT had a higher sensitivity.

The findings from these previous reviews in H&NC should not be generalized to HPV+ OPSCC, however we do acknowledge the research has helped form some guidance for the following review including certain aspects of the methodology. Therefore, due to the need for updated research, and a review focusing on the distinct and rising HPV+ OPSCC, the following review was undertaken.

In this chapter, an introduction to the health problem (HPV+ OPSCC), the topic of interest (accuracy of conventional imaging modalities at detecting ECS in HPV+ OPSCC), an overview of the current research, along with a rationale for undertaking a systematic review on the topic is provided. In the next chapter, the development and origins of EBHC, the process of evidence synthesis and systematic review methodology are introduced.

CHAPTER 2: METHODOLOGY

Overview of chapter 2

In chapter two, an overview of evidence-based healthcare (EBHC) and the Joanna Briggs Institute (JBI), including its model of EBHC is provided. Following this, there is a discussion on evidence synthesis and systematic review methodology, Levels of Evidence and the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, and the methodological basis of the chosen approach to synthesis.

Evidence-based healthcare

Evidence-based healthcare (EBHC)

Although no sole individual has been identified as responsible for the movement of evidence-based healthcare (EBHC), evidence-based practice has been reported from as early as the mid-18th century. Florence Nightingale is one such example, who through her time as a nurse in military hospitals during the Crimean War (1853-1856), identified extremely unsanitary conditions and the potential to improve patient outcomes through changes in hand hygiene. Supporting her findings with evaluations of changes to hygiene practices and patient outcomes across the neighbouring hospitals.[40] Although the term EBHC was not known at the time, the origins of EBHC seem to date back to the mid-18th and 19th century, although much of the literature will trace the inception of EBHC to the 1970s through prominent work by researchers such as Professor Archie Cochrane.[40]

In 1972, Professor Archie Cochrane a medical doctor and researcher from the United Kingdom has been credited for being a pioneer in highlighting a prevalent issue amongst health practitioners at the time. Professor Cochrane was under the impression that many of his colleagues and predecessors were practicing with a lack of scientific evidence to justify decision-making. Professor Cochrane began to propose the idea that researchers should collaborate on an international scale to systematically review the best and most up-to-date evidence for each health-related discipline to help guide the process of decision-making.[41] With this new scientific approach gaining momentum, in 1991 the term 'evidence-based medicine' (EBM) was introduced by David Sackett and his team.[42] The increasing awareness of weaknesses in clinical practice and the negative impact on patient care was seen to be the impetus for EBM. The term EBM was controversial at the time, with many practitioners left

to feel their clinical decision-making was assumed to be based mostly on opinion and personal experience although most likely true.[42,43] Nevertheless the term EBM was introduced and has become the mainstay of how many health practitioners are trained and practice today.

Evidence-based medicine (EBM) has been defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of the patient."[44(p.71)] The practice of EBM involves integrating individual clinical expertise with the best available external clinical evidence from systematic research. Then utilising all the available evidence to assist in the final decision-making process so that recommendations are scientifically supported and appropriate for each individual patient in real life clinical practice.[44] As one could postulate from the name, EBM was originally grounded in the domain of medicine, however the movement of EBM then evolved and became entrenched in other surrounding health disciplines such as pharmacy, physiotherapy and nursing. A prominent example is the transition from hospital-based teaching for nursing students to the university-based education programme. With the success and uprise of EBM, an evolution of EBM into the EBHC model resulted where EBM has been accepted and utilised by numerous health disciplines with the aim to provide healthcare that is supported by evidence.[45] However, despite being more than thirty years since its conception, EBM continues to invoke a polarised debate amongst researchers and clinicians. Clinicians not only argue that the usefulness of applying EBM to individual patients is limited because individual circumstances and values vary, but relying on EBM can also reduce autonomy amongst doctor-patient relationships by limiting a patients right to choose from interventions that might yet to be 'proven' effective.[46] Criticisms such as these described suggest that although EBM can be a useful tool, it needs to be used with caution particularly in individual patient care and with uncommon conditions where higher levels of evidence will rarely exist.[46]

JBI's approach to EBHC

In order to further help clinicians in the transparent development and dissemination of evidence-based research and reviews, a number of organisations have been established. Organisations such as JBI, Cochrane Collaboration and Campbell Collaboration have been formed to assist health professionals in seeking and utilising the best available evidence to help inform clinical decision-making.[47]

The JBI model of EBHC is based on the Institute's approach to translating the best available evidence into best practice in the appropriate healthcare setting. JBI considers evidence-based healthcare as clinical decision-making that considers the feasibility, appropriateness, meaningfulness and effectiveness (FAME) of healthcare practices.[47] The JBI model was first developed in 2005 and then updated in 2016 (see Fig. 1). Diagrammatically within the JBI Model of EBHC, the inner circle represents the 'pebble of knowledge', a conceptualisation of EBHC that seeks evidence from the literature to answer questions on the FAME of a specific intervention for a particular condition. The five 'inner wedges' represent the organisation's conceptualisation of steps involved in an evidence-based approach to clinical decision-making. Global health care needs are those identified by clinicians, patients or consumers. Evidence that incorporates the FAME approach is then generated to address the health needs identified. That evidence is then collated where the results are appraised, synthesised and transferred to the healthcare setting where it can be utilised by health professionals to improve health outcomes, health systems and professional practice. The 'outer wedges' operationalise the component parts of the model and reflect how each 'inner wedge' can be actioned in a practical way. Whilst bi-directional arrows are present to indicate that the entire process of the JBI Model of EBHC is not a linear process, whilst users can engage with the model from a starting point that best meets their needs.[47]

Figure 1: The current JBI Model of EBHC.[47]

Types of evidence in EBHC

Accessing the best available evidence to assist in clinical decision-making is the fundamental basis of EBHC, however evidence is a complex concept with multiple meanings.[47] Through the evolution of academic sociology, there are three prevailing philosophical paradigms in the way western health care research is conducted, all encompassing a diversity of research methodologies and methods.[48] The first is the positivist who prefers scientific quantitative methods, while the other two, the interpretive and critical paradigms are largely associated with humanistic qualitative methods.[48]

Quantitative evidence is generated by research based on traditional scientific methods that generate numerical data. It has been suggested that quantitative evidence in medicine originated in the eighteenth century, with physicians and surgeons using statistical methods to assess the effectiveness of therapies for conditions such as scurvy, palsies and syphilis.[49] From there, quantitative research has expanded to encompass aspects other than effectiveness, including (but not limited to) incidence, prevalence, aetiology of disease and psychometric properties. The strength of quantitative evidence lies in its validity and reliability; results must be repeatable and consistent, yielding the same results or answers time after time.[50]

Qualitative evidence on the other hand allows researchers to analyse human experience and cultural and social phenomena.[51] The term 'qualitative' refers to various research methodologies including action research, discourse analysis, ethnography, phenomenology, grounded theory and qualitative inquiry. Research methods in qualitative research include interviews (whether group or individual) and observation (either direct or indirect). Researchers who use qualitative methodologies seek a deeper understanding, aiming to "study things in their natural setting, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them."[52] Relevant to EBHC, qualitative evidence has a particular role in exploring why interventions are or are not effective from a patient centred perspective, or why an intervention is not adopted despite evidence for its effectiveness.[53] The strength of qualitative research lies in its credibility (i.e. close proximity to the truth), using selected data collection strategies that "touch the core of what is going on rather than just skimming the surface."[54(p.740)]

Evidence Synthesis

According to the JBI model, evidence synthesis is defined as "the evaluation or analysis of research evidence and opinion on a specific topic to aid in decision-making in healthcare."[50]

Evidence synthesis has a long history in the health sciences. The earliest example of a systematic review was written by James Lind in 1753, who produced a thesis on scurvy utilising the then-available published evidence on the disease.[55]

In present time, systematically searching, appraising and analysing up-to-date and reliable research has become the cornerstone of EBM and EBHC. With systematic reviews and RCTs continuing to lead the hierarchy of evidence in clinical research.[56]

Numerous types of reviews for evidence synthesis have been identified so far, however the most common include literature reviews, scoping reviews and systematic reviews.

Systematic reviews remain the gold standard to search for, collate, critique and summarize the best available evidence regarding a clinical question, conducted using standardised and transparent methodological processes.[50,57] Literature reviews provide an examination of recent or current literature that covers a wide range of subjects at various levels of completeness and comprehensiveness. Literature reviews are designed to provide an overview of sources a researcher has explored while reviewing a particular topic without the strict methodological guidance as seen in systematic reviews. Literature reviews are therefore considered to be largely subjective, unreproducible and lack transparency.[58] Scoping reviews, which are often undertaken prior to a systematic review, aim to provide an overview or map of the evidence using a transparent methodological approach. However similar to literature reviews, scoping reviews do not aim to produce a critically appraised and synthesised result/answer to a particular question.[58]

The systematic review

Systematic reviews can be broadly defined as a type of research synthesis that are conducted by people with specialized skills, who set out to identify and retrieve all the available evidence that is relevant to a particular question (or questions) and appraise and synthesize the results of this search to inform practice, policy and in some cases, further research.[57]

These types of reviews are often considered as the pillar of EBHC and are widely used to inform the development of trustworthy clinical guidelines.[57,59,60] Frequently a systematic review may be conducted to confirm or refute whether or not current practice is based on relevant evidence, and therefore addressing whether available evidence should be relied upon or whether further research is recommended.[59]

The nature of the clinical question will determine the systematic review type that is required to be undertaken. Currently there are numerous review types that exist including (but not

limited to) those focusing on: effectiveness, aetiology and/or risk, cost/economic evaluation, prognostic, prevalence and/or incidence, diagnostic test accuracy (DTA), qualitative, mixed-methods and text/opinion.[61] Relevant to the following thesis is the DTA review type, as the primary review objective is to evaluate the accuracy of imaging modalities and their ability to detect a phenomena of interest.

According to the Cochrane handbook, a systematic review "uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made."[62] As such, regardless of the review type there is a general consensus around the key steps that are required to conduct a systematic review and to distinguish itself from a literature review.[50,62] The recommended steps include:

1. Clearly formulating a research question and study objectives. This process will usually be undertaken following a brief review of the literature, to ensure there is a clinical need for the proposed review and that no concurrent review is being undertaken of similar methodology.

2. Defining the inclusion and exclusion criteria. For DTA studies the inclusion criteria is recommended to consider the domains of population, index test, reference standard, and diagnosis of interest (PIRD).[50]

3. A comprehensive search to identify all relevant studies, both published and unpublished. JBI undertakes this step using a three-stage search process to aid study screening and selection. Two independent reviewers are recommended to complete the study screening and selection phase to minimise the risk of missing any relevant studies.[50]

4. Critical appraisal of the quality of included studies (risk of bias) and assessment of the validity of their results/findings/conclusions. The step of appraisal is undertaken by two independent reviewers with quality assessment based on the chosen appraisal tool. As per the most recent Cochrane guidelines, the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) appraisal tool is recommended for all DTA review types.[62]

5. Data extraction from the included studies (including characteristic type data and outcome data). Two independent reviewers and the use of a standardised extraction tool will aim to minimise errors in data extraction.

6. Presentation and synthesis of the results/findings (data synthesis). This step can either involve a descriptive (narrative summary) or statistical (meta-analysis) analysis. For DTA studies, a meta-analysis can be conducted when results of similar (homogeneous) individual studies are combined to allow for assessment of overall performance for a particular diagnostic tool. The meta-analysis permits a summary about the performance of the particular diagnostic test compared to a reference standard. However, when there are limited studies or there is variation (heterogeneous) between the primary studies a narrative summary is provided. The narrative synthesis will include the reasons of the heterogeneity and the inappropriateness of combining the data statistically.

7. Interpretation of the methodology, results, establishing the certainty of the body of evidence (through systems such as GRADE) and making implications for practice and future research.[50]

An essential step in the early development of a systematic review is the development of a review protocol. The protocol is a completely separate document to the review report, which will usually be developed prior to the screening and selection phase of the conducted review. The protocol will pre-define the systematic reviews objectives and methods to allow for transparency of the review process and allow readers to see how reported findings and recommendations were arrived at.[50]

Although systematic reviews are often referred to when decisions around clinical practice are required, the process can be extensive and timely, and unfortunately by completion of the review the findings and recommendations might already be outdated. Nevertheless, systematic reviews are considered the gold-standard for evidence synthesis in EBHC and when conducted meticulously represent the highest level of evidence in the hierarchy of evidence pyramid.[63]

Hierarchy of evidence pyramid

As practising physicians in the early 1990's started to appraise and apply evidence to their practice, it became a common finding that not all evidence is the same. As a result, a compelling rationale for creating a hierarchy of evidence was formed. Subsequently in 1995, Guyatt and Sackett published the first hierarchy of evidence pyramid to allow for transparent use amongst clinicians when undertaking a systematic search, appraisal or analysis of evidence.[63]

Although there have been various versions of the evidence pyramid to date,[56,63-65] all tend to focus on showing study designs with the highest level of evidence (systematic reviews and meta-analysis) at the top of the pyramid, with randomised controlled trials (RCTs) immediately below, cohort and case–control studies in the middle, then study designs with the lowest level of evidence (case series/reports) at the bottom of the pyramid (see Fig. 2 for example).[56]

Figure 2: Hierarchy of evidence pyramid.[56]

Traditionally, systematic reviews have been considered the highest form of evidence, providing an unbiased and comprehensive analysis of all available evidence, while RCTs are considered to be the 'gold standard' approach to generating evidence of effectiveness, utilising 'randomisation' to minimise risks of bias in cause-and-effect relationships.[50,56]

As it is not appropriate to distinguish between different qualitative study designs (i.e. ethnographic verse phenomenological) via a hierarchy, qualitative studies are assessed

instead using a ranking scale (high, moderate, low to very low) with research studies starting off as 'high' and expert opinion pre-ranked at 'low.'[50]

Levels of evidence and the grading of recommendations

Although hierarchy systems such as hierarchy of evidence pyramids are considered to be valuable in providing an overview of study design rateability in quantitative research, these hierarchy models do not investigate numerous other factors which may impact on the overall quality of evidence.

In the early 2000s, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group was established. Subsequently in 2014, the GRADE working group developed a grading of evidence and recommendation system (referred to as the GRADE approach). Unlike hierarchy of evidence systems, the GRADE approach is not solely focused on study design, however takes into account numerous other factors such as inconsistency of results, risk of bias, directness, heterogeneity, precision and publication bias to provide a more accurate assessment of quality of evidence. The GRADE approach assists in collating the results of quantitative research, rating the quality of evidence for outcomes and clearly presenting the results in an evidence table, such as a Summary of Findings (SoF) table.[66]

The included evidence in the systematic review is ranked out of a possible four levels (High, Moderate, Low and Very Low)(Table 2). The evidence is initially pre-ranked according to the study design; high quality for RCTs and low quality for observational studies.[66]

Symbol	Quality	Interpretation
⊕⊕⊕⊕	High	We are very confident that the true effect lies close to that of the estimate effect.
⊕⊕⊕O	Moderate	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
⊕⊕OO	Low	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of effect.
⊕OOO	Very low	We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Table 2: GRADE ratings and their interpretation (from the GRADE Handbook, available at https://gdt.gradepro.org/app/handbook/handbook.html)

The evidence can then be either downgraded or upgraded depending on factors including: inconsistency of results, indirectness of evidence, imprecision, publication bias, large magnitude effect, dose-response gradient and confounding aforementioned factors (see Tables 3 and 4 for the process of grading the evidence).[66]

Factor	**Consequence**
Limitations in study design or execution (risk of bias)	Downgrade one or two levels
Inconsistency of results	Downgrade one or two levels
Indirectness of evidence	Downgrade one or two levels
Imprecision	Downgrade one or two levels
Publication bias	Downgrade one or two levels

Table 3: Factors that can reduce the quality of evidence (from the GRADE Handbook, available at https://gdt.gradepro.org/app/handbook/handbook.html)

Factor	**Consequence**
Large magnitude of effect	Upgrade one or two levels
All plausible confounding would reduce the demonstrated effect or increase the effect if no effect was observed	Upgrade one level
Dose-response gradient	Upgrade one level

Table 4: Factors that can increase the quality of evidence (from the GRADE Handbook, available at https://gdt.gradepro.org/app/handbook/handbook.html)

In comparison to grading evidence from therapeutic intervention studies, grading the strength of evidence in diagnostic studies involves additional challenges. For example, regarding study design, cohort or cross sectional studies with appropriate tests can be considered high quality studies in DTA research. Whilst risk of bias and judging indirectness involve a separate approach to grading evidence in DTA studies compared to intervention research. The factors of inconsistency, imprecision, publication bias and upgrading for dose effect, large estimates of accuracy and residual plausible confounding are less well established in DTA studies. As such, GRADE is in the process of updating a version for grading the quality of evidence in DTA studies. However at present, the recommendation is to use the standard GRADE approach, not a "modified" version to avoid confusion.[66]

Regarding qualitative research, SoF tables are also created to give an overall rating of confidence in the qualitative synthesized findings, which JBI have termed 'ConQual' for short.[50] Research studies are initially pre-ranked as 'high' and then downgraded (to moderate, low to very low) based on the factors of dependability and credibility using the relevant appraisal tool. Expert opinion is pre-ranked as 'low'.[50]

Methodological basis of the chosen approach to synthesis

Diagnostic test accuracy studies are frequently undertaken to compare a diagnostic test of interest (index test) to an existing diagnostic test (reference test) which is regarded as the gold standard or best test for currently detecting the presence or absence of the condition of interest. The outcomes of the index and reference tests are then compared to one another to evaluate the accuracy of the test in question (index test). Regarding DTA study types, two main designs exist; the first is the diagnostic case-control design, where people with the condition (cases) come from one population (e.g. a health care centre for people known to have the condition), while people without the condition come from another. The second study design is cross-sectional, which involves all patients who are suspected of having the condition of interest undergoing both the index and the reference test. Patients that test positive on the reference test for the condition can be considered to be the cases, whilst patients that test negative are considered to be the controls. This latter study design is thought to mimic actual practice better than the case-control design and therefore provide a more valid estimate of diagnostic accuracy.[62]

Systematic reviews of DTA provide an overview of test performance based on all available evidence, taking into the account the quality of published studies, and differences in findings between studies.[62,67] The overall estimates of test accuracy will frequently vary between studies, often as a result of differences in participants characteristics, study design, diagnostic test thresholds, and flow and timing of the index or reference test in the diagnostic pathway.[62] Diagnostic accuracy is commonly reported by two measures, sensitivity (SN) and specificity (SP); however sometimes other measures including predictive values, likelihood ratios, odds ratios, summary receiver operating characteristic (ROC) curves and interobserver agreements are used.[68]

Relevant to the following review; sensitivity refers to the probability of a person with the condition of interest having a positive test result (also referred to as the true positive [TP]),

while specificity is the probability of a person without the condition of interest having a negative test result (also referred to as the true negative [TN]). A test that is 100% SN is consistent with correctly identifying all people with the disease as having the disease (i.e. no false negatives), whilst 100% SP is consistent with correctly identifying all healthy individuals as healthy (i.e. no false positives). A specific test is therefore used for helping ruling in a disease, as it rarely misclassifies those without a disease as being sick. Whilst a sensitive test is used for helping excluding a disease, as it rarely misclassifies those with a disease as being healthy. While SN and SP measure the accuracy of a diagnostic test, they do not provide the probability of the diagnostic value of the result of the test.[68]

The positive predictive value (PPV) refers to the amount of participants with positive test results who were correctly diagnosed, while the negative predictive value (NPV) refers to the amount of participants with negative test results who were correctly diagnosed. As the calculations for PPV and NPV includes individuals with and without the disease, the values are affected by the prevalence of the disease in question. As the prevalence decreases, the PPV decreases because there will be more false positives for every true positive. Whilst the NPV will increase because there will be more true negatives for every false negative. Conversely, as the prevalence increases the PPV will increase whilst the NPV will decrease.[68]

Receiver Operating Characteristic (ROC) curve analysis are useful to evaluate the performance of diagnostic tests that classify individuals according to those with or without a condition. The data obtained from a diagnostic test will often exist on a scale, and a decision will need to be made on a diagnostic threshold for whether the condition is present (positive test) or absent (negative test). As SN and SP depend on the selection of a diagnostic threshold, ROC analysis is used to plot the sensitivity (y-axis) against 1-specificity (x-axis) as the threshold value changes. This in turn creates a visual representation of the relationship between SN and SP of a diagnostic test as the threshold value changes. The ROC outcomes can be measured quantitatively by measuring the area under the curve (AUC). The AUC for a perfect test is 1.0, whilst a test with no distinction between disorder and no disorder has an AUC of 0.5.[69]

Interobserver agreements (K value [Cohen's kappa coefficient]) are useful for evaluating the inter-rater reliability when two observers are used to independently detect a condition of interest. Based on Kappa statistics by Landis and Koch, K values and their interpretation include: <0 - poor agreement, 0.0 - 0.2: slight agreement, 0.2 - 0.4: fair agreement, 0.4 - 0.6:

moderate agreement, 0.6 - 0.8: substantial agreement, 0.8 - 1.0: almost perfect agreement between observers.[70]

In this chapter the origins of EBHC (including the JBI model), the process of evidence synthesis and systematic review, Levels of Evidence and the GRADE approach for assessing the certainty of evidence, and the methodological basis of the chosen approach to synthesis is provided. The next chapter outlines the systematic review methods including review objective, eligibility criteria, search strategy, study selection, critical appraisal, data extraction, data synthesis and assessment of the certainty in the findings.

CHAPTER 3: SYSTEMATIC REVIEW METHODS

Overview of chapter 3

In chapter three, the systematic review methods including the review objective, eligibility criteria, search strategy, study selection process, critical appraisal, data extraction, data synthesis and assessment of the certainty in the findings are provided. This systematic review was conducted in accordance with the JBI methodology for systematic reviews on diagnostic test of accuracy studies[50] and in accordance with a priori protocol.[71] (PROSPERO Registration Number: CRD42021250626).

Review objective

To evaluate the accuracy of conventional imaging modalities at detecting extracapsular spread (ECS) of cervical lymph node metastases in human papilloma virus positive (HPV+) oropharyngeal squamous cell carcinoma (OPSCC).

Inclusion criteria

The following inclusion criteria were based on the PIRD (population, index test, reference test, diagnosis of interest) mnemonic recommended by JBI for systematic reviews involving diagnostic test of accuracy studies[50]:

P) Population: only studies explicitly involving participants with a diagnosis of HPV+ OPSCC (histopathology confirmed), and a suspected diagnosis of cervical lymph node metastases and ECS (on clinical assessment) were included in the review. Participants were not excluded due to age, sex, race or education status. Participants must have completed a pre-treatment staging scan of any imaging modality (CT, MRI, PET, PET/CT, US) before receiving H&NC treatment.

Participants with recurrent disease, H&NC other than HPV+ OPSCC and participants without nodal disease were excluded.

I) Index test: studies that utilised a conventional imaging modality to detect ECS in HPV+ OPSCC including CT (ceCT or plain), MRI, PET, PET/CT, and US with use of a medical specialist (radiologist) for interpretation were included.

Machine learning methods and studies using only indirect measures that are not consistently utilised in standard radiologic assessment of ECS (i.e. lymph node size) were excluded.

R) Reference test: studies that utilised surgical histopathology as confirmation for a diagnosis of ECS were included (this is the gold standard in current clinical practice).

D) Diagnosis of interest: studies where the diagnosis of interest was ECS of cervical lymph node metastases were included. The primary outcomes were sensitivity and specificity measures (with 95% confidence intervals) for each imaging modality used to detect ECS in HPV+ OPSCC. Secondary outcomes included PPV, NPV, area under the ROC curve (AUC) and interobserver agreement (K)(where applicable) values for the different imaging modalities.

Type of studies: the review considered only published studies that examined the diagnostic accuracy (including sensitivity and specificity) of an imaging modality to detect ECS in HPV+ OPSCC with reference tests performed. Diagnostic cohort studies were the preferred study design for inclusion into the review, whilst diagnostic case-control studies were also considered.
Review articles, case studies and letters to the editor were excluded.

Search strategy

A three-step search strategy aimed to locate all published studies on the topic. An initial limited search of MEDLINE (PubMed) was undertaken to identify articles on the topic. The text words contained in the titles and abstracts of relevant articles, and the index terms used to describe the articles were then used to develop a full search strategy in consultation with a scientific librarian for each information source. On May 18th 2021 the electronic databases MEDLINE (PubMed), Embase (Ovid), Cochrane Central Register of Controlled Trials (via Cochrane Library), Web of Science and Scopus were searched with no restriction on publication date (see Appendix I for the full search strategy). Only studies in English were included in the review. Finally, the reference lists of all included studies were screened for additional studies. Grey literature was not searched due to time restraints.

Study selection

Following the search, all identified citations were collated and uploaded into EndNote X9 (Clarivate Analytics, PA, USA) and duplicates were removed manually. Following a pilot test of four studies, titles and abstracts were screened for assessment against the inclusion criteria for the review. Potentially relevant studies were retrieved in full and their citation

details imported into the JBI System for the Unified Management, Assessment and Review of Information (JBI SUMARI) (JBI, Adelaide, Australia). The full text of selected citations were assessed in detail against the inclusion criteria. Title and abstract screening was performed by the primary author (TM) and verified by another reviewer (CN), whilst full text screening was performed independently by two reviewers (TM, CN). Any disagreements that arose during the study selection process were resolved through discussion between the two reviewers. The results of the search and the study inclusion process is reported in full and presented in a Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram (see Fig. 3).

Assessment of methodological quality

As originally outlined in the review protocol,[71] eligible studies were critically appraised by two independent reviewers (TM, CS) for methodological quality using the critical appraisal checklist QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2).

The QUADAS-2 tool was conceived in 2003, developed specifically to assess the methodological rigor of diagnostic test accuracy studies in systematic reviews.[72] At present, the QUADAS-2 tool is recommended for use in all Cochrane DTA reviews[62] and was utilised in the three previous reviews on imaging and ECS in H&NC. [29,38,39] Other appraisal tools for DTA studies include the JBI critical appraisal checklist for DTA,[50] the SIGN (Scottish Intercollegiate Guidelines Network) critical appraisal checklist for diagnostic study,[73] and the CASP (Critical Appraisal Skills Programme) diagnostic check list.[74] Of those listed, the JBI and SIGN tools are both based on the QUADAS-2 tool.[50,73]

The QUADAS-2 checklist for primary diagnostic accuracy studies is structured around four domains; Patient Selection, Index Test, Reference Standard, and Flow and Timing. Each domain is assessed in terms of risk of bias, and the first three domains are also assessed regarding applicability. Signalling questions with 'yes', 'no' or 'unclear' responses are included to help determine the risk of bias. The risk of bias is then assessed to be either ‘A’ low risk of bias, ‘B’ unclear risk of bias, or ‘C’ high risk of bias based on the signalling questions for each domain (see Table 5). Concerns regarding applicability is also based on signalling questions with 'high', 'low', 'unclear' responses to ascertain whether the domain (Patient Selection, Index Test, Reference Standard) is relevant to the review question.

Assessment for overall rating of individual studies is not a feature of the QUADAS-2 tool. Per QUADAS-2, to assess risk of bias for each domain; if any of the three signalling questions are rated as ‘no,’ it should be considered there is high risk of bias. Whilst if all

three questions are rated as 'yes' then it is considered there is low risk of bias for that particular domain. If any of the questions are reported as 'unclear', then there is an unclear risk of bias and a judgement should be made on whether or not there is enough information to make a decision about the risk of bias.[72] In the conducted review it was decided amongst the two reviewers that a conservative approach be taken, and any unclear risk of bias in the signalling questions automatically resulted in that domain having unclear risk of bias (provided there was no high risk of bias to downgrade the level of bias). Otherwise the recommendations for assignment of low or high risk of bias were allocated per the QUADAS-2 tool guidelines.[72]

DOMAIN	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting):	Describe the index test and how it was conducted and interpreted:	Describe the reference standard and how it was conducted and interpreted:	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard:
Signalling questions(yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias: High/low/unclear	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/low/unclear	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Table 5: Summary of the QUADAS-2 tool.[72]

Concerns regarding applicability for each QUADAS-2 domain (Patient selection, Index test, Reference standard) was not assessed per JBI guidance.[50] This is due to the strict inclusion criteria in the following review which was based on the PIRD mnemonic recommended by JBI for DTA type reviews[50] and therefore included studies feature the appropriate patient population, index test and reference standard. Overall quality for each study was not assessed per the QUADAS-2 tool and Cochrane recommendations.[62,72] Any disagreements that arose for assessment of methodological quality were resolved through discussion between the two reviewers.

In the following review, all eligible studies were included regardless of methodological quality. Due to the differing treatment protocols that exist for managing patients with H&NC and ECS, there is always a risk of verification bias amongst the selected studies. A select group (i.e. ECS negative patients) will often undergo surgery and receive the gold standard (histopathology), whilst in some cases patients who are ECS positive go on to receive chemoradiotherapy without undergoing the gold standard.[33,34] Verification bias amongst all other relevant biases will be explained further in the results section of the review.

Data extraction

Following a pilot test, and utilising the JBI data extraction tool for DTA studies,[50] data was extracted from papers included in the review by the primary author (TM) and verified by a secondary author (AF). The data extracted included: basic information from the study (i.e. year of publication, authors, study design, location, number of participants), participant characteristics (i.e. age, sex, tumour type, tumour location), details about the imaging modalities used (i.e. CT, MRI, PET/CT), details about the index test and reference standard (i.e. diagnostic criteria, blinding) and outcomes (including sensitivity, specificity, PPV, NPV and inter-rater agreement (Cohen's Kappa (K)) for studies using two observers (radiologists)). For studies using two observers to detect radiologic ECS, the outcomes for each observer were extracted. All included studies reported their outcomes with sensitivity, specificity, PPV and NPV values except for the PET/CT study by Snyder et al.[75] With aid from a local statistician, calculations were required to convert the Snyder et al.[75] 'ECS misclassification analysis' to our primary outcomes. These calculations can be found in Appendix III.

The CT study by Faraji et al.[76] investigated the highest performing characteristics to aid radiologic ECS detection with no overall assessment for ECS being reported. Therefore in the following review the highest overall performing characteristic for ECS detection in the Faraji et al. study[76] ('absence of perinodal fat plane') was used in the CT meta-analysis.

The standards for the reporting of diagnostic accuracy studies (STARD) checklist and flow diagram were utilised to provide guidance during data extraction.[77] Any disagreements that arose between the two reviewers were resolved through discussion. Efforts to elicit further information by contacting authors of papers (n=4) directly to request missing or additional data were unsuccessful after five attempts.

Data synthesis

All statistical data analysis was performed using Stata version 15.1 (StataCorp LLC, Texas, USA) and Meta-DiSc version 1.4 (the Unit of Clinical Biostatistics team of the Ramón y Cajal Hospital, Madrid, Spain) with assistance from a local statistician.

The main outcomes were sensitivity and specificity measures with 95% confidence intervals for each imaging modality used to detect ECS in HPV+ OPSCC. Secondary outcomes included PPV, NPV, area under the ROC curve (AUC) and interobserver agreement (K)(where applicable) values for the different imaging modalities. If two observers were present in a study, the mean of both observers was used for analysis purposes.

A meta-analysis was conducted on four CT studies using a random-effects model.[78] The results for pooled CT sensitivity and specificity are displayed on paired forest plots and a summary receiver operating characteristic curve (SROC). The I^2 statistic was used to evaluate heterogeneity (with I^2 >50% indicating significant heterogeneity) as was Cochran's Q P value (with p value <0.05 indicating significant heterogeneity). In view of the heterogeneity found for both sensitivity and specificity, a random-effects model was used throughout.[78] Statistical tests were considered significant against the null hypotheses if p-values <0.05.

The remaining two studies (investigating combination PET/CT and 'CT and MRI')[75,79] underwent a narrative synthesis. Publication bias was unable to be assessed (as per Cochrane Guidelines)[62] due to the limited number of included studies (<10) in the review.

Assessing certainty in the findings

Three 'summary of findings' tables were created using GRADEpro GDT (Guideline Development Tool, McMaster University). The GRADE approach for assessing the certainty of evidence for diagnostic test accuracy studies was followed throughout.[66] The following details are included in the 'summary of findings' tables: sample size, imaging modality, accuracy estimates (SN, SP, true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN)) and certainty of the evidence.

In this chapter the methods used in the underlying systematic review were outlined, including the eligibility criteria, search strategy, study selection process, how the studies were critically appraised, data extraction, data synthesis and assessment of the certainty in the findings. In the next chapter, the search results, study selection process, assessment of the methodological quality, an overview of the results of the six papers included in the systematic review, and assessment of the certainty in the findings (using GRADEpro GDT) are provided.

CHAPTER 4: RESULTS

Overview of chapter 4

In chapter four, the findings of the systematic review conducted to evaluate the accuracy of conventional imaging modalities at detecting ECS of cervical lymph node metastases in HPV+ OPSCC are provided. A detailed description of the search results, the study selection process and the assessment of methodological quality is presented which is followed by the characteristics of the included studies. Furthermore, a meta-analysis of the four CT studies and a narrative synthesis of the PET/CT and 'CT and MRI' study is provided organised by imaging modality. Finally an assessment of the certainty in the findings using GRADEpro GDT is provided.

Study selection

A total of 1772 citations across four databases and one clinical trials register were retrieved in the initial search on May 18th 2021 (see Appendix I for the search strategies developed for each database). Following removal of duplicates (n=306), 1466 citations underwent title and abstract screening. Of the 1466, 110 papers were assessed for full-text eligibility and 104 were excluded (reasons for exclusion are presented in Appendix II). The primary reasons for exclusion were ineligible participants characteristics (i.e. no HPV or ECS status) and study design. Six studies were included in the review, with four eligible for meta-analysis[76,80-82] and two undergoing a narrative synthesis.[75,79] See Fig. 3 for the study selection process.[59]

Study characteristics

A total of six studies with 463 participants were included in the review.[75,76,79-82] All studies were of retrospective cohort study design and conducted between 2006-2017. Four hundred and three participants identified as male. Age was not consistently reported across individual studies, however the mean age across four studies was 57.8 years,[75,79,81,82] the study by Faraji et al.[76] had a median age of 56.7 years, the study by Geltzeiler et al.[80] assessed age as a continuous variable with no mean/median reported. All studies involved only participants with HPV+ OPSCC and a mixture of ECS positive and ECS negative diagnoses. The most common site of primary tumour was tonsillar (n=298) followed by base of tongue (n=127). Four of the studies were conducted in the United States of America,[75,76,80,82] one in Australia[81] and one in Korea.[79] Four of the studies assessed contrast-enhanced computed

tomography (ceCT),[76,80-82] one study assessed PET/CT,[75] and one study assessed 'CT and MRI'[79] (with no separation of index test to outcomes) for radiological detection of ECS. Five studies[75,76,79,81,82] contained assessments from two different radiologists to detect radiologic ECS, whilst the remaining study by Geltzeiler et al.[80] used only one observer. There was significant variability between studies for the radiological diagnostic criteria for ECS. Three studies did not report on the timing between index test and reference standard,[75,76,82] whilst the remaining three studies[79-81] involved a majority of participants who underwent an index test and reference standard within a six week time interval (recommended protocol).[33] A total of four studies (all ceCT)[76,80-82] with 280 participants were included in a meta-analysis. See Appendix IV for a detailed description of the included studies.

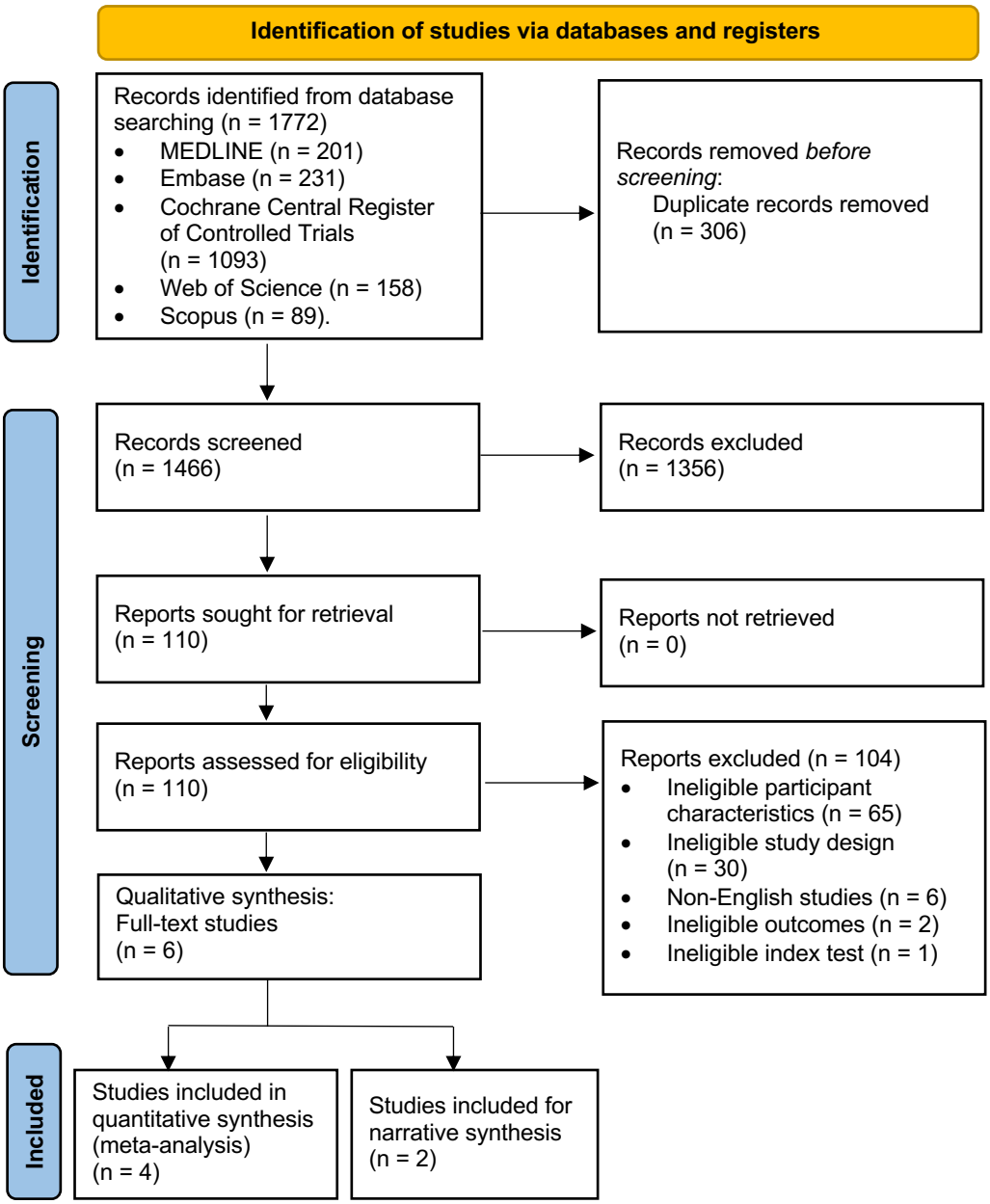


Figure 3: PRISMA Flow diagram of study selection.[57]

Risk of bias

Two independent reviewers carried out critical appraisal of the six included studies using the QUADAS-2 tool.[72] In the systematic review, none of the included studies were considered to be of high risk of bias in the QUADAS-2 domains of patient selection, index test, reference standard, and flow and timing.

Unclear risk of bias was found in the majority of studies in the domains of reference standard and flow and timing. This was mostly as a result of insufficient information regarding the blinding status of the pathologist to the index test, and a lack of information regarding timing between index test and reference standard. Clinically if there is more than a six week delay between imaging and surgical histopathology it is believed tumor characteristics could have evolved,[33] and therefore the reference standard in this situation is not a true representation of the index test leading to possible bias. This was apparent in three of the included studies.[75,76,82] Additionally, as discussed in the assessment of methodological quality section, another type of bias, verification bias, will exist in a large portion of studies where a select group of patients (i.e. radiologic ECS positive) may undergo definitive chemoradiotherapy without undergoing the reference standard (surgical histopathology). Subsequently influencing their risk of bias in the QUADAS-2 domain of reference standard, and the overall credibility of studies not using a reference standard for all participants regardless of ECS status.

Low risk of bias was found in the majority of studies in the domains of patient selection and index test. The included studies utilised the appropriate population of HPV+ OPSCC, avoided inappropriate exclusions, and utilised radiological techniques with pre-specified diagnostic thresholds with blinding to the reference standard.

Applicability of the QUADAS-2 domains (patient selection, index test, reference standard) and overall quality of individual studies was not assessed per the QUADAS-2 tool and Cochrane recommendations.[62,72] However the study by Snyder et al.[75] featured the most low risk of bias across the four domains (n=3), whilst the study by Lee et al.[79] featured the most unclear risk of bias across the four domains (n=3). A summary of results for the risk of bias assessment is presented in Table 6.

STUDY	RISK OF BIAS			
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Geltzeiler et al.[80]	A	A	B	B
Patel et al.[82]	A	B	A	B
Lee et al.[79]	A	B	B	B
Noor et al.[81]	A	A	B	A
Faraji et al.[76]	B	A	B	B
Snyder et al.[75]	A	A	A	B

A Low Risk B Unclear Risk C High Risk

Table 6: QUADAS-2 critical appraisal results.

Review findings

In the following section, the diagnostic performances of CT, PET/CT and 'CT and MRI' for detecting ECS in HPV+ OPSCC will be reported.

As discussed in chapter two, the measures of SN, SP, PPV, NPV, AUC and K (interobserver agreement) values are common to DTA studies, and reflect the primary and secondary outcomes of the following review.

SN, SP, PPV and NPV outcomes are typically reported in the range of 0-100% with 100% being considered a perfect diagnostic test. Although variable for many clinicians, in general, clinically acceptable diagnostic tests tend to range from greater than 70%. High specificity and/or PPV values are consistent with accurately detecting a phenomenon of interest i.e. ECS, whilst high sensitivity and/or NPV values are consistent with reliably excluding a diagnosis of interest (ECS).[83]

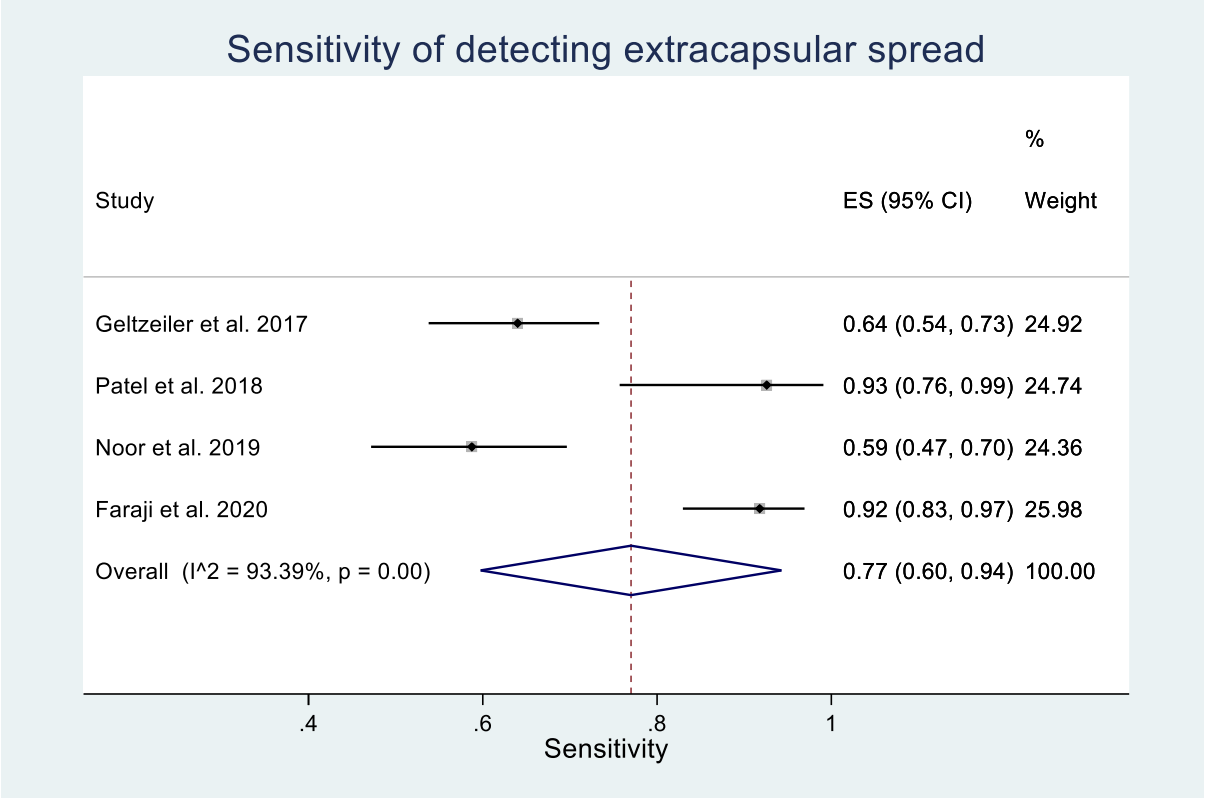
AUC values are representative of the overall performance of a diagnostic test with outcomes between 0.6 - 0.7 regarded as 'acceptable,' 0.7 - 0.8 'good,' 0.8 - 0.9 'very good,' and 0.9 - 1.0 'excellent.'[84]

K values (Cohen's kappa coefficient) are useful for evaluating the inter-rater reliability when two observers (i.e. radiologists) are used to independently detect a condition of interest (i.e. ECS). Based on Kappa statistics by Landis and Koch, K values and their interpretation include: <0 - poor agreement, 0.0 - 0.2: slight agreement, 0.2 - 0.4: fair agreement, 0.4 - 0.6: moderate agreement, 0.6 - 0.8: substantial agreement, 0.8 - 1.0: almost perfect agreement between observers.[70]

Evaluation of diagnostic ability:

CT

The results of the meta-analysis on the diagnostic value of CT for detecting ECS showed a pooled sensitivity of 77% [95% confidence intervals (CI) 60-94%], and specificity of 60% [95% CI 47-73%]. A paired forest plot of the meta-analysis can be seen in Fig. 4. The area under the SROC curve (AUC) was 0.72 which equates to a 'good' diagnostic test (Fig. 5). Cochran's Q test and Higgins I^2 statistics demonstrated substantial heterogeneity in the CT meta-analysis in terms of sensitivity and specificity (I^2 93.4%, P <0.0001 for Q test for sensitivity; I^2 80.7%, P <0.0001 for Q test for specificity). These findings were adjusted for using a random-effects model. Individual SN, SP, PPV and NPV values for each of the four CT studies can be found in Appendix IV.



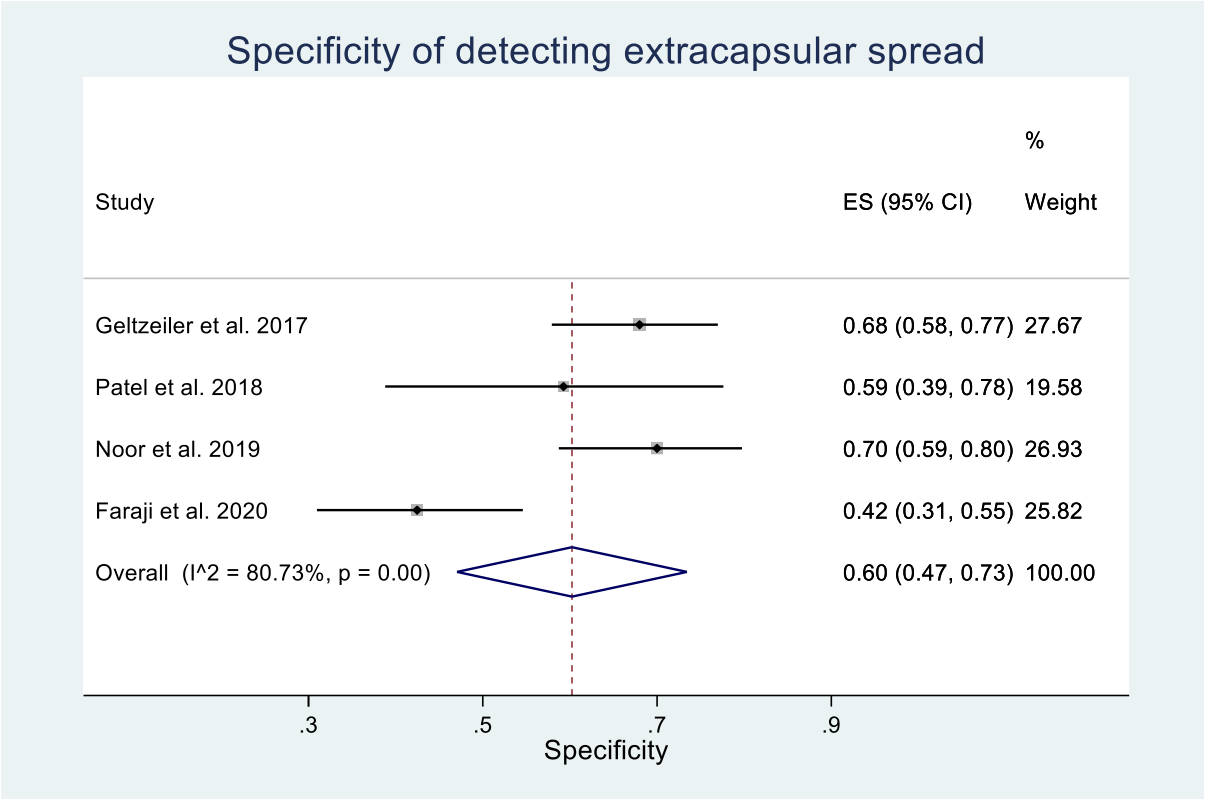


Figure 4: Coupled forest plot of sensitivity and specificity of CT for detecting ECS.

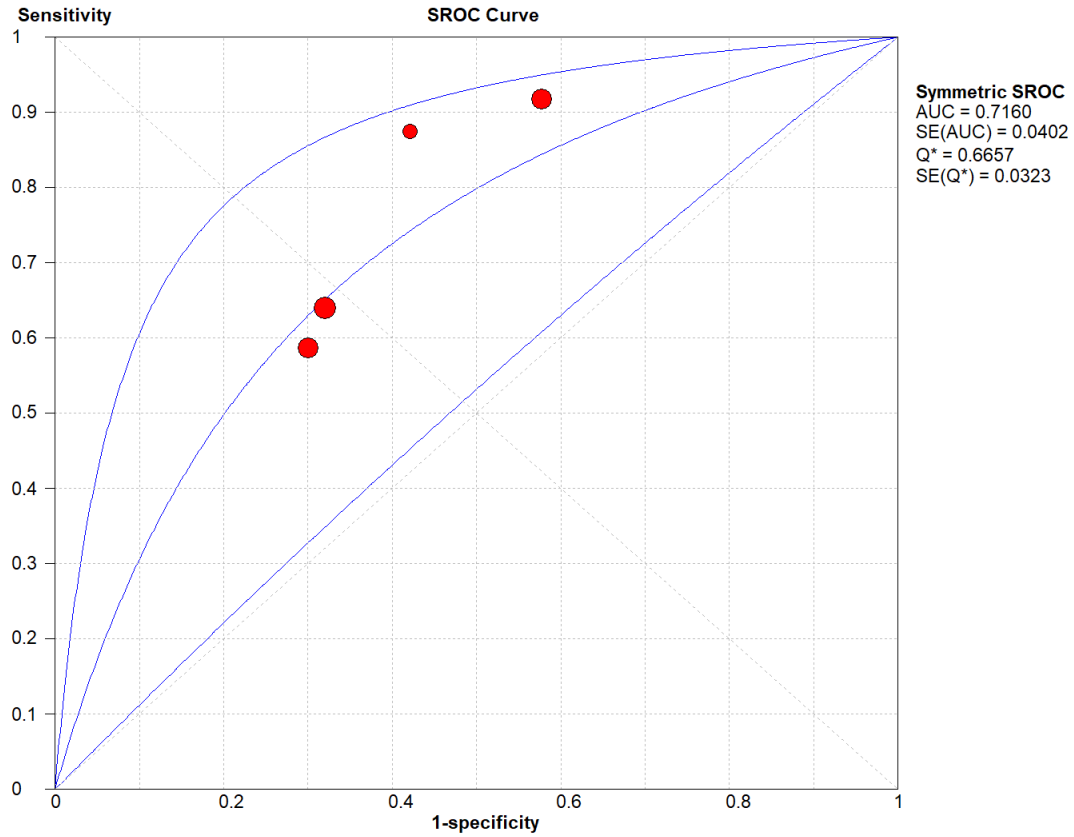


Figure 5: Summary receiver operating characteristic (SROC) curve of diagnostic performance of CT. The area under the SROC curve (AUC) was 0.72.

PET/CT

The sole PET/CT study[75] reported their outcomes in terms of an 'ECS misclassification analysis' (misclassified, under-staged, over-staged). With input from a local statistician to convert outcomes (see Appendix III), PET/CT had a calculated sensitivity of 86% [95% CI 73-94%], specificity of 76% [95% CI 61-87%], PPV of 37.5% and NPV of 97%. No SROC curve was reported or performed for an AUC value.

'CT and MRI'

The sole 'CT and MRI' study[79] had a sensitivity of 62% [95% CI 53-70%], specificity of 78% [95% CI 70-84%], PPV of 62%, NPV of 78% and AUC of 0.70. The results were unable to be separated to individual CT and MRI outcomes.

Interobserver agreement

Five of the studies[75,76,79,81,82] contained assessments from two different observers to detect radiologic ECS. In these studies, inter-rater agreements were reported using the K value (Cohen's kappa coefficient). Based on Kappa statistics by Landis and Koch:[70] three studies[75,76,81] were considered to have 'moderate' inter-rater agreements (K 0.4-0.5), whilst the remaining two studies[79,82] were considered to have 'substantial' agreement between observers (K 0.7) for detecting ECS (see Appendix IV).

Assessing certainty in evidence

Three 'summary of findings' tables were created using GRADEpro GDT (Guideline Development Tool, McMaster University). The GRADE approach for assessing the certainty of evidence for diagnostic test accuracy studies was followed.[66] The following details are included in the SoF tables: sample size, imaging modality, accuracy estimates (SN, SP, TP, FP, TN, FN) and certainty of the evidence.

Summary of Findings

Table 7: Summary of findings table for studies utilising CT as the index test (n=4).

Diagnostic accuracy of CECT in detection of radiologic ECS in HPV+ OPSCC

Patient or population: HPV+ OPSCC with suspected ECS of cervical lymph node metastases
Setting: Department of Radiology, John Hopkins Hospital/Winschip Cancer Institute/Oregon Health & Science University/Royal Adelaide Hospital
Index test: CECT | **Cut-off value**: presence/absence of radiologic ECS
Reference test: Surgical histopathology | **Threshold**: presence/absence of pathologic ECS
Pooled sensitivity: 0.77 (95% CI: 0.60 to 0.94) | **Pooled specificity**: 0.60 (95% CI: 0.47 to 0.73)

Test result	Number of results per 100 patients tested (95% CI)	Number of participants (studies)	Certainty of the Evidence (GRADE)
	Prevalence 45%		
True positives	**35** (27 to 42)	122 (4)	⨁◯◯◯ **Very low**[a,b,c]
False negatives	**10** (3 to 18)		
True negatives	**33** (26 to 40)	158 (4)	⨁◯◯◯ **Very low**[a,b,c]
False positives	**22** (15 to 29)		

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns regarding blinding of tests, time interval between index test and reference standard, and presence of partial verification bias.
b. Concerns regarding unexplained heterogeneity in the CT meta-analysis on sensitivity and specificity.
c. Concerns regarding small sample size across four studies (n=280). Publication bias suspected.

Table 8: Summary of findings table for the single PET/CT study.

Diagnostic accuracy of PET/CT in detection of radiologic ECS in HPV+ OPSCC

Patient or population: HPV+ OPSCC with suspected ECS of cervical lymph node metastases
Setting: Department of Radiology, University of Pittsburgh Medical Centre
Index test: PET/CT | **Cut-off value**: presence/absence of radiologic ECS
Reference test: Surgical histopathology | **Threshold**: presence/absence of pathologic ECS
Single study sensitivity: 0.86 (95% CI: 0.73 to 0.94) | **Single study specificity**: 0.76 (95% CI: 0.61 to 0.87)

Test result	Number of results per 100 patients tested (95% CI)	Number of participants (studies)	Certainty of the Evidence (GRADE)
	Prevalence 14%		
True positives	**12** (10 to 13)	7 (1)	⨁◯◯◯ **Very low**[a,b,c]
False negatives	**2** (1 to 4)		
True negatives	**65** (52 to 75)	42 (1)	⨁◯◯◯ **Very low**[a,b,c]
False positives	**21** (11 to 34)		

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns regarding a lack of reporting on time interval between index test and reference standard.
b. Concerns regarding wide confidence intervals for the outcomes of sensitivity and specificity.
c. Concerns with small sample size (n=49).

Table 9: Summary of findings table for the single 'CT and MRI' study.

Diagnostic accuracy of 'CT and MRI' in detection of radiologic ECS in HPV+ OPSCC

Patient or population: HPV+ OPSCC with suspected ECS of cervical lymph node metastases
Setting: Department of Radiology, University of Ulsan College of Medicine
Index test: 'CT and MRI' | **Cut-off value**: presence/absence of radiologic ECS
Reference test: Surgical histopathology | **Threshold**: presence/absence of pathologic ECS
Single study sensitivity: 0.62 (95% CI: 0.53 to 0.70) | **Single study specificity**: 0.78 (95% CI: 0.70 to 0.84)

Test result	Number of results per 100 patients tested (95% CI)	Number of participants (studies)	Certainty of the Evidence (GRADE)
	Prevalence 52%		
True positives	**32** (28 to 36)	70 (1)	⨁◯◯◯ **Very low**[a,b,c]
False negatives	**20** (16 to 24)		
True negatives	**37** (34 to 40)	64 (1)	⨁◯◯◯ **Very low**[a,b,c]
False positives	**11** (8 to 14)		

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Concerns regarding blinding during interpretation of both the index test and reference standard.
b. Concerns regarding reporting of separate index tests ('CT and MRI') as one test.
c. Concerns regarding small sample size (n=134).

In this chapter the underlying systematic review conducted to evaluate the accuracy of conventional imaging modalities at detecting ECS in HPV+ OPSCC was provided. A detailed description of the search results, the study selection process and the assessment of methodological quality was presented which was followed by the characteristics of the included studies. Furthermore, a meta-analysis and narrative synthesis of the results organised by imaging modality was provided. Finally, an assessment of the certainty in the findings using GRADEpro GDT was provided. In the next chapter, the findings and limitations of the systematic review will be discussed, along with the conclusions and the implications for practice and future research in this area.

CHAPTER 5: DISCUSSION, CONCLUSIONS AND IMPLICATIONS FOR PRACTICE AND RESEARCH

Overview of chapter 5

Two objectives are addressed in the final chapter of this thesis. The first objective is to provide an overview of the findings of the review, highlighting the diagnostic accuracy of CT, PET/CT, 'CT and MRI' at detecting ECS in HPV+ OPSCC and aligning it to the existing evidence base. The second objective is to discuss the strengths and limitations of the review, before finally concluding with remarks regarding implications of this review for clinical practice and future research.

Overview of findings

To our knowledge, this is the first systematic review to investigate the accuracy of conventional imaging modalities used to detect ECS explicitly in the HPV+ OPSCC population. By performing this review we were able report on the diagnostic value of CT and PET/CT to try aid clinical practice in assessing radiologic ECS in the future. Unfortunately, the included study by Lee et al.[79] reported their outcomes for 'CT and MRI' therefore precluding MRI outcomes from any analysis or comparative findings. Whilst no studies involving US to detect ECS in HPV+ OPSCC were detected during the search phase of the review.

The pooled sensitivity, specificity and AUC values for our CT meta-analysis were 77%, 60% and 0.72 respectively. Unfortunately, due to limited numbers and the Lee et al. study design,[79] we were unable to perform a comparative meta-analysis amongst CT, PET/CT and MRI. However, on direct comparison, the findings from our CT meta-analysis of four studies (n=280) appear inferior to the single PET/CT study[75] (n=49), with higher sensitivity and specificity values for ECS detection with PET/CT (86% vs. 77%, 76% vs. 60%, respectively). Although the PET/CT study by Snyder et al.[75] featured the lowest risk of bias across the QUADAS-2 domains (n=3), the small sample size, study design and single study comparison to the meta-analysis means this direct comparison needs to be interpretated with caution. Furthermore in the PET/CT study by Snyder et al.,[75] although the PPV for detecting nodal disease in the neck was reported to be 95%, our calculated PPV from their study for ECS detection is too low (37.5%) to suggest superior diagnostic value for ECS assessment compared to CT. Interestingly however, a calculated NPV of 95% for PET/CT and ECS

suggests PET/CT may be of more clinical value in aiding the exclusion rather than detection of ECS, although further large-scale high-quality studies are required to support these findings.

For both CT sensitivity and specificity there was substantial heterogeneity in the meta-analysis (I^2 93.4 vs. 80.7%, respectively). Again, due to the limited study numbers, no meta-regression for analysis on possible causes was able to be performed.

The findings of our CT meta-analysis reflect that similar to the single PET/CT study,[73] CT might not be a reliable tool for detecting radiologic ECS in HPV+ OPSCC. However, the superior sensitivity to specificity values in the CT meta-analysis (77% vs. 60%) suggest CT like PET/CT may also be of more clinical use in helping exclude rather than diagnose the presence of ECS although further research is ultimately required.

Comparison to previous studies

In comparison to the three previous reviews involving CT in the broader population of H&NC,[29,38,39] the CT outcomes in our study are inferior to all three, likely reflecting the different and more complex nature of HPV+ OPSCC. As discussed earlier, HPV+ SCC has a propensity for peritumoral desmoplasia (fibrosis) which can mimic ECS radiologically, making the diagnosis of ECS more complicated in HPV+ SCC disease.[25] Of the three previous reviews,[29,38,39] one review by Park et al.[38] published on the accuracy of CT and MRI to detect ECS in any SCC-related H&NC. A subgroup analysis involving five studies on CT in HPV+ OPSCC was performed which reported CT sensitivity to be similar to our findings (73% vs. 77%), however our pooled CT specificity was inferior (60% vs. 74%). The subgroup analysis by Park et al.[38] appears to be the only study that has performed an analysis on radiological techniques and their accuracy at detecting ECS in HPV+ OPSCC. The subgroup analysis however included the 'CT and MRI' study by Lee et al.[79] which likely explains our differences in results. The remaining four CT studies in the Park et al.[38] subgroup analysis are also included in this review and meta-analysis. Therefore our findings regarding CT outcomes are more of a true representation of the diagnostic value of CT to detect ECS in HPV+ OPSCC. Whilst comparing the single PET/CT study by Snyder et al.,[75] the sensitivity values are comparable to the previous findings in H&NC by Su et al.[29] and Abdel-Halim et al.[39] (86% vs 86% vs 80% respectively), however the specificity values for PET/CT are inferior (76% vs 86% vs 83% respectively) in our review in the HPV+ OPSCC population.

Diagnostic criteria and inter-observer agreement

In terms of radiological diagnostic criteria for ECS, this area remains an ongoing contentious issue. In four of the included studies,[76,80-82] observers utilised five or more different diagnostic features to detect radiologic ECS, however with variability between studies in their choice of diagnostic features and overall approach to detecting ECS. The combination of central node necrosis, irregular nodal margins, matted nodes and perinodal stranding featured as part of ECS assessment in four of the included studies.[76,80-82] One study[75] used an 'overall impression' for ECS as their assessment, whilst the other remaining study[79] assessed for irregularity of nodal rim or infiltration of adjacent soft tissues to predict the presence or absence of ECS.

This review alongside previous studies on ECS in H&NC reflect the ongoing subjective nature of radiologic ECS assessment. To highlight this further, amongst our included studies five[75,76,79,81,82] utilised two radiologists to independently predict ECS and reported their inter-observer agreements using Kappa values. Three of the five studies[75,76,81] reported only 'moderate' agreements (K 0.4-0.5) were reached between two observers for their assessment of radiologic ECS.

Notable findings from individual CT studies

Although not the focus of this review, it was interesting to note that across the four CT studies, apart from overall diagnostic value each study also assessed specific radiological diagnostic features with their association to ECS, all with significant findings.

The CT study by Faraji et al.[76] found irregular margins and/or absence of perinodal fat plane were significantly associated with ECS for both observers ($P < 0.05$). Similarly, the CT study by Noor et al.[81] found perinodal fat stranding was significantly associated with ECS for both observers ($P < 0.03$). Whilst the study by Geltzeiler et al.[80] found both border irregularity or number of radiographically suspicious nodes were significantly associated with ECS ($P < 0.02$). Whilst lastly the study by Patel et al.[82] found only lymph node size >3cm (4.7-5.4 odds ratio 95% CI 1.3-44, $P < 0.02$) demonstrated significant correlation with major pathological ENE (PENE >2mm) for both observers.

Interestingly, in the Geltzeiler et al.[80] study, it was reported that when counting the number of radiographically suspicious nodes (RSN), if greater than three RSN's were detected, the PPV could be improved from 71% per the radiologists overall assessment for ECS to 91% (see Appendix IV). Whilst the study also found significant associations regarding clinical nodal

classification and the rate of ECS. Participants with HPV+ OPSCC with no clinical evidence of nodal disease in their neck (known as cN_0) were found to have a low ECS rate of 3.3% (P <0.001), whilst those with clinical evidence of nodal disease involving one lymph node (<6cm) on the ipsilateral side of the primary tumor (known as $cN2_b$) had an ECS rate of 69.2% (P <0.001).

Although these studies offer encouraging findings, further research is required to validate these radiological features, clinical findings and their association with ECS.

Impact of ECS on treatment

Interestingly, new research is looking into degrees of nodal extension (microscopic versus macroscopic ECS >2mm) and their associated prognostic implications. A national trial by ECOG-ACRIN 3311 suggests that only macroscopic ECS is likely to require chemotherapy and full-dose radiotherapy (66-Gy), whilst surgery and low-dose adjuvant radiotherapy (50 or 60-Gy) might be an appropriate treatment regimen for patients with microscopic ECS <1mm.[85]

If these findings are accepted clinically, and research continues to focus on treatment de-intensification protocols, imaging modalities will need high precision to accurately differentiate between microscopic and macroscopic ECS. Currently this distinction requires surgical dissection and histopathology for confirmation.

Only the CT study by Patel et al.[82] identified degree of ECS extension and reported their outcomes for macroscopic ECS >2mm. The remaining studies did not indicate degree of ECS extension in their ECS positive cohorts which may have implications for future treatment protocols.

The future of radiological techniques

More recently, newer imaging techniques such as texture analysis and machine learning methods are currently being investigated in attempts to improve ECS detection and reduce the current subjective nature of its assessment amongst radiologists.[86,87] Machine learning methods are in the field of artificial intelligence that use layered neural networks to analyse data and predict outcomes such as ECS.

Recent findings by Kann et al.[86] in SCC-related H&NC suggest the utilisation of a CT machine learning algorithm showed superior diagnostic advantage at detecting ECS with an achieved AUC of 0.90 (88.6% accuracy), outperforming the radiologists reported AUC of 0.60 and 0.82 ($P < 0.0001$ and $P = 0.16$ respectively). The highest sensitivity and specificity

values for the CT machine learning method were 82% and 91% respectively. Whilst the variable sensitivity and specificity values for either radiologist ranged from 24-71% and 75-96% respectively for detection of ECS in SCC-related H&NC. In a HPV+ SCC-related H&NC subgroup analysis, the machine learning method also outperformed either radiologist with an AUC of 0.81 compared to 0.75 and 0.56.[86]

Although these findings were identified in SCC-related H&NC, machine learning methods may have superior diagnostic value in the HPV+ OPSCC population and further research is required.

Biomarkers

Given the unsatisfying diagnostic accuracy of conventional imaging modalities and their abilities to detect ECS in H&NC, attempts have been made to identify specific biomarkers to aid ECS assessment. A study by Michikawa et al.[88] found the presence of cyclin D1 gene (CCND1), a protooncogene, and the epidermal growth factor gene (EGFR), a transmembrane tyrosine kinase receptor, were significantly associated with ECS. Whilst at the protein level, a study by Brennan et al.[89] found that the expression of inducible nitric oxide synthase (iNOS) in the metastatic lymph node was significantly associated with ECS. Although promising findings and with potential to improve diagnostic value to routine clinical work up and radiological techniques, biomarkers for ECS detection have yet to make their way into clinical practice. Nevertheless, these molecules represent an exciting new development in the aim to detect ECS in a less invasive method without reliance on surgical histopathology (current gold standard).

Limitations

Despite every effort throughout the methodological steps to minimize limitations in the review, several limitations do exist.

Regarding the review methods, although the search was conducted both electronically and manually with the aid of a scientific librarian, the limited timeframe and exclusion criteria resulted in six non-English studies being excluded (see Appendix V) and only published studies being retrieved. It is therefore uncertain whether all relevant studies on the topic have been included in the review.

In terms of critical appraisal, unclear risk of bias was found in the majority of studies in the domains of reference standard and flow and timing. Insufficient information regarding the

blinding status of the pathologist to the index test has implications for introducing 'information bias', possibly leading to an overestimation of diagnostic accuracy. Whilst the lack of information regarding the timing interval between index test and reference standard could have introduced 'disease progression bias,' possibly leading to an under-or-overestimation of diagnostic accuracy. Whilst using the QUADAS-2 tool for critical appraisal, it is not advised to assign an overall risk of bias for each of the included studies leaving appraisal results up for some reader interpretation.

Regarding the included studies, all studies were of retrospective study design which has an impact on selection bias amongst the included participants. Whilst only a small number of studies (n=6), predominately from Westernized countries were included in the review. The majority of the included participants identified as Male between 56-60 years of age, however this is in keeping with prevalence and incidence statistics with middle-aged men being the demographic mostly affected by HPV+ OPSCC.[9-12]

In the CT meta-analysis involving four studies,[76,80-82] substantial heterogeneity was reported limiting the applicability of the meta-analysis findings. Multiple factors such as participant demographics, radiologist experience level, imaging protocol and use of different radiological features for the diagnosis of ECS could have contributed to this however no meta-regression was able to be performed to explore this further.

In terms of the reference standard, five of the six studies utilised pathologists that did not routinely differentiate between microscopic and macroscopic ECS, which may be of critical importance moving forward as trials look to differentiate between the management of the two entities.

In terms of individual studies, the included study by Lee et al.[79] reported their findings as 'CT and MRI' therefore precluding this study from any analysis or comparison to CT and PET/CT. The study by Snyder et al.[75] reported their PET/CT findings in terms of an 'ECS misclassification analysis,' and therefore calculations were required for estimates of SN, SP, PPV and NPV outcomes. These calculations were performed by a professional statistician and confirmed by the primary author (TM). Whilst being the only PET/CT study, no meta-analysis was able to be performed. Lastly, the CT study by Faraji et al.[76] investigated the highest performing characteristics to aid radiologic ECS detection with no overall assessment for ECS being reported. Therefore the highest overall performing characteristic for ECS detection ('absence of perinodal fat plane') was used in the review and CT meta-analysis. Given the diagnosis of radiologic ECS typically involves a combination of radiologic

features, the outcome values for radiologic ECS detection in the Faraji et al.[76] study are likely to be underestimated for the purpose of this review.

Implications for practice

Choice of imaging modality to assess ECS

Given that the radiological or histopathological diagnosis of ECS continues to warrant treatment with chemo-radiotherapy (primary or adjuvant),[33,34] there is clinical concern that a certain percentage of patients may undergo unnecessary treatment and side-effects related to chemoradiotherapy when the radiological diagnosis of ECS may be a false-positive. Conversely, patients who are presumed ECS negative on imaging may undergo surgery when the result could be a false-negative and the patient better served with primary chemoradiotherapy. The pre-treatment radiological diagnosis of ECS therefore plays an imperative role in the work up of patients with HPV+ OPSCC.

When assessing the findings from this review, it could appear the use of PET/CT or CT alone may be of greater value in helping exclude rather than detect the presence of ECS in HPV+ OPSCC. Clinically in this instance, if the use of PET/CT or CT suggests the absence of radiologic ECS (along with routine clinical work up), these patients may be better served with surgery and radiotherapy as opposed to treatment involving chemoradiotherapy. However these findings need to be interpretated with caution.

Three summary of findings (SoF) tables were created using the grade approach to rate the certainty of the evidence of the included studies and are presented in Tables 7-9. Utilising GRADE provided a transparent and structured process where the evidence was rated on the following: study design, inconsistency of results, indirectness of evidence, imprecision, publication bias, large magnitude effect and dose-response gradient.[66] Based on the assessment of the SoF tables, the six included studies in the review were all rated to be of 'very low' certainty in their evidence for diagnostic test accuracy outcomes. According to GRADE, 'very low' certainty equates to very little confidence in the effect estimate, with the true effect likely being significantly different from the estimate of effect.[66] As such, the findings of this review reflect the need for further large-scale high-quality studies before validating the findings of PET/CT and CT and their accuracy for radiologic ECS detection amongst patients with HPV+ OPSCC.

Unfortunately no studies exist on the remaining conventional imaging modalities (MRI and ultrasound) in patients with HPV+ OPSCC and ECS to allow for a true comparative analysis amongst CT, PET/CT, MRI and US.

Implications for research

Further research on the topic

Ideally future large-scale, high-quality, multi-centre RCTs are required in the area of imaging modalities and their abilities to detect ECS in which reliable and consistent diagnostic criteria are routinely applied.[90] To avoid verification bias, all participants should ideally undergo both the index test and reference standard regardless of radiologic ECS status. However given the current reference standard of surgical histopathology, this approach would be considered unethical and therefore current observational studies are likely to remain the preferred study design for this type of research.

To improve future research, whilst although the demographic mostly affected by HPV+ OPSCC is middle-aged men,[9-12] for more generalizable results, studies should consider recruiting a more balanced male-female ratio and wider range of participants from all different backgrounds. Whilst pertaining to study design, it is important that radiologists and pathologists are blinded to the patients ECS status, with an index test-reference standard interval of less than six weeks to reduce the risk of bias amongst the index test, reference standard, and flow and timing of the study.

Unfortunately given the lack of data on MRI and US, studies assessing these imaging modalities for ECS in HPV+ OSPCC are also required.

In lieu of high-quality RCTs, future reviews should consider the inclusion of high-quality observational studies. This would provide clinicians with balanced reports on both the benefits and limitations of imaging modalities and their abilities to detect radiologic ECS based on unbiased evidence.

Standardized criteria for ECS

For radiologists and pathologists, currently there is no consistent diagnostic criteria for the detection of radiologic or pathologic ECS. As such, there is an ongoing dispute into the precise definition and assessment criteria for ECS. Unfortunately, this leads to ongoing inconsistency and heterogeneity across a majority of studies where the radiologic and pathologic diagnosis of ECS may be different depending on subjective or institution based criteria. Research is ongoing into forming an internationally accepted and reproducible criteria for ECS detection to try improve homogeneity across studies alike.[91]

Conclusion

By performing this review, we were able report on the diagnostic accuracy of CT and PET/CT for assessing radiologic ECS in HPV+ OPSCC. In the CT meta-analysis, the pooled specificity values (60%) appear too low to suggest clinical value for CT in detecting ECS in HPV+ OPSCC. Pending further research, the use of CT or PET/CT however might have clinically acceptable capabilities at helping exclude the diagnosis of radiologic ECS in HPV+ OPSCC. This may be of benefit to patients who are presumed ECS negative on imaging and recommended a pathway of surgery and radiotherapy rather than chemoradiotherapy, however further large-scale high-quality studies are required to validate these findings.
No studies on MRI or US were identified for assessment of ECS in HPV+ OPSCC. Therefore at this stage, further diagnostic studies are required on CT, PET/CT, MRI and US and their abilities to detect ECS in HPV+ OPSCC to allow for a true comparative analysis between imaging modalities. Recent research has suggested new radiological techniques such as machine learning methods,[86-87] and the use of biomarkers[88-89] have potential for high diagnostic value for ECS assessment in H&NC and this may be the way of the future compared to current diagnostic methods.

Funding and Ethics

Funding: the authors state that this work has not received any funding.

Ethic approvals: not required.

Competing Interests

Two supervisors of the Master of Clinical Science (JBI) project (Dr. Foreman and Dr. Hodge) have published an article on the diagnostic value of CT in identifying ECS in P16+ OPSCC[81] which was analysed and discussed in the following review.

REFERENCES

1. Almulla A, Noel CW, Lu L, Xu W, O'Sullivan B, Goldstein DP, et al. Radiologic-pathologic correlation of extranodal extension in patients with squamous cell carcinoma of the oral cavity: implications for future editions of the TNM classification. Int J Radiat Oncol Biol Phys. 2018 May;102:698–708
2. Park CH, Song CM, Ji YB, Pyo JY, Yi KJ, Song YS, et al. Significance of the Extracapsular Spread of Metastatic Lymph Nodes in Papillary Thyroid Carcinoma. Clin Exp Otorhinolaryngol. 2015 Sep;8(3):289-94.
3. Willis RA. Epidermoid carcinoma of the head and neck, with special reference to metastasis. J. Pathol. 1930 33:501–26.
4. Bennett SH, Futrell JW, Roth JA, Hoye RC, Ketcham AS. Prognostic significance of histologic host response in cancer of the larynx or hypopharynx. Cancer. 1971 Nov;28(5):1255-65.
5. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, et al. Defining risk levels in locally advanced head and neck cancers: a comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (# 9501). Head Neck. 2005 Oct;27(10):843-50.
6. Cooper JS, Zhang Q, Pajak TF, Forastiere AA, Jacobs J, Saxman SB, et al. Long-term follow-up of the RTOG 9501/intergroup phase III trial: postoperative concurrent radiation therapy and chemotherapy in high-risk squamous cell carcinoma of the head and neck. Int J Radiat Oncol Biol Phys. 2012 Dec 1;84(5):1198-205.
7. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al. American Joint Commission on Cancer. AJCC Cancer Staging Manual, 8th edn. Springer International Publishing; 2017.
8. Mifsud M, Eskander A, Irish J, Gullane P, Gilbert R, Brown D, et al. Evolving trends in head and neck cancer epidemiology: Ontario, Canada 1993-2010. Head Neck 2017;39(9):1770–8.
9. Australian Institute of Health and Welfare. Cancer survival and prevalence in Australia: period estimates from 1982 to 2010. Cancer series. No 69. Cat. no. CAN 65. 2012 [cited 2021Apr 7]. Available from: https://www.aihw.gov.au/reports/cancer/cancer-survival-and-prevalence-in-australia-perio/contents/table-of-contents.
10. Australian Institute of Health and Welfare. Cancer in Australia: an overview 2014. Cancer series. No 90. Cat. no. CAN 88. 2014 [cited 2021 Apr 7]. Available from: https://www.aihw.gov.au/reports/cancer/cancer-in-australia-anoverview-2014/contents/table-of-contents.
11. Australian Institute of Health and Welfare. Australian Cancer Incidence and Mortality (ACIM) books. 2021 [cited 2022 Feb 04]. Available from: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/notes.
12. Australian Institute of Health and Welfare. Cancer in Australia 2021. Cancer series no. 119. Cat. no. CAN 122. 2021 [cited 2022 Feb 04]. Available from: https://www.aihw.gov.au/getmedia/8c9fcf52-0055-41a0-96d9-f81b0feb98cf/aihw-can-123.pdf.aspx.
13. Sanderson RJ, Ironside JA. Squamous cell carcinomas of the head and neck. BMJ 2002;325(7368):822–7.
14. Siegel R, Miller K, Jemal A. Cancer statistics, 2017. CA Cancer J Clin 2017;67(1):7–30.

15. Marur S, Forastiere A. Head and neck cancer: changing epidemiology, diagnosis, and treatment. Mayo Clin Proc 2008;83(4):489–501.
16. Sturgis E, Cinciripini P. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? Cancer 2007;110:1429–35.
17. Chaturvedi A, Engels E, Pfeiffer R, Hernandez B, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 2011; 29:4294–301.
18. Hong A, Lee CS, Jones D, Veillard A, Zhang M, Zhang X, et al. Rising prevalence of human papillomavirus-related oropharyngeal cancer in Australia over the last 2 decades. Head Neck 2016;38(5):743–50.
19. Pan C, Issaeva N, Yarbrough WG. HPV-drive oropharyngeal cancer: current knowledge of molecular biology and mechanisms of carcinogenesis. Cancers Head Neck 2018;3:12.
20. Jung AC, Briolat J, Millon R, Reynies A, Rickman D, Thomas E, et al. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. Int J Cancer 2010;126:1882–94.
21. Rischin D, Young RJ, Fisher R, Fox S, Le Q, Peters L, et al. Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. J Clin Oncol 2010;28(27): 4142–8.
22. Posner MR, Lorch JH, Goloubeva O, Tan M, Schumaker L, Sarlis N, et al. Survival and human papillomavirus in oropharynx cancer in TAX324: a subset analysis from and international phase III trial. Ann Oncol 2011;22(5):1071–7.
23. Kumar B, Cipolla MJ, Old MO, Brown N, Kang S, Dziegielewski P, et al. Surgical management of oropharyngeal squamous cell carcinoma: survival and functional outcomes. Head Neck 2016;38(Suppl 1):E1794.
24. Morani, A.C.; Eisbruch, A.; Carey, T.E.; Hauff, S.J.; Walline, H.M.; Mukherji, S.K. Intranodal cystic changes: A potential radiologic signature/biomarker to assess the human papillomavirus status of cases with oropharyngeal malignancies. J. Comput. Assist. Tomogr. 2013;May-Jun 37(3):343–345.
25. Haave H, Gulati S, Brekke J, Lybak S, Vintermyr OK, Aarstad HJ. Tumor stromal desmoplasia and inflammatory response uniquely predict survival with and without stratification for HPV tumor infection in OPSCC patients. Acta Otolaryngol. 2018 Nov;138(11):1035-1042.
26. Jamal Z, Anjum F. Oropharyngeal Squamous Cell Carcinoma. StatPearls 2022.
27. Rosenzweig MQ, Gardner D, Griffith B. The History and Physical in Cancer Care: A Primer for the Oncology Advanced Practitioner. J Adv Pract Oncol. 2014 Jul-Aug;5(4):262-8.
28. Monnet O, Cohen F, Lecorroller T, Vidal V, Jacquier A, Gaubert JY, Bartoli JM, Moulin G. Adénopathies cervicales [Cervical lymph nodes]. J Radiol. 2008 Jul-Aug;89(7-8 Pt 2):1020-36.
29. Su Z, Duan Z, Pan W, Wu C, Jia Y, Han B, et al. Predicting extracapsular spread of head and neck cancers using different imaging techniques: a systematic review and metaanalysis. Int J Oral Maxillofac Surg 2016;45(4):413–21.
30. Vishwanath V, Jafarieh S, Rembielak A. The role of imaging in head and neck cancer: An overview of different imaging modalities in primary diagnosis and staging the disease. Journal of contemporary brachytherapy. 2020;12(5):512-518.
31. Fischer CA, Zlobec I, Green E, Probst S, Storck C, Lugli A, et al. Is the improved prognosis of p16 positive oropharyngeal squamous cell carcinoma dependent of the treatment modality? Int J Cancer. (2010) 126:1256–62.

32. De Almeida JR, Li R, Magnuson JS, Smith RV, Moore E, Lawson G, et al. Oncologic outcomes after transoral robotic surgery. JAMA Otolaryngol Neck Surg. (2015) 141:1043
33. National Comprehensive Cancer Network. NCNN Guidelines for Cancer of the Oropharynx (version 1.2022). Retrieved from https://www.nccn.org/patients/guidelines/content/PDF/hn-oropharyngeal-patient.pdf
34. Machiels JP, René Leemans C, Golusinski W, Grau C, Licitra L, Gregoire V. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020 Nov;31(11):1462-1475.
35. Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. N Engl J Med 2004 May 6;350:1945– 52.
36. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Radiation Therapy Oncology Group 9501/Intergroup. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. N Engl J Med. 2004 May 6;350(19):1937-44.
37. Moore ZR, Pham NL, Shah JL, Nedzi L, Sumer BD, Day A, et al. Risk of unplanned hospital encounters in patients treated with radiotherapy for head and neck squamous cell carcinoma. J Pain Symptom Manage 2019 Apr 57:738–45.e3.
38. Park S, Guenette J, Suh C, Hannah G, Chung S, Baek J, et al. The diagnostic performance of CT and MRI for detecting extranodal extension in patients with head and neck squamous cell carcinoma: a systematic review and diagnostic meta-analysis. Eur Radiol 2021 April 31:2048–2061.
39. Abdel-Halim C, Rosenberg T, Dyrvig A, Høilund-Carlsen P, Sørensen J, Rohde M, et al. Diagnostic accuracy of imaging modalities in detection of histopathological extranodal extension: A systematic review and meta-analysis. Oral Oncol. 2021 Mar 114:105169.
40. Mackey A, Bassendowski S. The History of Evidence-Based Practice in Nursing Education and Practice. Journal of Professional Nursing. 20017 Jan-Feb 33(1):51-55.
41. Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust 1972.
42. Guyatt GH. Evidence-based medicine. ACP J Club. 1991: 114;A–16.
43. Sur RL, Dahm P. History of evidence-based medicine. Indian J Urol. 2011;27(4):487-489.
44. Sackett D, Rosenberg W, Gray M, Haynes B, Richardson S. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan 312:71.
45. Satterfield JM, Spring B, Brownson RC, Mullen EJ, Newhouse RP, Walker BB, Whitlock EP. Toward a transdisciplinary model of evidence-based practice. Millbank Q 2009;87(2):368-90.
46. Cohen AM, Hersh WR. Criticisms of evidence-based medicine. Evidence-based Cardiovascular Medicine. 2004 Sep 1;8(3):197-198.
47. Pearson A, Wiechula R, Court A, Lockwood C. The JBI Model of Evidence-based Healthcare. Int J Evid Based Healthc 2005; 3:207–215.
48. Kivunja C, Kuyini AB. Understanding and Applying Research Paradigms in Educational Contexts. The International Journal of Higher Education 2017 6; 26-29.
49. Trohler U. The 18th century British Origins of critical approach. Edinburgh: Royal College of Physicians, 2000.
50. Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. JBI, 2020. Available from: https://reviewersmanual.joannabriggs.org/

51. Jordan Z, Pittman P. A short history of a big idea. Melbourne, The Joanna Briggs Institute, 2006.
52. Denzin N, Lincoln Y. Eds. Handbook of qualitative research. California, Thousand Oaks; SAGE publications, 2005.
53. Black, N. Why we need qualitative research. Journal of Epidemiological Community Health 1994, 48:425-426.
54. Greenhalgh T. Papers that go beyond numbers (qualitative research). BMJ 1997, 315 (7110):740-743.
55. Lind J. A Treatise of the Scurvy. In Three Parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: printed by Sands, Murray and Cochran for A Kincaid and A Donaldson, 1753.
56. Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. BMJ Evidence-Based Medicine 2016;21:125-127.
57. Moher D, Liberati A, Tetzlaff J, Altman DG. the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. PLoS Med 2009;6(7):e1000097.
58. Grant M and Booth. A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. Health Information & Libraries Journal. 2009;26:91-108.
59. Aromataris E, Pearson A. The systematic review: an overview. AJN. Am J Nurs. 2014;114(3):53–58.
60. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. JBI Reports. 2004;2:45–64.
61. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.
62. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010.
63. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. JAMA. 1995 Dec 274(22):1800–1804.
64. Oxford Levels of Evidence Working Group. Levels of evidence. Oxford: Oxford Centre for Evidence-Based Medicine, 2011.
65. Haynes RB. Of studies, syntheses, synopses, summaries and systems: the “5S" evolution of services for evidence-based health care decisions. ACP J Club. 2006 Nov-Dec;145(3):A8-9.
66. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013.
67. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. Bmj, 2001;323(7305):157-62.
68. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect. 2014;20(2):105-13.
69. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007;115(5):654-7.
70. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.
71. Morey T, Stern C, Foreman A, Hodge JC. Accuracy of imaging modalities at detecting extracapsular spread of cervical lymph node metastases in HPV-associated

oropharyngeal cancer: a systematic review protocol of diagnostic test accuracy. JBI Evidence Synthesis. 2022 Jan; 20(1):189-195.
72. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.
73. Scottish Intercollegiate Guidelines Network (SIGN). SIGN critical appraisal checklists for diagnostic study. Edinburgh: Scottish Intercollegiate Guidelines Network; 2006 Available from: https://www.sign.ac.uk/what-we-do/methodology/checklists/
74. Critical appraisal skills programme (CASP). CASP diagnostic checklist [online]. 2018 Available from: https://casp-uk.net/casp-tools-checklists/
75. Snyder V, Goyal LK, Bowers EMR, Kubik M, Kim S, Ferris RL, et al. PET/CT Poorly Predicts AJCC 8th Edition Pathologic Staging in HPV-Related Oropharyngeal Cancer. Laryngoscope. 2021 Jul;131(7):1535-1541.
76. Faraji F, Aygun N, Coquia SF, Gourin CG, Tan M, Rooper LM, et al. Computed tomography performance in predicting extranodal extension in HPV-positive oropharynx cancer. Laryngoscope. 2020 Jun;130(6):1479-1486.
77. Gatsonis C. Do we need a checklist for reporting the results of diagnostic test evaluations? The STARD proposal. Acad Radiol 2003;10:599–600.
78. Salkind NJ. Encyclopedia of Research Design (Vols. 1-0). 2010.
79. Lee B, Choi YJ, Kim SO, Lee YS, Hong JY, Baek JH, et al. Prognostic Value of Radiologic Extranodal Extension in Human Papillomavirus-Related Oropharyngeal Squamous Cell Carcinoma. Korean J Radiol. 2019 Aug;20(8):1266-1274.
80. Geltzeiler M, Clayburgh D, Gleysteen J, Gross ND, Hamilton B, Andersen P, et al. Predictors of extracapsular extension in HPV-associated oropharyngeal cancer treated surgically. Oral Oncol. 2017 Feb;65:89-93.
81. Noor A, Mintz J, Patel S, Bajic N, Boase S, Sethi N, et al. Predictive value of computed tomography in identifying extracapsular spread of cervical lymph node metastases in p16 positive oropharyngeal squamous cell carcinoma. J Med Imaging Radiat Oncol. 2019 Aug;63(4):500-509.
82. Patel MR, Hudgins PA, Beitler JJ, Magliocca KR, Griffith CC, Liu Y, et al. Radiographic Imaging Does Not Reliably Predict Macroscopic Extranodal Extension in Human Papilloma Virus-Associated Oropharyngeal Cancer. ORL J Otorhinolaryngol Relat Spec. 2018;80(2):85-95.
83. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56(1):45-50.
84. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009;19(4):203-211. Published 2009 Jan 20.
85. Ferris R, Flamand Y, Weinstein G, Li S, Quon H, Mehra R, et al. Phase II Randomized Trial of Transoral Surgery and Low-Dose Intensity Modulated Radiation Therapy in Resectable p16+ Locally Advanced Oropharynx Cancer: An ECOG ACRIN Cancer Research Group Trial (E3311). Journal of Clinical Oncology 2022 40(2):138-149.
86. Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, et al. Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma. J Clin Oncol. 2020;38(12):1304-1311.

87. Frood R, Palkhi E, Barnfield M, Prestwich R, Vaidyanathan S, Scarsbrook A. Can MR textural analysis improve the prediction of extracapsular nodal spread in patients with oral cavity cancer? European Radiology. 2018;28(12): 5010-18.
88. Michikawa C, Uzawa N, Sato H, Ohyama Y, Okada N, Amagasa T. Epidermal growth factor receptor gene copy number aberration at the primary tumour is significantly associated with extracapsular spread in oral cancer. Br J Cancer. 2011;104(5):850-855.
89. Brennan PA, Dennis S, Poller D, Quintero M, Puxeddu R, Thomas GJ. Inducible nitric oxide synthase: correlation with extracapsular spread and enhancement of tumor cell invasion in head and neck squamous cell carcinoma. Head Neck. 2008 Feb;30(2):208-14.
90. Rodger M, Ramsay T, Fergusson D. Diagnostic randomized controlled trials: the final frontier. Trials. 2012;13:137.
91. Lodder WL, van den Brekel MW. Re: Extracapsular tumor extension in cervical lymph nodes: reconciling the literature and seer data. Head Neck. 2011 Dec;33(12):1809.

APPENDICES

Appendix I: Search strategy

MEDLINE (via PubMed). Search conducted on 18th of May 2021.

Search	Query
#1	(("Head and neck neoplasms"[mh] OR head and neck neoplasm*[tiab] OR oropharyngeal squamous cell carcinoma[tiab] OR squamous cell carcinoma of the head and neck[tiab] OR ("oropharynx"[mh] AND "carcinoma, squamous cell"[mh]) OR (head and neck[tiab] OR head[tiab] OR neck[tiab] OR oropharynx[tiab] OR oral[tiab] OR throat[tiab] OR mouth[tiab])) AND (neoplasm*[tiab] OR tumour*[tiab] OR tumor*[tiab] malignan*[tiab] OR metastas*[tiab] OR cancer*[tiab] OR carcinoma*[tiab] OR P16[tiab] OR HPV[tiab] OR human papillomavirus[tiab] OR nodal[tiab])
#2	"Diagnostic imaging"[mh] OR imaging[tiab] OR tomography[tiab] OR computed tomography[tiab] OR CT[tiab] OR magnetic resonance imaging[tiab] OR MRI[tiab] OR positron emission tomography[tiab] OR PET[tiab] OR Ultrasound[tiab] OR PET/CT[tiab] OR specificity[tiab] OR sensitivity[tiab]
#3	"Pathology"[mh] OR "pathology, surgical"[mh] OR histopatholog*[tiab] OR histolog*[tiab] OR patholog*[tiab] OR histo*[all] OR patholog*[all]
#4	"Extranodal extension"[mh] OR extracapsular spread[tiab] OR extracapsular extension[tiab] OR ENE[tiab] OR ECS[tiab] OR ECE[tiab]
#5	#1 AND #2 AND #3 AND #4
Total	The number of records retrieved using this strategy was 201.
Limitations	No limitations on date. Language limited to English.

Elsevier Embase. Search conducted on 18th of May 2021.

Search	Query
#1	(‘Head and neck neoplasms'/exp OR ‘head and neck neoplasm*':ti,ab OR ‘oropharyngeal squamous cell carcinoma’:ti,ab OR ‘squamous cell carcinoma of the head and neck’:ti,ab OR ('oropharynx'/exp AND 'carcinoma, squamous cell'/exp) OR (‘head neck’:ti,ab) OR head:ti,ab OR neck:ti,ab OR oropharynx:ti,ab OR oral:ti,ab OR throat:ti,ab OR mouth:ti,ab) AND (neoplasm*:ti,ab OR tumour*:ti,ab OR tumor*:ti,ab malignan*:ti,ab OR metastas*:ti,ab OR cancer*:ti,ab OR carcinoma*:ti,ab OR p16:ti,ab OR hpv:ti,ab OR ‘human papillomavirus’:ti,ab OR nodal:ti,ab)
#2	‘Diagnostic imaging'/exp OR imaging:ti,ab OR tomography:ti,ab OR ‘computed tomography’:ti,ab OR ct:ab OR ‘magnetic resonance imaging’:ti,ab OR mri:ti,ab OR ‘positron emission tomography’:ti,ab OR ultrasound:ti,ab OR pet:ti,ab OR ct:ti,ab OR specificity:ti,ab OR sensitivity:ti,ab
#3	'Pathology'/exp OR ‘pathology, surgical’ OR histopatholog*:ti,ab OR histolog*:ti,ab OR patholog*:ti,ab OR histo* OR patholog*
#4	‘Extranodal extension’/exp OR ‘extracapsular spread’:ti,ab OR ‘extracapsular extension’:ti,ab OR ene:ti,ab OR ecs:ti,ab OR ece:ti,ab

#5	#1 AND #2 AND #3 AND #4
Total	The number of records retrieved using this strategy was 231.
Limitations	No limitations on date. Language limited to English.

Cochrane Central Register of Controlled Trials (via Cochrane Library). Search conducted on 18th of May 2021. Screening titles, abstracts and keywords.

Search	**Query**
#1	("Head and neck neoplasms" OR "head and neck neoplasm*" OR "oropharyngeal squamous cell carcinoma" OR "squamous cell carcinoma of the head and neck" OR (oropharynx AND "carcinoma, squamous cell")) OR ("head and neck" OR head OR neck OR oropharynx OR oral OR throat OR mouth) AND (neoplasm* OR tumour* OR tumor* malignan* OR metastas* OR cancer OR carcinoma OR P16 OR HPV OR "human papillomavirus" OR nodal)
#2	"Diagnostic imaging" OR imaging OR tomography OR computed tomography OR CT OR "magnetic resonance imaging" OR MRI OR "positron emission tomography" OR PET OR Ultrasound OR specificity OR sensitivity
#3	"Pathology" OR "pathology, surgical" OR histopatholog* OR histolog* OR patholog* OR histo* OR pathology*
#4	"Extranodal extension" OR "extracapsular spread" OR "extracapsular extension" OR ENE OR ECS OR ECE
#5	#1 AND #2 AND #3 AND #4
Total	The number of records retrieved using this strategy was 1093.
Limitations	No limitations on date. Language limited to English.

Web of Science. Search conducted on 18th of May 2021. Screening: title, abstract, author keywords, and keywords plus.

Search	**Query**
#1	("Head and neck neoplasms" OR "head and neck neoplasm*" OR "oropharyngeal squamous cell carcinoma" OR "squamous cell carcinoma of the head and neck" OR ("oropharynx"AND "carcinoma, squamous cell") OR ("head and neck" OR head OR neck OR oropharynx OR oral OR throat OR mouth) AND (neoplasm* OR tumour* OR tumor* malignan* OR metastas* OR cancer OR carcinoma OR P16 OR HPV OR "human papillomavirus" OR nodal)
#2	"Diagnostic imaging" OR imaging OR tomography OR computed tomography OR CT OR "magnetic resonance imaging" OR MRI OR "positron emission tomography" OR PET OR Ultrasound OR specificity OR sensitivity
#3	"Pathology" OR "pathology, surgical" OR histopatholog* OR histolog* OR patholog* OR histo* OR patholog*

#4	“Extranodal extension” OR “extracapsular spread” OR “extracapsular extension” OR ENE OR ECS OR ECE
#5	#1 AND #2 AND #3 AND #4
Total	The number of records retrieved using this strategy was 158.
Limitations	No limitations on date. Language limited to English.

Scopus. Search conducted on 18th of May 2021. Screening article title, abstract and keywords.

Search	**Query**
#1	("Head and neck neoplasms" OR "head and neck neoplasm*" OR "oropharyngeal squamous cell carcinoma" OR "squamous cell carcinoma of the head and neck" OR ("oropharynx" AND "carcinoma, squamous cell") OR ("head and neck" OR head OR neck OR oropharynx OR oral OR throat OR mouth) AND (neoplasm* OR tumour* OR tumor* malignan* OR metastas* OR cancer* OR carcinoma* OR P16 OR HPV OR "human papillomavirus" OR nodal)
#2	"Diagnostic imaging" OR imaging OR tomography OR "computed tomography" OR CT OR "magnetic resonance imaging" OR MRI OR "positron emission tomography" OR PET OR Ultrasound OR PET/CT OR specificity OR sensitivity
#3	"Pathology" OR "pathology, surgical" OR histopatholog* OR histolog* OR patholog* OR histo* OR patholog*)
#4	"Extranodal extension" OR "extracapsular spread" OR "extracapsular extension" OR ENE OR ECS OR ECE
#5	#1 AND #2 AND #3 AND #4
Total	The number of records retrieved using this strategy was 89.
Limitations	No limitations on date. Language limited to English.

A hand search for primary studies published in the references of included studies was undertaken with nil further studies identified for retrieval.

Appendix II: Studies excluded on full text (n=104).

Aiken AH, Poliashenko S, Beitler JJ, Chen AY, Baugnon KL, Corey AS, et al. Accuracy of Preoperative Imaging in Detecting Nodal Extracapsular Spread in Oral Cavity Squamous Cell Carcinoma. AJNR Am J Neuroradiol. Jul 2015.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Almulla A, Noel CW, Lu L, Xu W, O'Sullivan B, Goldstein DP, et al. Radiologic-Pathologic Correlation of Extranodal Extension in Patients With Squamous Cell Carcinoma of the Oral Cavity: Implications for Future Editions of the TNM Classification. International Journal of Radiation Oncology Biology Physics. 2018; 102(4);698-708.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Anzai Y, Brunberg JA, Lufkin RB. Imaging of nodal metastases in the head and neck. J Magn Reson Imaging. 1997; 7(5);774-783.
Reason for exclusion: ineligible study design. This study is a review.

Ariji Y, Sugita Y, Nagao T, Nakayama A, Fukuda M, Kise Y, et al. CT evaluation of extranodal extension of cervical lymph node metastases in patients with oral squamous cell carcinoma using deep learning classification. Oral Radiol. 2020; 36(2);148-155.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Baik SH, Seo JW, Kim JH, Lee SK, Choi EC, Kim J, et al. Prognostic Value of Cervical Nodal Necrosis Observed in Preoperative CT and MRI of Patients With Tongue Squamous Cell Carcinoma and Cervical Node Metastases: A Retrospective Study. 2019; 213(2):437-443.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Bhattasali O, Thompson LDR, Schumacher AJ, Iganej S. Radiographic nodal prognostic factors in stage I HPV-related oropharyngeal squamous cell carcinoma. Head Neck. 2019; 41(2):398-402.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Billfalk-Kelly A, Yu E, Su J, O'Sullivan B, Waldron J, Ringash J, et al. Radiologic Extranodal Extension Portends Worse Outcome in cN+ TNM-8 Stage I Human Papillomavirus-Mediated Oropharyngeal Cancer. Int J Radiat Oncol Biol Phys. 2019; 104(5);1017-1027.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Burusapat C, Jarungroongruangchai W, Charoenpitakchai M. Prognostic factors of cervical node status in head and neck squamous cell carcinoma. World journal of surgical oncology. 2015; 13;51.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Carlton, J. A; Maxwell, A. W.; Bauer, L. B.; McElroy, S. M.; Layfield, L. J.; Ahsan, H. Computed tomography detection of extracapsular spread of squamous cell carcinoma of the head and neck in metastatic cervical lymph nodes. The Neuroradiology Journal. 2017;30(3);222-229.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Chai RL, Rath TJ, Johnson JT, Ferris RL, Kubicek GJ, Duvvuri U, et al. Accuracy of Computed Tomography in the Prediction of Extracapsular Spread of Lymph Node Metastases in Squamous Cell Carcinoma of the Head and Neck. JAMA Otolaryngol Head Neck Surg. 2013; 139(11);1187-94.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants who have P16+ OPSCC. Some patients had primary disease of the larynx or unknown primary neoplasm.

Choi KH, Song JH, Park EY, Hong JH, Yoo IR, Lee YS, et al. Analysis of PET parameters as prognosticators of survival and tumor extent in Oropharyngeal Cancer treated with surgery and postoperative radiotherapy. BMC Cancer. 2021; 21;317.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Chung MS, Cheng KL, Choi YJ, Roh JL, Lee YS, Lee SS. Interobserver reproducibility of cervical lymph node measurements at CT in patients with head and neck squamous cell carcinoma. Clin Radiol. 2016; 71(12);1226-1232.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants. Whilst participants were not exclusively limited to those with OPSCC, with some patients having primary disease of the hypopharynx or larynx.

Cole I, Hughes L. The relationship of cervical lymph node metastases to primary sites of carcinoma of the upper aerodigestive tract: A pathological study. Aust N Z J Surg. 1997; 67(12);860-5.
Reason for exclusion: ineligible study design. This study is not a diagnostic test of accuracy study; the study was conducted to find an association between types of neck dissection and primary tumour surgery.

Corby MR, Mukherjee S, Fedder K, Jameson MJ. A novel radiographic grading system to evaluate extranodal extension in regional nodal metastases from head and neck squamous cell carcinoma. International Journal of Radiation Oncology Biology Physics. 2018; 100;1359.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Daniels CP, Liu HYH, Bernard A, Williams C, Foote MC, Ladwa R, et al. The declining role of post-treatment neck dissection in human papillomavirus-associated oropharyngeal cancer. Radiother Oncol. 2020; 151;242-248.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Davis RJ, Rettig E, Aygun N, Rooper L, D'Souza G, Eisele DW, et al. From presumed benign neck masses to delayed recognition of human papillomavirus-positive oropharyngeal cancer. Laryngoscope. 2020; 130(2);392-397.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

De Paz D, Kao HK, Huang Y, Chang KP. Prognostic Stratification of Patients With Advanced Oral Cavity Squamous Cell Carcinoma. Curr Oncol Rep. 2017; 10;19(10);65.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Dequanter D, Shahla M, Aubert C, Deniz Y, Lothaire P. Prognostic value of FDG PET/CT in head and neck squamous cell carcinomas. Onco Targets Ther. 2015; 8;2279-2283.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants. Whilst participants were not exclusively limited to those with OPSCC.

Dhanda J, Shaw R, Hanlon B, Lloyd B, Risk J, Woolgar J, et al. A molecular signature to aid in the clinical diagnosis of ECS in OSCC. British Journal of Oral & Maxillofacial Surgery. 2011; 49(1);16.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Dhanda J, Triantafyllou A, Liloglou T, Kalirai H, Lloyd B, Hanlon R, et al. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. Br J Cancer. 2014; 111;2114–2121.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Duenne A, Barth P, Budach V, et al. Neck dissection after radiochemotherapy for head-neck cancer. Onkologe. 2007; 13;129–138.
Reason for exclusion: non-English study (German).

Dunphy L, Sood V, Hislop WS. Accuracy of MRI in prediction of tumour thickness and nodal stage in oral carcinoma. British Journal of Oral & Maxillofacial Surgery. 2012; 50(1);46.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Frood R, Palkhi E, Barnfield M, Prestwich R, Vaidyanathan S, Scarsbrook A. Can MR textural analysis improve the prediction of extracapsular nodal spread in patients with oral cavity cancer? Eur Radiol. 2018; 28(12);5010-5018.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Fujita A, Buch K, Truong MT, Qureshi MM, Mercier G, Jalisi S, et al. Imaging characteristics of metastatic nodes and outcomes by HPV status in head and neck cancers. Head and Neck. 2015; 126(2);392-398.
Reason for exclusion: ineligible study design. This study was not a diagnostic test of accuracy study; the study investigated the link between nodal recurrence and association with HPV status.

Furukawa M, Dillon JK, Futran ND, Anzai Y. The prevalence of lymph node metastases in clinically N0 necks with oral cavity squamous cell carcinoma: is CT good enough for nodal staging? Acta Radiologica. 2014;55(5);570-578.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Gandikota N, Ng SA, Cotter R, Akin Y, Som PM, Genden E, et al. A combined evaluation strategy of FDG PET/CT and contrast-enhanced diagnostic CT (ceCT) for lymph nodes in the neck. Journal of Clinical Oncology. 2012; 30(15);5578-5578.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

García J, López M, López L, Bagué S, Granell E, Quer M, et al. Validation of the pathological classification of lymph node metastasis for head and neck tumors according to the 8th edition of the TNM Classification of Malignant Tumors. Oral Oncol. 2017; 70;29-33.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Giancarlo T, Palmieri A, Giacomarra V, Russolo M. Pre-operative evaluation of cervical adenopathies in tumours of the upper aerodigestive tract. Anticancer Res. 1998; 18(4B);2805-9.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Golusinski P, Di Maio P, Pehlivan B, Colley S, Nankivell P, Kong A, et al. Evidence for the approach to the diagnostic evaluation of squamous cell carcinoma occult primary tumors of the head and neck. Oral Oncol. 2019; 88;145-152.
Reason for exclusion: ineligible study design. This study is a review.

Greenberg JS, El Naggar AK, Mo V, Roberts D, Myers JN. Disparity in pathologic and clinical lymph node staging in oral tongue carcinoma. Implication for therapeutic decision making. Cancer. 2003; 98(3);508-515.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Hanasoge S, Prabhu R, Magliocca KR, Hudgins PA, Aiken A, Chen AY, et al. Predictive Accuracy of CT Imaging for Nodal Extracapsular Extension in Patients With Laryngeal or Oral Cavity Cancer Managed With Initial Surgery. International Journal of Radiation Oncology, Biology, Physics. 2013; 87;191-192.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Hao SP, Ng SH. Magnetic resonance imaging versus clinical palpation in evaluating cervical metastasis from head and neck cancer. Otolaryngol Head Neck Surg. 2000; 123(3);324-7.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Hararah MK, Stokes WA, Jones BL, Oweida A, Ding D, McDermott J, et al. Nomogram for preoperative prediction of nodal extracapsular extension or positive surgical margins in oropharyngeal squamous cell carcinoma. Oral Oncol. 2018; 83;73-80.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Haughey BH, Sinha P. Prognostic factors and survival unique to surgically treated p16+ oropharyngeal cancer. Laryngoscope. 2012; 122;13-33.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Hogan K., Hoskins K., Noorsaeed A., Wood S. Histopathologic processing and examination of extranodal extension in head and neck squamous cell carcinoma. Mod. Pathol. 2020; 33(3);1859-1860.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Hosal AS, Carrau RL, Johnson JT, Myers EN. Selective neck dissection in the management of the clinically node-negative neck. Laryngoscope. 2000;110(12);2037-2040.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Huang SH, O'Sullivan B, Su J, Bartlett E, Kim J, Waldron JN, et al. Prognostic importance of radiologic extranodal extension in HPV-positive oropharyngeal carcinoma and its potential role in refining TNM-8 cN-classification. Radiother Oncol. 2020; 144;13-22.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Ishibashi N, Maebayashi T, Nishimaki H, Okada M. Computed Tomography of Lymph Node Metastasis Before and After Radiation Therapy: Correlations With Residual Tumour. In Vivo. 2020; 34(5);2721-2725.
Reason for exclusion: ineligible participant characteristics. Included participants had recurrent disease.

Joo YH, Yoo IR, Cho KJ, Park JO, Nam IC, Kim MS. Extracapsular spread and FDG PET/CT correlations in oral squamous cell carcinoma. Int J Oral Maxillofac Surg. 2013; 42(2);158-63.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Kanakamedala MR, Giri SP, Albert AA, Abraham RS, Vijayakumar S. Outcome analysis of extracapsular extension in cervical lymph nodes for locally advanced squamous cell carcinomas of head and neck. Int. J. Radiat. Oncol. Biol. Phys. 2014; 90(1);552-553.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, Decker RH, et al. Pretreatment Identification of Head and Neck Cancer Nodal Metastasis and Extranodal Extension Using Deep Learning Neural Networks. Sci Rep. 2018; 8(1);14036.
Reason for exclusion: ineligible study design. This study used deep learning neural networks rather than conventional imaging modalities to identify radiological ECS.

Kann BH, Buckstein M, Carpenter TJ, Bakst R, Misiukiewicz K, Genden E, et al Radiographic extracapsular extension and treatment outcomes in locally advanced oropharyngeal carcinoma. Head Neck. 2014; 36(12);1689-1694.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Kann BH., Buckstein M., Carpenter T.J., Golchin A., Bakst R.L., Misiukiewicz K., et al. Radiographic extracapsular extension (ECE) and treatment outcomes in locally advanced oropharyngeal carcinoma (OPC). J. Clin. Oncol. 2013; 31;15.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, et al. Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. J. Clin. Oncol. 2020; 38(12);1304-1311.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Kann BH, Hicks DF, Payabvash S, Mahajan A, Gupta V, Burtness B, et al. External Validation and Radiologist Comparison of a Deep Learning Model (DLM) to Identify Extranodal Extension (ENE) in Head and Neck Squamous Cell Carcinoma (HNSCC) with Pretreatment Computed Tomography (CT) Imaging. Int. J. Radiat. Oncol. Biol. Phys. 2019; 105(1);71.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Karaman Z.F., Cagli S., Yuce I., Ozturk M., Guney E., Ozcan N. Characterization of cervical lymph nodes with 16 slice multislice computed tomography and histopathologic correlation. Erciyes Tip Derg. 2009; 31(2);169-175.
Reason for exclusion: non-English study (Turkish).

Kelly HR, Curtin HD. Chapter 2 Squamous Cell Carcinoma of the Head and Neck-Imaging Evaluation of Regional Lymph Nodes and Implications for Management. Seminars in Ultrasound, CT and MRI. 2017; 38(5):466-478.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Kendi AT, Corey A, Magliocca KR, Nickleach DC, Galt J, Switchenko JM, et al. 18F-FDG-PET/CT parameters as imaging biomarkers in oral cavity squamous cell carcinoma, is visual analysis of PET and contrast enhanced CT better than the numbers?. Eur. J. Radiol. 2015; 84(6);1171-1176.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Kimura Y, Sumi M, Sakihama N, Tanaka F, Takahashi H, Nakamura T. MR imaging criteria for the prediction of extranodal spread of metastatic cancer in the neck. Am. J. Neuroradiol. 2008; 29(7);1355-1359.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

King AD, Tse GMK, Yuen EHY, To EWH, Vlantis AC, Zee B, et al. Comparison of CT and MR imaging for the detection of extranodal neoplastic spread in metastatic neck nodes. Eur. J. Radiol. 2004; 52(3);264-270.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Kubicek GJ, Champ C, Fogh S, Wang F, Reddy E, Intenzo C, et al. FDG-PET staging and importance of lymph node SUV in head and neck cancer. Head Neck Oncol. 2010; 2;19.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Lau CK, Cameron MG, Thompson MK. DoesCTnodal involvement correlate with pathology findings?. Br. J. Oral Maxillofac. Surg. 2011; 49;75.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Lee JR, Choi YJ, Roh JL, Kim JS, Lee JH, Cho KJ, et al. Preoperative Contrast-Enhanced CT Versus ^{18}F-FDG PET/CT Evaluation and the Prognostic Value of Extranodal Extension for Surgical Patients with Head and Neck Squamous Cell Carcinoma. Ann Surg Oncol. 2015; 22(3);1020-7.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Liao CT, Chang JTC, Wang HM, Ng SH, Hsueh C, Lee LY, et al. Preoperative [18F]Fluorodeoxyglucose Positron Emission Tomography Standardized Uptake Value of Neck Lymph Nodes Predicts Neck Cancer Control and Survival Rates in Patients With Oral Cavity Squamous Cell Carcinoma and Pathologically Positive Lymph Nodes. Int. J. Radiat. Oncol. Biol. Phys. 2009; 74(4);1054-1061.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Liu JT, Kann BH, De B, Buckstein M, Bakst RL, Genden EM, et al. Prognostic value of radiographic extracapsular extension in locally advanced head and neck squamous cell cancers. International Journal of Radiation Oncology Biology Physics. 2014; 90(1);509-S510.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Liu JT, Kann BH, De B, Buckstein M, Bakst RL, Genden EM., et al. Is radiographic extracapsular extension prognostic in human papillomaviruse-related oropharyngeal cancers?. Int. J. Radiat. Oncol. Biol. Phys. 2014;90(1);S514.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Liu JT., Kann BH, De B, Buckstein M, Bakst RL, Genden EM, et al. Prognostic value of radiographic extracapsular extension in locally advanced head and neck squamous cell cancers. Oral Oncol. 2016; 52;52-57.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Lodder WL, Lange CAH, Van Velthuysen MLF, Hauptmann M, Balm AJM, Van Den Brekel MWM, et al. Can extranodal spread in head and neck cancer be detected on MR imaging. Oral Oncol. 2013; 49(6);626-633.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Lodder WL, Vogel WV, Lange CAH, Hamming-Vrieze O, Van Velthuysen MLF, Pameijer FA, et al. Detection of extranodal spread in head and neck cancer with [18F]FDG PET and MRI: Improved accuracy?. Q. J. Nucl. Med. Mol. Imaging. 2015; 59(3):327-335.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Lu HJ, Tseng SW, Peng CY, Tseng HC, Hsin CH, Chen HL, et al. Predictors of early progression after curative resection followed by platinum-based adjuvant chemoradiotherapy in oral cavity squamous cell carcinoma. Postgrad. Med. 2021;133(3);377-384.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Lwin CT-TJW, Hanlon R, Lowe D, Shaw RJ, Rogers SN, Brown JS, et al. Accuracy of MRI in prediction of tumour thickness and nodal stage in oral carcinoma. Br. J. Oral Maxillofac. Surg. 2011; 49(1);24-25.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Ma DJ, Price KA, Moore EJ, Patel SH, Hinni ML, Garcia JJ, et al. Phase II evaluation of aggressive dose de-escalation for adjuvant chemoradiotherapy in human papillomavirus-associated oropharynx squamous cell carcinoma. J. Clin. Oncol. 2019; 37(22);1909-1918.
Reason for exclusion: ineligible outcomes. The primary end point was locoregional tumour control, and secondary end points involved progression-free survival, overall survival, toxicity and swallow function.

Mair MD, Baker A, Vaidhyanath R. Combined accuracy of computed tomography and magnetic resonance imaging in detecting level wise metastatic neck nodes and extracapsular spread. Ann. Oncol. 2020; 31(4);677.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Maxwell JH, Rath TJ, Byrd JK, Albergotti WG, Wang H, Duvvuri U, et al. Accuracy of computed tomography to predict extracapsular spread in p16-positive squamous cell carcinoma. Laryngoscope. 2015; 125(7);1613-1618.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

McMullen CP, Garneau J, Weimar E, Ali S, Farinhas JM, Yu E, et al. Occult nodal disease and occult extranodal extension in patients with oropharyngeal squamous cell carcinoma undergoing primary transoral robotic surgery with neck dissection. JAMA Otolaryngol. Head Neck Surg. 2019; 145(8);701-707.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Mellor S, Chow K, Colver GB. What is the role of ultrasound evaluation of lymph nodes in patients with high-grade squamous cell carcinoma of the head and neck?. Br. J. Dermatol. 2014; 171(1);76-77.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Meyer MF, Meinrath J, Seehawer J, Lechner A, Odenthal M, Quaas A, et al. The relevance of the lymph node ratio as predictor of prognosis is higher in HPV-negative than in HPV-positive oropharyngeal squamous cell carcinoma. Clin. Otolaryngol. 2018; 43(1);192-198.
Reason for exclusion: ineligible study design. This study utilises a non-conventional indirect measure (lymph node ratio) for assessment of radiological ECS.

Moon H, Choi YJ, Lee YS, Lee SW, Kim SB, Roh JL., et al. Value of extranodal extension detected by computed tomography for predicting clinical response after chemoradiotherapy in head and neck squamous cell cancer. Acta Oto-Laryngol. 2018; 138(4);392-399.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Moreno KF, Cornelius RS, Lucas FV, Meinzen-Derr J, Patil YJ. Using 3 Tesla magnetic resonance imaging in the pre-operative evaluation of tongue carcinoma. J. Laryngol. Otol. 2017; 131(9);793-800.
Reason for exclusion: ineligible study design. This study utilised a non-conventional imaging tool (3 Tesla MRI) for assessment of radiological ECS.

Ochoa E, Stanford-Moore G, Fakhry C, Ryan WR. Predicting Adverse Histopathology and Need for Postsurgical Adjuvant Therapy for Human Papilloma Virus–Associated Oropharynx Carcinoma. Otolaryngology–Head and Neck Surgery. 2021; 165(2);309-316.
Reason for exclusion: ineligible index test. This study utilised clinical nodal staging as opposed to conventional imaging modalities to assess for presence of ECS.

Ozer E, Naiboglu B, Karapinar U, Agrawal A, Ozer HG, Schuller DE. Clinicopathological determinants of positron emission tomography computed tomography fluorodeoxyglucose standardised uptake value in head and neck carcinoma. J. Laryngol. Otol. 2013; 127(7);676-680.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Ozer F, Ozer C, Erkan AN, Yavuz H. [The therapeutic role and effectiveness of selective neck dissection in the management of N0 neck]. Kulak Burun Bogaz Ihtis Derg. 2009; 19(4);192-197.
Reason for exclusion: non-English study (Turkish).

Pauzie A, Gavid M, Dumollard JM, Timoshenko A, Peoc'h M, Prades JM. Infracentimetric cervical lymph node metastasis in head and neck squamous cell carcinoma: Incidence and prognostic value. Eur. Ann. Otorhinolaryngol. Head Neck Dis. 2016; 133(5);307-311.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Persky M.J., Albergotti W.G., Rath T.J., Kubik M.W., Abberbock S., Geltzeiler M, et al. Positive Margins by Oropharyngeal Subsite in Transoral Robotic Surgery for T1/T2 Squamous Cell Carcinoma. Otolaryngology - Head and Neck Surgery (United States), 158(4), 660-666.
Reason for exclusion: ineligible study design. This study was not a diagnostic test of accuracy study; the study investigated the link between different types of surgery and positive margin rates post operatively.

Pilar A, Yu E, Su J, Bartlett E, O'Sullivan B, Waldron JN, et al. Validating and Refining the 8th Edition TNM N-Classification for HPV Negative Oropharyngeal Cancer. Int. J. Radiat. Oncol. Biol. Phys. 2020; 108(3);835.
Reason for exclusion: ineligible participant characteristics. This study did not involve participants with P16+ OPSCC.

Prabhu RS, Magliocca KR, Hanasoge S, Aiken AH, Hudgins PA, Hall WA, et al. Accuracy of computed tomography for predicting pathologic nodal extracapsular extension in patients with head-and-neck cancer undergoing initial surgical resection. Int. J. Radiat. Oncol. Biol. Phys. 2014; 88(1);122-129.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Quinlan-Davidson SR, Mohamed ASR, Myers JN, Gunn GB, Johnson FM, Skinner H, et al. Outcomes of oral cavity cancer patients treated with surgery followed by postoperative intensity modulated radiation therapy. Oral Oncol. 2017; 72;90-97.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Randall DR, Lysack JT, Hudon ME, et al. Diagnostic utility of central node necrosis in predicting extracapsular spread among oral cavity squamous cell carcinoma. Head Neck. 2015; 37(1);92-96.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Rassekh CH, Johnson JT, Myers EN. Accuracy of intraoperative staging of the NO neck in squamous cell carcinoma. Laryngoscope. 1995; 105(12);1334-1336.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Rath TJ, Narayanan S, Hughes MA, Ferris RL, Chiosea SI, Branstetter BF. Solid lymph nodes as an imaging biomarker for risk stratification in human papillomavirus-related oropharyngeal squamous cell carcinoma. Am. J. Neuroradiol. 2017; 38(7);1405-1410.
Reason for exclusion: ineligible study design. This study was not a diagnostic test of accuracy study; the study investigated the presence of cystic lymph nodes on CT to predict treatment failure in patients with HPV-OPSCC.

Roh JL, Lee H, Choi SH, Nam SY, Kim SY. Tumor-related leukocytosis predictive of recurrence and survival in patients with oral cavity squamous cell carcinoma. Oral Dis. 2019; 25(6);1511-1518.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Saindane A.M. Pitfalls in the Staging of Cervical Lymph Node Metastasis. Neuroimaging Clin. North Am. 2013; 23(1);147-166.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Sajeeda S, Panda N, Mann SBS, Katariya S, Kalagara S. The role of ultrasonography in the management of tumors of the neck. Ear Nose Throat J. 2000; 79(8);586-589.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Sharma A, Jaiswal AA, Umredkar G, Barle R, Sharma N, Banerjee PK, et al. Lymph Node Central Necrosis on the Computed Tomography as the Predictor of the Extra Capsular Spread in Metastatic Head and Neck Squamous Cell Carcinoma. Indian J Otolaryngol Head Neck Surg. 2017; 69;323-332.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Shaw RJ, Lowe D, Woolgar JA, Brown JS, Vaughan ED, Evans C, et al. Extracapsular spread in oral squamous cell carcinoma. Head Neck. 2010; 32(6);714-722.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Sheppard SC, Giger R, Bojaxhiu B, Sachpekidis C, Dammann F, Dettmer MS, et al. Multimodal Imaging With Positron Emission Tomography/Computed Tomography and Magnetic Resonance Imaging to Detect Extracapsular Extension in Head and Neck Cancer. Laryngoscope. 2021; 131(1);163-169.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Shonka DC, Shoushtari AN, Thomas CY, Moskaluk C, Read PW, Reibel JF, et al. Predicting residual neck disease in patients with oropharyngeal squamous cell carcinoma treated with radiation therapy: Utility of p16 status. Arch. Otolaryngol. Head Neck Surg. 2009;135(11);1126-1132.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Spurr M, Hallett E, Berry S, Rhys R. Are radiologists aware of their boundaries?. Ultrasound. 2017; 25(2);10-11.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Steinkamp HJ, Beck A, Werk M, Felix R. Extracapsular spread of cervical lymph node metastases: Diagnostic value of magnetic resonance imaging. RoFo Fortschr. Geb. Rontgenstr. Bildgebenden Verfahren. 2002; 174(1);50-55.
Reason for exclusion: non-English study (German).

Steinkamp HJ, Beck A, Werk M, Rademaker J, Felix R. Extracapsular Spread of Cervical Lymph Node Metastases: Diagnostic Relevance of Ultrasound Examinations. Ultraschall Med. 2003; 24(5);323-330.
Reason for exclusion: non-English study (German).

Steinkamp HJ, van der Hoeck E, Böck JC, Felix R. [The extracapsular spread of cervical lymph node metastases: the diagnostic value of computed tomography]. Rofo. 1999; 170(5);457-62.
Reason for exclusion: non-English study (German).

Subramanian HE, Park HS, Barbieri A, Mahajan A, Judson BL, Mehra S, et al. Pretreatment predictors of adjuvant chemoradiation in patients receiving transoral robotic surgery for squamous cell carcinoma of the oropharynx: a case control study. Cancers Head Neck. 2016; 1;7.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Sumi M, Nakamura T. Extranodal spread in the neck: MRI detection on the basis of pixel-based time-signal intensity curve analysis. J. Magn. Reson. Imaging. 2011; 33(4);830-838.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Tam K, Ghazizadeh S, Kwan K, St John M. Prediction of extranodal extension using pretreatment imaging. Otolaryngol. Head Neck Surg. 2020; 163(1);196-197.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Tian S, Ferris MJ, Switchenko JM, Magliocca KR, Cassidy RJ, Jhaveri J, et al. Prognostic value of radiographically defined extranodal extension in human papillomavirus-associated locally advanced oropharyngeal carcinoma. Head Neck. 2019; 41(9);3056-3063.
Reason for exclusion: ineligible study design. This study is a prognostic study not a diagnostic test of accuracy study.

Tirelli G, De Groodt J, Sia E, Belgrano MG, Degrassi F, Boscolo-Rizzo P, et al. Accuracy of the Anatomage Table in detecting extranodal extension in head and neck cancer: A pilot study. J. Med. Imaging 2021; 8(1);014502.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Tirelli G, Palmieri A, Giacomarra V, Russolo M. Pre-operative evaluation of cervical adenopathies in tumours of the upper aerodigestive tract. Anticancer Res. 1998; 18(4 B);2805-2809.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Toya R, Saito T, Matsuyama T, Kai Y, Shiraishi S, Murakami D, et al. Diagnostic value of FDG-PET/CT for the identification of extranodal extension in patients with head and neck squamous cell carcinoma. Anticancer Res. 2020; 40(4);2073-2077.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Url C, Schartinger VH, Riechelmann H, Gluckert R, Maier H, Trumpp M, et al. Radiological detection of extracapsular spread in head and neck squamous cell carcinoma (HNSCC) cervical metastases. Eur. J. Radiol. 2013; 82(10);1783-1787.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Woolgar JA, Beirne JC, Vaughan ED, Lewis-Jones HG, Scott J, Brown JS. Correlation of histopathologic findings with clinical and radiologic assessments of cervical lymph-node metastases in oral cancer. Int J Oral Maxillofac Surg. 1995; 24(1);30-37.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Yadav VS, Anehosur VSR, Adirajaiah S, Krishnamurthy K. Is Site-Specific Assessment of Neck Nodes Relevant for Neck Dissection. J Maxillofac Oral Surg. 2021; 20(4);566-572.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Zheng J, Flaman A, Yegendorf D, Purgina B, Chakraborty S, Gaudet M, et al. Metastatic lymph node features with and without extracapsular extension in head and neck Squamous Cell Carcinoma. Radiother. Oncol. 2019; 133(1);664-665.
Reason for exclusion: ineligible outcomes. This study investigated the link between ECS status and the number of metastatic lymph nodes and P16 status.

Zimmermann I, Durisin M, Raab P, Tschammer JD, Lenarz T, Helmstadter V. Detection of extracapsular lymph node involvement in computed tomography (CT) in patients with squamous cell carcinoma of the head and neck-a retrospective analysis. Laryngo- Rhino-Otol. 2018; 97(2);147.
Reason for exclusion: ineligible participant characteristics. This study did not specify HPV or P16 status amongst their participants.

Zoumalan RA, Kleinberger AJ, Morris LGT, Ranade A, Yee H, Delacure MD, et al. Lymph node central necrosis on computed tomography as predictor of extracapsular spread in metastatic head and neck squamous cell carcinoma: Pilot study. J. Laryngol. Otol. 2010; 124(12);1284-1288.
Reason for exclusion: ineligible participant characteristics. This study did not exclusively involve participants with P16+ OPSCC.

Appendix III: Calculations for the Snyder et al PET/CT study.

Initial outcome data from the Snyder et al. study[75] presented in the form of an 'ECS misclassification analysis' amongst two observers for radiologic detection of ECS:

Inter-Rater Reliability Extracapsular Extension.

		Radiologist B		Kappa	95% CI
		No	Yes		
Radiologist A	No	32	10	0.403	0.139–0.667
	Yes	1	6		

This data was then inputted into a 'confusion matrix' to convert outcomes to our review primary outcomes (SN and SP values):

	Disorder	No Disorder
Positive Test Result	True Positive (TP)	False Positive (FP)
Negative Test Result	False Negative (FN)	True Negative (TN)

Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)
PPV = TP/(TP+FP)
NPV = TN/(FN+TN)

Table with Snyder et al.[75] data in a 'confusion matrix':

	Disorder	**No Disorder**
Positive test result	6 (TP)	1 (FN)
Negative test result	10 (FP)	32 (TN)

Calculated new outcomes:

- **SN**: 6 / 7 = 0.86 x 100 = 86%
- **SP**: 32 / 42 = 0.76 x 100 = 76%
- **PPV**: 6 / 16 = 0.375 x 100 = 37.5%
- **NPV**: 32 / 33 = 0.97 x 100 = 97%

Appendix IV: Characteristics of included studies.

Study	Country	Study Design	Participants / Tumour	Tumour Location	Reference Test	Blinding	Imaging Modality	Diagnostic Scoring System	Index-Reference Test Interval	Observers	Outcomes (%)
2020 Faraji et al.[76]	USA	RS 2006-2015	n=73 Median age 56.7 67M : 6F 73 HPV+ OPSCC 32 ENE+ : 41 ENE-	Oropharynx - 43 Palatine tonsils - 28 BOT or lingual tonsils - 2 Unknown	Pathology	Yes	CECT	1) Capsular contour 2) Node margins 3) Perinodal stranding 4) Perinodal fat plane 5) Necrosis 6) Intranodal cyst 7) Matted nodes	N/A	2	Absence of perinodal fat plane: **SN**: 87.1, 96.4 **SP**: 50, 34.5 **PPV**: 58.8, 52.9 **NPV**: 61.7, 63.3 Interobserver agreement: **K**= 0.5
2017 Geltzeiler et al.[80]	USA	RS 2010-2015	n=100 Age not reported 80M : 20F 100 HPV+ OPSCC 39 ENE+ : 61 ENE-	Oropharynx - 71 Tonsil - 29 BOT	Pathology	Yes	CECT	1) Largest node </> 20mm 2) Density 3) Borders (3 point scale) 4) Standing 5) Level 4 Adenopathy 6) Matted notes 7) # of RSN	≤ 6 weeks	1	**SN**: 64 **SP**: 68 **PPV**: 71 **NPV**: 61 Combination of Severe Irregular Borders and >3 Radiographically Suspicious Nodes: **SN**: 61 **SP**: 94 **PPV**: 92 **NPV**: 67
2019 Lee et al.[79]	Korea	RS 2006-2016	n=134 Mean age 59.9 118 M : 16F 134 HPV+ OPSCC 70 ENE+ : 64 ENE-	Oropharynx - 115 Palatine Tonsil - 14 BOT - 1 Others - 4 Unknown	Pathology	N/A	CECT and MRI	1) Enhancement, thickening, irregularity of nodal rim 2) Infiltration of the adjacent fat or other soft tissue planes	≤ 8 weeks (n=114) ≥ 8 weeks (n=20)	2	For CT and MRI: **SN**: 62 **SP**: 77.8 **PPV**: 61.9 **NPV**: 77.8 Interobserver agreement: Overall: **K**=0.7 CT: K= 0.7 MRI: K= 0.7

2019 Noor et al.[81]	Australia	RS 2010-2016	n=80 Mean age 58 68M : 12F 80 HPV+ OPSCC 43 ECS+ : 37 ECS-	Oropharynx - 56 Tonsil - 23 BOT - 1 Unknown	Pathology	Yes	CECT	1) Capsule smooth 2) Internal characteristics 3) Invasion 4) Diameter 5) Matted nodes 6) Perinodal stranding	≤ 6 weeks	2	**SN**: 56.5, 60.9 **SP**: 73.3, 66.7 **PPV**: 68.4, 65.1 **NPV**: 62.3, 62.5 Interobserver agreement: **K**= 0.4
2018 Patel et al.[82]	USA	RS 2014-2016	n=27 Mean age 57 27M : 0F 27 HPV+ OPSCC	Oropharynx - 14 BOT - 13 Tonsil	Pathology	Yes	CECT	1) Necrosis 2) Lobular contours 3) Perinodal stranding 4) Matted nodes 5) Gross invasion of adjacent structures	N/A	2	For detecting major pENE (>2mm extension): **SN**: 88, 100 **SP**: 52.6, 63.2 **PPV**: 43.8, 53.8 **NPV**: 90.9, 100 Interobserver agreement: **K**= 0.7
2021 Snyder et al.[75]	USA	RS 2011-2017	n=49 Mean age 56.4 43M : 6F 49 HPV+ OPSCC	Oropharynx - 27 Tonsil - 19 BOT - 3 Overlapping	Pathology	Yes	PET / CT	Presence or absence of ECS	N/A	2	**SN**: 85.7 (mean) **SP**: 76.2 (mean) **PPV**: 37.5 (mean) **NPV**: 97.0 (mean) Interobserver agreement: **K** = 0.4

RS, Retrospective study; M, male; F, female; BOT, Base of tongue; RSN, Radiographically suspicious nodes; N/A, not available; SN, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value; K, Cohen's Kappa value.

Appendix V: Non-English studies excluded at full text.

Duenne A, Barth P, Budach V, et al. Neck dissection after radiochemotherapy for head-neck cancer. Onkologe. 2007; 13;129–138.
Language: German.

Karaman Z.F., Cagli S., Yuce I., Ozturk M., Guney E., Ozcan N. Characterization of cervical lymph nodes with 16 slice multislice computed tomography and histopathologic correlation. Erciyes Tip Derg. 2009; 31(2);169-175.
Language: Turkish.

Ozer F, Ozer C, Erkan AN, Yavuz H. [The therapeutic role and effectiveness of selective neck dissection in the management of N0 neck]. Kulak Burun Bogaz Ihtis Derg. 2009; 19(4);192-197.
Language: Turkish.

Steinkamp HJ, Beck A, Werk M, Felix R. Extracapsular spread of cervical lymph node metastases: Diagnostic value of magnetic resonance imaging. RoFo Fortschr. Geb. Rontgenstr. Bildgebenden Verfahren. 2002; 174(1);50-55.
Language: German.

Steinkamp HJ, Beck A, Werk M, Rademaker J, Felix R. Extracapsular Spread of Cervical Lymph Node Metastases: Diagnostic Relevance of Ultrasound Examinations. Ultraschall Med. 2003; 24(5);323-330.
Language: German.

Steinkamp HJ, van der Hoeck E, Böck JC, Felix R. [The extracapsular spread of cervical lymph node metastases: the diagnostic value of computed tomography]. Rofo. 1999; 170(5);457-62.
Language: German.